



The Center for International Cooperation in E-Business
China University of Geosciences, Wuhan 430074, P.R.China,
Tel: 0086-27- 67883357, Fax: 0086-27- 87801763, www.whiceb.com

The Center for International Cooperation in E-Business

The Center for International Cooperation in E-Business is a leading research and development unit of China University of Geosciences, Wuhan, with strong ties to industries, universities and research institutes. Its mission is to advocate an integrated and comprehensive approach to E-Business and E-Commerce development in China, and its strategic, managerial and organizational impacts of relative IT technology. Emphasis is on collaboration with international partners in public or private institutions and other research centers. Our aim is to provide thoughtful and practical guidance to organizations in the format of research projects, education/training program, consultancy and international/domestic trade support services. Our network comprises researchers and specialists from other leading business schools and center in Europe, North America and also Asia. The center is charged with the advancement of knowledge in E-Business and building links with government, industry and commerce.

ISBN 978-0-88953-449-0



University of Calgary Press

The Proceedings of Twentieth Wuhan International Conference on E-Business

Wuhan, P. R. China
May 28-30, 2021

The Proceedings of Twentieth Wuhan International Conference on E-Business



Wuhan, P. R. China

May 28-30, 2021

Conference hosts

The Center for International Cooperation in E-Business, China University of Geosciences, Wuhan, China
School of Economics and Management, China University of Geosciences, Wuhan, China
Baden-Wuerttemberg Cooperative State University Heidenheim, Heidenheim, Germany
College of Business, Alfred University, Alfred, New York, USA

Organizers

The Center for International Cooperation in E-Business, China University of Geosciences, Wuhan, China
School of Economics and Management, China University of Geosciences, Wuhan, China

Conference Co-Chairs

Jing Zhao, Professor, CICEB, School of Economics and Management, China University of Geosciences
Wuhan, China (zhao5563@outlook.com)
Juergen Seitz, Professor, Baden-Wuerttemberg Cooperative State University Heidenheim, Germany
(seitz@dhbw-heidenheim.de)
Doug Vogel, Professor, AIS President 2012-2013, Harbin Institute of Technology, China
(isdoug@hit.edu.cn)

Conference Honorary Chair

Wilfred V. Huang, Professor, Alfred University, USA (fhuang@alfred.edu)

Publication chair & Proceedings Editor

Yiliu (Paul) Tu, Professor, University of Calgary, Canada (paultu@ucalgary.ca)
Maomao Chi, Professor, China University of Geosciences, China (chimaomao@aliyun.com)

Program Committee

Chairs:

Weiguo(Patrick) Fan, Professor, University of Iowa, USA
Zhen Zhu, Professor, China University of Geosciences, Wuhan, China

Members:

Yukun Bao, Professor, Huazhong University of Sciences & Technology, China
Zhao Cai, Assistant Professor, University of Nottingham Ningbo, China
Cong Cao Associate Professor, Zhejiang University of Technology, China
Meng Chen, Assistant Professor, Soochow University, China
Xiayu Chen, Associate Professor, Hefei University of Technology, China
Xusen Cheng, Professor, Renmin University of China, China
Maomao Chi, Professor, China University of Geosciences, Wuhan, China
Zhaohua Deng, Professor, Huazhong University of Sciences & Technology, China
John Qi Dong, Professor, University of Dublin, Ireland
Rong Du, Professor, Xidian University, China
Qiang Gong, Professor, Zhongnan University of Economics and Law, China
Dongxiao Gu, Associate Professor, Hefei University of Technology, China
Tailai Wu, Assistant Professor, Huazhong University of Science and Technology, China
Juho Hamari, Professor, Tampere University, Finland
Zhongyi Hu, Assistant Professor, Wuhan University, China

Yi Jiang, Associate Professor, China University of Geosciences, China
Yuanchun Jiang, Associate Professor, Hefei University of Technology, China
Xiaoling Jin, Associate Professor, Shanghai University, China
Hongxiu Li, Assistant Professor, Tampere University, Finland
Jia Li, Associate Professor, East China University of Science and Technology, China
Mengxiang Li, Associate Professor, Hong Kong Baptist University, China
Xiaoling Li, Professor, Chongqing University, China
Hefu Liu, Professor, University of Science and Technology of China, China
Yaobin Lu, Professor, Huazhong University of Sciences & Technology, China
Jian Mou, Assistant Professor, Pusan National University, Korea
Rohit Nishant, Associate Professor, University of Laval, Canada
Xiaoliang Shen, Professor, Wuhan University, China
Si Shi, Associate Professor, Southwestern University of Finance and Economics, China
Yongqiang Sun, Professor, Wuhan University, China
Yuan Sun, Professor, Zhejiang Gongshang University, China
Yiliu (Paul) Tu, Professor, University of Calgary, Canada
Kanliang Wang, Professor, Renmin University, China
Shan Wang, Associate Professor, University of Saskatchewan, Canada
Lu Wang, Assistant Professor, Zhongnan University of Economics and Law, China
Yi Wang, Professor, Southwestern University of Finance and Economics, China
Christopher Westland, The University of Illinois at Chicago, USA
Qiang Wei, Associate Professor, Tsinghua University, China
Shaobo Wei, Associate Professor, University of Science and Technology of China, China
Hong Wu, Assistant Professor, Huazhong University of Science & Technology, China
Jiang Wu, Professor, Wuhan University, China
Tailai Wu, Assistant Professor, Huazhong University of Science & Technology, China
Nannan Xi, Postdoctoral researcher Tampere University, Finland
Huosong Xia, Professor, Wuhan Textile University, China
Jinghua Xiao, Professor, Sun Yat-Sen University, China
Wen-Lung Shiau, Professor, Zhejiang University of Technology, China
Xiaobo (Bob) Xu, Professor, Xi'an Jiaotong-Liverpool University, China
Jun Yan, Associate Professor, University of Wollongong, Australia
Xiangbin Yan, Professor, University of Science and Technology Beijing, China
Junjie Zhou, Associate Professor, Shantou University, China
Zhongyun Zhou, Associate Professor, Tongji University, China
Ling Zhao, Associate Professor, Huazhong University of Sciences & Technology, China
Weiyong Zhang, Associate Professor, Old Dominion University, USA

Session Chairs

Kanliang Wang, Professor, Renmin University, China
Jinghua Xiao, Professor, Sun Yat-Sen University, China
Rong Du, Professor, Xidian University, China
Yi Jiang, Associate Professor, China University of Geosciences, China

Organization Committee

Chairs: Shuwang Yang, Professor, China University of Geosciences, Wuhan China

Members:

Yao Zhang, Assistant Professor, China University of Geosciences, Wuhan China, orangeyao@126.com

Fei Wang, Assistant Professor, China University of Geosciences, Wuhan China

Jing Wang, Assistant Professor, China University of Geosciences, Wuhan China

Xiaochuan Wang, Associate Professor, China University of Geosciences, Wuhan China

Qian Zhao, China University of Geosciences, Wuhan China

Rui Guo, Professor, China University of Geosciences, China

Jianzhong Xiao, Professor, China University of Geosciences, China

Guangmin Wang, Professor, China University of Geosciences, China

Jundong Hou, Professor, China University of Geosciences, China

Sheng Cheng, Professor, China University of Geosciences, China

Secretary General:

Yao Zhang, China University of Geosciences, Wuhan China

International Advisory Board

Chair: Joey George, Professor, AIS President 2010-2011, Iowa State University, USA

Robert Kauffman, Professor, Copenhagen Business School, Denmark

J. Christopher Westland, The University of Illinois at Chicago, USA

-Pacific Asian

Patrick Chau, Professor, University of Nottingham Ningbo China, China

Guoqing Chen, Professor, Tsinghua University, China

Wei Kwok Kee, Professor, City University of Hong Kong, Hong Kong, China

Ting-Peng Liang, Professor, National Sun Yat-Sen University, Taiwan, China

Feicheng Ma, Professor, Wuhan University, China

Jiye Mao, Professor, Renmin University, China

Michael D. Myers, Professor, AIS President 2006-2007, University of Auckland, New Zealand

Bernard Tan, Professor, AIS President 2009-2010, National University of Singapore, Singapore

Kanliang Wang, Professor, Renmin University, China

Nilmini Wickramasinghe, Professor, Deakin University, Australia

Kang Xie, Professor, Sun Yat-Sen University, China

Qiang Ye, Professor, Harbin Institute of Technology, China

J. Leon Zhao, Professor, City University of Hong Kong, Hong Kong, China

-North American

Bob Carasik, Wells Fargo Bank, USA

Weiguo Fan, Professor, Virginia Polytechnic Institute and State University, USA

Joey George, AIS President 2010-2011, Iowa State University, USA

Zhangxi Lin, Professor, Texas Tech University, USA

Ning Nan, Assistant Professor, University of British Columbia, Canada

Paul A. Pavlou, Professor, Temple University, USA

Arun Rai, Professor, Editor in Chief of MISQ, Georgia State University, USA

Richard Watson, Professor, University of Georgia, USA

Christopher Yang, Professor, Drexel University, Philadelphia, USA

Han Zhang, Professor, Georgia Institute of Technology, USA

Zhongju Zhang, Professor, Arizona State University, USA

-European

David Avison, Professor, ESSEC, France

Niels Bjorn-Andersen, Professor, Copenhagen Business School, Denmark

Marco De Marco, Professor, Università Cattolica, Italy

Reima Suomi, Professor, Turku School of Economics, Finland

Yao-Hua Tan, Professor, Vrije University Amsterdam, Netherlands

Hans-Dieter Zimmermann, Professor, FHS St. Gallen, University of Applied Sciences, Switzerland

Editorial Board of the Proceedings

Yiliu (Paul) Tu, University of Calgary, Canada, Editor

Maomao Chi, China University of Geosciences, China, Editor

- Artificial intelligence and new IS research

Yaobin Lu, Huazhong University of Science & Technology, China

Ling Zhao, Huazhong University of Science & Technology, China

Jiang Wu, Wuhan University, China

-Big Data and Business Analytics

John Qi Dong, University of Groningen, The Netherlands

Rohit Nishant, University of Laval, Canada

-Bright and Dark Sides in E-Business

Zhongyun (Phil) Zhou, Tongji University, China

Xiao-Liang Shen, Wuhan University, China

Yongqiang Sun, Wuhan University, China

Xiao-Ling Jin, Shanghai University, China

-Digital Inclusive Society and ICTs Innovation for Senior Citizens in the New Normal COVID-19

Junjie Zhou, Shantou University, China

Xusen Cheng, Renmin University of China, China

- Digital Technologies, Digital Transformation, and Business Value

Xiaobo (Bob) Xu, International Business School, Xi'an Jiaotong-Liverpool University, China

Weiyong Zhang, Old Dominion University, VA, USA

- E-business Strategy & Online Marketing

Xiaoling Li, Chongqing University, China

Lu Wang, Zhongnan University of Economics and Law, China

- Emerging digital technologies in the workplace

Yi Wang, Southwestern University of Finance and Economics, China

Yuan Sun, Zhejiang Gongshang University, China

Si Shi, Southwestern University of Finance and Economics, China

- Engaging Technologies

Nannan Xi, Tampere University, Finland

Juho Hamari, Tampere University, Finland

Hongxiu Li, Tampere University, Finland

- ICTs in fighting against the COVID-19 pandemic

Zhaohua Deng, Huazhong University of Science & Technology, China

Hong Wu, Huazhong University of Science & Technology, China

- Information Management and Health Outcomes

Zhaohua Deng, Huazhong University of Science & Technology, China

Tailai Wu, Huazhong University of Science & Technology, China

Dongxiao Gu, Hefei University of Technology, China

Jia Li, East China University of Science and Technology, China

- Information Systems and Operations Management

Hefu Liu, University of Science and Technology of China, China

Cai Zhao, University of Nottingham Ningbo, China

Meng Chen, Soochow University, China

-Social Commerce and Immersive Interaction

Xiayu Chen, Hefei University of Technology, China

Yuanchun Jiang, Huazhong University of Science & Technology, China

Shaobo Wei, University of Science and Technology of China, China

-Trust in Digital Economy

Cong Cao, Zhejiang University of Technology, China

Jun Yan, University of Wollongong, China

Mengxiang Li, Hong Kong Baptist University, China

Best Paper Award and Journal Publication Committee

Chair: Yiliu (Paul) Tu, Professor, University of Calgary, Canada

Maomao Chi, Professor, China University of Geosciences, China

Members:

Alain Chong, Professor, University of Nottingham - Ningbo China

Chris Yang, Professor, Drexel University, United States

Chris Westland, Professor, The University of Illinois at Chicago, USA,

Doug Vogel, Professor, Harbin Institute of Technology, China

Jun Wei, Professor, University of West Florida, USA

John Qi Dong, Associate Professor, University of Groningen, The Netherlands

Patrick Y.K. Chau, Professor, University of Nottingham - Ningbo China

Weiguo (Patrick) Fan, Professor, University of Iowa, USA

Wen-Lung Shiau, Professor, Zhejiang University of Technology

Sponsors

Association for Information Systems (AIS)

China Association for Information Systems (CNAIS)

China Information Economics Society

University of Calgary, Canada

Deakin University, Australia

University of North Dakota, USA

New Jersey Institute of Technology, New Jersey, USA

University of Turku, Finland

Harbin Institute of Technology, China

Huazhong University of Sciences & Technology, China

Wuhan University, China

Wuhan University of Technology, China
Zhongnan University of Economics and Law, China
Huazhong Normal University, China
Wuhan Textile University

Sponsoring journals (alphabetical order)

Electronic Commerce Research and Applications (SSCI),
Electronic Markets-The International Journal on Networked business (SSCI),
Electronic Commerce Research (SSCI),
Internet Research (SSCI),
Journal of Database Management (SCIE),
Journal of Organizational and End User Computing (SCI & SSCI),
International Journal of Mobile Communication (SSCI)
International Journal of Networking and Virtual Organizations (EI),
Journal of Systems and Information Technology (EI).

Preface

The annual Wuhan International Conference on E-Business (WHICEB) is an AIS affiliated conference. WHICEB 2021 is held in Wuhan, P R China on May 28 to 30, 2021. WHICEB promotes intellectual research and facilitates academic and corporate networking in e-business and related fields. The intent is to encourage academic research and business development through exchanging ideas about the e-business, global and corporate financial issues, and necessity for continuous innovation. The conference aims at presenting innovative research findings, solutions and approaches to make the Internet a productive and efficient vehicle for global commerce. Whether running an e-business or transforming a business into e-business, we constantly encounter challenges ranging from technological to behavioral issues, from marketing to data analysis issues, and from effectiveness to security issues. The past years all over the world initiatives have been started for the next step of development. Some people have already talked about the fourth industrial revolution. After consumer oriented mass production we focus nowadays on personalized products and services, which demands cyber physical systems, cloud computing and big data. There are integration issues for management of technology, management of supply chains, management of human resources and management of knowledge and intelligence that are being resolved in an e-business environment. Organizations, regardless of its locations and sizes, should consider having a strategic decentralized planning effort that includes e-business as a pillar for sustainable competitive advantage.

Proceedings of the Twentieth WHICEB document the breadth and depth of research from different aspects of business and from different disciplines that have major implications for e-business. There are twelve tracks in the proceedings and the proceedings will be listed in the appropriate indexes. The selected best papers from the proceedings will be recommended to international academic journals including but not limited to the following: Electronic Commerce Research and Applications (SSCI), Electronic Commerce Research (SSCI), International Journal of Networking and Virtual Organizations (EI), and International Journal of Services Technology and Management (EI).

The research papers in the proceedings went through a double blind peer review process. Papers are accepted based upon a clear research methodology and contributions to the knowledge of e-business including but not limited to case study, experiment, simulation or survey. The efforts made by our track chairs in reviewing submissions are really appreciated, which ensures the quality of the proceedings. I personally thank them for their professional diligence. They are: Yaobin Lu, Ling Zhao and Jiang Wu, Artificial intelligence and new IS research; John Qi Dong and Rohit Nishant, Big Data and Business Analytics; Zhongyun (Phil) Zhou , Xiao-Liang Shen, Yongqiang Sun and Xiao-Ling Jin, Bright and Dark Sides in E-Business; Junjie Zhou and Xusen Cheng, Digital Inclusive Society and ICTs Innovation for Senior Citizens in the New Normal COVID-19; Xiaobo (Bob) Xu and Weiyong Zhang, Digital Technologies, Digital Transformation, and Business Value; Xiaoling Li and Lu Wang, E-business Strategy & Online Marketing; Yi Wang, Yuan Sun and Si Shi, Emerging digital technologies in the workplace; Nannan Xi, Juho Hamari and Hongxiu Li, Engaging Technologies; Zhaohua Deng and Hong Wu, ICTs in fighting against the COVID-19 pandemic; Zhaohua Deng, Tailai Wu, Dongxiao Gu and Jia Li, Information Management and Health Outcomes; Hefu Liu, Cai Zhao, and Meng Chen, Information Systems and Operations Management; Xiayu Chen, Yuanchun Jiang and Shaobo Wei, Social Commerce and Immersive Interaction; Cong Cao, Jun Yan and Mengxiang Li, Trust in Digital Economy.



Yiliu (Paul) Tu

Editor, Proceedings of 20th Wuhan International Conference on E-Business
University of Calgary, Canada,

Welcome Message from the President of China University of Geosciences

The world today is in the wave of economic globalization and informatization. The Internet and Information Technology (IT) have provided new opportunities to the world economy with tremendous achievements. In this globalizing information era, IT and the Internet are silently changing the way people work, live and learn. In recent years, many scholars have conducted research on e-business from many different disciplines such as computer science, telecommunications, economics, management, human resources, law and sociology. As a result of these efforts, e-business has created a bright perspective for further development in the world and is expected to continue to contribute to global growth and stability.

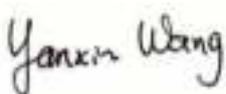
The Wuhan International Conference on E-business is jointly organized by the International Cooperation Center for E-business and the College of Economy and Management, China University of Geosciences (Wuhan), and by the College of Business, Alfred University in the United States and Baden-Wuerttemberg Cooperative State University Heidenheim, Heidenheim, Germany.

This conference has become an AIS Affiliated Conference In 2011, and the Proceedings of the Twentieth Wuhan International Conference on E-Business (WHICEB 2021) will be included into AIS library. The focus of this conference is related to realizing the full potential of technology in the globalization era. The conference will demonstrate the latest achievement in the fields of Internet economy and organizations. Conference tracks include Artificial intelligence and new IS research, Big Data and Business Analytics, Bright and Dark Sides in E-Business, Digital Inclusive Society and ICTs Innovation for Senior Citizens in the New Normal COVID-19, Digital Technologies, Digital Transformation, and Business Value, E-business Strategy & Online Marketing, Emerging digital technologies in the workplace, Engaging Technologies, ICTs in fighting against the COVID-19 pandemic, Information Management and Health Outcomes, Information Systems and Operations Management, Social Commerce and Immersive Interaction, Trust in Digital Economy. It will providing scholars from home and abroad an academic exchange platform for the promotion of technological innovation and international cooperation, and hence the healthy development of e-business in the world.

CUG, a state key university in China with geosciences as its world-famous academic programs, always attaches great importance to the studies of the interaction among population, economy, society, resources and the environment, in an attempt to promote the harmonious development of both human and nature, and pays great attention to the development of international academic exchange platforms for promoting the international collaboration, and facilitating the process of internationalization of the university.

In recent years, great progress has been made on e-business and the development of management science, which not only promotes the development of management science in CUG, but also provides an important support for the construction of a complete disciplinary system based on geosciences for CUG.

The prodigious economic growth in China will provide essentials and supports for the studies of e-business and related fields. E-business will have a major impact on our future prosperity in all facets of life, business and government. Let us work together to strive for a more dynamic e-environment and a more enjoyable life for mankind in this magic e-era!



Yanxin Wang

PhD, Professor

President

China University of Geosciences, Wuhan, China

Welcome Message from the AIS President

I am delighted to welcome you to the 20th Wuhan International Conference on E-Business (WHICEB) conference. The conference provides an excellent forum for faculty and students to build friendships and share creative ideas and findings. Local conferences are an important part of our community because they provide opportunities for faculty and students to come together, form friendships, and exchange ideas. By fostering such relationships and ideas, WHICEB cultivates high quality scholarship in China, in Asia, and around the globe.

On behalf of the Association for Information Systems (AIS), I would like to thank the conference organizers, Jing Zhao, from the China University of Geosciences, Juergen Seitz from Baden-Wuerttemberg Cooperative State University in Heidenheim, Germany, and Doug Vogel from Harbin Institute of Technology for their service to the conference and to the Information Systems community. I would like to extend a special thanks to Wilfred V. Huang, honorary conference chair, for his many years of service. The conference organizers have assembled a world-class conference committee, attracted a strong set of globally recognized journal outlets, and attracted significant global sponsorship to ensure that the conference, and its delegates is an ongoing success.

The AIS is honored to count WHICEB as an affiliated conference. The past two years have been very different ones for this conference, and the world as whole. The conference provides an exemplary example of how a group of scholars can come together and have a lasting impact on the quality of research and networking that is necessary to advance the Information Systems discipline.

As AIS President, I also want to personally invite you to join the Association and participate in AIS activities. In addition to WHICEB, AIS offers many high-quality conferences, including the International Conference on Information Systems (ICIS), the Pacific Asia Conference on Information Systems (PACIS), the Americas Conference on Information Systems (AMCIS) and the European Conference on Information Systems (ECIS). The latter three conferences are available as a Virtual Conference Series which can be accessed through a single cost-effective registration which will open in May.

As AIS President, I personally invite each of you to familiarize yourself with the Association. We offer a growing set of digital services, such as the eLibrary, where the WHICEB conference proceedings will be available to download, webinars on research methods and career management, and access to workshops that can help develop your skills for research and teaching. In particular, I would like to bring your attention to our Research Exchange initiative which we introduced as a series of virtual events that are of interest to the AIS community. Previous events have been recorded and are available at the Research Exchange web link.

For AIS members and non-members alike, I urge you to visit www.aisnet.org to learn more about the benefits of membership.

Best wishes for your conference.



Brian Fitzgerald

President, Association for Information Systems

Director, Lero, the Science Foundation Ireland Research Centre for Software (www.lero.ie)

Frederick Krehbiel Chair in Innovation, University of Limerick

Keynotes Speakers



Patrick Y.K. CHAU is Vice Provost for Research and Knowledge Exchange and Li Dak Sum Chair Professor in Information Systems and Operations Management at the University of Nottingham Ningbo China (UNNC). Prior to joining UNNC in March 2019, he was Padma and Hari Harilela Professor in Strategic Information Management at The University of Hong Kong. Dr. Chau conducts research in IT adoption/implementation and electronic commerce. He has close to 100 journal publications with many of them published in various toptier and highly reputed journals, including MIS Quarterly, Journal of Management Information Systems, Journal of the Association for Information Systems, European Journal of Information Systems, Decision Support Systems, Information & Management, International Journal of Electronic Commerce, Decision Sciences and others. As of April 2021, the total number of citations of his papers in Google Scholar reaches over 22,700 with an H-index of 56. He is currently the Editor-in-Chief of Information & Management and a Senior Editor of Journal of the Association for Information Systems. He received the AIS Fellow Award in 2013 and is currently President-Elect of the Association for Information Systems



Kevin Hong is a Full Professor of MIS, Director of the Bauer College PhD Programs, and a Bauer Senior Fellow in the C. T. Bauer College of Business at University of Houston. Kevin currently serves as a Senior Editor of Production and Operations Management, and an Associate Editor at Information Systems Research and the Journal of the Association for Information Systems. He is also currently guest co-editing a POM Special Issue on Social Technologies in Operations. Kevin's research interests are in the areas of Future of Work, Digital Platforms, User-generated Content, and Human-AI Interactions. His research has been published in premier journals such as Management Science, Information Systems Research, MIS Quarterly, Production and Operations Management, Journal of Management Information Systems, Journal of the Association for Information Systems, and Journal of Consumer Psychology. His research papers have won a number of best paper awards at major conferences, including the Workshop on Information Systems and Economics (2018), the International Conference on Information Systems (2012, 2018, 2020), Hawaii International Conference on System Sciences (2017), America's Conference on Information Systems (2012), and the China Summer Workshop on Information Management (2018). In 2017, Kevin was awarded the college-wide W. P. Carey Faculty Research Award. Further, he was awarded the Association for Information Systems Early Career Award (2018) and the INFORMS Information Systems Society Sandy Slaughter Early Career Award (2019). And recently, he was awarded the Associate Editor of the Year Award (2018) from Information Systems Research. According to the AIS Research Rankings, Kevin was ranked #3 in the world based on publications in the top four information systems journals (MISQ, ISR, JMIS, JAIS). And most recently, Kevin received the Outstanding PhD Alumni Award (2020) from the Fox School of Business, Temple University.

CONTENTS

Artificial intelligence enabled Business Innovation

Enterprise Adoption of User Ideas in the Online User Innovation Community: An Empirical Analysis based on IAM <i>Ning Zhang, Wenfei Zhao, Zhiliang Pang, Lifeng He, Weiguo Fan</i>	1
Exploring the Collaborative Enabling and Value Innovation Mechanism in Platform Enterprises:A Case Study <i>Xueyan Dong, Zidie Chen, Xu Liu, Tienan Wang</i>	3
An AdaBoost-DT Model for Credit Scoring <i>Jiali Zhao, Zengyuan Wu, Bei Wu</i>	15
Robotic Workforces’ Value and Relationships with Human Employees: The Impacts on Human Work Efficacy <i>Meng Zhang, Yi Wang</i>	22
Behavior Analysis and Recognition of Hidden Populations in Online Social Network Based on Big Data Method <i>Minglei Li, Guoyin Jiang, Wenping Liu, Junli Lei</i>	29
Investigating the Time-varying Effect of Search Index in Predicting Tourism Volume Using Dynamic Model Averaging <i>Yuan Chen, Yiping Ma, Zhongyi Hu, Tao Xiong, Jiang Wu, Tingshun Xiong</i>	38
Big Data and Business Analytics	
Research Progress of Tie-Generative Mechanism in Network Based on ERGM <i>Bo Jia, Jiahua Jin, Zhiyuan Ge, Xiangbin Yan</i>	46
Empirical Study of the Regional Diffusion of a Cloud Ecosystem Using Institutional Theory <i>Xin Zhang, Bingqing Ji, Gaoyong Li, Ge Zhang</i>	56
A Study on the Characteristics of Regional Social Network for Patent Technology Transfer of China’s “985 Project” universities <i>Xiaoli Li, Zihan Peng, Hao Zhou</i>	68
Leveraging Linguistic Signaling to Prompt Feedback in Open Innovation Communities <i>Suya Hu, Di Xu, Yan Li</i>	79
Research on Early Warning of Non-performing Loans of Small and Medium-sized Micro-enterprises Under the Background of COVID-19— Taking XX Branch of N Bank as an Example <i>Xinying Guo, Cui Guo, Junjie Zhou, Guoxin Wang, Huanlian Yang</i>	81
The Relationship Amongst Big Data Capability, Supply Chain Dynamic Capability, And Dynamic Innovation Capability <i>Guangqian Peng, Yang Zhang, Ping Gao , Junfeng Yin</i>	93

Bright and Dark Sides in E-Business

A Meta-analysis of Online Impulse Buying and The Moderating Effect of Economic Development Level <i>Yang Zhao, Bin Yang, Ning Wang</i>	95
The Content Influence Mechanism of the Behavior of Poverty Alleviation Crowdfunding Users: An Empirical Study of A Chinese Crowdfunding Platform <i>Xiaoping Liu, Shuang Chen, Yi Yang</i>	102
Research on Continuous Use of B2B Platform in Chinese Intelligent Engineering Companies Based on the Theory of Resource Complementary <i>Zhichao Zhang, Hongbo Jiang, Ting Shao</i>	110
Research on the Development of Regional E-commerce Based on the Perspective of Industrial Conjunction: A Case Study of Hubei Province <i>Xiangbing Chen, Haiying You</i>	122
An Empirical Research on Online Negative Comments Control Based on Consumer Satisfaction <i>Haiying You, Xiangbing Chen</i>	132
Digital Inclusive Society and ICTs Innovation for Senior Citizens in the New Normal COVID-19	
A Study on the Characteristics of the Publishers of Rumor-Refuting Information about COVID-19 <i>Xiaping Xin, Yan Tu, Tianmei Wang</i>	140
Public Negative Emotions Regulation During the COVID-19 Emergency: GIR's Content Features and Lingual Form Do Matter <i>Xie Yu, Xuefeng Li, Wei Zhang, Mei Zhang, Yanchun Zhu</i>	152
From Liking to Following: Research on the Antecedents of Middle-aged and Elderly Douyin Users' Fans <i>Jin Yan, Ningyue Cao, Changyu Wang</i>	163
An Analysis of the Characteristics and Evolution of China's Pension Policy Based on Text Mining <i>Liang Yang, Meng Zhao, Miyan Liao, Liangqiang Li, Xiaolin Li</i>	174
Digital Technologies, Digital Transformation, and Business Value	
Digital Technology-driven Business Model Innovations:A Bibliometric Analysis <i>Jiangping Wan, Mengyan Liu, ChunxiaoYang, Hexin Zhang</i>	186
Enterprise Technology Innovation and Customer Concentration under the Uncertainty of Market Environment <i>Wendi Wang, Wenbin Qu, Xiaofeng Wang</i>	194
Customer cooperative capability, perceived value and customer loyalty: a perspective of smart retailing <i>Jinfeng Wu, Zhiyuan Zhou, Delin Hou</i>	206
Digital Platform Expectation Disconfirmation and Customer Satisfaction in Sharing Economy: Moderations of Direct and Indirect Network Effects <i>Xiangcheng Yang, Yi jiang</i>	212
Construction and Verification of Night-time Economy Development Evaluation Index System <i>Xiaoying Li, Yan Tu</i>	224

What Drives Knowledge Payment in the Knowledge Payment Platform? An Empirical Study on the Purchase Behavior in the Emerging Market <i>Min Zhang, Jingyi Zhao, Weiwei Yan</i>	232
E-business strategy & online marketing	
Research on Effects of Paid Search Advertising Based on Customer loyalty Segmentation <i>Jiaxin Chen, Li Li, Weiwei Liao</i>	244
The Influence of Live Streamer Morality on Consumer Purchase Intentions from the Perspective of Identity <i>Cuiping Wang, Li Wang</i>	252
A Study on Purchasing Optimisation for Cross-border E-commerce Enterprises under Uncertainties <i>Jin Lin</i>	264
Research on the Influencing Factors of Knowledge Sharing Behavior among Makers in IT-enabled Open Innovation Communities <i>Lijun Li, Jian Tian</i>	271
The influence mechanism of product attributes on product demand in online reviews <i>Ying Chen, Li Li, Yulan Lin</i>	284
Does Online Product Presentation Matter: An Empirical Analysis of the Effects of Listing Photos on Short-term Rental Purchases <i>Jiang Wu, Xinru Zheng, Jingxuan Cai</i>	290
Research on the Influencing Factors of User's Adoption of Paid Information in Social Q&A Community <i>Bohong Li, Yanmei Zhang, Zhibin Cui</i>	294
Research on the Influence of Content Presentation Mode of Online Video Comments on Consumer Interaction Behavior <i>Hong Peng, Junfang Cao</i>	303
A New Analytic Framework for Identifying Tour Routes Service Innovation Using Knowledge Graph <i>Shuang Zhang, Minghao Li, Zhen Zhu</i>	315
Online Customer Experience at China's Double 11 Online Shopping Festival: An Empirical Model of Antecedents and Consequences <i>Wen-Lung Shiau, Mengru Zhou, Chang Liu</i>	317
Emerging digital technologies in the workplace	
Understanding Barriers to Service Robot Usage: A Qualitative Study in Hotels <i>Xiaojiang Zheng, Shixuan Fu, Ge Liu</i>	325
A Study on the Impact of Short-Video Product Placement Advertising on Viewers Ad Adoption Intention: A Perspective of ELM and Social Learning Theory <i>Yuzhen Liu</i>	333
An Empirical Study of the Impacts of Digital Marketing Contents on User Engagement in Social E-commerce Platform <i>Ronglin Hou, Fei Wang, Zhen Zhu, Bingchen Guo, Jing Zhao</i>	345

Why Do Employees Choose to Share Information Publicly? The Evidence from Enterprise Social Networks <i>Yu Wang, Daqing Zheng</i>	347
Engaging Technologies	
The Benefit Coordination Mechanism of Cloud Platforms for Rural Cultural Tourism Based on a Case Analysis of TikTok, AAuto, and Bilibili <i>Yueying Li, Guoyin Jiang, Wanqiang Yang</i>	354
How does Knowledge Association of Online Travel Reviews Affect Product Sales? Based on the Knowledge-based Theory and Social Network Theory <i>Chang Wang, Shengyou Liu, Ping Wang, Zhiping Hou, Wei Luo</i>	365
Online Marketing of Cultural Tourism: A Case Study of a 5A Ancient Town <i>Jiali Li, Bangchen Pan</i>	367
An Empirical Study of the Impacts of Digital Marketing Contents on User Engagement in Social E-commerce Platform <i>Ronglin Hou, Fei Wang, Zhen Zhu, Bingchen Guo, Jing Zhao</i>	377
Augmented Reality and Marketing: A Systematic Review <i>Jun Liu, Zhao Du</i>	388
ICTs in fighting against the COVID-19 pandemic	
Formation Mechanism of Regional Stigma —— Impact of Major Public Health Emergencies <i>Jiayi Xu, Keyi Zhang, Cuicui Liu, Sijia Guo</i>	390
Research on the Location of Nucleic Acid Detection Institutions Based on Coverage Model <i>Jiangping Wan, Xuejian Li, Chunxiao Yang, Hexin Zhang</i>	398
Research on the Competitiveness and Trade Potential of China-India Pharmaceutical Trade under the One Belt and One Road Initiative <i>Lin Ni, Shihui Shuai, Wei Li</i>	406
Analysis on Public Opinion Sentiment Evolution of COVID-19 Based on Weibo Data <i>Jiangping Wan, Xu Liu, Yihang Zuo, Jianfeng Luo</i>	416
What Motivates People to Share Online Rumors? Deconstructing the Ambiguity of Rumors from a Perspective of Digital Storytelling <i>Xiangcheng Yang, Yi Jiang, Wenqian Xu, Siqi Guo</i>	424
Information Management and Health Outcomes	
An Empirical Study on Inter-firm Knowledge Governance, Knowledge Transfer and Innovation Performance <i>Xiang Yang, Zhao Li</i>	436
Meta-analysis of COVID-19 Clinical Symptoms, Prevalence of Comorbidities and Influencing Factors of Severity and Mortality Cases <i>Yuting yang, Liyi Zhang</i>	448
Mediating Effect of ICT Impacts on ICT and Health Goals of MDGs: A Cross-national Panel Analysis	

<i>Jing Fan, Zishan Huang, Mingxing Shao, Hongquan Li, Yan Wan</i>	459
Understanding the Status of Reemployment of Elderly Talents in Digital Society: Evidence from Aged Job Search Website in China <i>Junjie Zhou, Guoxin Wang, Huiying Wei, Chengcheng Cai, Xiaomin Zhan, Xinling Lao, Wenhao Wang</i> ..	461
Verifying Online Health Rumors on Social Media: An Empirical Research Based on the Stimulus-Organism-Response Framework <i>Ying Liu, Yi Jiang, Shuang Zhang, Zhenkun Wei</i>	472
Information Systems and Operations Management	
Task Decomposition for Knowledge-intensive Crowdsourcing: Managing Dependency and Structural Complexity <i>Xiaojie SHI</i>	484
E-commerce Platform Resources, Sequences of Digital Strategic Actions and Competitive Advantage: An Empirical Study from Online Tourism Industry <i>Lewei Hu, Jing Zhao, Yi Jiang</i>	496
The Mechanism and strategy of Digital Logistics Park Cross Supply Chain Collaboration based on Evolutionary Game <i>Zhou Xingjian, Cai Lihua, Xu Gechen, Huang Yingpei</i>	498
Research on Location Optimization of Waste Transfer Center in Harbin <i>Jiangping Wan, Xuejian Li, Pengjie Wu, Zhaotao Chen, Zhiwei Du</i>	500
Research on Authors' Co-authorship Network in Supply Chain Finance in China Based on Social Network Analysis <i>Yang Zhao, Ning Wang</i>	508
Understanding Users' Response to Open Government Data Portal: an Exploratory Study <i>Ting Shao, Xiaofei Cui, Hongbo Jiang, Xiangling Huang</i>	518
Dual-channel Service Operations with the BOPS Option Considering Infection Risk Aversion Behavior <i>Xinyi Li, Yue Xu, Hui Li, Xi Wang</i>	529
Social Commerce and Immersive Interaction	
Institutional Entrepreneurship and Acquiring Legitimacy of Social Commerce Platform <i>JingYang, Ruoyu Sun, Zeguo Yang, Qianxi Zhang</i>	541
Interactive Effects of Order Assignment Strategy and Information Disclosure on Courier's Continuance Participation Intention Arising in the On-Demand Logistics <i>Quanwu Zhao, Jiamin Yuan, Yuqing liu</i>	549
The Influence of Customer Co-creation Experience Value on Brand Equity in the context of Social Media <i>Yunjun Jiang, Hong Zhang, Guiming Lian, Siqian Wang</i>	551
Exploring the Influence of Live Streaming in Social Commerce on Impulse Buying from a Affordance Perspective <i>Yan Mo, Qi Wang</i>	563

Visualizing Research Evolution in E-Commerce Field: Year 2001 to 2020 <i>Yang Bai, Hongxiu Li, Jianyao Chen</i>	575
Understanding User Contributions in Smoking cessation Online Health Communities <i>Chenglong Li, Hongxiu Li, Yuting Jiang, Reima Suomi</i>	583
Are Live Streamers Similar or Different? Grid Typology Model of Live Streamers in E-commerce <i>Dingtian Zhang, Yi Wang, Qinyi Zhou</i>	591
Trust in Digital Economy	
Traffic Monetization on Community E-commerce Platforms: Trust in Virtual Communities and Its Effect on Actual Purchase <i>Yi Li</i>	597
Influence of Social Network Integration on the Online Review Helpfulness <i>Jiaxuan Wu, Shengli Li</i>	605
The Influence of Perceived Trust and Time Moderator on the Purchase Intention of Consumers in the Context of E-Commerce Livestreaming <i>Kun Jia</i>	613
Solution Exemplars and Sales Performance of Crowdsourcing Solvers: the Moderating Role of Reputation and Competence <i>Quanwu Zhao, Zhiyuan Zhou, Xi Yang</i>	620
Research on Influencing Factors of User Satisfaction of Knowledge Paid Products -- from the Aspect of Heuristic-Systematic Model <i>Yufei Han, Chuyu Jiang, Yanping Zeng</i>	622

Full Research Paper (Extended Abstracts)**Enterprise Adoption of User Ideas in the Online User Innovation****Community: An Empirical Analysis based on IAM**Ning Zhang^{1*}, Wenfei Zhao¹, Zhiliang Pang^{1†}, Lifeng He², Weiguo Fan³¹Business School, Qingdao University, Qingdao, 266100, China²School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai, 200433, China³University of Iowa Tippie College of Business, University of Iowa, Iowa, United States

Even though a large amount of user-generated content in the innovation community reduces the cost of searching for user ideas by enterprises, the information overload caused by it increases the cost of cognition of effective ideas. Previous studies usually equate enterprises' adoption willingness with adoption behavior of user ideas, lacking the discussion on the adoption mechanism from willingness to behavior. By employing Information Adoption Model, the present study establishes a model of the influencing factors of user idea adoption, and investigates the mechanism of enterprises adoption of user ideas in two stages.

Information Adoption Model was proposed by American scholars Sussman and Siegal in 2003. Through the integration of the classic Elaboration Likelihood Model (ELM) and Technology Acceptance Model (TAM) dual process model, the information adoption process was explained from the central route and the peripheral route. The central route is mainly argument quality, and the peripheral route is mainly source credibility. The perceived usefulness of information recipients is an intermediate variable between information characteristics and information adoption, which has a positive impact on information adoption.

On the basis of IAM, we constructs a model of the influencing factors of user idea adoption. The model is divided into two stages. The first stage mainly analyzes the user ideas text characteristics that affect the designer perceived usefulness. The central route of the model includes information relevance, readability and sentiment features, and the peripheral route includes user authority degree and idea popularity. The second stage mainly studies the process of transforming the designer perceived usefulness into the enterprise's adoption of user ideas. This stage analyzes how the designer perceived usefulness affects the adoption of user ideas, and discusses the moderating effect include degree of feature attention, degree of referencing, and innovation forms on the designer perceived usefulness and user idea adoption.

Through the Python scripting language, we collected the online user ideas data of the MIUI system circle from January 15, 2020 to December 17, 2020. At the same time, a total of 50 update logs in this time period were manually tracked. We collected 432936 pieces of online user ideas. After eliminating the missing personal information (user's score) and no designers' response in the original data, a total of 30001 valid research data were obtained.

The first stage of the study found that the information quality and source credibility of user ideas affect the perceived usefulness of product designers. Relevant information can provide a more detailed description of product innovation, readable text can increase the interest of readers, and positive sentences can improve the willingness of designers to deal with information, so it is conducive to improve the perceived usefulness of designers. Influenced by the authority effect, the higher the user's score, the more credible the designer will consider the suggestions be, so it is easier for designers to accept them. Idea's popularity reflects the consistency of users' needs and implies the potential success rate of the innovative idea in the market, so it can positively

* Corresponding author. Email: zhang_ning1980@126.com(Ning Zhang), qdupangzhiliang@163.com(Zhiliang Pang)

affect the perceived usefulness of designers.

The above conclusions provide management significance for enterprise operation innovation community. First of all, designers in the enterprise community should guide users to publish their ideas in an appropriate form. Specifically, they should advocate concise and clear forms of expression and encourage positive expressions. Secondly, enterprises should provide more rewards for leading users in the community and establish incentive mechanism, so that leading users can provide more high-quality innovative ideas. For example, when purchasing a brand product, users can deduct their user scores accordingly.

The second stage further verified the view of information adoption model, that is, perceived usefulness is an intermediate variable between information characteristics and adoption behavior, which has a positive impact on information adoption. Product feature attention and reference degree can enhance the impact of designer perceived usefulness on user idea adoption. The more attention paid to the product features, the more users' needs can be met and the higher economic benefits can be brought to the enterprise. The higher the degree of reference, the higher the feasibility of the user ideas. Product designers can refer to related products to achieve improvement. The innovation form has a moderating effect on the designer's perceived usefulness and the user idea adoption. For the improvement suggestions of the original function, the designer's perceived usefulness has a greater impact on the user idea adoption. In other words, enterprises are more inclined to gradual, stable and low-cost innovation.

The above conclusions reveal the mechanism of enterprises from ideas adoption intention to behavior, and suggest that enterprises need to consider market demand, competitive products and R&D difficulty factors in the process of adopting user ideas. When a designer takes a useful user idea as an innovation proposal, he should first examine the market attention of the proposal. High attention may also mean that adopting the user idea for product innovation can obtain more support and profits. Secondly, when enterprises carry out product innovation, they should not be limited to their own products. They should understand the related products, especially the products of competitors, maintain the advantages of their own products, and prevent them from being surpassed by competitive products. Finally, enterprises should consider their own R&D ability in product innovation, and avoid blind investment, so as to avoid the negative effects of excessive investment.

This paper attempts to carry out innovative research from the following aspects: First, use IAM based on the enterprise perspective to expand the application scope of this model in network text analysis. The designer's response is used to measure perceived usefulness, and the text similarity between ideas and product update logs is used to measure the results of user idea adoption, which better solves the measurement problem of two variables. Second, distinguish between the designer's perceived usefulness and the enterprise's adoption behavior of user ideas. And further Exploring the factors that may affect this process to enrich the theoretical content of IAM from the designer's perspective.

The following presents the theoretical significance of this study: (1) the information adoption model is applied to the analysis of user ideas from the perspective of enterprises, expanding the application scope of the model. (2) This paper distinguishes the perceived usefulness of designers and the adoption behavior of enterprises to user ideas, discusses the mechanism of moderating variables from adoption intention to adoption behavior, and improves the research system of enterprise adoption mechanism. The practical significance of the present study is to help enterprises effectively adopt user ideas to support product innovation: (1) it provides management suggestions for enterprise innovation community operation, such as standardizing the form of ideas publishing and establishing leading user incentive mechanism. (2) It is suggested that enterprises should pay more attention to market demand, competitive products and R&D difficulty factors in user idea adoption.

Full Research Paper**Exploring the Collaborative Enabling and Value Innovation****Mechanism in Platform Enterprises:A Case Study***Xueyan Dong^{1*}, Zidie Chen¹, Xu Liu¹, Tienan Wang^{2*}*¹School of management, Northwestern Polytechnical University, Xi'an,710000 , China² School of management, Harbin Institute of Technology, Harbin, 150001, China

Abstract: The digital innovative evolution of platform enterprises has shown obvious advantages in promoting the development of value creation, and encouraging business transformation in current digital economy era. However, we still know little about the value innovation mechanism in platform enterprises. Thus, based on grounded theory, this paper explores the characteristics of collaborative enabling and extracts three attributes -openness, symbiosis and collaboration- in driving the internal development of platform enterprises, by coding the relevant information of Shi Wai Tao Yuan (SWTY). Through the case analysis on SWTY, this paper proposes three enabling functions in the field of culture, technology and data, and builds the multi-level path of value innovation including value co-creation, resource sharing and idea symbiosis. Finally, this paper constructs a theoretical framework on collaborative enabling and value innovation in platform enterprises based on nine core categories. Overall, this paper contributes to research on the formation process and action mechanism of collaborative enabling and value innovation in platform enterprises, also shed light on the exploration on platform enterprises, and business ecosystems evolution. In doing so, this paper also provides novel insights into the sustainable development of platform enterprises in practice.

Keywords: Platform enterprise, Collaborative empowerment, Value innovation

1. INTRODUCTION

The evolution of digital technology has subverted people's lifestyles, triggered changes in organizational contexts, made multiple subjects connect and participate in organizational activities, and promoted the platform-development mode of inter-organizational connections^[1]. Theoretically based on this reality and environment of Internet, platform enterprises will develop by paying more attention to embedding diversified products or services in virtual electronic trading spaces, gaining value by constructing network connections between multiple participants. In real life, the effective integration of digital technology and traditional business formats has triggered a trend of innovation in platform business, which has given birth to a large number of platform enterprises such as Alibaba, Didi, and Meituan. These platform enterprises have fierce competition, showing obvious growth in speed and scale, and are becoming a typical example for innovation in new economic mode of Internet. To give a summary, business model of platform enterprises, as a new business form, is attracting widespread attention from the commercial and academic community.

In academic, studies on platform economy, platform organization and platform model can be traced back to the 1990s, and then research on these topics becomes more and more popular^[2]. Wang and Chen^[3] found that the characteristics of "de-intermediation, decentralization, and de-boundaryzation" of the platform business model accurately hit the uncertain features of new business era, which help companies find ways to develop, clarify core advantages and reshaping business form according to real situation. It not only creates symbiotic and synergistic business ecosystem while profiting, but also provides support for multiple interaction, balancing of demand and supply as well as value co-creation of platform users. No matter building platform organization

* Corresponding author. Email: xueyandong@nwpu.edu.cn, wtn@hit.edu.cn

inside the system or joining specific platform through enterprise alliance, the core target of a company always lies in giving full play to the complementary advantages of bilateral or multilateral platforms, stimulating vitality, and flexible efficiency.

Nowadays, platform economy often involves multiple subjects with different value appeals. As a result of that, when considering design of platform organization and innovation mechanism, it is necessary to solve integrational problem of different value appeals and encourage all participants to help realize overall creation and promotion of platform. Based on this condition, scholars introduced the research perspective of enabling, trying to answer how platform enables different entities as to further improves value creation and innovation of business model^[4]. Some studies show that enabling is a significant concept in macro and micro management of organization, which aims to provide capacity and capability for related subjects to realize their goals^[5]. And from the perspective of enabling mechanism on platform enterprises, different enabling models, such as structural, leadership, psychological, resource and cultural enabling, have become the key driving factors to enhance value of a platform^[6]. However, what is the specific connotation of enabling for platform enterprises? What characteristics does it exhibit? How does enabling promote value creation and realization of platform enterprises? These questions still need further discuss and explore. Particularly, from the perspective of system integration, there is still lack of explanations on how and why platform enterprises can promote value creation through enabling mechanisms.

Therefore, this study selects case of Ningxia Shi Wai Tao Yuan Commercial Management Company (here it is shortened as SWTY) to explore, and conducts deep analysis through grounded theory, aims to answer the following research questions: (1) What type of enabling mechanism exists in platform enterprises and how do they make effect? (2) How do platform enterprises realize the goal of value creation through enabling?

2. LITERATURE REVIEW

2.1. Platform business model

The word “platform” can be traced back to theory of bilateral market of Rochet and Tirol^[7]. Bilateral market refers to operators who provide services to platform and users who implement transactions through platform. From perspective of demand, products or services provided by users and platforms can be interconnected and interdependent. And from the perspective of supply, the existence of a platform is necessary, it should fully consider needs of bilateral users so as to provide services and create value. Besides, Zhang^[8] stated that platform is a space for transaction, an interface to create and exchange value between different users; Wang^[3] defined it as a multi-cooperation business model based on different target of enterprises. Therefore, the role of platform is to actively guide and enhance multi-customer transactions and finally maximize benefits of them^[9]. Other researches on platform can start from perspective of product development platforms to exploration of bilateral market and strategic innovation platform. All in all, researches focus on product development platforms mainly devoted to explore design and implementation of products, structural connotation of the platform; researches focus on bilateral market platforms concentrate more on related strategy or rules about competition and governance of platforms; As researches on strategic innovation mainly cares about formation of business model, ecosystem and influence mechanism of platform^[10].

Platform ecosystem has already realized an efficient and dynamic value creation network under guidance of technologies like Internet and big data, which motivate value symbiosis and mutual benefit of different entities. Among them, construction of business model based on platform has been the focus of researches. Previous studies have shown that the business model is an emerging model based on network effects and multilateral construction to realize value innovation^[11]. They have found that, for one thing, companies need pioneering products to achieve accumulation of initial users to develop; for another, they can make

achievements by obtaining data resources, extending industrial chain, aggregating all parties and build an effective platform for sustainable operations. Also, Xiao ^[12] found that platform business model has different method on value creation compared to traditional ones. During process of upgrade and evolution, it connects different entities like customers, products, stakeholders, government, different core relations such as supply chain, value chain, innovation chain, service chain to form a multi-dimensional ecological environmental system.

2.2. Innovation mechanism of platform

With the rapid growth of Internet technology, platform business model is getting more and more popular, and the innovation of it can be summarized as transformation and subversion of existing thinking patterns of value creation. Here we can also consider it as an innovation of elements and links of business model. So far, while playing its basic role in the market, platform business model is also trying to change, upgrade and refactor.

By summarizing results of some studies, we found that innovation of platform business model mainly focuses on value model, competition model, operation model and profit model. More specifically, in terms of value model, platform business aims to realize value co-creation by re-discovering and re-thinking value of users, integrating multi-agent resources, empowering platform participants, reducing costs and eliminating multilateral information asymmetry^[13]. For competition model, platforms will choose single or multi-homing mode according to environmental changes and use target pricing strategies and differentiate service strategies to qualify its system^[14]; For operation model, platforms aim to achieve the optimal allocation of resources and establish a more effective operation mechanism. To reach this goal, they will use a hierarchical platform to form high- and low-level collecting systems and distribute products and services^[4]. Besides, more and more platforms are using online and offline integrated to continuously expand growing space and stimulate innovation ability^[15]. For profit model, some platforms either gradually cancel commission or use differentiated method to get it, they may also use customized mode and advertisements to win more money^[13]. After analyzing these four models, we know that studies before have already discussed innovation mechanism of various business model of platforms, however, we still don't know how they interact and collaborate with each other, which in turn make some difficulty to better understand overall mechanism of elements in platforms.

2.3. Empowerment and enabling of platform business model

In recent years, more and more studies are concentrating on how to explain innovation mechanism of platform business from the perspective of enabling, which seems a totally new one in the field. To a certain degree, this new perspective take place of traditional one, which is called empowerment. “Enabling” and “empowerment” have different meanings in this content. “Empowerment” is a traditional one to research on platform enterprises with a long history of study, it mainly concentrates on giving internal driving force to participants of platforms. In this circumstance, participants have more decision-making power and autonomy, and the platforms, will have more time and energy to determine how to lead their participants. Then internet business and platform enterprises develop and more attention is paid to mechanism of “enabling”, which is an improved pattern of “empowerment”. Different from “empowerment”, “enabling” stress more on how business model contributes to improvement of ability to create more value. Luo^[2] has shown double effects of enabling, it is not only an internal driving force to platform enterprises, but also a main mechanism to push platforms for upgrading and value creation. Luo^[2] believe that platform enabling refers to the ability of a platform to share and integrate information and technical resources at different stages of development to help enterprises or stakeholders win better place in competition. Besides, some studies also divide process of enabling as stage, object, subject to do more specific and detailed research, which will explain working patterns of empowerment of platforms more vividly.

Zhou et al. ^[6] analyzed that enabling of entrepreneurial platforms to promote performance from five

dimensions: structural enabling, leadership enabling, psychological enabling, resource enabling and cultural enabling. Additionally, some scholars have also conducted further analysis on empowerment/enabling and its significance stakeholders. Related studies have found that platform enterprises can make use of automatic data from platforms to provide basis for decision making and enabling, this will in turn improve regulatory capability and level of government^[11]. Moreover, they can use technology enabling by providing technical access to other companies. Also, they play the role as value enabling to other entities in evolution of business ecology. Besides, platform leaders in a favorable position in platform ecosystem can also empower various enterprises through creation and coordination mechanism^[14]. We can tell from the above that different patterns of enabling are vital for innovation mechanism of value creation of platform enterprises.

However, we can conclude from studies above: firstly, researches on enabling of platform enterprises are still relatively limited, they still concentrate more on simple description on how different empowerment works and lack deeper research on effect, interaction and improving mechanism among different kinds of enabling, so there's still a failure to form a closed loop of dynamic effect and synergistic influence of platform enabling. Secondly, some researches about ecological and value enabling merely explain them as a certain dimension, which lack detailed analysis to overall dynamic and collaborative enabling. In this situation, it is not enough to efficiently explain and clarify inner relation of platform enabling and dynamic mechanism of stimulating enabling of platform. Therefore, this paper adopts a single case study method to explore the relationship between enabling and value co-creation, and to clarify and explain how platform collaborative enabling affect the value innovation.

3. RESEARCH METHOD

3.1. Case sample selection

Single case study is often used to explain phenomenon which is not fully understood^[16] and this method, with vivid description, is helpful for us to get more significant situation about the target research object^[17]. We use single case study to conduct the whole work for several reasons. First, the purpose of this study is to explore how do platform enterprises enable and make value innovation, this is under the question scope of “how”. Second, given the limited theory evidence, our main goal is to further develop theory of enabling and value innovation by recognizing key attributes and actions of platform enterprises, this is particularly relevant for exploratory question. Furthermore, nowadays, researches on how do platform enterprises try enabling and value innovation are still in exploratory stage, it is still necessary to use scenario and vivid description to identify key factors during process of enabling and value innovation. Therefore, single case study is well qualified for this exploratory research.

According to Eisenhardt^[18], standard of selecting a single case is, on the one hand, that the source of the case should be rare and unique to conform the principle of theoretical sampling, and on the other hand, that the case should have many characteristics as enlightening one. Based on this reality, we choose Ningxia Shi Wai TaoYuan Business Management Company (abbreviate SWTY) to study. It is a comprehensive community business service platform, which takes living community as service radius, takes AFA (all for all) as core business to serve all citizens. Gradually, it will cover a wide range of entities and users of platforms such as community, users, sellers, governments and so on. Therefore, it is more appropriate to choose SWTY as a case to understand the question of enabling and value innovation in platform enterprises.

3.2. Introduction of case

SWTY is obviously a typical one in developing period of digital platforms. It is a single service enterprise, and gradually becomes an innovative marketing enterprise. During this process, it keeps improving its own business and finally transforms into an platform enterprise focusing on community service for all. Under the

transformation of developing and business model, SWTY finally broke traditional system of single service and formed service system for all citizens based on AFA service system of all citizens like smart service station in community, home service station, which shows construction of integrated smart service. Based on growing history of this company, Table 1 lists the whole developing history of key incidents and turning point of business model of SWTY.

Table 1. Transformation history of business mode of Shi Wai Tao Yuan

time	key incidents (set up)	key points of business transformation
2002	Ningxia Zhongchen safety and technology company	Single service enterprise concentrating on business
2011	Ningxia Shi Wai Tao Yuan Agricultural Development Company	Innovative marketing enterprise focusing on products
2015	Ning Xia Shi Wai Tao Yuan business management company (SWTY)	Community service platform focusing on data of community
2015	Ning Xia Yi Dao Tong advertising company	Media service platform focusing on data from media
2017	Ning Xia health management company	Community service platform based on data of health
2018	Western (Ning Xia) smart industry developing company	Industrial service platform based on smart data

Data resource: internal data of the enterprise

3.3. Data sources

In order to ensure the reliability and validity of the study, we use triangle verification method composed of multi-source data^[18]. And our data mostly comes from internal materials and internal interviews. Meanwhile, research record, official interviews and informal interviews, information from websites, news reports, WeChat platforms as well as chairman's report on government's reform and improvement in construction of smart city are also included. To have a deeper understanding, our team have a three-year long period of interview and data collection to SWTY. The first interview was in March, 2019, and after that we did more and more deeper interviews. People we interviewed covered from founders, manager of operation to monitor of marketing. Our interview and data collection mainly focus on three things, the first one is evolution and innovation of business model in community service; The second one is status and experience of community service content, service mode and profit mode, and the third one is construction process of value chain system in resource integration of SWTY.

3.4. Data analysis steps

This paper mainly adopts grounded theory when studying case of SWTY, using classical steps of case study for reference and analyzing data through three steps. Firstly, referring to process method of data research, the study sorts our key events of SWTY which is helpful to recognize and understand turning point of the company to know the whole process of its development better. Secondly, this paper focuses on key factors and mechanism of platform enterprises on enabling and value creation. Then we decode and analyze the collected data according to this logic. We divide our 6 research members into 3 groups and give them the standard training of decoding. After that, they code and analyze the collected case data sentence by sentence, and then integrate and revise previous round of coding by iteration. After three rounds of independent analysis, they form a consensus through comprehensive discussion, and determine the initial coding pool. Thirdly, in order to understand the relationship among the coded constructs, and recognize the changes of the constructs themselves in the research context, the whole work is carried out in the following three periodic steps: refine the decoding, emerging thinking, and refine theoretical mechanism. Finally, we end the analysis when our theoretical model can fully and appropriately used as a theoretical interpretation model^[18].

4. CASE ANALYSIS

4.1. Open decoding

Open decoding is a process to have an overall arrangement for data we collected and give simplified and abstract analysis to common phenomenon, including labeling, conceptualization and categorization, which means, decomposing existing materials, refining information and finally shortening categorization entries with the method of comparative analysis and summary. This paper uses coding methods sentence by sentence to deal with relative information according to its content, and then finds sentences concerning empowerment of platform organization and value creation. Firstly, we give label (a) to them and then, we form 36 conceptualized sentences (A) according to these labels. By continuous comparison and refinement, these conceptions gradually grow into 23 categorical expression (AA). After method of open decoding, we finally get 36 labels and 23 categories to describe empowerment of platform organizations and value creation.

4.2. Axial coding

Axial coding aims to build internal relations based on expressions from period of open decoding and gradually digs and understands materials deeper and better. We follow the logic of canonical coding paradigm of Glaser^[19], which is “condition-action or interaction strategy-outcome”. Condition is the reason and out come is what it forms, these two factors use “action and interaction strategy” to realize complete logic train, which is “axis”. This paper firstly categorizes 23 initial categories in the open decoding and forms 19 sub-categories. Then we integrate these 19 categories with all information as axis. For example, “Platform has accumulated huge number of users’ data”, “deeply dig of data”, “finding potential demand of users” are three sub-categories, and their internal logic is as follows: The platform has accumulated a large amount of data of users, so the next step is to have deeper analysis on these data and acquire information hiding in them, which is potential demand of users. These three sub-categories can be induced and concluded as “data enabling”. After process of axial decoding, we finally get 9 main categories (see table 2).

Table 2. Results of axial coding

Main categories	Sub-categories		
	Condition	Action or interaction strategy-outcome	Results
Openness attribute	the existence of market environment, multilateral competition and cooperation relationship, and network externality	make the best of the situation, seize market opportunities, break barriers with the advantage of the platform, and promote cooperation	from "Single-win" to "Multi-win" to "Win-win"
Symbiosis attributes	from "Single-win" to "Multi-win" to "Win-win"	cross-boundary cooperation stimulates network effect, clarifies market positioning and profit model, and formulates platform ecological rules	forming a ecosphere including multi-subject and multi-chain
Collaboration attributes	forming a ecosphere including multi-subject and multi-chain	integrate resources, break the boundaries of information dissemination and communication, and extend community functions	achieve industrial linkage and mutual benefit
Culture enabling	achieve industrial linkage and mutual benefit	honesty is the foundation; strengthen cultural soft power	focus on research and development input on brand and technology
Technology enabling	focus on research and development input on brand and technology	technology innovation, introduction and application	accumulated massive user data
Data enabling	accumulated massive user data	dig deep into the data	discover the potential needs of users

Main categories	Condition	Sub-categories	
		Action or interaction strategy-outcome	Results
Idea sybiosis	discover the potential needs of users	provide users with extended services	constantly innovating products and services, extending to the whole industry chain, from online to offline
Resource sharing	constantly innovating products and services, extending to the whole industry chain, from online to offline	provide one-stop service	constantly upgrading and evolving business models
Value co-creation	constantly upgrading and evolving business models	enterprises of all sizes interact and cooperate to support each other and empower each other across borders, to improve people's lives	ecosphere eable all subjects to create value

4.3. Selective decoding

The aim of selective decoding is to tell a good story, which means by describing phenomenon, we can stress key categories after clarifying relations between main categories. To increase theory saturation, we uses method of theoretical reference and comparison between literatures, we establish a solid and sufficient model by comparing original data and the on-going built theory model, The story line is as follows.

Due to open nature which is also called market-oriented environment, multilateral competition and cooperation relationship and network externality, platform organizations should make full use of the situation and seize opportunities in market, break barriers with help of advantages of platform to promote cooperation. This can be considered as a way from “single-win” to “multi-win” and then to “win-win”. Under this mode of “symbiosis”, it is necessary to have accurate market position and profiting mode, make ecological rules of platform to further form ecosphere including multi-entities and multi-chain. By doing this, platforms can successfully break through boundary of information transformation and communication, and then extending function of community to realize industrial linkage and mutual benefit. Under function of station of platforms, we can use cultural enabling, technical enabling and data enabling to dig potential demand of users and make more innovation on products and services. From online to offline, our production chain will extend gradually and provide high quality of one stop service. During whole process of platform enabling, business model of enterprises will be constantly updated and upgraded, each entity in this ecological circle will cooperate with each other, help each other, and gradually realize enabling cross boundary, finally, resource sharing and enabling for value creation is achieved on the whole.

5. CONCLUSION AND DISCUSSION

5.1. Conclusion

Through normative case study procedure, this paper explores the logic relationship between collaborative enabling and value innovation. We construct a theoretical model to reveal the mechanism about how platform enterprises achieve value innovation through their collaborative enabling (See Figure 1). Also we conceptualize three attributes that drive the internal development of platform enterprises, they are openness, symbiosis and collaboration; we summarizes three types of enabling in the field of culture, technology and data; and conclude three paths in value innovation,they are value co-creation, resource sharing and idea symbiosis. We explain the constructed model as follows.

Nowadays, competitions in internet industry are always fierce, industrial chain of platform including platform, customers, sellers has already formed initially, which in turn brings out phenomenon of homogeneous competition^[20]. Based on this marketing reality and in order to adapt to rapid development of digital technology,

avoid homogeneous competition and reach demand of sustainable development of company, platform enterprise, as a new business mode, should transform its mode of value creation into interaction between enterprises and value co-creation of customer and company^[21]. By efficiently integrating internal and external resources, enterprises can make fulfilling multi-demand of users and providing characteristic service a reality. According to our study, fundamental innovation development of SWTY is “openness, symbiosis, collaboration”, and these three attributes are also the driven forces to reach the goal of platform’ users. Therefore, this paper considers “openness, symbiosis, collaboration” as three attributes of enabling, which are the detailed explanation for the enabling connotation in platform enterprises.

When it refers to enabling, the use and development of digital technology has made platform enterprises have the stronger ability of identification, judgement and control to its environment and objective condition. However, changes in ways of production and living from information age, the mode of platforms will turn from “empowerment” to “enabling”, and better show their value creation. Therefore, according to the analysis results we stressed above, enabling of platforms now can be divided into culture, technology and data three aspects.

Cultural heritage, as endless vitality of platform enterprises, links future with history. However, since rapid growth of technology on mobile internet and blockchain, cultural enabling is growing toward digitalization, informatization and intellectualization, and gradually makes full use of innovation mechanism of platforms^[22]. Core ideas of cultural enabling is to put culture into innovation of business model of platforms. By enhancing soft power of company culture and concentrating on researches of brand culture and technology, this kind of enabling will bring some value for platform enterprises’ innovation to a certain degree as well as deeper cultural connotation, which further realizes innovation in content and mode^[4]. For example, SWTY specially developed a kind of rice that can listen to music, giving specialized “rice culture” to Ningxia rice and broaden the market relying on this. In the case company itself, cultural atmosphere for more concentration on researches on brand and technology is formed and plays as a foundation for technical enabling.

Technology can enable the innovation of business model of platforms^[1] and can push platform enterprises to better accumulate resources and realize value co-creation. Relying on technical innovation, technical import and technical application, enterprises will acquire good economic benefit and lead themselves to update. This mode will bring business model positive influence and enhance competitive power of platform enterprises as well as strengthen effective integration among platform ecosphere^[23]. According to our case analysis, we can find that in model of AFA community platform, technical enabling has played a very significant role in accumulating large sum of users’ data and further improving the process of business innovation for SWTY. For example, it introduced smart medical examination and health management technology to help basic construction on internet medical popularization; it uses “community housekeeper” to acquire data of high frequency and high viscosity, and accumulate about 300 thousand precise users in a very short time.

However, it is still not enough for platform enterprises to have sustained achievements by just having large sum of users’ data. Platform enterprises should also try deeper dig to digital trajectory of existing users. Only after that they’ll find potential demand of users and try precision marketing, and this is main significance and value of data enabling. Also, data enabling can tell economic effects from big data to traditional business analysis^[3].

Above all, under this three kinds of enabling, platform enterprises can have better interaction with users and maximize benefits of them to further improve innovation of business model, and gradually they work as keys effects to value innovation mechanism. Platform enterprise is a key factor in platform industrial chain, it leads and forms multi-entity ecosphere including users, developers, suppliers and so on. However, the multi-entity as the participants object of platform enabling, will always realize “common ideas, resource sharing, value so-creation” relying on culture, technology and data. Also, it is a way to express the importance of

enabling behind platform mode to development of enterprises.

To make a summary, common ideas, resource sharing and value co-creation has combined action to objects of platform enabling. They contribute to well-being of ecological platforms on innovation of products and service, extending industrial train, online and offline extension, evolution of business model and value creation of entities.

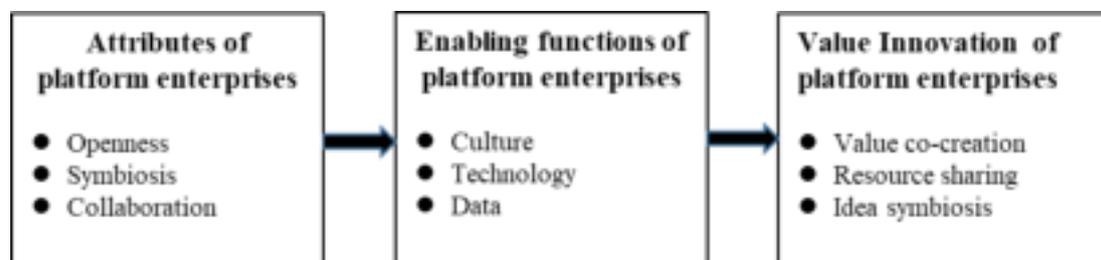


Figure 1. Theoretical model about enabling and value innovation in platform enterprises

5.2. Theoretical contribution

This paper chose SWTY as a case and conducts further exploration on enabling and mechanism of value creation for platform enterprises, providing research ideas of value co-creation for enterprises on carrier, method and logic. Our study introduced the concept of enabling, using culture, technology and data as driving factors to clarify the path mechanism of value innovation, given theoretical brace for practice of platform enterprises, which not only enrich and polish research system for platform enterprises in innovation field, but also contribute to systematicness and depth of related researches. Theoretical contribution can be concluded into the following three aspects.

Firstly, this paper fully clarifies concept of enabling of platform enterprises. Based on objects of enabling, this paper extracts three attributes as “openness, symbiosis, collaboration”, specifically explains the connotation of enabling for platform enterprises. Also, since previous research has fuzzy statement on what is platform enabling, this paper helps make up for this shortcoming^[23]. Platform economy is a typical representative of innovation of business mode, but construction of platform is not the destination of enabling and value innovation. Aattributes of platforms is the premise of development of platforms^[21], this research has distinguished characteristic of SWTY. In detail, openness reflects in constructing thoughts of AFA community business platform. From the perspective of vertical development and transformation of platform, AFA community business platform is transformed from a traditional community fire service company to a smart community service platform through development of community agricultural products company. The open nature of platform enables SWTY to integrate needs of multiple participants, and with the help of platform thinking, expands externality of platform network. Symbiotic attribute is embodied in business ecosystem of AFA community. In this system, through embedding of internet technology and interaction of new media information, traditional subjects such as residents, communities, neighborhood committees, streets, government, property companies and so on are incorporated into a complete system, which makes symbiotic cooperation of multiple subjects possible. Collaborative attribute is reflected in the fact that AFA community business platform, as a professional community service function platform, has integrated platform resources of enterprises, suppliers, communities, property, government, finance, colleges and universities, providing smart community platform for community neighborhood committees, digital property management for community property management, and services for community businesses to meet their own industry management and service needs. Meanwhile, the platform also integrates payment platform of public utilities (e.g., water, electricity, gas, telecommunications) and bank payment gateway to meet needs of different participants. Compared with

previous research on platform, this paper can explain essential characteristics of platform enterprises in more detail, which provides a new idea for further exploring commonness and heterogeneity of community business platform.

Secondly, this paper clarifies the process of collaborative enabling, and also helps to expand explanation on different platform enterprises' enabling differences. Difference of operation effect of platform enterprises is directly reflected in enabling of different platforms, that is, different enabling means will produce different collaborative results. Enabling of AFA community business platform is reflected in the whole process of its operation. In this process, platform meets needs of multi participants in its ecosystem from three aspects (technology, data and culture). First of all, technology is foundation of business development of platform enterprises. The case enterprises use new technologies such as social media, mobile Internet, Internet of things and big data to promote communication of participants and the integration of internal resources, provide extreme and closed-loop customer experience for platform participants, and promote interaction between platform and participants. At the same time, this study also found that in order to break through traditional business model, platform enterprises need to first use technology as support to try to combine traditional services and mobile Internet point-to-point, and realize connection of internet platforms. Different from other scholars' view that sharing economy platform is a core-periphery value creation system, this paper holds that community business platform is a hierarchical complex system composed of multiple participants, which is not only an information platform^[23], but also a platform for value creation and symbiosis. Among them, interaction of information among participants is reflected in the process of collaboration, data and cultural enabling. The creation and symbiosis of value is reflected in the fact that each subject does not use value consumedly, but pursues creation of new value in the way of mutual cooperation. Therefore, this paper reveals and discusses empowerment and means of community business platform in detail, which is an in-depth excavation of business logic behind platform economy, which, to a certain extent, drives development of theory of empowerment for platform under the background of sharing economy.

Thirdly, this paper establishes community platform innovation mechanism model, which has a certain contribution to research on value creation process of platform enterprises and value innovation mechanism of platform economy under background of digital economy. Nowadays, community business platform enterprises are developing rapidly, but discussion on value innovation mechanism of platform enterprises is not enough. For example, interaction, process in value innovation of platform, and the way of user's co-creation and win-win mode still needs to be further discussed. Value innovation mechanism of community business platform no longer presents a one-dimensional static mode dominated by platform enterprises, but a multi-dimensional dynamic mode with multi participants^[20]. Based on theoretical logic deduction of "platform attribute-enabling-value innovation", this paper constructs a theoretical model of collaborative enabling and value innovation of platform enterprises, and explores value innovation in dimensions of value co-creation, resource sharing and common idea, which depicts motivation and process of value creation of platform enterprises and describes their leading law of value creation. Also, research process and results of value creation of platform enterprises deepen and enrich current research on digital transformation of enterprises driven by new technologies^[5].

5.3. Practical enlightenment

First of all, platform enterprises should pay attention to construct efficient enabling mechanism, and this mechanism needs benefit sharing beyond normal condition and stresses the importance of integration of cultural ideas. Foundation of platform community is to fulfill different demand of different participants, which is, accurately acquire, scientifically recognize and match demand of sharing. Therefore, it is vital for platform enterprises to build enabling mechanism and have "win-win" culture. Also, they can break boundary of

informative transformation and communication, extend community function of enterprises by constructing community service platform. Based on this, value sharing is getting smaller and cultural thoughts are gradually put into ecosystem of platforms. By using online and offline path, it provided convenient service for different participants in platform enterprises.

Then platform enterprises have to fully concentrate on creation, collection, sharing and application of data, innovatively apply information and optimize business model and value creation according to digital information. Sharing of digital information is the driving force of participants to participate in community platforms, so these platforms should fully analyze and build ways of information sharing and applying. Therefore, value of information can be applied innovatively. “community housekeeper” is an app of SWTY to acquire qualified data of users, so this app gives a channel for information sharing by platform enterprises. As a result of that, platform enterprises can have a deeper research to data of users and give accurate information of users in a very short time. This helps quickly understand interests of users, and enterprises can design innovation mode accordingly to raise innovative ability under the condition of information sharing.

Finally, under the background of smart community and smart city, based on wide application field of business mode of community platform, enterprises can stick to characteristic of “win-win” and collaborate with different entities including government, company, social organizations. Also, they have to concentrate more on mechanism of value innovation to have more vitality of sustainable development. Shi Wai Tao Yuan has deep relation with community, advertisements and government, they form “value alliance” and try to make value creation among ecological platforms. Besides, in period of COVID-19, communities form a tiny system to help each other to a certain degree, and this also brings some suggestions for smart business platform of smart city in the coming future.

5.4. Limitation and future direction

Though this study has given some valuable suggestions for development of theory and practice, there still exists some limitations, which needs to be further discussed in the future. Firstly, this paper didn’t conduct analysis by categorizing multi-participants according to their different characteristics and needs when exploring enabling and value creation of SWTY AFA community business. Community platform and other entities (governments, communities, citizens) have form symbiotic business ecosystem, and characteristics or demands of these participants have great impact on enabling and value creation. Future research can consider multi participants’ characters, systematically understand the process and outcomes of enabling and mechanism of value creation. Secondly, the case lacks consideration on community platform business mode. Due to lack of ways to collect information and limit of reference sample, this study conduct exploration by using single case. However, in real practice, there are many kinds of community platforms and different kinds of operation modes, current analysis may be not fully appropriate for the discussion on enabling and innovation mechanism in similar platforms, thus, comparative studies of multi-cases can help enhance the generalability of the proposed theoretical model in the future.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71902158 and 71972061; and Fundamental Research Expenses for the Central Universities under Grant 3102019XJS04.

REFERENCES

- [1] Wang Wei-nan, Yan Zichun, Mei Liang, Chen Jin. The Evolution of Platform Enterprises Based on Data Resources: Lessons from Variflight Company [J]. *Economic Management Journal*, 2020,42(6):96-115 (in Chinese)
- [2] Luo Zhongwei, Li Xianjun, Song Xiang, Li Yaguang. Evolution of Enterprise Organization Structure Based on the

- Hypothesis of "Empower" to "Enable": Based on the Case Study of Handu Group's Practice [J]. *China Industrial Economics*,2017,(9):174-192 (in Chinese)
- [3] Wang Ying, Chen Weiru. [J]. *Tsinghua Business Review*,2016,(4):74-80 (in Chinese).
- [4] Wang Fengbin, Wang Xiaopeng, Zhang Chi. Ultra-modular Architecture in Platform and Customized Support for Intrapreneurship-An Embedded Case Study of Haier's Transformation to Platform Organization. [J]. *Journal of Management Word*, 2019,35,(2):121-150+199-200. (in Chinese).
- [5] Cai Chunhua, Lliu Wei, Zhang Chi. Effect of Business Model Contextualization on Value Creation: A Case Study of Rainbow's Digital Transformation from 2007 to 2018[J]. *Nankai Business Review*,2020,23,(3):98-108(in Chinese).
- [6] Zhou Wenhui, Li Bing, Zhou Yifang, Chen Lingzi. The Impact of Entrepreneurial Platform's Empowerment on Entrepreneurial Performance:A Case Study Based on "Haier + Thunderobot"[J]. *Management Review*,2018,30,(12):276-284 (in Chinese).
- [7] Jean C. R., and T. Jean. Defining Two-sided Markets[D]. Working Paper, IDEI University of Toulouse,2004.
- [8] Zhang Xiao-ning, A Literature Review on Platform Strategy[J]. *Economic Management*.2014,36,(3):190-199(in Chinese).
- [9] Zhang Yunsheng, Wang Xuemei. The Pricing Model of the Internet Value-added Services Platform[J]. *Beijing:Forum on Science and Technology in China*,2010(10):94-99 (in Chinese).
- [10] Wang Jiexiang, CAI Ning. Schools, Tendencies and Theory Frameworks of Platform Research: Interpretation Based on Bibliometric and Content Analysis Method[J]. *Hangzhou: Journal of Business Economics*, 2018,(3):20-35(in Chinese).
- [11] Zhou Xiang, Luo Shunjun, Su Yu Feng. From Product to Platform: A Multiple Case Study of Platform Strategy. [J]. *Quarterly Journal of Management*, 2016,1,(3):102-127 (in Chinese).
- [12] Xiao Hongjun. Shared Value, Business Ecosphere and the Transformation of Competitive Paradigms for Business[J]. *Chongqing: Reform*,2015,(7):129-141(in Chinese).
- [13] Yi Famin, Zhu Jie. The Business Model Innovation of ICT-enabled Poverty Alleviation Platform [J]. *Management Review*, 2019,31,(7):123-132 (in Chinese).
- [14] Xing Xiaoqiang, Zhou Pinglu, Tang Xinhui. Digital Technology, BOP Business Model Innovation and Inclusive Market Construction. [J]. *Journal of Management Word*, 2019,35,(12):116-136 (in Chinese).
- [15] Hao Jin-lei, YIN Meng. Sharing Economy: Empowerment, Value Co-creation and Business Model Innovation: A Case Study of ZhuBajie[J].*Harbin: Commercial Research*,2018,(5):31-40(in Chinese).
- [16] Marshall, C., and G. B. Rossman. *Designing Qualitative Research*[M]. Los Angeles:Sage Publications,2014.
- [17] Eisenhardt, K. M. Building Theories from Case Study Research. *Academy of Management Review* [J].1989,14,(4): 532-550.
- [18] Eisenhardt, K. M. and E. M. Graebner. Theory Building From Cases: Opportunities And Challenges[J]. *Academy of Management Journal*, 2007,50(1),pp.25-32.
- [19] Glaser, B. G., A. L. Strauss, and E. Strutzel. Discovery of Grounded Theory[J]. *Nursing Research*,1968,17,(4):364.
- [20] Wang Shuilian, Li Zhigang, Du Yingying. Research on Value Creation Process Model of Shared Economic Platform——A Case Study of Didi, Airbnb and Douyin[J]. *Management Review*,2019,31,(7):45-55(in Chinese).
- [21] Bai Jing-kun, Zhang Zhen-zhen, Xue Liuyang. Organizational Mechanism of Enterprise Innovation Based on Platform in The Context of Internet: A Case Study of Handu Group[J]. *China Soft Science Magazine*,2019,(2):181-192(in Chinese).
- [22] Qi Yong, Zhang Ming, Ding Gang. Strategy on Resource-Sharing among Collaborative Innovation Entities Based on Game Theory[J]. *China Soft Science Magazine*, 2013,(1):149-154(in Chinese).
- [23] Yang Xuecheng, Tu Ke. Research on the Dynamic Value Co-creation in the Sharing Economic Background: A Case Study of the Travel Platform[J]. *Management Review*, 2016,28,(12):258-268(in Chinese)

Short Research Paper**An AdaBoost-DT Model for Credit Scoring***Jiali Zhao¹, Zengyuan Wu¹, Bei Wu^{2*}*¹School of Economics & Management, China Jiliang University, China²School of Management and E-business, Zhejiang Gongshang University, China

Abstract: Credit scoring for loan applicants is an essential measure to reduce the risk of personal credit loan. Due to low percentage of non-performing loans, credit scoring is typically considered as an imbalanced classification problem. It is difficult to address this kind of problem using a single classifier. In order to settle the problem of imbalanced samples in credit scoring system, an ensemble learning classification model named AdaBoost-DT is proposed. In this model, we employ adaptive boosting (AdaBoost) to cascade multiple decision trees (DT). The weights of the base classifier can be adjusted automatically by enhancing the learning of misclassified samples. In order to verify the effectiveness empirically, we use data from the Kaggle platform. Ten-fold cross-validation is carried out to evaluate and compare the performance among the AdaBoost-DT model, DT, and Random Forest. The empirical results show that the AdaBoost-DT model has higher accuracy. This model is valuable for banks and other financial institutions to evaluate customers' credit efficiently.

Keywords: credit scoring, ensemble learning, imbalanced classification

1. INTRODUCTION

The rapid development of the economy and the change in consumer attitudes have driven the development of personal credit loans. According to data from the People's Bank of China, China's overall personal credit consumption balance rose from \$18.95 trillion in 2015 to \$43.97 trillion in 2019. However, with the increasing amount of credit transactions, the non-performing loan rate has also increased year by year. According to the data released by the China Banking Regulatory Commission, the non-performing loan rate of commercial banks in China reached 1.86% at the end of the fourth quarter of 2019, and the non-performing loan balance reached 2.41 trillion yuan^[1]. Non-performing loans not only affect the normal operation of banks, but also induce social moral risks and cause a series of adverse reactions. Credit scoring for loan applicants is an important tool to reduce credit risk and non-performing loan rates.

Credit scoring refers to classifying customers into "good credit" and "bad credit" customers according to their default risk. The idea of credit scoring appeared in the United States, where David Durand^[2] proposed firstly the application of statistical methods in this field in 1941 to determine the goodness of loan customers. In the late 1960s, the emergence and development of credit cards made banks and companies with credit operations aware of the importance of credit scoring, and more and more experts began to study it. Altman^[3], Meyer^[4], Tam^[5], Lundy^[6] and other scholars used multivariate discriminant analysis, regression analysis, k-nearest neighbor discriminant analysis, and cluster analysis to evaluate individual credit. Leong^[7] used a Bayesian network model to solve the truncated sample, real-time implementation problem in credit risk scoring. Comparing to logistic regression and neural networks, this model performs better in several dimensions such as accuracy and sensitivity. Fang & Chen^[8] propose a credit scoring model based on semi-supervised generalized additive (SSGA) logistic regression to use both labeled and unlabeled sample information.

In recent years, with the development of big data and Internet finance, some artificial intelligence methods have been widely applied in credit scoring, including ANN^[9], decision trees^[10], and SVM^[11]. Tony & Jonathan^[12] conducted a comparison study between SVM and traditional methods such as logistic regression. They found that

* Corresponding author. Email: wuzengyuan@cjlu.edu.cn(Zengyuan Wu), zjl13396717601@163.com(Jiali Zhao)

SVM can be used as a feature selection method to discriminate the most important features that determine the magnitude of default risk. Hussain^[13] used an artificial neural network approach to provide technical support for commercial banks' lending decisions. Kambal^[14] used decision trees (DT) and artificial neural networks (ANN) to build credit scoring models and performed comparative analysis. He found that the ANN approach mostly outperformed DT, but the results of DT were more explanatory. Artificial neural networks would improve the efficiency of credit decisions and help financial institutions save analysis time and cost. With the deepening of theoretical and practical research, the imbalance of samples was noticed, i.e., the number of customers with good credit is not the same as the number of customers with bad credit. And the current single classifier cannot obtain good classification results when processing unbalanced data. Therefore, Corchado^[15] used ensemble learning algorithm by cascading several weak classifiers to overcome the limitations of a single classifier. Wang^[16] proposed an ensemble algorithm with decision trees to reduce the effects of data noise and redundancy of features, and confirmed that relatively high classification accuracy could be obtained. Finlay^[17] used a variety of Boosting and Bagging for modeling in the credit scoring, and the results showed that the ensemble learning method achieved better classification results than single classifier.

According to the shortcoming of a single classifier to process imbalanced data, we employ the adaptive boosting algorithm for credit scoring. As a typical ensemble algorithm, the AdaBoost algorithm can automatically adjust the weights of the base classifier and improve the classification accuracy by enhancing the learning of misclassified samples. Decision tree is employed as the base classifier. An ensemble learning classification model is proposed for credit scoring, where we employ adaptive boosting algorithm by cascading multiple decision trees (DT), named AdaBoost-DT model. Area Under Curve (AUC) and G-mean are selected as performance evaluation metrics. Furthermore, we empirically test our model using the data from Kaggle platform. In order to verify the effectiveness, we compare the performance of our proposed model with Decision Tree and Random Forest.

2. ADABOOST-DT MODEL

2.1 Decision tree classifier

A decision tree is a tree-like structure that divides a set of input samples into several smaller sets based on certain features of their attributes, and it is a fundamental classification method in machine learning. Unlike traditional statistical classifiers, decision trees use a multi-stage or sequential approach to the label assignment problem. The labelling process is considered as a simple decision chain based on successive test results rather than a single complex decision. In general, decision tree structures include tree nodes, bifurcation paths, and leaf nodes. The root node represents the object, while each branch fork path represents the value of an attribute of the object, and the leaf node represents the value of the object as represented by the path experienced from the root node to that leaf node.

Decision trees were chosen as the base classifier for three main reasons. Firstly, the resulting classification model is easier to explain and illustrate due to its intuitive presentation^[18]. Secondly, unlike statistical models, decision trees require fewer assumptions in terms of data distribution^[17]. Finally, they are relatively fast to construct compared to other techniques.

2.2 AdaBoost algorithm

The AdaBoost^[19] algorithm is a classical ensemble algorithm proposed by Yoav Freund and Robert Schapire in 1995 to achieve better prediction by cascading several weak classifiers. The basic idea is that at the beginning, if there are N samples, each training sample is given the same weight $1/N$. If a sample fails in training during the training process, a larger weight is given, which can make the classifier in the next iteration will focus on learning those failed samples. However, for accurately classified samples, their weights are reduced to obtain a new sample distribution. The AdaBoost algorithm training process is as follows.

Input: training sample set

$S = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_i, Y_i)\}, i = 1, 2, \dots, n, Y_i \in \{0, 1\}$, M is the number of iterations, and H is the base classifier.

(1) Initialize the weight distribution of each sample in the training data

$$D_1 = (u_{11}, u_{12}, \dots, u_{1i}, \dots, u_{1N}), u_{1i} = \frac{1}{N}, i = 1, 2, \dots, N \quad (1)$$

(2) Perform M iterations

(a) The training sample set with the weight distribution H_m is learned to obtain the weak classifier.

$$H_m(x): \mathcal{X} \rightarrow \{-1, +1\} \quad (2)$$

(b) Calculate the classification error rate e_m , and discard the weak classifier if e_m is greater than 50%.

$$e_m = \frac{N}{\sum_{i=1}^N P(H_m(x_i) \neq y_i)} = \frac{N}{\sum_{i=1}^N u_{mi} I(H_m(x_i) \neq y_i)} \quad (3)$$

(c) Calculate the importance of the weak classifier in the final classifier

$$\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \quad (4)$$

(d) Update the weight distribution of the training sample set for the next round of iterations.

$$D_{m+1}(i) = \frac{D_m(i)}{Z_m} \exp(-\alpha_m y_i H_m(x_i)), i = 1, 2, \dots, N \quad (5)$$

where Z_m is the normalization factor and is the sum of all samples corresponding to weights of 1.

$$Z_m = \sum_{i=1}^N u_{mi} \exp(-\alpha_m y_i H_m(x_i)) \quad (6)$$

(3) Combining weak classifiers to output strong classifiers

$$H(x) = \text{sign}(f(x)) = \text{sign} \left[\sum_{m=1}^M \alpha_m H_m(x) \right] \quad (7)$$

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Experimental data set

To verify the effectiveness of the model in this paper, the experiment uses the customer credit dataset from the public dataset provided by Kaggle. We define the customers with two months or more overdue loan repayments in Status as the default sample (positive sample) and the rest as the compliance sample (negative sample), where the number of positive samples is 422 and the number of negative samples is 24,712, with an imbalance ratio of 1:58. In addition to the repayment status, each record also contains 17 attributes, as shown in Table 1.

Table 1 Sample properties

Property Name	Meaning	Property Type
ID	Customer Number	Continuous type
CODE_GENDER	Gender	Discrete type
FLAG_OWN_CAR	Is there a car	Discrete type
FLAG_OWN_REALTY	Whether there is a property	Discrete type
CNT_CHILDREN	Number of children	Continuous type
AMT_INCOME_TOTAL	Annual income	Continuous type
NAME_INCOME_TYPE	Income Category	Discrete type
NAME_EDUCATION_TYPE	Education level	Discrete type
NAME_FAMILY_STATUS	Marital Status	Discrete type
NAME_HOUSING_TYPE	Living Style	Discrete type
DAYS_BIRTH	Birthday	Continuous type
DAYS_EMPLOYED	Start date	Continuous type
FLAG_MOBIL	Availability of cell phones	Discrete type
FLAG_WORK_PHONE	Availability of working telephone	Discrete type
FLAG_PHONE	Availability of telephone	Discrete type
FLAG_EMAIL	Is there email	Discrete type
OCCUPATION_TYPE	Career	Text type
CNT_FAM_MEMBERS	Family size	Continuous type
MONTHS_BALANCE	Month of recording	Continuous type
STSTATUS	Repayment Status	Discrete type

3.2 Data pre-processing

Customer credit data is characterized by large volume, missing data and anomalies, which are not conducive to finding the required information quickly. Therefore, the above-mentioned characteristics of credit datasets need to be pre-processed before data mining to provide clean and more targeted high-quality data for data mining algorithms, thus improving data mining efficiency.

Firstly, features with more than half of the missing values are removed and the rest are filled with the missing values using the mean value. Secondly, the continuous values of income, age and years of work are discretized to increase the robustness to abnormal data. Finally, the income categories of customers are aggregated, and all job categories are divided into "lab work", "office work", "high-tech work". This is used to analyze the relationship between the customer's income category and the default or non-default.

3.3 Evaluation indicators

The current evaluation metrics for binary classification problems usually use the classification correctness^[20] (Accuracy, Acc), but Acc ignores the performance of recognition of a few classes. For example, the prediction of a certain disease, even if the accuracy reaches 99%, but does not identify the people who are really sick, such a high accuracy is meaningless. The effective identification of a few classes in unbalanced data classification is more practically meaningful, so Acc is not sufficient as a performance metric for the model. In order to better evaluate the accuracy of the model, the geometric mean criterion (G-means metric) and AUC (area under the ROC curve) are used as evaluation metrics in this paper. Most of the above evaluation methods are represented by confusion matrix (Table 2). Among them, TP indicates that the positive class sample prediction is still positive, TN indicates that the negative class sample prediction is still negative, FP indicates that the negative class sample misclassification is positive, and FN indicates that the positive class sample misclassification is negative.

Table 2 Confusion matrix

Actual category	Predicted results	
	Positive Class Sample	Negative Class Sample
Positive Class Sample	TP	FN
Negative Class Sample	FP	TN

(1) G-mean value

$$G - mean = \sqrt{\frac{TP}{TP + FP} \times \frac{TP}{TP + FN}} \tag{15}$$

The G-mean is based on the positive class classification accuracy and the negative class classification accuracy, so it can better measure the comprehensive performance of the classification method on the unbalanced data set. And its value is in the range of [0,1], the larger the G-mean value is, the better the comprehensive performance of the model is. The G-mean value is large only when both positive and negative classes achieve high accuracy.

(2) AUC

The ROC curve is a curve plotted with the false positive rate as the horizontal coordinate and the true positive rate as the vertical coordinate, depicting the changes of the false positive rate and true positive rate under different parameter variations. The classifier’s performance will be better if the curve is close to the upper left corner. The performance of the classifier is generally evaluated by the area AUC between the ROC curve and the coordinate axis. The higher the value of AUC, the better the performance of the classifier.

3.4 Analysis of experimental results

The experiments were conducted using a ten-fold cross-validation method to divide the data set into 10 parts equally, and the ratio of the training set to the test set was 1:9. The average value was finally used as an estimate of the accuracy of the classification algorithm. In this paper, AUC and G-mean values are used as model evaluation indexes, and they are compared and analyzed with decision tree(DT) and random forest. The classification performance results of the three models are shown in Table 3.

Table 3 Performance comparison of various algorithms

Models	AUC mean value	G-mean average value
DT	0.5218	0.2072
Random Forest	0.5500	0.5125
AdaBoost-DT	0.6854	0.7495

From the experimental results, it is seen that the AdaBoost-DT model has larger values on both AUC and G-mean than the other two models. Therefore, the proposed model in this paper has a higher accuracy rate.

According to Table 3, it can be seen that the average AUC value of the AdaBoost-DT algorithm proposed in this paper is greater than the other two algorithms, so the classification prediction effect of the model in this paper is better than the other two models. Generally, the closer the AUC value is to 1, the better the classification performance of the model is. And the AUC value of the algorithm proposed in this paper reaches 0.69, which has a better performance.

G-mean value is used to measure the accuracy of the positive and negative classes of the sample, and high G-mean value indicates that the discrimination between the two classes is accurate. The AdaBoost-DT algorithm proposed in this paper achieves 0.75 in the G-mean value. Thus, the model in this paper can effectively discriminate not only the good credit customers but also the bad credit customers, and can fully maintain the classification balance between the two.

In summary, the AdaBoost-DT model proposed in this paper has higher accuracy for customer credit evaluation in unbalanced data classification.

4. CONCLUSIONS

In the era of big data where the financial industry is moving towards information technology, it has become an inevitable trend to apply machine learning to credit scoring models. In this context, the AdaBoost ensemble algorithm is proposed to build a credit scoring model in order to improve customer credit prediction. According to the shortcoming of a single classifier to process imbalanced data, we use the AdaBoost-DT algorithm to build a credit scoring model. This model uses the AdaBoost ensemble method with DT as the base classifier, and solve the propensity problem for most samples by using the AdaBoost algorithm to increase the weight of the sample automatically. The empirical results show that the AdaBoost-DT model has higher accuracy than DT and random forest. The study provides a new credit scoring model for banks and credit companies, which can contribute to predicting the credit level of customers and reducing default loan generation.

Although the AdaBoost-DT model constructed in this paper has better classification effect than other models, the parameter optimization methods and the selection of base classifiers are not comprehensive enough. Future research should try more parameter optimization methods and try to implement more traditional classifiers as base classifiers. In addition, the credit scoring model proposed in this paper only targets the imbalance of data without considering the coexistence of data multidimensionality. Although the traditional classification method can obtain better classification results in low-dimensional data, it is more difficult to handle in high-dimensional data. Therefore, future research will consider classification models for high-dimensional, unbalanced data to improve the generalizability and application value of the methods in real life.

ACKNOWLEDGEMENT

This research was supported by Zhejiang Philosophy and Social Science Project under Grant 19NDJC203YB.

REFERENCES

- [1] Li Runfa. (2020). CBRC: China's commercial banks' non-performing loan ratio 1.86% at the end of 2019. http://www.gov.cn/shuju/2020-02/17/content_5480190.html (in Chinese).
- [2] Durand D. Risk Elements in Consumer Installment Financing[M]. New York: National Bureau of Economic Research, 1941: 78-129.
- [3] Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy[J]. *Journal of Finance*, 1968, 23: 589-609.
- [4] Meyer P A, Pifer H. Prediction of bank failures[J]. *Journal of Finance*, 1970, 25: 853-868.
- [5] Kar Y T, Melody Y K. Managerial Applications of Neural Networks: The Case of Bank Failure Predictions[J]. *Management Science*, 1992, 38(7): 927-947.
- [6] Lundy M. Credit Scoring and Credit Control[M]. New York: Oxford University Press, 1993: 25-36.
- [7] Leong C K. Credit Risk Scoring with Bayesian Network Models[J]. *Computational Economics*, 2016, 47(3):423-446.
- [8] Fang K N, Chen Z L. A credit scoring method based on semi-supervised generalizable additive logistic regression[J]. *Systems Engineering Theory and Practice*, 2020,40(02):124-134. (in Chinese).
- [9] Sustersic M, Mramor D, Zupan J. Consumer Credit Scoring Models with Limited Data[J]. *Expert Systems with Applications*, 2009,36(3):4736-4744.
- [10] Pathak A N, Sehgal M, Francolin C. A study on fraud detection based on data mining using decision tree[J]. *International Journal of Computer Science Issues*, 2011, 8(3): 258.
- [11] Pawiak P, Abdar M, Acharya U R. Application of New Deep Genetic Cascade Ensemble of SVM Classifiers to Predict

- the Australian Credit Scoring[J]. *Applied Soft Computing*, 2019, 84: 105740.
- [12] Bellotti T, Crook J. Support vector machines for credit scoring and discovery of significant features[J]. *Expert Systems with Applications*, 2009, 36(2):3302-3308.
- [13] Bekhet H A, Eletter S F K. Credit risk assessment model for Jordanian commercial banks: Neural scoring approach[J]. *Review of Development Finance*, 2014, 4(1):20-28.
- [14] Kambal E, Osman I, Taha M, et al. Credit scoring using data mining techniques with particular reference to Sudanese banks[C]// *International Conference on Computing*. IEEE, 2013.
- [15] Michał W, Manuel G, Emilio C. A survey of multiple classifier systems as hybrid systems[J]. *Information Fusion*, 2014,16:3-17.
- [16] Wang G, Ma J, Huang L, et al. Two credit scoring models based on dual strategy ensemble trees[J]. *Knowledge Based Systems*, 2012, 26:61-68.
- [17] Finlay S. Multiple classifier architectures and their application to credit risk assessment[J]. *European Journal of Operational Research*, 2011, 210(2): 368-378.
- [18] Xia Y, Zhao J, He L, et al. A novel tree-based dynamic heterogeneous ensemble method for credit scoring[J]. *Expert Systems with Applications*, 2020, 159:113615.
- [19] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to Boosting[J]. *Journal of Computer and System Sciences*, 1999, 55(01): 119-139.
- [20] Razia S, Rao A S. Machine learning techniques for thyroid disease diagnosis- A review[J]. *Indian Journal of Science & Technology*, 2016, 9(28): 1-9.

Short Research Paper

Robotic Workforces' Value and Relationships with Human Employees: The Impacts on Human Work Efficacy

Meng Zhang¹, Yi Wang²

¹School of Business Administration, Southwestern University of Finance and Economic, Chengdu, 611130, China

²School of Business Administration, Southwestern University of Finance and Economic, Chengdu, 611130, China

Abstract: As artificial intelligence (AI) and robotic process automation (RPA) develop rapidly, robotic workforces are playing increasingly prominent roles in enterprises as an emerging human resource force. This development is affecting the psychological state of human employees. Drawing upon social comparison perspective, this research examines how human employees perceive the value of robotic workforces and the human-robot relationship. The study also analyzes how these perceptions jointly influence employee work efficacy. The following two unexplored issues are addressed: First, the study focuses on the interactions between the perceived value of robotic workforces and the perceived human-robot relationships by human employees. Those interactions can be divided into four categories. Second, the study examines the social comparison process in human-robot interaction. This research enriches the meaning of social comparison and analyzes the changes of human employees' psychological states. A mixed design is developed to empirically test the proposed model and hypotheses.

Keywords: robotic workforce; value; human-robot relationship; social comparison; work efficacy

1. INTRODUCTION

In the age of the digital economy, companies increasingly deploy advanced digital technologies, including artificial intelligence and robotic process automation, to engage in digital transformation[1]. The term “robotic workforce” essentially refers to robots that are capable of conducting tasks in a human-like way in the workplace. These tasks mainly consist of automatically running software robots and service-oriented robots with physical characteristics^[1].

Representing an emerging type of workforce, the robot workforce is bound to greatly influence the existing human resource ecosystem^{[2][3]}. According to McKinsey's forecast, automation investment in the United States will reach \$145 billion by 2030, and the modeled unemployment displacement will be 23 percent^[4]. The World Economic Forum report predicts that 42% of workers will be replaced by machines^[5]. With high standards in efficiency, competence, controllability and low costs, a robot workforce can play a momentous role, especially in the fields of manufacturing, finance, accounting, healthcare and other industries^[1]. This degree of automation in the workplace will arguably change human employees, psychologically speaking, and will eventually transform their behavior. Therefore, exploring how human employees perceive and respond to the presence of their robot colleagues is critical.

Robotic workforces have attracted considerable academic attention. Ljungblad et al. (2012) examines how human workers perceive their robotic peers^[6]. Yogeewaran et al. (2016) focuses on the impact of certain features of robot workforces on human employees^[7]. Muthusamy et al. (2020) looks at the contribution of a robot workforce to corporate operations^[8]. Prior studies contribute to our understanding of the value of robot workforces. However, little attention has been paid to how a robot workforce influences specific psychological changes among human employees. Because this phenomenon has only appeared over a relatively short time, people are still in the initial stages of figuring out what a robot workforce is, rather than focusing on why and how it is. This study

* Corresponding author. Email: wangyi63@swufe.edu.cn(Yi Wang)

highlights two research gaps that deserve further examination.

First, few studies to date have examined how human employees perceive their robot peers; the impacts of these perceptions on the psychological states of human workers have also not been explored. Despite the fact that some studies argue that the features of digital workers can have effects on human workers [7], the underlying mechanism is unclear. Furthermore, various features of digital workers could jointly, rather than separately, influence how human employees perceive and behave in human-robot interactions [7]. These joint effects are far from clear.

Second, prior literature tends to treat robots as a tool or instrument in the workplace. Little research focuses on possible social comparisons during human-robot interaction. With the development of AI and RPA, human employees tend to treat robots with in-built social cues as colleagues, and social comparisons may emerge in human-robot interaction. Therefore, social comparison, which is typically used to understand human-human relationship/interaction, may be a meaningful perspective to understand this issue.

Motivated by these research gaps, this research explores the effects of perceived value and human-robot relationships on human employees' work efficacy from a social comparison perspective. Overall, this study's model proposes to disclose how the perceived features of robots in the workplace influence human employees' psychological states.

2. THEORY AND HYPOTHESES DEVELOPMENT

2.1 Robot value and human-robot relationship

The perceived value of a robot workforce (by human employees) indicates the extent to which robots in the workplace are perceived valuable in terms of performing tasks. Human perception of robot value relies on the functionality of robots and tasks these robots are supposed to fulfill. Robots may be perceived to present high value in performing routine, repetitive, and error-prone tasks [1].

Human employees may perceive robots in the workplace as either competitors or collaborators. Thus, human-robot relationships can be divided into cooperative and competitive categories. Human employees that hold a perception of cooperative human-robot relationships usually view robots as assistants in performing tasks. Srnicek and Williams (2015) argued the positive role of robots by proposing that robots can save people from tedious chores, rather than replacing humans [9]. On the contrary, other human employees may perceive that robots can equally or better deal with the same tasks and that they (the human employees) may be replaced by these superior robots. This pessimism is also supported by Ford (2015), who noted that robots could threaten human employees' job security [10]. Various perceptions of human-robot relationships may generate different emotions among human employees. In contrast to competitive robots, perceived cooperative robots likely bring about positive emotions in human employees [11]. Overall, human employees feel differently when they perceived different human-robot relationships.

Robot value and human-robot relationships can jointly influence human workers' psychological states and behaviors [7]. Based on the levels of perceived robot value and the types of human-robot relationships, this research classifies workplace human-robot interaction into four categories.

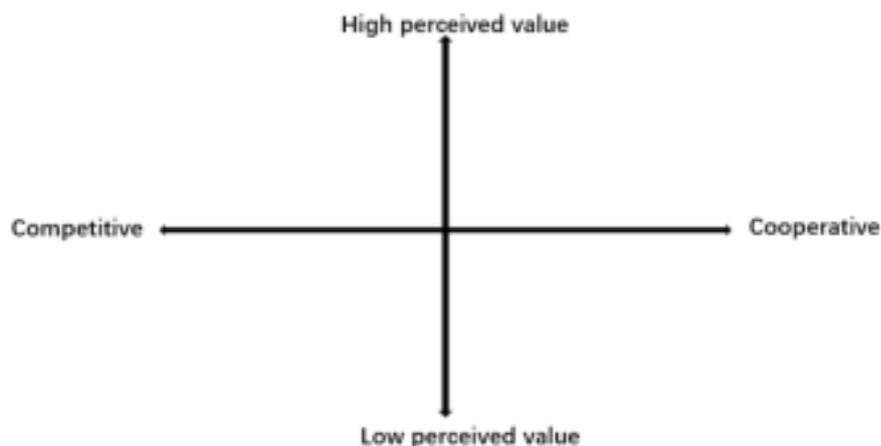


Figure 1. The category of robot workforce interaction

As Figure 1 shows, four corresponding types of robot relationships exist. (1) The high value and cooperative relationship represent a robot that is perceived by human employees to have high value and be cooperative during human-robot interaction. This type of robot can be a great workplace partner and helps reinforce human employees' efficiency. (2) The high value and competitive relationship represent a robot that is perceived by human employees to have high value and be competitive during human-robot interaction. This type of robot is a strong competitor for human employees' jobs and could even replace human beings in the future. (3) The low value and competitive relationship represent a robot that is perceived by human employees to have low value and be competitive. This type of robot is an inferior competitor that actually brings confidence to humans. (4) The low value and cooperative relationship represent a robot that is perceived by human employees to have low value and be cooperative. This type of robot is an inefficient collaborator.

2.2 Social comparison in human-robot interaction

Festinger (1954) proposed the social comparison process among individual situations, status, characters, etc., in order to subjectively realize individual evaluations ^[12]. Social comparison produces either a contrast effect or an assimilation effect, which engenders diverse perceptions, depending on the specific situation ^[13]. Social comparison appears so universal that it has been used to explain various social behaviors, such as human self-evaluation, self-improvement, and self-satisfaction.

Social comparisons can be divided into three types: parallel comparison, upward comparison and downward comparison. Parallel comparison means to compare with others who are less divergent from himself ^[12]. Downward comparison means to compare with others who are inferior; conversely, upward comparison means to compare with others who are superior ^{[14][15]}. Although early studies believed that individuals can achieve a good understanding of themselves through parallel comparison, later studies noted that downward comparison and upward comparison work better in self-evaluation. Thus, this study focuses more on the latter two. In particular, humans tend to choose inferior comparison targets that can induce positive emotions and desirable conclusions. Prior literatures examine how downward comparison is likely to produce positive emotions, whereas upward comparison is likely to produce negative emotions ^{[15][16]}. Conversely, it was noted that self-expectation influences comparison results. In other words, if human expectation is lower than the expectations related to the comparison objects, negative emotions and self-evaluation will emerge. Alternatively, if human expectation is equal to or higher than the expectations related to the comparison objects, positive emotions and self-evaluation emerge ^[18]. In particular, upward comparison can produce positive assimilation effects when the intimate comparison object

becomes an important part of the human's self-concept ^[17].

Humans evaluate themselves not only by comparing themselves with others, but also by comparing themselves ^[18]. Social comparison with others represents an inter-person comparison, which includes parallel comparison, upward comparison and downward comparison. Social comparison with selves represents an intra-person comparison, which includes comparisons with selves in the past, present and future. This type of comparison can also be called temporal comparison ^[19]. The purpose of intra-person social comparison is to obtain satisfied emotions, rather than to perform an accurate self-evaluation ^[18].

According to Reeves et al. (1996) and Yi Mou (2020), if technologies present social cues, humans may unconsciously treat technologies as humans ^[20] ^[21]. This implies that the understanding of social interactions among humans can be applied to human-robot contexts. Ljungblad et al. (2012) argued that human employees regard robots as workplace partners ^[6]. Particularly, physical robots are given certain human features, such as human-like figures, specific genders and human names ^[12] ^[21]. Besides, there exist three basic conditions for the formation of social comparison: (1) Humans are motivated to evaluate their own value and capabilities ^[22]; (2) Comparisons with others enhance self-understanding ^[22]; (3) A comparison is useful for evaluating individual value and capabilities ^[22]. These assumptions of social comparison also apply to the human-robot interaction. Overall, workplace robots could be the targets of social comparison for human employees.

When human workers perceive that a robotic workforce presents high value and cooperativeness, they may experience upward comparison. Since no substitution relationship exists, human employees treat robots as workplace partners, and they can collaborate with robots more efficiently. This situation reinforces human employees' self-awareness and positive emotions. Moreover, human employees can achieve self-improvement by observing and learning from their robotic peers. Therefore, this robotic workforce can create a sense of support ^[13]. Additionally, human employees feel more confident when they conduct an intra-person comparison and find that they can perform tasks better with the help of robots than they can in situations with no robots. Thus, we propose:

Hypothesis 1a: When human employees perceive that robotic workforce has high value and is cooperative, they perceive supports from robots.

When human employees perceive a high value of robots and a competitive human-robot relationship, they experience an upward comparison. Competition between humans and robots implies that robots can replace human employees and therefore decrease the benefits of human employees. The competition between humans and robots becomes fierce when robots are believed to have high value. In this circumstance, human employees may interpret robots as a highly-competitive substitute and as a source of threats to the humans' employment ^[15]. Thus, we propose:

Hypothesis 1b: When human employees perceive that a robotic workforce has high value and is competitive, they perceive threats from robots.

When perceiving low value robots and a competitive human-robot relationship, human employees may go through a downward comparison. Comparing themselves with inferior competitors, human employees may generate positive emotions and enhanced self-evaluation. Therefore, this comparison brings about a sense of support ^[15].

Hypothesis 1c: When human employees perceive that a robotic workforce has low value and is competitive, they perceive support from the robots.

When perceiving low value robots and a cooperative human-robot relationship, human employees may go through a downward comparison. A cooperative human-robot relationship may cause the development of positive feelings toward human-robot interaction. However, the perceived low value of robots may mitigate this feeling. Thus, we posit:

Hypothesis 1d: When human employees perceive that a robotic workforce has low value and is cooperative, they have neutral feelings toward the robots.

2.3 Work efficacy

Work efficacy refers to the subjective evaluation of confidence in human ability as related to performance [23]. Work efficacy is most important for motivation, attitude, and performance in the workplace. Self-judgment and social comparison can form work efficacy. When humans conduct various types of social comparisons, they develop a clearer evaluation of themselves and further influence how they evaluate their capacity in the workplace. Therefore, we propose:

Hypothesis 2a: If human employees perceive supports, their work efficiency would be high.

Hypothesis 2b: If human employees perceive threats, their work efficiency would be low.

Hypothesis 3: Human employees' perceptions (i.e., perceived supports/threats) mediate the integrative effect of robot value and the human-robot relationship on work efficacy.

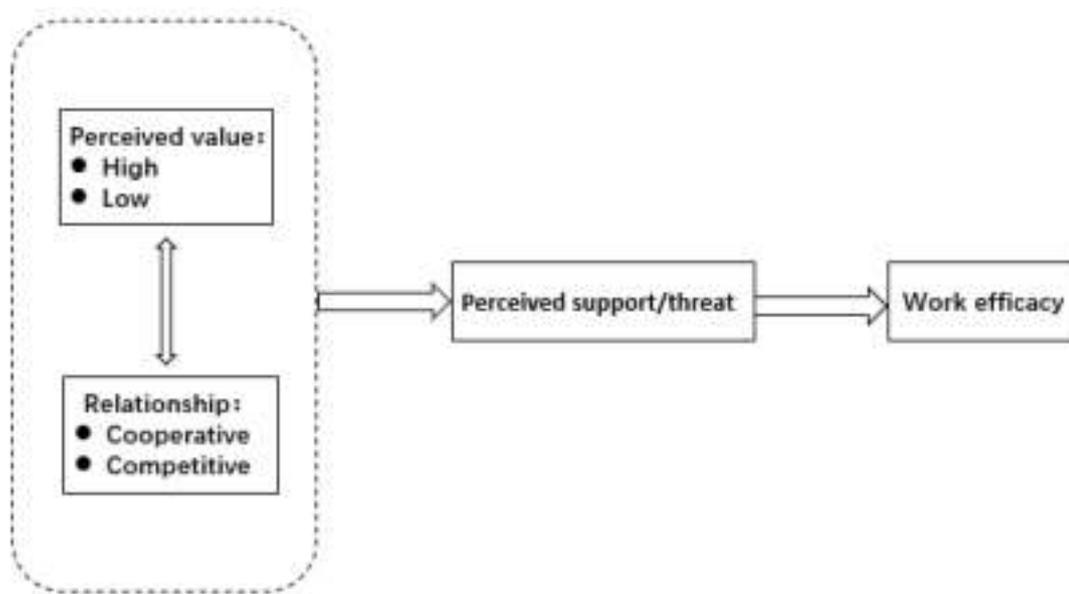


Figure 2. The process of social comparison in human-robot interaction

In summary, the conceptual model is shown in Figure 2.

3. RESEARCH DESIGN

This research intends to use a mixed research design to empirically test the study's propositions. Two studies are planned: Study 1 would conduct a scenario-based experiment to verify specific hypotheses. Study 2 purposes to conduct a survey, in order to further verify the hypotheses.

3.1 Study 1: An experiment

We plan to recruit college students to participate in this experiment. The participants will be invited to take part in a spelling check game. They will be told that robots will be used to work with them to identify and correct misspellings. The time limit for each round will be 15 seconds. If all the misspellings are found and corrected in the given time limit, the round ends automatically. Robots will be designed to have both low-level and high-level performance in terms of speed of spelling check. Human-robot relationships will be manipulated by designing different game rules. Specifically, a game rule is set for a cooperative relationship, in that the total commission for success is shared by participants and robots, with 80 percent for participants and 20 percent for robots. A competitive relationship is manipulated by setting a rule that stipulates that participants who perform better get

all the commissions. Participants will be informed that the supervisor expects them to complete more than 50 accurate corrections per round if they are to receive special rewards (to make them participate seriously).

The experiment will be conducted with an inter-group design of 2 (robot value: high vs. low) x 2 (human-robot relationship: cooperative vs. competitive). First, participants will be asked to fill in their demographic information. Second, the task scenario will be introduced to them. Third, the novice task will be completed first. This indicates that the participants who complete tasks alone and receive the full commission represent the benchmark. Fourth, the experiment will formally start. Participants will be randomly divided into four groups that match the four types of robot (high value cooperator, high value competitor, low value competitor and low value cooperator). The participants will be reminded of how to calculate their commission. Then, the respective challenge begins. Finally, participants will be asked to answer the following questions regarding: (1) social comparison results. For example, “Do you think a robotic workforce can help (or threaten) you a lot if the game continues?” (2) They will also be asked about work efficacy. For example, “If you continue to work with the robot, are you more confident in getting a higher commission, compared with the novice task?”

3.2 Study 2: A survey study

The authors further plan to develop a questionnaire combined with existing scales to collect data from employees working with robots. The questionnaire consists of four parts: (1) demographic information; (2) human employees' perception of robot value and human-robot relationships; (3) human employees' perceived support, threat, and neutral feelings, and (4) work efficacy. A structural equation modeling method will be used to analyze data and empirically test the proposed model.

4. SUMMARY

This study's purpose is to make the following theoretical contributions: (1) The study applies social comparison perspective to examine human-robot interaction. (2) The study further explores how various perceptions of robot characteristics jointly affect human employees' psychological states.

We expect this study to contribute to practice by: (1) encouraging managers to assess human employees' psychological changes when introducing robots in the workplace and to take a few measures in advance, and (2) by offering helpful insights regarding how to respond to the coming age of robots.

REFERENCES

- [1] Minfeng Lu, Jin Ma. (2020). Research on Digital Employee Management Under the Digitization of Commercial Banks. *Financial theory and teaching*, 163(5),1-6+11. (in Chinese)
- [2] Daugherty, Paul R., Wilson H J. (2018). *Human+ machine: Reimagining work in the age of AI*. Harvard University: Harvard Business Press, 85-101.
- [3] Madakam S, Holmukhe R M, Jaiswal D K. (2019). The future digital work force: robotic process automation (RPA). *Journal of Information Systems and Technology Management*, 16(16), 1-16.
- [4] Manyika J, Lund S, Chui M, et al. (2017). Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages. www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages
- [5] World Economic Forum. (2020). The Future of Jobs Report 2020. <https://www.weforum.org/reports/the-future-of-jobs-report-2020>.
- [6] Ljungblad, Sara, et al. (2012). Hospital robot at work: something alien or an intelligent colleague? *ACM*, Washington, 2012. Washington: ACM, 177-186.

- [7] Yogeewaran, Kumar, et al. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, 5(2), 29-47.
- [8] Muthusamy, Vinod, et al. (2020). Do's and Don'ts for Human and Digital Worker Integration. <https://arxiv.org/pdf/2010.07738.pdf>
- [9] Srnicek N. and Williams A. (2015), *Inventing the Future. Postcapitalism and a World Without Work*. London: Verso, 78-109.
- [10] Ford, M. (2015). The rise of the robots: Technology and the threat of mass unemployment. *International Journal of HRD Practice Policy and Research*, 1(1), 111-112.
- [11] Johnson D W, Maruyama G, Johnson R, et al. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological bulletin*, 89(1), 47.
- [12] Festinger L. (1954). A theory of social comparison processes. *Human relations*, 7(2), 117-140.
- [13] Koomen W. (2001). I, we, and the effects of others on me: How self-construal level moderates social comparison effects. *Journal of personality and social psychology*, 80(5), 766-781.
- [14] Hakmiller K L. (1962). *Social comparison processes under differential conditions of ego-threat*. Doctoral dissertation. Minnesota: University of Minnesota.
- [15] Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological bulletin*, 90(2): 245-271.
- [16] Aspinwall L G, Taylor S E. (1993). Effects of social comparison direction, threat, and self-esteem on affect, self-evaluation, and expected success. *Journal of personality and social psychology*, 64(5): 708-722
- [17] McFarland C, Buehler R, MacKay L. (2001). Affective responses to social comparisons with extremely close others. *Social cognition*, 19(5): 547-586.
- [18] Jiongwen Sun, Quanquan Zheng. (2004). Self-cognition in Social Comparison and Temporal Comparison. *Advances in Psychological Science*, 12(2): 240-245. (in Chinese)
- [19] Albert S. (1977). Temporal comparison theory. *Psychological review*, 84(6): 485-503.
- [20] Yi Mou. (2020). From Hermeneutic to Alterity Human-Technology Relations: The Social and Cognitive Changes Brought by AI Media Technology. *Journal of Shanghai Normal University*, 49 (1): 92-101. (in Chinese)
- [21] Reeves B, Nass C. (1996). *The media equation: How people treat computers, television, and new media like real people*. Cambridge: Cambridge university press, 120-124.
- [22] Shufen Xing, Guoliang Yu. (2005). A Review on Research of Social Comparison. *Advances in Psychological Science*. 13(1), 78-84. (in Chinese)
- [23] Rigotti T, Schyns B, Mohr G. (2008). A short version of the occupational self-efficacy scale: Structural and construct validity across five countries. *Journal of Career Assessment*. 16(2): 238-255.

Full Research Paper**Behavior Analysis and Recognition of Hidden Populations in Online****Social Network Based on Big Data Method***Minglei Li^{1,2*}, Guoyin Jiang², Wenping Liu¹, Junli Lei¹*¹School of Information Management and Statistics, Hubei University of Economics, China²School of Public Administration, University of Electronic Science and Technology of China, China

Abstract: Hidden populations refer to the minority groups that not well-known to the public. Traditional statistical survey methods are difficult to apply in the study of hidden populations because of that the hidden populations individuals are very troublesome to be found and they are not willing to share the inner opinion with the others. On the other hand, with the development of the Web 2.0, the hidden populations gather and share their views in online social networks due to the openness and anonymity of the Internet. So, this paper analyzes the behavioral characteristics of the hidden populations based on their data in online social networks. This paper uses the lesbian population as an example and analyzes the behavioral characteristics of lesbian by analyzing the data of the lesbian population in Douban Group. First, the activity data on lesbian are collected from Douban Group. Second, behavior characteristics of lesbian are analysed, the regional characteristic, temporal characteristic and text characteristic are mined out by big data method. Third, a lesbian recognition model is proposed based on the above analytical characteristics, and the effectiveness of the recognition model is varified by experiment study. The research of this paper is helpful to understand the behavioral characteristics of hidden populations deeply, and provides decision-making basis of management and service for hidden populations.

Keywords: Hidden population, Online social network, Lesbian, Behavior analysis, Lesbian recognition

1. INTRODUCTION

Hidden populations^[1-3] refer to the minority groups that not well-known to the public. Generally speaking, hidden population is the population that hard to be contacted, and the people in these groups are unwilling to expose themselves for many various reasons, such as: HIV/AIDS, LGBT, patients with depression. Hidden populations are a small proportion of the overall population. However, there is a large population in China, so the absolute quantity of hidden population is big. Due to long misunderstanding, the hidden populations are very sensitive and self-protective, and they are unwilling to communicate with the mainstream society or express their true ideas in the real society^[4]. At the same time, the people involved in hidden population are more complicated, and it is easy to generate events that affect social security such as extreme thoughts, spread of diseases, and fraud, which bring great hidden dangers to society^[5]. Therefore, it is very meaningful for stable and harmonious development of society to improve the level of investigation and analysis of hidden populations and manage hidden populations reasonably and effectively.

The traditional analysis methods of hidden population are mainly based on statistical survey, such as snowball sampling^[6,7] and respondent-driven sampling^[8,9]. However, these traditional methods have many limitations. First, the hidden population is difficult to find, and sufficient survey samples cannot be obtained easily. Second, the hidden population is unwilling to participate in the survey. The last but not all, the cost of traditional method is high. With the development of the Web 2.0, especially the online social networks, many people share their comments on the Internet. Due to the anonymity and openness of the Internet, many hidden populations are more willing to gather and share their ideas in online social networks. And the hidden populations leave a large amount

* Corresponding author. Email: liminglei@hbue.edu.cn(Minglei Li)

of data in online social networks. So we can understand the behavioral characteristics of the hidden population through the analysis of these data [10,12]. The amount of these data materials is very large, and the value density of them is low, so the traditional statistical analysis methods are difficult to find valuable information. Big data method [13,14] is a technology that specializes in processing large-scale data, and is widely used in online data mining, social management and other fields. In this paper, we use big data method to analyze the online social network data of the hidden population, and gain insight into the behavior characteristics of the hidden population.

This paper uses lesbian as an example. We collect the data of lesbian in Douban Group, and analyse the online behavior characteristics of lesbians based on big data method, then propose a recognition model to identify the lesbian population based on machine learning model. The paper is organized as follows. Section 2 presents the overall method of this paper. Section 3 shows the behavior characteristics analysis based on big data method, specifically, we aim to extract the text feature of lesbian based on text analysis technology. By using the behavior characteristics features, Section 4 designs lesbian recognition model based on machine learning model. In Section 5, experimental study is used to demonstrate the effectiveness and practicability of the recognition model. Finally, the conclusions and future works are given in the Section 6.

2. OVERALL METHOD

2.1 Overall structure of the method

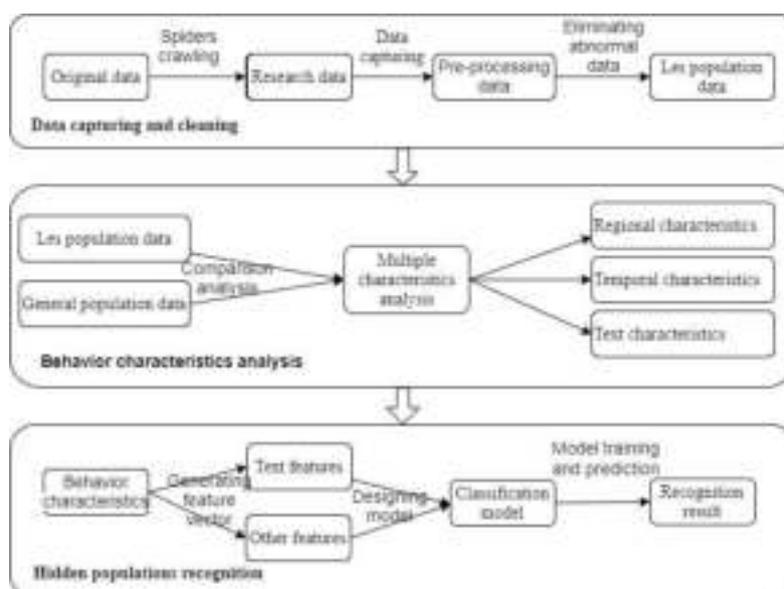


Figure 1. The overall structure of this method

Aiming to analyze the behavior characteristics of hidden populations and recognize the hidden populations based on their data in online social networks. The overall structure of the method proposed in this paper is shown as Figure 1. There are three steps in the method, including data capturing and cleaning, behavior characteristics analysis and hidden populations recognition. Specifically, first, the data in online social networks are collected by Python crawler, data cleaning is used to improve the quality of the data, and irrelevant data, such as spammer and advertisement, are eliminated in this step. Second, the comparative analysis between hidden population and general population is used to study the behavior characteristics, we use the big data technology to analyze the regional characteristics and temporal characteristics, and use text analysis technology to analyze the test characteristics of hidden populations. And finally, we construct the feature vector based on the above characteristics, and design a recognition model based on machine learning model to distinguish the hidden population from general population.

2.2 Data sources

As one of the world's largest Chinese online social network, Douban Group is provided by Douban, and there are more than 400 000 douban groups. The Douban Group involves many fields such as film and television, reading, photography, society, and many hidden populations such as homosexuality and depression. In the Douban group, like-minded users open their hearts and express their opinions freely, so we can get a lot of real data from it.

In this paper, we use lesbian as an example of hidden population, and collect the hidden population data from "Les Sky" Group ('https://www.douban.com/group/lala/'). And a general population data from "Readingmania" Group ('https://www.douban.com/group/readingmania/') is collected as comparative analysis. The collected data involve user information, post and comment data as shown in Table 1. We collect 50 000 comments from "Les Sky" Group and 50 000 comments from "Readingmania" Group.

Table 1. Data fields from Douban Group

User_information	Comment_information	Reply_information
User_id	Comment_id	Reply_id
User_name	Comment_title	Comment_id
User_residence	Comment_text	Author_id
User_url	Author_id	Reply_to_whom
User_join_time	Creation_time	Reply_floor
	Reply_number	Reply_text
		Reply_time

3. BEHAVIOR CHARACTERISTICS ANALYSIS

In this Section, we will find that the hidden population's behavior patterns in online social networks, by comparing and analyzing the behavioral characteristics of the hidden population and the general population in online social networks.

3.1 Temporal characteristic

It can be found that when the different populations are active on the online social network through temporal characteristic analysis. The result is shown as Figure 2. It can be found that the active period of users in Douban Group is night, and it reaches the peak at night, especially around midnight, and the low period is 1:00 to 8:00. In addition, it can be clearly found that between 8:00 to 20:00, the number of comments posted by general population is higher than that of the hidden population. However, after 21:00, the number of comments posted by the hidden population is higher than that of the general population. It can be seen that lesbians prefer to be in online social networks at night.

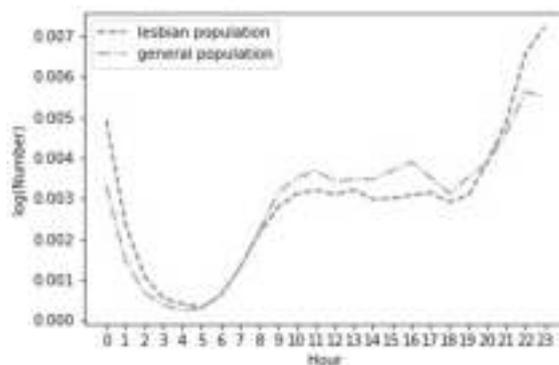


Figure 2. Comments submission time of different populations

3.2 Distribution of number of comments

In Figure 3, it shows the number distribution of comments under each post. As we can see from Figure 3, there are small number of comments under many posts of both the hidden population and general population. However, the hidden population has an obvious aggregation phenomenon, that is, there is a lot of discussion under certain posts.

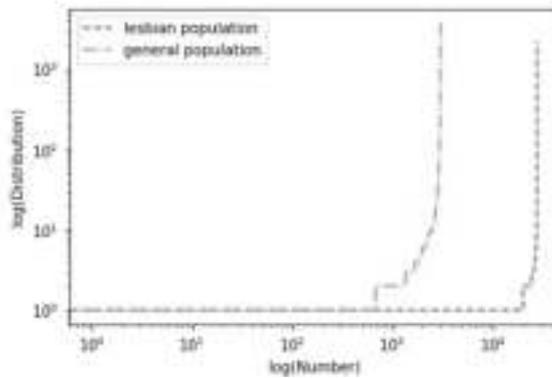


Figure 3. Number distribution of comments under each post of different populations

Figure 4 shows the number of comments posted by each user. It can be found from Figure 4, the number of comments posted by many users is small in Douban Group, and the hidden population is more likely to generate active users, who submit a large number of comments.

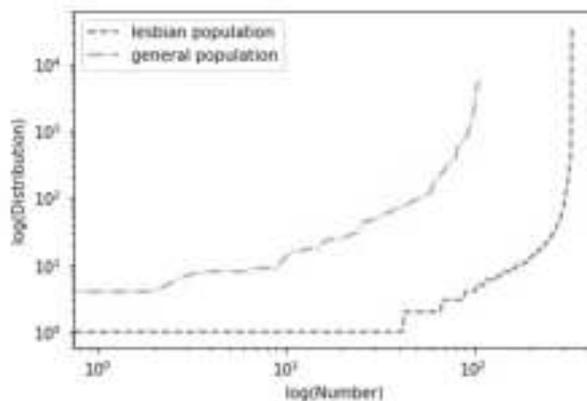


Figure 4. Number distribution of comments from each user of different populations

3.3 Textual characteristic



Figure 5. Comments text wordcloud of different populations

By textual characteristic analysis, we try to discover the topics that population care about and discuss generally. We generate the wordcloud of the comments text each from the lesbian and the general populations. The most of comments text in Douban Group are in Chinese, so we segment the comments text by Jieba module and remove irrelevant words, then the 100 most frequent words are used to generate the wordcloud. The wordcloud results are shown in Figure 5, the left wordcloud is from the lesbian population and the right is from the general population. It can be found from Figure 5, the part of high-frequency words from the hidden population and the general population are same, indicating that the lesbian population is also interested in the topic that the general population cares about. In addition, many unique words, such as "sister" and "feeling", have appeared in the high-frequency words from lesbian population.

4. HIDDEN POPULATION RECOGNITION

From the behavior characteristic analysis result in Section 3, it can be found that there are many differences between the hidden population and the general population. So, we can use these different characteristics to identify the hidden population.

4.1 Features for hidden population recognition

In this paper, the temporal characteristic, number of comments characteristic and textual characteristic are used as the feature vector of the hidden population recognition model. For temporal characteristic, based on the results in Section 3.1, the time is divided three periods: 8:00 to 20:00, 20:00 to 4:00 and the rest of time period. The model uses in which time period the user submits the most comments as the temporal feature value of this user. For number of comments characteristic, this model uses the number of comments submits by the user as the feature value of this author. For textual characteristic, the IF-IDF^[15] is used to vectorize the comments text of the user. Specifically, based on the word segmentation results, we can calculate the TF-IDF value of each word by:

$$TF - IDF(w, u) = TF(w, u) \times IDF(w) \quad (1)$$

where, $TF(w, u)$ is the number of word w appears in the comments text of the user u , $IDF(w) = \log \frac{M}{N}$ (N is the number of word w appears in the text, and M is the number of all text). $TF - IDF(w, u)$ is the feature vector value of word w in comments text of the user u .

4.2 Recognition model

4.2.1 Model structure

The core of recognition model is XGboost^[16]. The XGboost is an extension of the gradient boosting decision tree algorithm. XGboost belongs to ensemble learning model, and its basic idea is to combine multiple Classification And Regression Trees (also known as CART).

Given data set $D = \{(x_i, y_i)\}$ (x_i is the feature vector mentioned in Section 4.1 and y_i is label of user). The integrated model is expressed as

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (2)$$

where K is the number of CARTs in the integrated model, $\mathcal{F} = \{f(x) = \omega_{q(x)}\}$ ($q: \mathcal{R}^m \rightarrow T, \omega \in \mathcal{R}^T$) is the space of CARTs. T represents the number of leaves in the decision tree. Each f_k corresponds to an independent tree structure q and leaf weights ω .

The objective function of XGboost is expressed as

$$\arg \text{Min} \mathcal{L}(\Theta) = \sum_i l(\hat{y}_i - y_i) + \sum_k \Omega(f_k) \quad (3)$$

The objective function consists of two parts, the first part $l(\cdot)$ is the training error. And the second part $\Omega(\cdot)$ is the regularization item that helps to smooth the final learnt weights to avoid over-fitting. $\Omega(\cdot)$ penalizes

the complexity of the model as

$$\Omega(f) = \gamma T + 1/2\lambda \|\omega\|^2 \quad (4)$$

where γ and λ are coefficients.

4.2.2 Training algorithm of the model

The training of XGboost is a boosting algorithm. This algorithm trains a base learner from the training data set firstly, and improves the model based on the base learner iteratively. That is to say, each iteration the model is retained, and a new item is added to the model as

$$\begin{aligned} \hat{y}_0^{(0)} &= 0, \\ &\dots \\ \hat{y}_i^{(t)} &= \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i). \end{aligned} \quad (5)$$

where $\hat{y}_i^{(t)}$ is the predicted value of the samples i in the round t . The choice of adding a new item in each iteration is to minimize the objective. And, the objective function Equation (3) can be turned as

$$\begin{aligned} \mathcal{L}^{(t)} &= \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \\ &= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + f_t(x_i) + \Omega(f_t) + C. \end{aligned} \quad (6)$$

The error function $l(\cdot)$ is squared error, and the objective function Equation (5) can be written as:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n [2(\hat{y}_i^{(t-1)} - y_i)f_t(x_i) + f_t(x_i)^2] + \Omega(f_t) + C. \quad (7)$$

Taylor expansion is used approximately and the objective function Equation (6) can be defined as

$$\bar{\mathcal{L}}^{(t)} \simeq \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + 1/2h_i f_t^2(x_i)] + \Omega(f_t) + C, \quad (8)$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$. The constant terms are removed and the simplified objective at step t is

$$\bar{\mathcal{L}}^{(t)} \simeq \sum_{i=1}^n [g_i f_t(x_i) + 1/2h_i f_t^2(x_i)] + \gamma T_t + 1/2\lambda \|\omega_t\|^2. \quad (9)$$

XGboost generates and updates the decision tree, and being trained by the samples data set, until the objective function satisfies the condition.

5. EXPERIMENTAL RESULTS AND ANALYSIS

5.1 Experimental data

We collect 50 000 comments from "Les Sky" Group and 50 000 comments from "Readingmania" Group. After removing the irrelevant user such as spammer and advertisement, we mark the users who have submitted comments in "Les Sky" Group as lesbian, and mark the users who have submitted comments in "Readingmania" Group as general user. And we choose randomly 3000 users in "Les Sky" Group and 3000 users in "Readingmania" Group. We will use the behavior features to identify whether a user in Douban Group is lesbian.

5.2 Experimental results and analysis

The recognition model in this paper is developed based on XGboost library[†], and we implement the recognition model by Python3 language.

In this paper, the data set is divided into two parts randomly. The one part, which is 80% of the total users, is the training set and is used to train the recognition model. The other one, which is the remaining 20% of total users, is the testing data and is used to verify the effectiveness of this recognition model.

[†] <https://github.com/dmlc/xgboost>

5.2.1 Performance comparison of different features

In this section, we study the impact of using different features on recognition model. We design two features sets: in the features set one (FS1), the features are all the features mentioned in Section 4.1; and in the features set two (FS2), the features are only the textual features. Each of FS1 and FS2 is used to train XGboost by the training data, and verify the performance of the recognition models by the testing data. The hidden population identification problem is a binary classifying problem, so the performance evaluation indicators for binary classifying model are used to measure the effectiveness of the recognition models, such as: accuracy, recall, and precision. The results is shown in Table 2.

Table 2. Performance comparison of different features

Features set	Accuracy	Precision	Recall
FS1	0.8503	0.8357	0.8713
FS2	0.7520	0.7618	0.7333

From the results in Table 2, it can be found that the recognition model using FS1 has better performance than using FS2. The result indicates that all the behavior characteristics analyzed in this paper are helpful to identify the hidden population. And it also reflects that more features and more implicit information. So, the features set FS1 is used in the next experiment.

5.2.2 Performance comparison of different models

In this section, we compare the recognition model XGboost to other familiar machine learning models: linear regression model (LR), decision tree model (DT), ANN model and SVM regression model. In ANN model, the number of hidden layer nodes is 120, the learning rate is 0.001, the maximum number of training time is 5000. In SVM, the kernel is gaussian function, which is vary popular in SVM. All the models are all implemented using Python's machine learning library sklearn. The comparison result is shown in Figure 6.

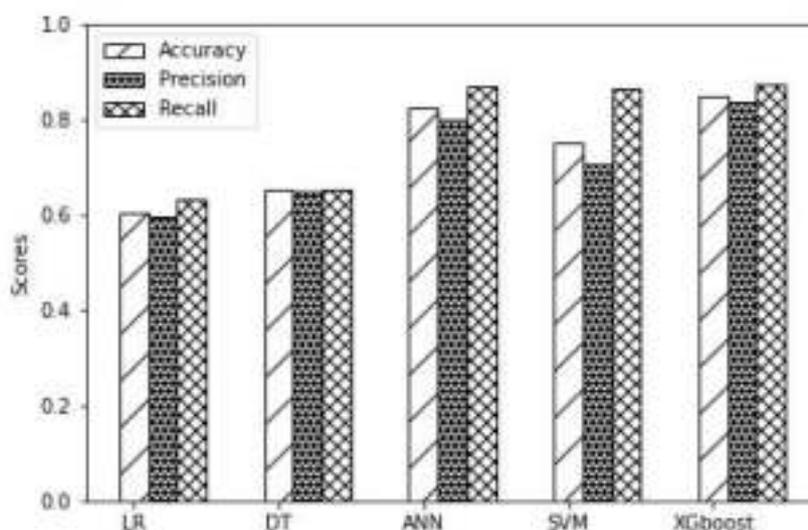


Figure 6. Performance comparison of different models

From the results in Figure 6, we can find that the recognition performance of XGboost is the best among all the models. The result illustrates the effectiveness of the method proposed in this paper.

6. CONCLUSIONS

In order to survey the hidden population deeply and widely, this paper analyzes the data in online social network of the hidden populations based on big data technology. This paper uses lesbian in Douban Group as an example. The main contributions of this paper are: (1) big data technology is used to analyze the data in online

social network of the hidden populations, so as to understand the behavior characteristics of these populations, (2) according to these behavior characteristics in online social network, a recognition model based on machine learning model is proposed to distinguish the hidden population from general population. Furthermore, the effectiveness and practicability of the recognition model are demonstrated with the data from Douban Group.

The research in this paper shows the potential and advantages of big data method in hidden population analysis. However, the methods and conclusions of this paper are still preliminary and limited, Our ongoing work is to consider the following aspects: (1) the more data and more types of hidden populations in online social network should be collected, and the methods of this paper will be applied for these data to get more information, (2) the group behaviors of hidden populations are also very important, we will use the social complex network analysis and other methods to analyze the group behaviors of the hidden populations, (3) this paper mainly analyzes the static characteristics of the hidden population, the next work is to study the dynamic characteristics of the hidden populations and design a model to predict the dynamic characteristics of the hidden populations.

ACKNOWLEDGEMENT

This work was partially supported by the National Natural Science Foundation of China under Grant 72071031 and Grant 62072163, and the Humanities and Social Sciences Research Youth Foundation of the Ministry of Education of China under Grant 20YJCZH072, the Fundamental Research Funds for the Central Universities of China (No. ZYGX2017KYQD185), and the Educational Commission of Hubei Province, China under Grant 19Q145.

REFERENCES

- [1] Spreen M. Rare populations, hidden populations, and link-tracing designs: What and why?[J]. *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique*, 1992, 36(1): 34-58.
- [2] Magnani R, Sabin K, Saidel T, et al. Review of sampling hard-to-reach and hidden populations for HIV surveillance[J]. *Aids*, 2005, 19: S67-S72.
- [3] Wesson P, Reingold A, McFarland W. Theoretical and empirical comparisons of methods to estimate the size of hard-to-reach populations: a systematic review[J]. *AIDS and behavior*, 2017, 21(7): 2188-2206.
- [4] Griffiths P, Gossop M, Powis B, et al. Reaching hidden populations of drug users by privileged access interviewers: methodological and practical issues[J]. *Addiction*, 1993, 88(12): 1617-1626.
- [5] Duncan D F, White J B, Nicholson T. Using Internet-based surveys to reach hidden populations: case of nonabusive illicit drug users[J]. *American journal of health behavior*, 2003, 27(3): 208-218.
- [6] Frank O, Snijders T. Estimating the size of hidden populations using snowball sampling[J]. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-*, 1994, 10: 53-53.
- [7] Rao A, Stahlman S, Hargreaves J, et al. Sampling key populations for HIV surveillance: results from eight cross-sectional studies using respondent-driven sampling and venue-based snowball sampling[J]. *JMIR public health and surveillance*, 2017, 3(4): e72.
- [8] Lu X, Bengtsson L, Britton T, et al. The sensitivity of respondent-driven sampling[J]. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2012, 175(1): 191-216.
- [9] Handcock M S, Gile K J, Mar C M. Estimating hidden population size using respondent-driven sampling data[J]. *Electronic journal of statistics*, 2014, 8(1): 1491.
- [10] Liu C, Lu X. Analyzing hidden populations online: topic, emotion, and social network of HIV-related users in the largest Chinese online community[J]. *BMC medical informatics and decision making*, 2018, 18(1): 1-10.
- [11] Rice E, Tulbert E, Cederbaum J, et al. Mobilizing homeless youth for HIV prevention: a social network analysis of the acceptability of a face-to-face and online social networking intervention[J]. *Health education research*, 2012, 27(2): 226-

236.

- [12] Young S D, Jaganath D. Online social networking for HIV education and prevention: A mixed methods analysis[J]. Sexually transmitted diseases, 2013, 40(2).
- [13] Ghani N A, Hamid S, Hashem I A T, et al. Social media big data analytics: A survey[J]. Computers in Human Behavior, 2019, 101: 417-428.
- [14] Oussous A, Benjelloun F Z, Lahcen A A, et al. Big Data technologies: A survey[J]. Journal of King Saud University-Computer and Information Sciences, 2018, 30(4): 431-448.
- [15] Kim D, Seo D, Cho S, et al. Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec[J]. Information Sciences, 2019, 477: 15-29.
- [16] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.

Full Research Paper

Investigating the Time-varying Effect of Search Index in Predicting Tourism Volume Using Dynamic Model Averaging

Yuan Chen^{1,2}, Yiping Ma^{1,2}, Zhongyi Hu^{1,2,*}, Tao Xiong³, Jiang Wu^{1,2}, Tingshun Xiong¹

¹School of Information Management, Wuhan University, Wuhan, 430072, China

²Center for Ecommerce Research and Development, Wuhan University, Wuhan, 430072, China

³College of Economics and Management, Huazhong Agricultural University, Wuhan, 430070, China

Abstract: Search index has been gradually used for tourism volume forecasting. While most literature aimed to improve the accuracy of a tourism prediction model, rarely considered the influencing characteristics of the search index in the prediction model. In this study, a dynamic model averaging (DMA) approach is applied to build tourism volume model and to investigate the time-varying effect of search indexes in predicting tourism volume. The tourism volume of Jiuzhai Valley and related search indexes from Baidu Index are considered for experimental purpose. According to the results, search index present time-varying characteristic, specifically, the search indexes of *Jiuzhai Valley hotel*, *Jiuzhai Valley*, *Jiuzhai Valley map* and *Jiuzhai Valley guide* have higher probabilities in supporting the prediction of tourism volume.

Keywords: tourism forecasting, search index data, time-varying effect, dynamic model averaging

1. INTRODUCTION

Tourism volume is an essential indicator of the tourism industry's stability and development. The accurate forecasting of tourism volume can help to allocate resources and formulate pricing strategies effectively [1]. Therefore, both tourism practitioners and government regulators pay much attention to the prediction of tourism volume.

Until now, many scholars have investigated tourism prediction and paid a lot of efforts to improve prediction accuracy. In terms of datasets, various online data is put forward to overcome the lag of simple historical tourist data, like search index and web traffic data [2,3]. Search index, which is produced by search engine companies, is a kind of index by integrating Internet users' searching query data in the search engine (e.g., Baidu, Google, and 360 search). Considering the widely use of search engines, the search index is useful in capturing Internet users' search behavior. In the area of tourism, many tourists would like to use search engines to search for related information about destinations to make travel plans. Therefore, the search index about a place to some extent reflects users' interests and intention to visit. Recently, it has been an emerging trend in using the search index for tourism forecasting [4].

To demonstrate the effectiveness of search engine data in forecasting tourist volume, many studies have applied search index in their prediction models. For example, Bangwayo-Skeete and Skeete^[5] compared simple Autoregressive (AR) method and the Seasonal Autoregressive Integrated Moving Average (SARIMA) method of historical tourist arrivals data with AR method using high-frequency Google search queries data. The results showed that AR model with Google search queries data gave superior forecasts to AR and SARIMA time series models. Apart from Google Trends, Baidu Index is widely adopted when the study is about China, and it turns out that model with Baidu index performs well [6,7]. Sun et al. built a prediction model for tourist arrivals of popular destinations in China using search indexes from Google and Baidu [3]. The experimental results verified the Granger causality and co-integration relationship between search indexes and tourist arrivals. The

* Corresponding author. Email: Zhongyi.hu@whu.edu.cn (Zhongyi Hu)

forecasting accuracy was also significantly improved while applying search indexes for predicting tourist arrivals. However, the existing studies mainly focused on the effectiveness of the search index in predicting tourism volume; the question about how the search index was involved in the prediction model remains unanswered. Especially, the coefficients of the search index in the prediction model can change over time [8,9].

To handle these issues, dynamic model averaging (DMA) approach was proposed in which both forecasting models and the coefficients of predictors can change over time. Raftery et al. [10] proposed a DMA model to forecast the output strip thickness for a cold rolling mill and pointed out that DMA performed better than single models. As a special case, if only the highest probability model at each point of time is selected to form the forecasting model, it turns to be a dynamic model selection (DMS) approach. DMA/DMS approach has been widely applied in different forecasting fields including housing market [11], stock market [12], and fossil fuels prices [13]. It turns out that the DMA/DMS approach not only has better forecasting performance but also identifies different factors' influence on prediction. As for tourism forecasting area, with the influencing factors being diverse, DMA method seems suitable to improve the forecasting performance and figure out factors' impact of future tourism arrivals. However, to our knowledge, among the existing studies on tourism prediction, while most literature focused more on improving forecasting accuracy, there has been no literature that considers the dynamically changing coefficients of predictors.

To fill the above research gap, this paper applies the DMA approach to forecast tourists' arrivals of Jiuzhai Valley with search index and historical tourist volumes. The predictors are selected from Jiuzhai Valley-related search indexes according to the Pearson correlation coefficient. The time-varying effect of search index is then demonstrated based on DMA. This paper contributes to the existing tourism prediction literature in three ways. First, we introduce the DMA approach for tourism forecasting, in which the forecasting model and coefficients of predictors are dynamically updated over time. To date, this is the first study that applies DMA in tourism forecasting. Second, this study investigates the time-varying effect of search index in predicting tourist volume. There are few, if any, studies have focused on the influencing of search index on tourist volume. Third, by taking Jiuzhai Valley as a case study, this study shows some interesting findings of the time-varying influences between search index and tourist volume.

The rest of this study is organized as follows. The methodology of the DMA approach is explained in Section 2. Section 3 describes datasets and some preliminary analysis. Section 4 presents the experimental results and discussions. Finally, we conclude this study with some future directions in Section 5.

2. METHODOLOGY

Allowing the regression coefficients and forecasting models to change over time simultaneously, the DMA framework suits the situation well when the optimal forecasting model evolves over time.

To illustrate the DMA framework, we suppose a set of K forecasting models with the subsets of z_t being predictors. And y_t refer to tourism volume at time t . The set of forecasting models can be written as follows:

$$y^t = z_t^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)} \quad (1)$$

$$\theta_{t+1}^{(k)} = \theta_t^{(k)} + \eta_t^{(k)} \quad (2)$$

where $z_t^{(k)}$ denote the subset of predictors for $k = 1, 2, \dots, K$, and errors $\varepsilon_t^{(k)} \sim N(0, H_t^{(k)})$ and $\eta_t^{(k)} \sim N(0, Q_t^{(k)})$

If z_t contains m predictors, the number of possible forecasting models comes to be $K = 2^m$, and at each point of time the models are different. At time $t - 1$, let $\pi_{t|t-1,k}$ denote the probability of model K being the optimal forecasting model. DMA method gets the final forecasting result by averaging forecasting values with $\pi_{t|t-1,k}$ as weights at each time point (see Equation 3). When the final forecasting value consists of the best performance model at each time point, it becomes DMS. Let $\pi_{t|t-1}^*$ denotes the maximum value of the probability of model selection at each time point, and the equation is shown as follows.

$$\hat{y}_t^{DMA} = \sum_{k=1}^K \pi_{(t|t-1,k)} \hat{y}_t^{(k)} \quad (3)$$

$$\hat{y}_t^{DMS} = \sum_{k=1}^K \pi_{(t|t-1)}^* \hat{y}_t^{(k)} \quad (4)$$

Therefore, the next step is to estimate the coefficients and probability of each forecasting model at each time point. By following Kalman filtering^[14], we can get Equation (5), where $\Sigma_{t|t-1}^{(k)} = \Sigma_{t-1|t-1}^{(k)} + Q_t^{(k)}$. After updating the above equation according to Kalman filter, Equation (6) can be used for recursive forecasting.

$$\begin{aligned} \theta_{t-1}^{(k)} | y^{t-1} &\sim N(\hat{\theta}_{t-1}^{(k)}, \Sigma_{t-1|t-1}^{(k)}) \\ \theta_t^{(k)} | y^{t-1} &\sim N(\hat{\theta}_{t-1}^{(k)}, \Sigma_{t|t-1}^{(k)}) \end{aligned} \quad (5)$$

$$y_t | y^{t-1} \sim N(x_{t-1} \hat{\theta}_{t-1}^{(k)}, H_t^{(k)} + x_{t-1} \Sigma_{t|t-1}^{(k)} x_{t-1}') \quad (6)$$

To estimate $Q_t^{(k)}$, a forgetting factor approach^[10] is applied, as shown in Equation (7):

$$\Sigma_{t|t-1}^{(k)} = \frac{\Sigma_{t-1|t-1}^{(k)}}{\lambda} \quad (7)$$

or $Q_t^{(k)} = (1 - \lambda^{-1}) \Sigma_{t-1|t-1}^{(k)}$, where $0 < \lambda \leq 1$. The forgetting factor λ^m indicates the weight of the observation value in the past m period. A small forgetting factor value implies a rapid changing of the coefficient.

Due to the fact that Markov switching process has quite heavy calculation burden when estimating $\pi_{t|t-1,k}$, we apply a forgetting factors approach^[10] and α is the forgetting factor:

$$\pi_{t|t-1,k} = \pi_{t|t-1,k}^\alpha / \sum_{l=1}^K \pi_{t|t-1,l}^\alpha \quad (8)$$

where $0 < \alpha \leq 1$, and the interpretation can be similar to λ .

Finally, the estimate of $H_t^{(k)}$ uses the Exponentially Weighted Moving Average (EWMA) approach as follows^[8]:

$$\hat{H}_t^{(k)} = \sqrt{(1 - \kappa) \sum_{j=1}^t \kappa^{j-1} (y_j - z_j^{(k)} \hat{\theta}_j^{(k)})^2} \quad (9)$$

where κ is a decay factor. The EWMA specification can get volatility forecasts approximation by a recursive form.

3. DATA PREPARATION

To verify the effectiveness of DMA approach with search indexes, this study takes Jiuzhai Valley's tourist volume and search index data as a case study. This section describes the dataset with preliminary analysis.

(1) Data description

Jiuzhai Valley, locally known as Jiuzhaigou, is a popular tourist spot in Sichuan province of China. Its outstanding view has attracted millions of tourists per year. Unfortunately, there was an earthquake in August 8, 2017, which resulted in shut down of the place until September of 2019. Therefore, the daily tourist arrival data, ranging from May 27, 2012 to August 7, 2017 (with 1899 observations), is considered as the explained variable. As for search index, according to the results of the CNZZ data center (www.cnzz.com), Baidu shares the largest market among the search engines in China. Thus, this paper collects search indexes about Jiuzhai Valley from Baidu Index (<http://index.baidu.com/>) with the same time period of tourist volume.

(2) Keywords selection

To select reasonable and comprehensive search indexes, it is vital to contain diverse and most related

searching keywords about Jiuzhai Valley. First, we set 12 initial keywords of six different aspects including eating, lodging, traffic, scenic spot, shopping and entertainment. Second, a keyword mining tool (<https://www.chinaz.com/>) was used to extend relevant keywords and finally resulted in 56 keywords. By collecting search indexes from Baidu Index based on each keyword, we obtained 35 keywords' search indexes, whereas the others were not tracked by Baidu Index. To filter out those search indexes that are more relevant to tourist volume, we further calculated the Pearson correlation coefficient between each keyword and tourist volume. Finally, eight keywords with the correlation coefficient above 0.65 were selected as final predictors, which are listed in Table 1. As shown in this table, the selected keywords are all highly relevant to travelling of Jiuzhai Valley. Figure 1 illustrates the trend of several search indexes and tourism volume. From this figure, we can find that these search indexes have a similar trend as tourism volume.

Table 1. Selected keywords (translated from Chinese)

Keywords	Correlation coefficients
Jiuzhai Valley map (九寨沟地图)	0.7794
Jiuzhai Valley Huanglong (九寨沟黄龙)	0.7521
Jiuzhai Valley guide (九寨沟攻略)	0.7408
Jiuzhai Valley scenic spot (九寨沟景点)	0.7164
Jiuzhai Valley hotel (九寨沟酒店)	0.7066
Jiuzhai Valley (九寨沟)	0.6743
Jiuzhai Valley scenic map (九寨沟景区地图)	0.6670
Jiuzhai Valley free travel (九寨沟自由行)	0.6541

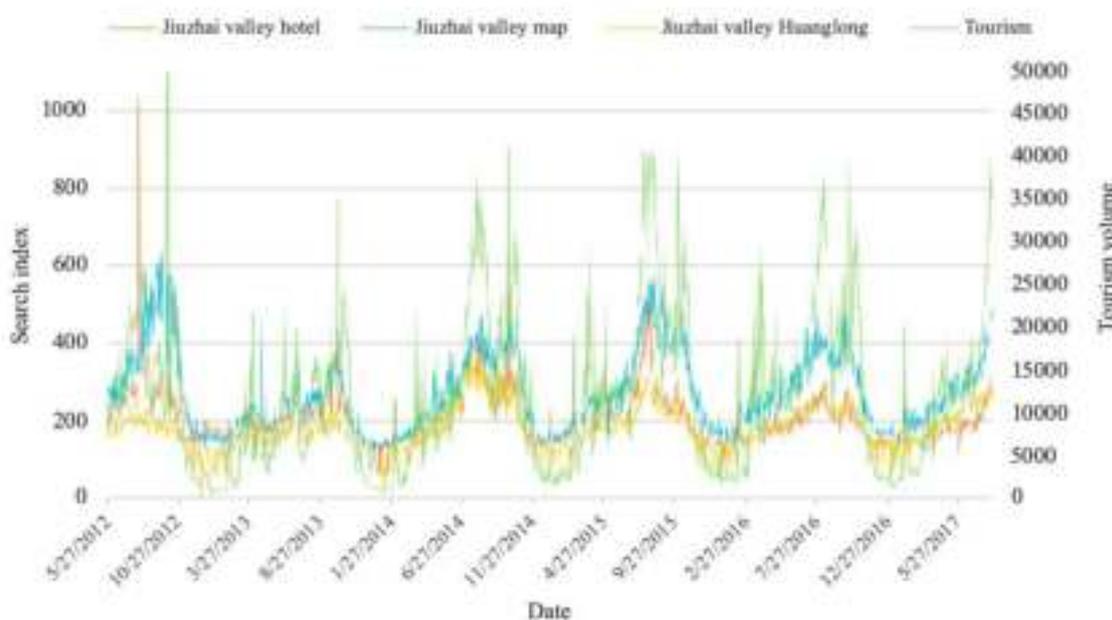


Figure 1. Visualization of the search index and tourism volume

(3) Stationary test

The augmented Dickey-Fuller (ADF) test method is first applied to examine the stationarity of both explanatory and explained variables, and the test results are shown in Table 2. It can be concluded that the original series of some variables like *Tourist volume*, *Jiuzhai Valley*, *Jiuzhai Valley hotel* are stationary, which do not need further transformation. On the other hand, the original non-stationary series like *Jiuzhai Valley map* and *Jiuzhai Valley scenic spot*, are stationary after transformation by taking their first order difference.

Table 2. Results of variables' stationary test

	Variables	ADF (original series)	ADF (first order difference series)	T code
Explained Variables	<i>Tourist volume</i>	0.003***	--	0
	<i>Jiuzhai Valley hotel</i>	0.001***	--	0
Explanatory Variables (predictors)	<i>Jiuzhai Valley map</i>	0.013	0.000***	1
	<i>Jiuzhai Valley free travel</i>	0.056	0.000***	1
	<i>Jiuzhai Valley guide</i>	0.655	0.000***	1
	<i>Jiuzhai Valley scenic spot</i>	0.091	0.000***	1
	<i>Jiuzhai Valley scenic map</i>	0.090	0.000***	1
	<i>Jiuzhai Valley Huanglong</i>	0.101	0.000***	1
	<i>Jiuzhai Valley</i>	0.000***	--	0

Note: ADF: the p-value statistics from the augmented Dickey-Fuller unit root tests. The asterisks ***: rejections of the null hypothesis at 1% significance level. T code: transformation code, 0 means no transformation; 1 means transformation with the first-order difference.

4. EXPERIMENTAL RESULTS

This section presents experimental results using DMA model with the aforementioned eight selected explanatory variables.

4.1 The average number of predictors in DMA

With the DMA approach allowing explanatory variables to change over time, the numbers of explanatory variables may be different at each time point according to the forecasting performance. The average number of predictors at each time point is presented in **Figure 2**. From this figure, it is clear that the average number of explanatory variables varies over time. At the beginning of 2012, less than four explanatory variables are used, which are under half of the whole eight alternative predictors. Then the number rises over four by the end of 2012. And from then on, the fluctuation about the average number of predictors at each time point is relatively stable, mainly between 3 to 6. Overall, the average number of predictors is 4.17, which is less than eight alternative predictors. The results indicate that not all the search indexes are useful in predicting tourism volume, and the DMA can select good predictors effectively.

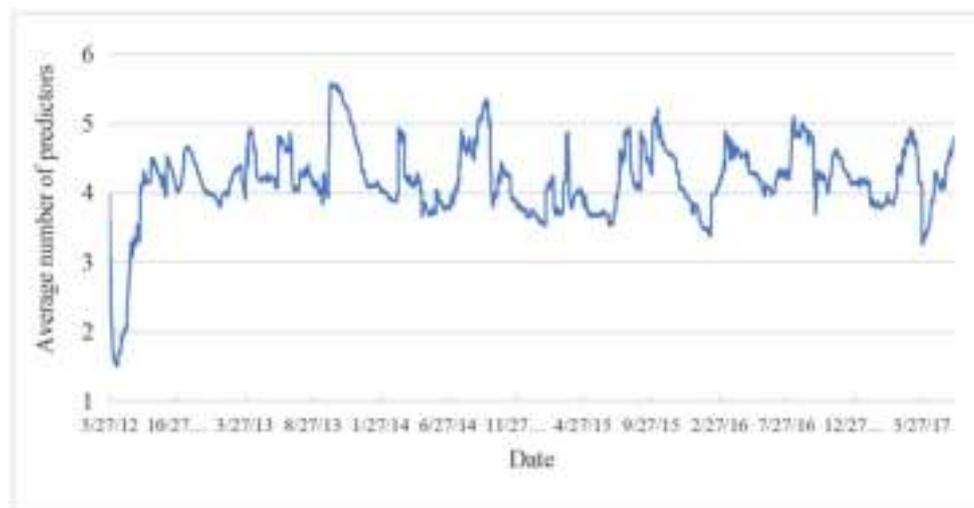


Figure 2. The average number of predictors in DMA

4.2 The number of predictors for the optimal forecasting model

According to the selection probability of the prediction model at each time point, the prediction model with the highest selection probability can be referred to as the optimal prediction model at each time point. The number of predictors for the optimal prediction model at each time point is presented in Figure 3. As indicated in the figure, most optimal models use only 2.8 predictors on average from 8 potential predictors. Besides, only few optimal models have more than six explanatory variables. The results further provide evidence that, at different time points, the search index variables play different roles. By applying the DMA approach, the time-varying effect is likely to be considered, resulting in less explanatory variables in the prediction modelling and less computation cost.

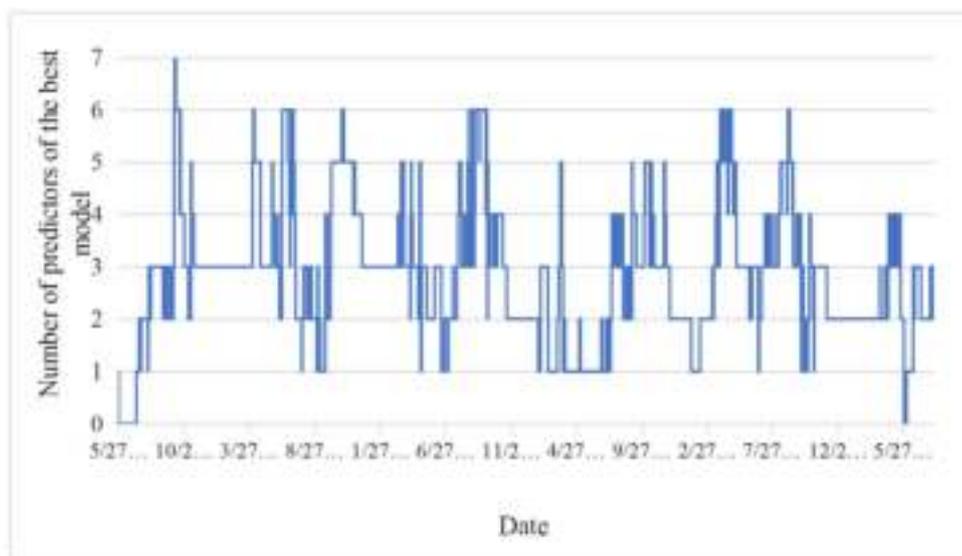


Figure 3. The number of predictors for the optimal forecasting model

4.3 Inclusion probability of predictors in DMA modelling

The inclusion probability of a predictor indicates the sum probabilities of the predictor employed among all potential predictors in the forecasting model. A predictor with high probability presents its strong prediction power for tourism volume. The probabilities of the eight predictors that were selected in the forecasting model are shown in Figure 4.

As Figure 4 presents, the predictors for tourism volume forecasting vary over the whole time period. At the year of 2012, *Jiuzhai Valley hotel* and *Jiuzhai Valley* present higher forecasting power than the other predictors with the inclusion probability over 80%. Then the inclusion probability of *Jiuzhai Valley hotel* drops rapidly. The forecasting power of *Jiuzhai Valley free travel* is at a low level from 2012. But around 2014 the inclusion probability of *Jiuzhai Valley free travel* rises to over 70% while at the same time *Jiuzhai Valley scenic spot*, *Jiuzhai valley guide*, *Jiuzhai Valley map*, *Jiuzhai Valley Huanglong* are at a low forecasting level. Before the year 2014, *Jiuzhai Valley map* appears an upsurge but gradually lose its forecasting power over time. After that, *Jiuzhai Valley scenic spot*, *Jiuzhai Valley guide*, *Jiuzhai Valley map* and *Jiuzhai Valley Huanglong* have similar fluctuations. At the middle of the year 2014, the inclusion probability of *Jiuzhai Valley scenic spot* increases significantly (up to 90%). Thus, *Jiuzhai Valley scenic spot* overtakes other predictors to become one of the best predictors. Generally speaking, *Jiuzhai Valley Huanglong*, *Jiuzhai Valley scenic spot*, *Jiuzhai Valley scenic map* do not show strong prediction power during the whole time period with the average inclusion probability around 42%. *Jiuzhai Valley hotel* and *Jiuzhai Valley* have the highest average inclusion probability of 78.4% and 76.1%, which overtakes the others. Therefore, *Jiuzhai Valley hotel* and *Jiuzhai Valley* show superior forecasting power to tourism volume.

In conclusion, search indexes do have time-varying effect in predicting tourist volume of Jiuzhai Valley.

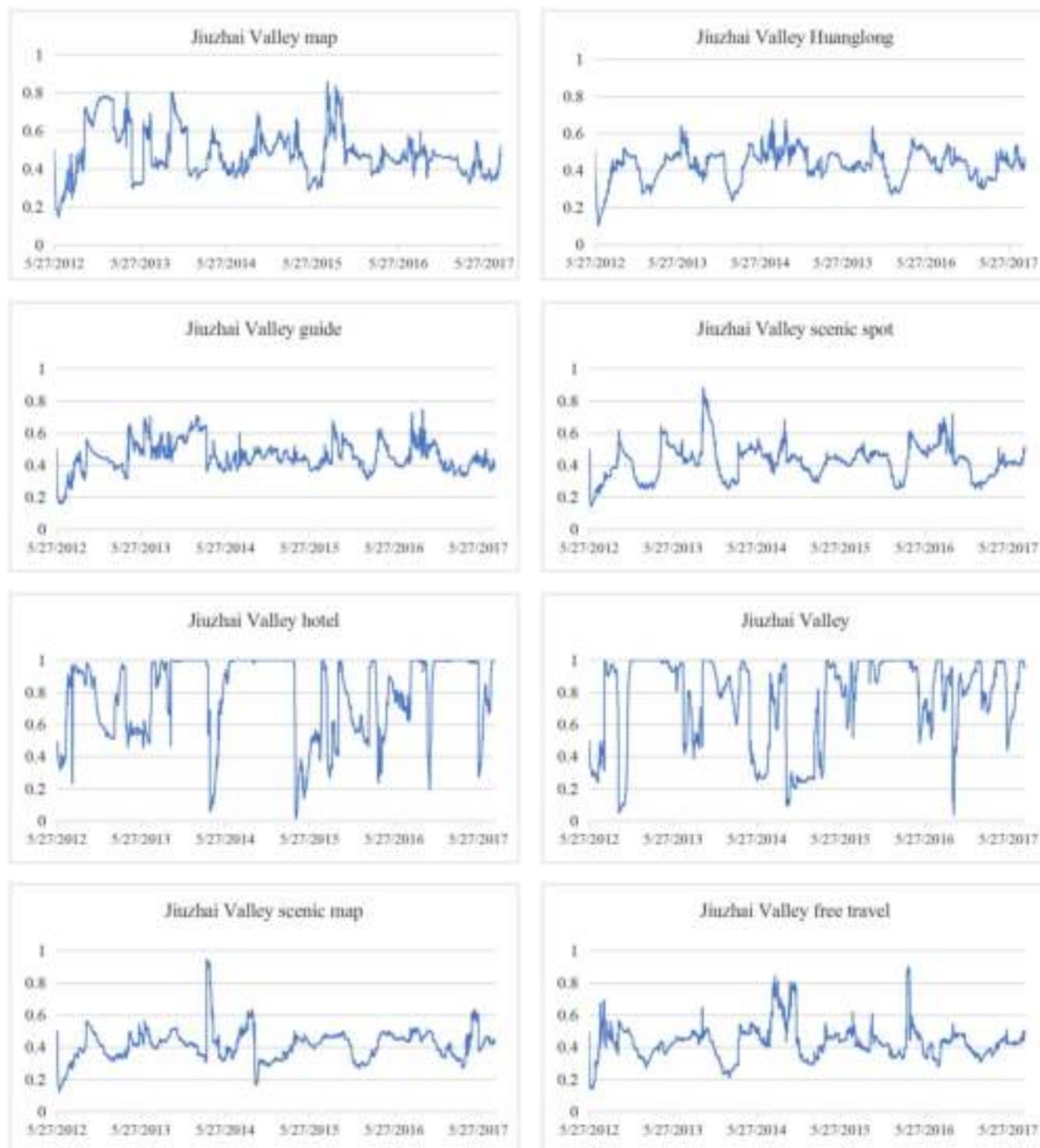


Figure 4. Inclusion probability of predictors

5. CONCLUSIONS

Although previous studies have provided sufficient evidence that the search index can improve the prediction of tourism volume, few studies investigated the time-varying effect of search index in building prediction models of tourism volume. In this study, we fill this gap by proposing a DMA-based model and exploring the variation of search indexes in predicting tourism volume of Jiuzhai Valley. Based on the analysis of the average number of predictors in DMA and the optimal prediction models, and the inclusion probability of

predictors at different time period, we find that these predictors have time-varying effect in predicting tourism volume of Jiuzhai Valley. In particular, some keywords-related search index such as *Jiuzhai Valley hotel*, *Jiuzhai Valley*, *Jiuzhai Valley map* and *Jiuzhai Valley guide* show higher probabilities in supporting the prediction of tourism volume. This study provides both scholars and industry practitioners a clear understanding of relations between search indexes and tourism volume.

In the future, this study will include more relevant search indexes and explore the forecasting power of more predictors. The prediction performance of DMA-based model and other counterparts will also be considered for further comparisons.

ACKNOWLEDGEMENT

This research was supported by the National Key Research and Development Program of China under Grant 2019YFB1405600 and the National Natural Science Foundation of China under Grants 71601147 and 71771101.

REFERENCES

- [1] Song H, Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. *Tourism Management*, 29 (2): 203–220.
- [2] Emili S, Figini P, Guizzardi A. (2020). Modelling international monthly tourism demand at the micro destination level with climate indicators and web-traffic data. *Tourism Economics*, 26 (7): 1129–1151.
- [3] Sun S, Wei Y, Tsui, K.-L, Wang S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management*, 70: 1–10.
- [4] Liu H, Liu Y, Wang Y, Pan C. (2019). Hot topics and emerging trends in tourism forecasting research: A scientometric review. *Tourism Economics*, 25 (3): 448–468.
- [5] Bangwayo-Skeete P.F, Skeete R.W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46: 454–464.
- [6] Huang X, Zhang L, Ding Y. (2017). The Baidu Index: Uses in predicting tourism flows -A case study of the Forbidden City. *Tourism Management*, 58: 301–306.
- [7] Wen L, Liu C, Song H. (2019). Forecasting tourism demand using search query data: A hybrid modelling approach. *Tourism Economics*, 25 (3): 309–329.
- [8] Koop G, Korobilis D. (2012). FORECASTING INFLATION USING DYNAMIC MODEL AVERAGING. *International Economic Review*, 53 (3): 867–886.
- [9] Wei Y, Cao Y. (2017). Forecasting house prices using dynamic model averaging approach: Evidence from China. *Economic Modelling*, 61: 147–155.
- [10] Raftery A.E, Kárný M, Ettl P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52 (1): 52–66.
- [11] Bork L, Møller S.V. (2015). Forecasting house prices in the 50 states using Dynamic Model Averaging and Dynamic Model Selection. *International Journal of Forecasting*, 31 (1): 63–78.
- [12] Risse M, Ohl L. (2017). Using dynamic model averaging in state space representation with dynamic Occam's window and applications to the stock and gold market. *Journal of Empirical Finance*, 44: 158–176.
- [13] Naser H, Alaali F. (2018). Can oil prices help predict US stock market returns? Evidence using a dynamic model averaging (DMA) approach. *Empirical Economics*, 55 (4): 1757–1777.
- [14] Kalman R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82 (1): 35–45.

Short Research Paper

Research Progress of Tie-Generative Mechanism in Network Based on ERGM

Bo Jia¹, Jiahua Jin^{2*}, Zhiyuan Ge³, Xiangbin Yan⁴

^{1, 2, 4}School of economics and management, Beijing University of science and technology, Beijing, 100083, China

³School of economics and management, Beijing University of technology, Beijing, 100124, China

Abstract: The research on the tie-generative mechanism in network is conducive to explore the factors affecting network evolution and provide theoretical supports and decision-making suggestions for promoting (restraining) the formation (disappearance) of network edges. Based on the analysis of the current situation and development trend of ERGM from 2015 to 2020, this paper presents the research progress and limitations of ERGM. It has made good progress in the fields of network dynamic evolution mechanism, capturing network node heterogeneity, multi-layer network formation mechanism and Research on large (small) scale network formation mechanism. However, it must be admitted that the problems such as extending ERGM to obtain connection weighted information, modeling the heterogeneity of nodes in the network, detecting the multicollinearity that may exist in ERGM, studying the dynamic evolution mechanism of multi-layer network and using ERGM to deal with network missing data have not been well solved, waiting for further exploration.

Keywords: ERGM, tie-generative mechanism, social network, research progress

1. INTRODUCTION

Social network analysis radically changes quantitative analysis by shifting the focus from individuals to their relationships and interactions^[1]. The traditional regression model is often used for social network analysis, which is based on the assumption of independence and unable to consider endogenous structural effect of the network^[2]. However, this effect, that is, the relationship between nodes can be formed by the self-organization of the network, will affect the probability of edge generation^[3]. Taiye Luo and Cuichang Ma also pointed out the formation of social network is the result of the joint influence of endogenous structure and exogenous variables.^[4] Therefore, the relevant conclusions obtained from the traditional regression model which can't take the endogenous structure effect into account will be biased to some extent. Compared with the limitations of the independence assumption set by the traditional regression model for the observation object, the advantage of ERGM is that it can simultaneously consider endogenous structure and exogenous variables of the network to more comprehensively study tie-generative mechanism and evolution process of the network^[5], which makes the analysis based on ERGM more rigorous and reliable. Guancan Yang, Tong Liu et al.^[6] pointed out that ERGM can be well applied to analyze the influencing factors of citation relationship formation. Ghosh A , Ranganathan R also recognized ERGM as a network analysis tool in their research and stated ERGM should be an important part of the standard toolkit for future network tie-generative mechanism research.^[7]

Understanding the formation mechanism of network could provide theoretical supports and decision-making suggestions for promoting (restraining) the generation of edges. Due to the advantages of ERGM, many scholars have applied ERGM to the research of network tie-generative mechanism in recent years. Helian Xu, Tianyang Sun et al.^[8] studied the influence of endogenous structural variables, actor attribute variables and network covariables on the formation of high-end manufacturing trade network along the "One Belt and One Road" through ERGM. Song H discussed the factors affecting the generation of personal political

* Corresponding author: Jiahua Jin(1985.12 -), Email: jinjiahua@ustb.edu.cn

discussion network's edges by using ERGM. ^[9] Anna Llupià, Puig J et al. used ERGM to evaluate the role of homophily and individual characteristics in social network. ^[10] Dang-Pham D, Pittayachawan S et al. also explored the formation mechanism of security consulting network by using ERGM. ^[11] Wenlong Yang and Debin Du studied the factors that affected the formation of investment network by comprehensively considering endogenous structural effect, actor-relationship effect and binary covariates. ^[12] Moreover, Guancan Yang, Zhanlin Liu et al. also analysed the tie-generative mechanism of the Nelarabine Drug patent citation network by using ERGM. ^[2]

However, although ERGM provides a series of flexible and highly scalable methods for network analysis, there is still a lack of an ERGM method that can take into account unobserved heterogeneity. [13] In addition, ERGM based on binary networks can't be used to simulate networks with weighted edges, which leads to the model unable to capture the network information of link strength between different nodes. [14] ERGM can only analyze static or cross-sectional data. The MCMCML algorithm for ERGM estimation and verification is very complex, which leads to the limited network scale that the model can simulate. This is consistent with the difficulty of directly estimating the ERGM parameters of very large scale networks proposed by Stivala A and Robins G. [15] These are new challenges that ERGM faces. From the perspective of literature, this paper will summarize the relevant concepts and development history of ERGM, as well as the present situation and future prospect of the research on the mechanism of network formation carried out by ERGM, so as to provide reference for future studies.

2. THE DEVELOPMENT OF ERGM AND RELATED CONCEPTS

The development of ERGM (Exponential Random Graph Model) can be traced back to the simple random graph model proposed by Erdos P and Renyi A ^[16] in 1959. The model contains the assumption that the relationships among network members are generated independently of the relationships among other members. Although the simple random graph model can not grasp the structural characteristics of the observed network well, it can provide a baseline for the comparison of other more complex models. In 1981, Holland P W and Leinhardt S developed a binary independent model to estimate the differences caused by reciprocation and differential attractiveness. ^[17] After that, Frank O, Strauss D introduced Markov dependence assumption into the model in 1986, and put forward a binary dependency model which assumes edges containing the same node are not independent of each other. ^[18] In 2006, Davidrh and Marksh extended the Markov model into p^* model, which has a broader conditional dependence relationship, that is, the probability of the simultaneous existence of any two lines in the network is not equal to the probability combination of the respective existence of two lines. ^[19] Most ERGMs based on dependency assumptions can be regarded as p^* models. However, p^* model still has a series of problems such as the problem of degeneration. In order to solve these problems, Hunter D R put forward a method including GWD (Geometrically Weighted Degree Distribution), GWESP (Geometrically Weighted Edge Shared Partners), GWDSP (Geometrically Weighted Dyadwise Shared Partners) and other statistical items replace the complex structure and dependency conditions in observation network. ^[20] The current p^* model, known as ERGM, has been able to incorporate high order dependence conditions into the analysis framework. In 2013, Wang P, Robins G extended ERGM to multilevel networks and demonstrated it with the collaboration network of the French Institute for Cancer Research. ^[21] Since then, multilevel networks have attracted the attention of many scholars. Brennecke J, Rank O studied the influence of enterprise knowledge network on inventor interaction network through the multi-level network that based on ERGM. ^[22] Smith M, Gorgoni S pointed out that a complete multi-level network includes networks at the micro, meso and macro levels, and applied multi-level ERGM to the study of complex interactions between activities at the enterprise level and international trade patterns. ^[23] At the same time, with the deepening of network research,

academics are no longer satisfied with only statically studying the formation mechanism of networks. As proposed by Linqing Liu and Ziruo Chen, existing studies are mainly completed by analyzing static or cross-sectional network data, which makes it difficult to effectively reveal the dynamic evolution mechanism of the network.^[24] People hope to analyze the evolution process of network dynamically so as to understand the formation mechanism of network more deeply. In 2014, Krivitsky P N , Hancock M S extended ERGM to STERGM (Separable Temporal Exponential Random Graph Mode).^[25] STERGM is able to distinguish between newly created and previously existing relationships in the network, and longitudinally study the evolution process of the network.^[26] Bjorklund P, Daly A J used ERGM and STERGM to study the influence of homogeneity and proximity dimension on the generation of social connections in pre-service teacher identification network.^[27] At present, ERGM is still in the period of rapid development, and its flexible expansibility endows it with strong vitality. As Peng, Tai-Quan said, ERGM has a bright future.^[28]

ERGM is a dynamic network model based on network binary relations. Different from traditional measurement models, ERGM emphasizes more on the dependence of relationships in the network^[14]. According to the observation network, ERGM generates a random network graph, and through the steps of estimation, diagnosis, simulation, comparison and improvement, the generated network is more and more close to the observation network, so as to check which factors significantly affect the generation of the network^[8].

As an innovative statistical inference method, ERGM allows a variety of deformation and expansion, and can include multiple factors that may affect network formation, including exogenous node attributes, endogenous structural effects and binary covariates, into the model^{[2],[3]}. King S, Lusher D et al. used ERGM to study the influence of endogenous network features, attribute based exogenous features and geographical proximity dimensions on network formation.^[29] Among them, exogenous node attribute refers to that the probability of linking between two nodes is affected by the node attributes itself. When the nodes have a certain attribute (refers to binary variables) or the larger the value of a node's attribute (refers to continuous variables), the higher the probability of link between nodes. The sender effect, receiver effect, homogeneous effect and heterogeneous effect all belong to node attribute effect. The sender effect refers to possess some properties of nodes are more likely to send a link to other nodes, correspondingly, the receiver effect refers to a certain attribute nodes are more likely to accept links from other nodes. Homophily (heterogeneity) means that the probability of forming a connection between nodes with the same (different) property is higher than the probability of forming between two nodes selected at random. In addition, as a complex system, the emergence of some relationship come from the internal process of the network relational system. Such self-organizing effects are usually called "endogenous structural effects". Cao Q, Liao L et al. pointed out that the social network among the residents of the treatment community of women was mainly generated from the endogenous structural factors of the treatment community itself.^[30] The transitive effect, connectivity effect, preferential attachment effect and sparsity effect all belong to the self-organizing effect of network. The transitive effect refers to the influence of transitive closure structure on the formation of relationships in a network and the connectivity effect refers to the influence of 2-path structure. Preferential attachment can be divided into two types: convergence and expansion which refers to the influence of the in-degree (out-degree) distribution of network nodes on the relationship formation. The research on the effects of endogenous structure is indispensable, because it's influence is stable and strong, and far beyond expectations.^[31] Network synergistic effect refers to that in addition to generative node variables and endogenous structure, duality attributes can also affect the generation of the network. For example, the similarity of patent content will affect the generation of patent reference relationship.

3. RESEARCH STATUS AND LIMITATION OF TIE-GENERATIVE MECHANISM IN NETWORK USING ERGM

3.1 Research on the formation mechanism of different topic networks.

From 2016 to 2020, there are a lot of researches on the formation mechanism of different topic networks including transnational student exchange network [33], terrorist organization alliance network [34], strategic literature alliance network [6], healthy community network based on users' replies [35], patent technology diffusion network [31], patent collaborative innovation network [4], academic social network [36], patent citation network [37], scientific cooperation network [38], pyramid selling organization network [39] and the urban network based on the perspective of China's top 100 electronic information companies [14].

In addition to discussing tie-generative mechanism in network from common perspectives such as reciprocation, preferential attachment and preferential selection, some scholars have also studied from the angles of similarity and social factors. Qingfeng Duan and Xiaohuan Pan discussed the influence of similarity in social attributes of literature on citation preference. [40] In addition, Lee S K , Kim H [41], Anna Llupià, Puig J [10] also pointed out that not only the purely academic factors, but also many complex and diverse factors such as political background, economic level, cultural atmosphere, individual psychology and intelligence may affect the formation of the network, which should be paid attention to in future research.

3.2 Studying the network tie-generative mechanism longitudinally.

In order to explore the evolution mechanism of the complex network, Peng, Tai-Quan [28] divided the journal citation network into four time series and skillfully studied the evolution of the network. Xiaoyan Liu, Jinpeng Li et al. also used a similar method to explore the evolution mechanism of technology trading network in the integrated circuit industry. [42] In 2014, Krivitsky P N , Hancock M S extended ERGM to STERGM. [25] The emergence of STERGM makes it more convenient to analyze the mechanism of network formation longitudinally. Linqing Liu and Ziruo Chen studied dynamic evolution mechanism of Chinese dominant industrial combination by TERGM. [43] Xiaoyan Liu and Jing Wang also discussed the different influences of exogenous node attributes and endogenous structure on the formation and dissolution of cooperation in different growth stages of OLED technology innovation network based on TERGM. [44]

3.3 Exploring network formation mechanism through multi-level ERGM.

The connections of different organizational levels in the network are not completely independent. They are interdependent at all levels and affect each other in a complex way. [45] The networks of different organizational levels and the connections between them constitute a multi-level network. The complex causal relationship in the multi-level network is not simply one-way. For example, the long-term transaction network between companies affects the relationship between companies, which in turn will bring new business opportunities and constraints to their companies. [46] Liu Xuan, Wang Linwei and others also pointed out in their research that there may be a great correlation among friend network, user access network and reply network in online health community, and it is of great significance to pay attention to the relationship and interaction between these networks. [35] ERGM was originally developed for single-level network. In 2013, Wang P , Robins G and others extended ERGM to multi-level network, and used the cooperative network of French cancer research elites and their affiliated institutions as an example to prove that a full understanding of the network requires cross level parameters. [21] This study was later seen as the beginning of multilevel network analysis using ERGM. Since multi-layer ERGM can capture the relationship that cannot be explored by a single network [47], and more fully reproduce the observation network, it has been widely concerned by the academic community since then. Brennecke J , Rank O used multi-level ERGM to research the influence of enterprises' knowledge network on inventors' interactive network [22], and Holloway J , Koskinen J also discussed how to embed bi-edge cluster into multi-level interdependent network system based on this model. [48] Smith M , Gorgoni S applied multi-level

ERGM to explore the complex interaction between enterprise-level activities and international trade. [23] In addition, Wang P, Robins G et al. also proposed SSM (social selection model) as an extension of multi-level ERGM in 2016 [49], this model integrates the information about nodes into the modeling framework, which may help to determine the degree of homogeneity and other attributes that may affect the affiliation within and between levels and the structure of the whole multilayer network, so that people can have a more detailed and complete understanding of the network structure and basic network process.

However, as a developing model, multi-level ERGM has many limitations. Multi-level ERGM can only deal with binary data, which means that the observation network fitted by multi-level ERGM loses the important information of edge weight. [23] This model can not simulate the dynamic evolution mechanism of observation network vertically, which is also one of the limitations of the model. As Brennecke J and Rank O mentioned, the future vertical research should investigate the co-evolution of social network and knowledge network within the enterprise. [22] In addition, Wang P., Robins G. also pointed out that they often find that the homogeneity assumption under ERGM may be too strong, especially for large empirical networks. [49]

3.4 Extending ERGM to capture the heterogeneity of network nodes.

To understand the "unobserved heterogeneity" and what impact this heterogeneity will have on the research, we use a passage from Box-Steffensmeier J M , Christenson D P [13] to explain:

"However, we may suspect that there are other, intangible factors specific to each individual that are difficult if not impossible to measure, such as "friendliness" or "charisma," that are also related to network structure (people that are friendlier are likely to have more friends, increasing the centrality of friendly individuals above what we would expect given their other, known attributes). In other words, the observed and measured characteristics are not sufficient for explaining the network we observe. Further, because these unobserved characteristics may be correlated with both the outcome (network structure) and the other explanatory variables, there is the potential for mistaken inferences when such heterogeneity is not accounted for."

In order to fit the observation network more accurately, researchers begin to study how to extend ERGM to capture the heterogeneity existing in the network. Thiemichen S , Friel N [50] ameliorated ERGM to capture the heterogeneity of network nodes. Henry T R , Gates K M [51] also modeled the unobserved heterogeneity. Box-Steffensmeier J M , Christenson D P [13] and Koskinen J , Wang P [52] all discussed how to extend ERGM to alleviate the dilemma caused by failure to model heterogeneity. In 2020, Henry T R , Gates K M [51] developed SRFM-ERGM (Sender/Receiver Finite Mixed Exponential Random Graph Model) that can capture heterogeneity.

3.5 Extending ERGM to accommodate networks of different sizes.

Monte carlo estimation of ERGM parameters is a computationally intensive process, which results in strict limits on the size of the network that ERGM can fit [53]. Due to the difficulty of parameter estimation, the practical application of this kind of model is limited to relatively small networks, up to several thousand nodes, usually only a few hundred nodes (such as online social networks) or fewer nodes (such as church social networks). [15] It is rarely applied to the research of small networks (6 or less nodes in directional networks) from team or home [54] and very large-scale networks.

Most of the researches on small networks are based on descriptive statistics. The main limitation of this work is the availability of network models. ERGM, which is often used to fit large and medium-sized networks, can not be used because the problem of maximum likelihood estimation is more obvious in small networks. [54] In addition, there are a series of problems in its application to the research of ultra large scale network. From a conceptual point of view, with the expansion of the network scale, the MCMLE (Monte Carlo maximum likelihood estimation) commonly used by ERGM requires that the nodes need to fully understand the

assumption that all other nodes are connected when looking for the nodes that establish the connection, which makes it impractical. Stivala A and Robinson G also pointed out that it is very difficult to directly estimate the parameters of ERGM for a very large network.^[15] From a technical point of view, a series of operations such as loading large-scale network data, estimation and testing have high requirements for computer memory, network analysis software and other related technologies. All of these restrict the application of ERGM in very large scale network.

Aiming at the problem, people have done some research. Yon G G V, Slaughter A et al. proposed the “ergmito” method based on ERGM extension to realize the fitting of small networks.^[54] Weihua also pointed out that when modeling large-scale network, the key is to achieve a good balance between accuracy / consistency and speed / stability.^[1] He elaborated two broad methods and seven large-scale network fitting methods based on ERGM, explained the advantages and disadvantages of each method, and provided suggestions for researchers to choose methods. AADS, BJHK demonstrated how to use snowball sampling and conditional estimation method to estimate ERGM parameters of large undirected networks, and verified the feasibility of this method.^[53] In 2019, Stivala A, Robins G et al. demonstrated the EE (equilibrium expectation) algorithm, which can estimate the ERGM parameters of a social directed network model with more than one million nodes, and applied it to an online social network with more than 1.6 million nodes.^[15] At present, the research on the application of ERGM in different scale networks is still in the development stage, and we believe that more scholars will join the research in the future.

4. FUTURE PROSPECTS OF TIE-GENERATIVE MECHANISM IN NETWORK BASED ON ERGM

Compared with the traditional regression analysis method, the advantage of ERGM is that it gets rid of the constraint of independent hypothesis and comprehensively considers the influence of network structure, node attributes and covariables on the network formation. At present, this research area has made good progress in the aspects of model extension and influence factors exploration, but there are still many troubles waiting for further research.

4.1 Extending ERGM to capture information on connection strength.

The weakness of ERGM is that it can only process binary network while can't capture the information on connection strength, which makes scholars have to convert the weighted network into the binary network for research, this operation leads to the loss of rich information in the weighted network.^[28] Extending ERGM to capture connection strength information in weighted networks is a topic worthy of further discussion.

4.2 Research on the dynamic mechanism of network formation.

Although there are TERGM and STERGM which can consider time dependence, there are still many gaps in this field. BraillyJ, Favreg et al. pointed out that the influences among different levels of networks are not unidirectional, in fact, they influence and depend on each other.^[46] However, current researches based on cross-sectional data can't enable them to explore complex causal relationships among multi-level networks, which limits people's understanding of the evolution process of networks. Shaohua Shi and Yehong Sun also proposed the interaction between networks of different levels would be the future research direction.^[55] In addition, the differences in the effects of influencing factors at different stages of network development also need to be further explored.^[44]

4.3 Extending ERGM to capture heterogeneity.

Heterogeneity refers to the intangible factors that are not observed or difficult to measure in the research but have an impact on the formation of the network, such as personal charm and corporate culture. Some scholars have shown that this unobserved heterogeneity may be the main reason for the inappropriate fitting of the network model.^[50] At present, although some academics have proposed SRFE-ERFM to alleviate the

dilemma caused by heterogeneity, this model still has some limitations. The research object of this model must be a directional network and the receiver and sender can't be modeled simultaneously. Furthermore, this model is only applicable to cross-sectional data, and can't be used for longitudinal research on the network. All of these limit the further application of SRFM-ERGM. How to extend the model so that it can broadly model various influencing factors while capturing the heterogeneity of nodes will be the focus of future research.

4.4 Developing methods to detect multicollinearity in ERGM.

Multicollinearity is when there is at least one independent variable in the model that changes as a function of the other independent variables. The undetected collinearity in ERGM would cause problems in the inference of model parameters, resulting in model degradation or non-convergence. In addition, the increasing number of structural items in the network also increases the possibility of collinearity^[56] and eventually leads to the model being unusable. Unfortunately, multicollinearity has not received enough attention and is often not examined effectively. Although Duxbury S W proposed a method proved to be valid in detecting multicollinearity in 2017^[56], in general, few studies have been conducted on how to detect and eliminate multicollinearity in ERGM, and further studies are needed in the future.

In addition to the above outlook, such as exploring how to estimate and impute the missing data of the network based on ERGM^[57], comparing ERGM with other models to discuss their respective advantages, disadvantages and applicable network types^[58], extending the concept of network formed from the formation of the edge to the formation and disappearance of the edge^{[41], [44]}, and so on are all research points that should be paid attention to in the future.

5. CONCLUSIONS

This paper takes the literature that discussing network formation mechanism based on the ERGM as the research object, mainly analyzes the research progress and limitations in this field from the perspective of the ERGM algorithm expansion (modeling node's heterogeneous, proposing longitudinally ERGM model, applying ERGM in different scale network) and the complex factors that affect the formation of the network (the research on multistage network, the influence of multiple social factors on the network) during the five years from 2015 to 2020, and puts forward the future prospects in this field, which provides reference for future research.

ACKNOWLEDGEMENT

This work is partly supported by the National Natural Science Foundation of PRC (No. 72025101, 71729001, 71601100), and the Fundamental Research Funds for the Central Universities (No. FRF-TP-20-045A2).

REFERENCES

- [1] Weihua An. (2016). Fitting ERGMs on Big Networks. *Social Science Research*, 59(59): 107-119
- [2] Yang Guancan, Liu Zhanlin, Li Gang. (2019). Research on the Formation Mechanism of Patent Citation Relationship Based on Exponential Random Graph Model—Taking Nelarabine As an Example. *Library and Information Service*, 63(10): 75-86(in Chinese)
- [3] Liu Huajun, Du Guangjie. (2018). Research on the Spatial Correlation of China's Haze Pollution. *Statistical Research*, 35(04): 3-15(in Chinese)
- [4] Luo Taiye, Ma Cuichang. (2018). Research on the Formation Mechanism of Collaborative Innovation Network Based on Exponential Random Graph Model. *Information Studies: Theory & Application*, 41(10): 143-146, 72(in Chinese)

- [5] Zhang Lupeng, Xue Lan, Zhou Yuan, Pan Meijuan. (2019). The Formation Mechanism and Empirical Research of Technology Diffusion Network of Emerging Industries Guided by Social Capital. *China Soft Science*, 339(03): 34-45(in Chinese)
- [6] Yang Guancan, Liu Tong, Chen Liang, Zhang Jing. (2018). Research on Influence Factors of Patent Citation Relationship Formation Based on ERGM. *Science Research Management*, 39(11): 125-134(in Chinese)
- [7] Ghosh A, Ranganathan R, Rosenkopf L. (2016). The Impact of Context and Model Choice on the Determinants of Strategic Alliance Formation: Evidence From a Staged Replication Study. *Strategic Management Journal*, 37(11): 2204-2221
- [8] Xu Helian, Sun Tianyang, Cheng Lihong. (2015). Study on Trade Pattern and Influencing Factors of "One Belt And One Road" High-end Manufacturing Industry -- Analysis of ERG Based on Complex Network. *Finance & Trade Economics*, (12): 74-88(in Chinese)
- [9] Song H. (2015). Uncovering the Structural Underpinnings of Political Discussion Networks: Evidence From an Exponential Random Graph Model. *Journal of Communication*, 65(1): 146–169
- [10] Anna Llupià, Puig J, Mena G, et al. (2016). The Social Network Around Influenza Vaccination in Health Care Workers: A Cross-Sectional Study. *Implementation Science*, 11(1): 152
- [11] Dang-Pham D, Pittayachawan S, Bruno V. (2017). Why Employees Share Information Security Advice? Exploring the Contributing Factors and Structural Patterns of Security Advice Sharing in the Workplace. *Computers in Human Behavior*, 67(67): 196-206
- [12] Yang Wenlong, Du Debin. (2018). The Structure and Influencing Factors of the Investment Network of Countries Along the "Belt and Road". *World Economy Studies*, (05): 80-94+136-137(in Chinese)
- [13] Box-Steffensmeier J M, Christenson D P, Morgan J W. (2018). Modeling Unobserved Heterogeneity in Social Networks with the Frailty Exponential Random Graph Model. *Political Analysis*, 26(01): 1-17
- [14] Sheng Kerong, Zhang Hongxia, Zhao Chaoyue. (2019). Influencing Factors of Urban Network Connectivity Pattern in China: From the Perspective of Electronic Information Enterprise Network. *Geographical Research*, 38(05): 32-46(in Chinese)
- [15] Stivala A, Robins G, & Lomi, A. 2020. Exponential random graph model parameter estimation for very large directed networks. *PloS one*, 15(1): e0227804.
- [16] Erdos P, Renyi A. (1959). On Random Graphs. *Publications Mathematicae*, 6: 290-297
- [17] Holland P W, Leinhardt S. (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373): 33-50
- [18] Frank O, Strauss D. (1986). Markov Graphs. *Journal of the American Statistical Association*, 81(395): 832-842
- [19] David R Hunter, Mark S Handcock. (2006). Inference in Curved Exponential Family Models for Networks. *Journal of Computational and Graphical Statistics*, 15(3): 565-583
- [20] Hunter D R. (2007). Curved Exponential Family Models for Social Networks. *Social Networks*, 29(2): 216-230
- [21] Wang P, Robins G, Pattison P, et al. (2013). Exponential Random Graph Models for Multilevel Networks. *Social Networks*, 35(1): 96-115
- [22] Brennecke J, Rank O. (2017). The Firm's Knowledge Network and the Transfer of Advice Among Corporate Inventors—A Multilevel Network Study. *Research Policy*, 46(4): 11787
- [23] Smith M, Gorgoni S, Cronin B. (2019). International Production and Trade in a High-Tech Industry: A Multilevel Network Analysis. *Social Networks*, 59: 50-60
- [24] Liu Linqing, Chen Ziruo, Tian Bifei. (2020). How does Structural Dependence Affect the Formation and Evolution of Trade Networks: A Case Study of "One Belt And One Road". *World Economy Studies*, (06): 106-120+137(in Chinese)
- [25] Krivitsky P N, Handcock M S. (2014). A Separable Model for Dynamic Networks. *Journal of the Royal Statistical Society*, 76(1): 29-46.

- [26] Bodin R , Nohrstedt D , Baird J , et al. (2019). Working at the “Speed of Trust”: Pre-Existing and Emerging Social Ties in Wildfire Responder Networks in Sweden and Canada. *Regional Environmental Change*, 19(8): 2353-2364
- [27] Bjorklund P, Daly A J. (2021). The Ties that Belong: Tie Formation in Preservice Teacher Identification Networks. *Teaching and Teacher Education*, 97(97): 103223
- [28] Peng, Tai-Quan. (2015). Assortative Mixing, Preferential Attachment, and Triadic Closure: A Longitudinal Study of Tie-Generative Mechanism in Journal Citation Networks. *Journal of Informetrics*, 9(2): 250-262
- [29] King S , Lusher D , Hopkins J , et al. (2020). Industrial Symbiosis in Australia: The Social Relations of Making Contact in a Matchmaking Marketplace for SMEs. *Journal of Cleaner Production*, 270(30): 122146
- [30] Cao Q, Liao L, Warren KL. (2020). Social and Programmatic Interactions in a Therapeutic Community for Women: an Exponential Random Graph Model Analysis. *Therapeutic Communities: the International Journal for Therapeutic and Supportive Organizations*, 41(3/4): 69-79
- [31] Duan Qingfeng, Ma Dandan. (2018). Empirical Study on Diffusion Mechanism of Patent Technology Based on ERGM. *Science & Technology Progress and Policy*, 35(22): 29-35(in Chinese)
- [32] Liu Yunong, Liu Minrong. (2018). The Social Network Structure and Influencing Factors of Virtual Knowledge Community -- Taking Zhihu as an Example. *Library and Information Service*, 62(04): 89-96(in Chinese)
- [33] V Gtle E M , Windzio M. (2016). Networks of International Student Mobility: Enlargement and Consolidation of the European Transnational Education Space?. *Higher Education*, 72(6): 1-19
- [34] Asal V H , Park H H , Rethemeyer R K , et al. (2016). With Friends Like These ... Why Terrorist Organizations Ally. *International Public Management Journal*, 19(1): 1-30
- [35] Liu Xuan, Wang Linwei, Li Jia, Zhang Pengzhu. (2017). Study on the influence mechanism of user response behavior in online health community. *Journal of Management Science*, 30(01): 62-72(in Chinese)
- [36] Duan Qingfeng, Feng Zhen. (2019). Research on the Formation Mechanism of Academic Social Network Based on ERGM. *Information Science*, 37(07): 84-89+145(in Chinese)
- [37] Yang Guancan, Chen Liang, Zhang Jing, Li Gang. (2019). An Explanatory Framework for Patent Citation Relationship Formation: An ERGM Perspective. *Library and Information Service*, 63(05): 100-109(in Chinese)
- [38] Liu Xuan, Wang Linwei, Li Jia, Zhang Pengzhu. (2019). Forming Mechanism of Scientific Research Cooperation Network: Based on Stochastic Exponential Graph Model. *Journal of Systems & Management*, 28(03): 520-527(in Chinese)
- [39] B P F A , B X L , C Z G , et al. (2020). A Case Study of the Pyramid Scheme in China Based on Communication Network. *Physica A: Statistical Mechanics and its Applications*, 565(565): 125548-
- [40] Duan Qingfeng, Pan Xiaohuan. (2018). An Empirical Study on the Effect of Literature Similarity on Scientific Citation Preference. *Library and Information Service*, 62(04): 97-106(in Chinese)
- [41] Lee S K , Kim H , Piercy C W. (2016). The Role of Status Differentials and Homophily in the Formation of Social Support Networks of a Voluntary Organization. *Communication Research*, 46(2): 208-235.
- [42] Liu Xiaoyan, Li Jinpeng, Shan Xiaohong, Yang Juan. (2020). The Impact of Multidimensional Proximity on Patent Technology Trading in Integrated Circuit Industry. *Studies in Science of Science*, 38(05): 834-842+960(in Chinese)
- [43] Liu Linqing, Chen Ziruo. (2020). Study on the Dynamic Evolution Mechanism of China's Dominant Industry Portfolio -- Based on TERGM Empirical Analysis. *Science & Technology Progress and Policy*, 37(11): 70-78(in Chinese)
- [44] Liu Xiaoyan, Wang Jing, Shan Xiaohong. (2020). Research on the Evolutionary Dynamics of Technological Innovation Network Based on TERGMS. *Science Research Management*, 41(04): 171-181(in Chinese)
- [45] Brennecke J , Rank O N. (2016). The Interplay Between Formal Project Memberships and Informal Advice Seeking in Knowledge-Intensive Firms: A Multilevel Network Approach. *Social Networks*, 44(1): 307-318
- [46] Brailly J , Favre G , Chatellet J , et al. (2016). Embeddedness as a Multilevel Problem: A Case Study in Economic Sociology. *Social Networks*, 44: 319-333

- [47] Locatelli B , Pramova E , Gregorio M D , et al. (2020). Climate Change Policy Networks: Connecting Adaptation and Mitigation in Multiplex Networks in Peru. *Climate Policy*, 20(3): 354-372
- [48] Hollway J , Koskinen J. (2016). Multilevel Embeddedness: The Case of the Global Fisheries Governance Complex. *Social Networks*, 44: 281-294.
- [49] Wang P, Robins G, Pattison P, & Lazega, E. 2016. Social selection models for multilevel networks. *Social Networks*, 44: 346-362.
- [50] Thiemichen S , Friel N , Caimo A , et al. (2016). Bayesian Exponential Random Graph Models with Nodal Random Effects. *Social Networks*, 46(4): 11-28.
- [51] Henry T R , Gates K M , Prinstein M J , et al. (2016). Modeling Heterogeneous Peer Assortment Effects Using Finite Mixture Exponential Random Graph Models. *Psychometrika*, 85(1): 8-34
- [52] Koskinen J , Wang P , Robins G , et al. (2018). Outliers and Influential Observations in Exponential Random Graph Models. *Psychometrika*, 83(4): 809-830
- [53] AADS, BJHK, ADAR , et al. (2016). Snowball Sampling for estimating exponential random graph models for Large Networks - ScienceDirect. *Social Networks*, 47: 167-188.
- [54] Yon G G V , Slaughter A , Haye K D L. (2020). Exponential Random Graph Models for Little Networks. *Social Networks*, 64: 225-238
- [55] Shi Shaohua, Sun Yehong. (2017). Research on Tourism Development Interest Network Governance of Heritage Sites: Based on ERGM, a Case Study of Hani Terraces in Yunnan, an Agricultural Cultural Heritage Site. *Business Management Journal*, 39(02): 147-162(in Chinese)
- [56] Duxbury S W. (2017). Diagnosing Multicollinearity in Exponential Random Graph Models. *Sociological Methods & Research*, 417
- [57] Wang C , Butts C T , Hipp J R , et al. (2016). Multiple Imputation for Missing Edge Data: A Predictive Evaluation Method with Application to Add Health. *Social Networks*, 45: 89-98
- [58] Block P , Stadtfeld C , Snijders T A B . (2016). Forms of Dependence: Comparing SAOMs and ERGMs From Basic Principles. *Sociological Methods & Research*, 48(1): 202-239

Full Research Paper

Empirical Study of the Regional Diffusion of a Cloud Ecosystem Using Institutional Theory

Xin Zhang¹, Bingqing Ji¹, Gaoyong Li^{1*}, Ge Zhang¹²

¹School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan, 250014, China

²Yanshan College of Shandong University of Finance and Economics, Jinan, 250202, China

Abstract: Cloud computing, a new type of digital infrastructure, can provide the necessary computing power to develop artificial intelligence, enhance the Internet of Things and the Industrial Internet to support their infrastructure layers, and reduce costs and increase efficiency for organizations. However, there has been little adoption or use of cloud computing in China, and it remains unknown what factors influence the diffusion of cloud computing in the region. This study establishes econometric models based on institutional theory and empirically analyzes the impact of three actors, namely, governments, enterprises that adopt cloud computing, and cloud computing suppliers, on the diffusion of cloud ecosystem using relevant data from Shandong Province, China. The results indicate that these participants have varying types of influence on the breadth and depth of cloud ecosystem diffusion in the region through their institutional pressure. This research also put forward some practical implications for the diffusion of cloud computing in the region.

Keywords: Cloud computing; cloud ecosystem; technology diffusion; institutional theory

1. INTRODUCTION

Cloud computing releases organizations from the burdens associated with the construction of complicated information systems, enabling them to focus more on their core business functions, reducing their costs, and increasing their efficiency. The technical features and service modes of cloud computing can also promote the implementation of new technologies, such as artificial intelligence and the Internet of Things ^[1]. Thus, the numerous advantages that cloud computing brings are leading more and more organizations to consider adopting it.

The extant literature largely explores the factors that affect the adoption of cloud computing at the organization level, which provides guidance to enhance the adoption of cloud computing. These studies have largely focused on establishing the impact of the characteristics of enterprises and cloud computing on the adoption of cloud computing. For example, the adoption of cloud computing could be improved by obtaining top management support ^[2] and improving the security and privacy of the technology used ^[3]. However, previous research has not been able to explain the diffusion of cloud computing at the regional level. Most of them ignore the influence of cloud computing as infrastructure and ecosystem in the region and neglect other participants' influence, like suppliers and core enterprises.

Cloud computing is a type of infrastructure with strong externalities, unlike other technological innovations. It does not affect a single enterprise alone but radiates its influence over an entire region. The diffusion of cloud computing is a multi-agent innovation diffusion process, which requires both the services provided by a supplier and the active participation of others in the region, such as the government and core enterprises. Therefore, cloud computing ecosystem should be developed to explore the factors that affect the diffusion of cloud computing at a regional level, which has not been established by previous research. Theories of adoption and diffusion, such as Technology-Organization-Environment (TOE) Framework, Technology Acceptance Model

* Corresponding author. Email: ligao Yong0706@163.com(Gaoyong Li)

(TAM), and Diffusion-of-Innovation (DOI), could partly explain environmental factors that affect the diffusion of cloud computing at the organizational level but cannot comprehensively explain the impact of the external environment on the diffusion of cloud computing ecosystem in the region. This study, thus, employs institutional theory as its theoretical basis because it is suitable for establishing the influence of the external environment on innovation diffusion.

Furthermore, there has generally been little adoption of cloud computing technology in China, especially in the northern region, and only a few core enterprises have seen in-depth use of them, due to the limitations of the levels of technological and economic development. As a typical representative of northern provinces, Shandong province is facing traditional challenges of economic transformation. Therefore, this study takes Shandong province as an example to explore the influence factors of the diffusion of the cloud ecosystem in the region.

This study explores the influence of government, core enterprises, and cloud computing suppliers on the diffusion of the cloud ecosystem in the region, using institutional theory and data from Chinese cloud computing enterprises to provide feasible suggestions for the promotion of the development of cloud ecosystems.

2. THEORETICAL BACKGROUND

2.1 Institutional theory

Institutional theory is a core theory of sociological institutionalism, first put forward by DiMaggio and Powell [4]. This perspective holds that organizations carry out certain behaviors to satisfy the requirements of government and industry to obtain legitimacy to evade environmental restrictions. Institutional theory is an important approach for effectively explaining technology adoption, especially for technologies influenced by external actors, other technological approaches, and general policies.

DiMaggio and Powell identify three types of institutional pressure: coercive, mimetic, and normative. Coercive pressure derives from formal and informal pressures exerted on organizations by other organizations upon which they are dependent and from institutional expectations. It mainly results from the requirements of other suppliers in the industry chain [5] and their supervision or from the government [6]. The government and industry associations greatly influence enterprises' adoption behavior, especially in developing countries [7]. For example, the Chinese government once forced enterprises to adopt accounting computerization to aid the normalization of financial and tax management. Moreover, extant literature has found that coercive pressure plays an important role in cloud computing adoption [8][9].

To avoid the adverse effects of the application of new technologies, organizations may observe and imitate enterprises in their region that have obtained legitimacy or competitive advantages to reduce risk, such behavior is influenced by mimetic pressure. It has been shown that mimetic pressure could promote the adoption of cloud computing [6], as can be seen with the example of ERP. Because the implementation of ERP is full of uncertainty, enterprises are inclined to imitate successful ERP implementers at a similar scale, in a similar line of business, or in the same region to reduce their chance of failure [10].

Normative pressure is associated with professionalization. Professionalization means that professional bodies in the field, such as suppliers, customers, and third-party organizations, establish the cognitive base and enable the occupational autonomy to define the conditions and methods of their work [4]. Further, this professionalization normalizes the behavior of the organizations that enter the field in the future through professionalization. Martins et al. [8] found that the normative rules for adopting and implementing cloud computing in the industry significantly influence the adoption of cloud computing with normative pressure.

This research explores the diffusion of cloud ecosystems within a region based on institutional theory. This approach entails a few characteristics: it assumes an institutional isomorphism that could reasonably explain convergence between organizations. Cloud computing is an infrastructure similar to the Internet that can affect

the whole society^[9]. So, with the growth of enterprises adopting cloud computing, a cloud ecosystem formed in the region, which at the same time developed isomorphisms. Second, as a theory used to explain organizational adoption behavior, the institutional theory is often used to explain the adoption of cloud computing in other contexts^{[6][8][11]}. This is because whether an enterprise adopts cloud computing depends not only on its own internal influences but also on the external pressures exerted by other organizations in the region^{[8][12]}. However, the extant literature largely uses institutional theory to explore the adoption of cloud computing at the organization level, while our research explores the impact of institutional pressure on cloud ecosystem diffusion at a regional level. Third, the influence of the government, core enterprises, and suppliers on the diffusion of the cloud ecosystem in the region can be explained through coercive, mimetic, and normative pressures in the way that these are assessed in institutional theory. This is another important reason for using institutional theory as our theoretical basis.

2.2 The application of institutional theory in the cloud ecosystem diffusion

The government, core enterprises, and suppliers affect the diffusion of a cloud ecosystem by means of their different institutional pressures. The government utilizes policies and fiscal subsidies to promote the diffusion of the cloud ecosystem and to guide the direction of its development^[9]. For example, the government guides the development of the technology ecosystem by formulating its own strategies and policies, reduces the cost of technology implementation by providing fiscal subsidies, and stimulates the diffusion of a technology ecosystem by rewarding benchmark enterprises. The instructions of higher levels of government are transmitted to industries through lower levels. These bodies may formulate different policies to accomplish their goals, and this could exert coercive pressure on enterprises in the region. Apart from administrative measures, the government could also use market forces to promote the adoption of cloud computing. For example, in China, the State Council and the Ministry of Industry and Information Technology (MIIT) have proposed publicizing typical cases and successful adoptions of cloud computing, as well as recommending cloud benchmarking enterprises to promote the development of cloud computing. These regulations allow enterprises to study how benchmarking enterprises can realize information transformation and improve their production efficiency through cloud computing, which can be illustrated by mimetic pressure and normative pressure.

The influence of core enterprises on the cloud ecosystem could also be characterized as institutional pressure. When enterprises come together, industrial clusters are formed in which a large number of suppliers begin to produce and sell around a core enterprise. These core enterprises generally have a dominant capability, so they may prefer to adopt new technologies such as cloud computing early^[13]. Other enterprises may be subject to mimetic pressure to imitate those successful core enterprises. Alternatively, some core enterprises may force their upstream and downstream suppliers to adopt their information system to facilitate logistics management, another form of coercive pressure. Furthermore, some core enterprises (e.g., Haier) may build cloud platforms that conform to the characteristics of the industry after cloud computing is implemented and may provide such a platform to other enterprises, such as the COSMO Plat built by Haier, which allows mass customization. For platform adopters, this type of empowerment may form coercive and normative pressure, while enterprises that do not have a restrictive relationship with core enterprises may experience mimetic pressure from the success of cloud platform developers.

As a provider of technology and services, a supplier is an important participant in an ecosystem and also provides an important impetus promoting the development of the business ecosystem^[14]. Suppliers often publicize the advantages of cloud computing to improve enterprises' willingness to adopt cloud computing. Participants in the activities organized by suppliers generally report their participation improves their understanding of cloud computing. Therefore, besides the improvement to service capabilities, cloud computing suppliers could also impose mimetic pressure through market publicity activities to promote the diffusion of

cloud ecosystems.

3. RESEARCH METHODOLOGY

3.1 Data

This study takes the county region as its sample unit, for the following reason. Ni ^[15] indicated that the industrial cluster in China takes counties as units, and there are often one or more core enterprises in a given cluster, enabling network effects. This affects the application of cloud computing, making it gradually form a cloud ecosystem. Therefore, the influence of core enterprises on the diffusion of a cloud ecosystem can be observed through the county region. Moreover, county administrative departments are grass-roots practitioners of policy implementation in China. The impact of policy on the diffusion of cloud ecosystems can be observed in the county region because the government distributes subsidies to enterprises through county finance departments. Therefore, we choose the county region as our sample unit.

This study acquires the data for 136 counties through the questionnaire data on the integration of information technology and industrialization of Shandong Province and statistical data from Shandong Province in relation to excellent enterprises adopting cloud computing, from which we obtained the subsidies invested to promote the adoption and use of cloud computing for each sample and information on the core enterprises and suppliers in each sample.

3.2 Research variables

To measure how these participants affect cloud ecosystem diffusion through institutional pressure more objectively, this study proposes variables and an integrated model of the diffusion of the cloud ecosystem according to the characteristics of each participant. Figure 1 illustrates the research model.

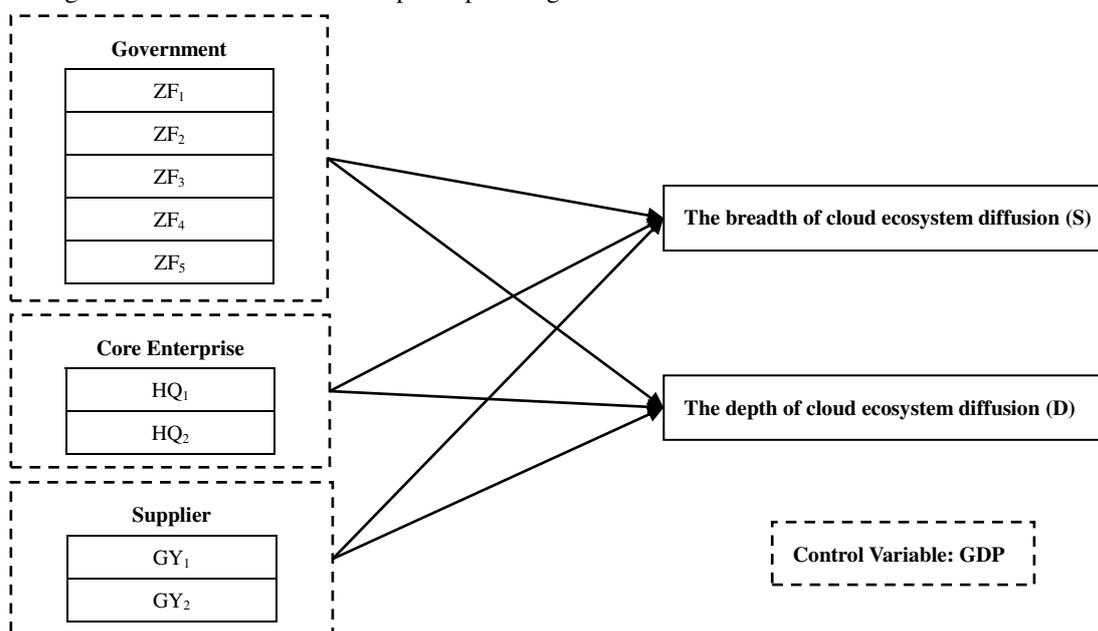


Figure 1. Research model

3.2.1 Government

The government fundamentally affects the diffusion of the cloud ecosystem through the use of three types of measures: fiscal subsidies, policy support, and publicity activities. This research proposes to investigate the amount of cloud service coupons (ZF_1) and the bonus of benchmarking enterprises (ZF_2) as variables to represent fiscal subsidies. Cloud service coupons are small, no more than 10000 RMB, while the bonus awarded to a benchmarking enterprise is large, with strict criteria. The cloud service coupons have largely been given to

SMEs, which can exchange coupons for fiscal subsidies. Due to the low standards, enterprises are easily able to obtain such subsidies. When SMEs in the region receive subsidies by adopting cloud computing, the herding effect leads other enterprises to adopt imitative behavior, which can accelerate the diffusion speed of the cloud ecosystem in the region. Thus, ZF_1 largely affects the diffusion of a cloud ecosystem through mimetic pressure.

The number of benchmarking enterprises in the region (ZF_3) is viewed as an independent variable for policy support. Benchmarking enterprises are chosen by the government according to criteria that are formulated in advance. Success in cloud computing by such an organization can form mimetic pressure in its region. Further, during their implementation of cloud computing, they can deepen their understanding of cloud computing and present a more better-grounded opinion, which can then be used by the government to revise its criteria. After being modified, the criteria can become more standardized and professional and can also, to a certain extent, reduce the difficulty of application, which could increase enterprises' enthusiasm for cloud computing. This means that larger numbers of benchmarking enterprises could also promote the diffusion of cloud computing through the addition of normative pressure.

Beyond providing fiscal subsidies, the government also produces publicity to advertise the benefits of cloud computing, such as reducing costs and improving efficiency. Such activities may stimulate enterprises to imitate other enterprises that have successfully adopted cloud computing to gain competitive advantages. Therefore, the number of publicity activities organized by the government (ZF_4), and the average number of participating enterprises in each (ZF_5) are proposed as non-fiscal subsidy variables.

3.2.2 Core enterprise

Core enterprises are those that have a dramatic influence on other enterprises in an industrial cluster. They are the leaders or the strategic core in the cluster and can affect the behavior or performance of other enterprises through their own methods, such as reducing or increasing the number of orders and advancing or postponing the time of payment.

The questionnaire provided by the MIIT divides the degree of implementation of cloud computing in an enterprise into three stages: initial construction stage, coverage of a single operational stage, and integration and innovation stage. We propose the number of integration and innovation enterprises (HQ_1), most of which are early adopters of cloud computing in the region as an independent variable. These are core enterprises and have great influence in their industrial cluster. The successful adoption of cloud computing by these enterprises can form a demonstration effect for other enterprises that have not taken up cloud computing, so HQ_1 influences the diffusion of the cloud ecosystem through mimetic pressure.

In addition, the number of enterprises that have formed a cloud platform (HQ_2) is another independent variable on the core enterprise level. On the one hand, enterprises that have formed a cloud platform are core enterprises with professional capability. The increase in the number of these enterprises can gradually standardize the use of cloud computing in the industry and form normative pressure. On the other hand, as core enterprises, they may force their partners to use their own cloud platforms, which represents coercive pressure.

3.2.3 Cloud computing supplier

The literature ^[16] indicates that regions or industries with a higher degree of industrial development and concentration tend to have a more extensive technological ecosystem, so we choose the number of cloud computing suppliers (GY_1) to represent the degree of development and concentration in the cloud computing industry in the region. The normative pressure caused by the increase of these professionals can affect other enterprises' attitudes and behaviors in the region. The competitive pressure caused by the increase in suppliers can stimulate them to enhance their service capabilities, which will directly influence enterprises' experience of using cloud computing and improve their willingness to use cloud computing ^[17].

Moreover, we choose the number of publicity activities organized by suppliers (GY_2) as an independent

variable. Unlike the activities organized by the government, suppliers' activities are not mandatory, and they are more inclined to introduce their products from the perspective of the market. They choose to publicize the advantages of cloud computing and excellent cases to attract other enterprises, and this promotes mimetic pressure.

3.2.4 Cloud ecosystem

This study integrates the effects of factors on different dimensions of the cloud ecosystem, including breadth (S) and depth (D). The breadth of the diffusion of the cloud ecosystem refers to the popularity of cloud computing among the enterprises in the region, which is calculated by the proportion of enterprises adopting cloud computing among the total number of enterprises. The depth of cloud ecosystem diffusion refers to the understanding and application degree of cloud computing by the enterprises that have adopted cloud computing, which is calculated by the average proportion of cloud computing applications in business systems of enterprises in the region that have adopted cloud computing. The combination of the two dimensions can comprehensively reflect the development of the cloud ecosystem.

3.2.5 Control variable: GDP of each region

First, as the adoption of cloud computing requires investment in enterprises, the government tends to choose to pilot programs in more developed regions. Second, enterprises in developed regions tend to have access to more advanced thinking and better information, which can help enterprises encounter and adopt the relevant concepts of cloud computing at an earlier stage. Therefore, the diffusion of the cloud ecosystem may strongly correlate with GDP, which is why GDP has been chosen as a control variable to control the impact of different economic levels on this study.

Finally, from these considerations, two econometric models can be written, including the disturbance terms γ and ε , as follows:

$$S = \alpha_0 + \alpha_1 ZF_1 + \alpha_2 ZF_2 + \alpha_3 ZF_3 + \alpha_4 ZF_4 + \alpha_5 ZF_5 + \alpha_6 HQ_1 + \alpha_7 HQ_2 + \alpha_8 GY_1 + \alpha_9 GY_2 + \lambda_1 GDP + \varepsilon \quad (1)$$

$$D = \beta_0 + \beta_1 ZF_1 + \beta_2 ZF_2 + \beta_3 ZF_3 + \beta_4 ZF_4 + \beta_5 ZF_5 + \beta_6 HQ_1 + \beta_7 HQ_2 + \beta_8 GY_1 + \beta_9 GY_2 + \lambda_2 GDP + \gamma \quad (2)$$

4. RESULT

4.1 Analysis of descriptive statistics

Table 1 presents the characteristic variables of this investigation. The standard deviations ($STD_S = 3.32$, $STD_D = 5.13$) and the differences between the maximum and the minimum values for variables S and D indicate the degree of popularity of cloud computing in Shandong Province is not high in either sample, around the middle of the range for both. Compared to the degree of popularity of cloud computing, the difference in average business coverage rate for cloud computing between the samples is large, representing different depths of cloud computing diffusion.

The standard deviations for fiscal subsidy indicators ZF_1 and ZF_2 are relatively large, indicating that the financial budget for cloud computing development is quite different across regions, which may be related to the local level of economic development. For the non-fiscal subsidy indicators, according to the standard deviations of ZF_4 and ZF_5 , although there is little difference in the number of publicity activities organized by the government in different regions, the number of participants in activities is obviously distinct in each region.

The number of integration and innovation enterprises (HQ_1) differs distinctly by region. The number of enterprises forming cloud platforms (HQ_2) is generally low, not more than 1 in most regions, and only 3 in the largest. Thus, we speculate that the impact of HQ_2 on the diffusion of the cloud ecosystem may be slight.

The value of the standard deviation for GY_1 is quite large. The area with the lowest value has only two small suppliers which are mainly product agents, while the largest region has 112 suppliers. One possible reason for this phenomenon is cloud computing technology requires greater talent and a higher economic level, so most

suppliers will choose regions where they have access to more educational resources and greater economic vitality to set up companies.

In addition, the disparity in levels of economic development between regions within Shandong province is large, although Shandong is a province with a high level of economic development in China. The region with the highest GDP is nearly 50 times larger than that of the lowest region, so we expect that the impact of GDP on the diffusion of the cloud ecosystem may be significant and needs to be controlled.

Table 1. Variable summary statistics

Variable name	Definition	Min.	Max.	Mean	S.D.
ZF_1	The amount of cloud service coupons	836	5635	2287.37	23.11
ZF_2	The bonus of benchmarking enterprises	532	1962	769.86	16.62
ZF_3	The number of benchmarking enterprises	5	36	9.51	2.37
ZF_4	The number of publicity activities organized by the government	0	23	5.23	1.59
ZF_5	The average number of enterprises participants in an activity	0	533	156.36	8.38
HQ_1	The number of Integration and Innovation enterprises	123	657	236.45	8.68
HQ_2	The number of enterprises forming cloud platforms	0	3	0.23	0.47
GY_1	The number of suppliers of cloud computing	2	112	24.63	3.92
GY_2	The number of publicity activities organized by the supplier	5	19	8.03	1.52
S	The proportion of enterprises adopted cloud computing (%)	53.20	80.65	60.32	3.32
D	The proportion of cloud computing applications in the business system of enterprises (%)	9.30	60.62	36.23	5.13
GDP		77.80	3554.44	514.42	480.88

4.2 Statistical test

First, the White test is used to check for heteroscedasticity, a problem that often exists in cross-sectional data. The result indicates that $Obs \cdot R\text{-squared} = 57.12$, which is less than the critical value when $\alpha = 0.05$. Therefore, the original hypothesis is accepted, indicating an absence of serious heteroscedasticity in the model.

Second, all values for the correlation coefficient between independent variables in Table 2 are less than 0.5, which indicates the absence of a serious multicollinearity problem in this study. And the correlation coefficient between independent and dependent variables almost larger than 0.5, confirming that they have a strong correlation and are suitable for regression tests.

The regression results are shown in Table 3, where Study 1 shows the influence of factors on the breadth of cloud ecosystem diffusion (S), and Study 2 shows the influence on the depth (D) of cloud ecosystem diffusion. According to the p values, the influence of each independent variable on the dependent variable are statistically significant, and the values of the F statistic indicate that the effects of all independent variables on the dependent variable are jointly significant in both models.

Table 2. The results of the correlation coefficients

	S	D	ZF_1	ZF_2	ZF_3	ZF_4	ZF_5	HQ_1	HQ_2	GY_1	GY_2
S	1.00										
D	0.82**	1.00									
ZF_1	0.60**	0.60**	1.00								
ZF_2	0.56*	0.71***	0.33**	1.00							
ZF_3	0.65**	0.50**	0.14*	0.09**	1.00						
ZF_4	0.67***	0.80*	0.21**	0.47***	0.65**	1.00					

ZF_5	0.58***	0.48*	0.32**	0.46***	0.57**	0.35***	1.00				
HQ_1	0.58**	0.58**	0.26***	0.57***	0.57**	0.17***	0.28***	1.00			
HQ_2	0.43**	0.65***	0.43***	0.06*	0.44**	0.80**	0.63**	0.52**	1.00		
GY_1	0.66***	0.74***	0.25**	0.44***	0.30***	0.31***	0.41***	0.43***	0.57**	1.00	
GY_2	0.74**	0.66***	0.39***	0.16**	0.02*	0.09***	0.25***	0.31**	0.84***	0.29**	1.00

Table 3. The results of multiple regression

	Study 1		Study 2	
	α	p	β	p
ZF_1	0.896	0.000	0.851	0.000
ZF_2	0.223	0.008	0.274	0.000
ZF_3	0.468	0.001	0.563	0.001
ZF_4	0.667	0.001	0.727	0.000
ZF_5	0.698	0.001	0.668	0.000
HQ_1	0.332	0.006	0.338	0.009
HQ_2	0.246	0.032	0.365	0.05
GY_1	0.835	0.009	0.795	0.000
GY_2	0.394	0.041	0.453	0.036
GDP	1.56	0.000	2.63	0.000
Adjust R-square	42.30%		33.68%	
F	3.365		2.479	

4.3 Multiple regression analysis

In this study, the regression results are analyzed from the perspective of the three participants that influence the diffusion of the cloud ecosystem. The government influences the diffusion of the cloud ecosystem through fiscal subsidies, market publicity, and other means. The regression indicates that state subsidies (ZF_1 and ZF_2) have different effects on the breadth and depth of cloud ecosystem diffusion. The amount of cloud service coupons obviously influences the diffusion of cloud ecosystem in both dimensions ($\alpha_1=0.896$, $\beta_1=0.851$), which could reduce the cost to enterprises of adopting cloud computing and make it easy for SMEs to obtain. However, few enterprises can meet the criteria for becoming benchmarking enterprises. Therefore, the bonus to benchmarking enterprises affects the diffusion of cloud ecology only negligibly ($\alpha_2=0.223$, $\beta_2=0.274$). For enterprises that are deeply engaged in working with cloud computing, applying for the status of a benchmarking enterprise can stimulate them to deepen their commitment to cloud computing and bring them a huge bonus. For enterprises that have adopted cloud computing and whose demand for cloud computing is dispensable, although the bonus is attractive, the harsh criteria will damage their enthusiasm for the in-depth application to cloud computing. Enterprises that have not adopted cloud computing may be even further deterred. They will respond better to a small subsidy at first, such as cloud service coupons. Moreover, continuous increases in market publicity activities (ZF_4 , $\alpha_4=0.667$, $\beta_4=0.727$) and in the number of enterprises participating in activities (ZF_5 , $\alpha_5=0.698$, $\beta_5=0.668$) can significantly stimulate the diffusion of the cloud ecosystem, both in breadth and depth.

The core enterprise is another important participant in the diffusion of cloud ecosystems. Standardized regression coefficients show that the breadth of cloud ecosystem diffusion (S) is positively correlated with HQ_1 ($\alpha_6=0.332$, $\beta_6=0.338$), a result that does not support previous research. According to Chen et al. [18], if half of all enterprises in a cluster adopt a technological innovation, the influence of the number of adopters on the

diffusion of the innovation decreases steeply, and when the adopters exceed threshold, their influence on the diffusion of the innovation is almost zero. One important reason for our finding is that the development of the cloud ecosystem is at an early stage. After analyzing each sample, we found no region in which the number of integration and innovation enterprises is more than 50% of the number of all local enterprises. Therefore, it is reasonable to conclude that the HQ_1 positively influences the diffusion of the cloud ecosystem in this study.

Furthermore, it is worth noting that the number of benchmarking enterprises (ZF_3) strikingly impacts the diffusion of the cloud ecosystem, both in breadth and depth ($\alpha_3=0.468$, $\beta_3=0.563$), while the number of enterprises forming cloud platforms (HQ_2) influences the cloud ecosystem only mildly ($\alpha_7=0.246$, $\beta_7=0.365$). The reasons for this result are as follows: (1) according to the descriptive statistics, compared to the number of other enterprises, such as ZF_3 , only a negligible number of enterprises are forming cloud ecosystem in the region, and (2) Cui and Shi ^[19] indicate that the coercive pressure of core enterprises on other suppliers is mainly concentrated within a single supply chain, while the proportion of enterprises in a certain industrial chain is not large over an entire region. For both reasons, although the number of enterprises forming cloud platforms is statistically significant, its impact on the diffusion of cloud ecosystem in the region is low.

The presence of suppliers of cloud computing is also an important factor for enhancing the diffusion of the cloud ecosystem in the region. On the one hand, the standardized coefficients of the regression indicate that the number of cloud computing suppliers (GY_1) significantly influences the diffusion of the cloud ecosystem ($\alpha_8=0.835$, $\beta_8=0.795$). We speculate that an increase in the number of suppliers could prompt suppliers to enhance competitiveness and capability to occupy greater market share, which could improve the breadth and depth of the cloud ecosystem in the region. On the other hand, market publicity activities organized by suppliers (GY_2) only mildly influences the diffusion of the cloud ecosystem ($\alpha_9=0.394$, $\beta_9=0.453$). The publicity is not mandatory, separating it from government publicity. Further, it is market-oriented and pays closer attention to whether enterprises can be transformed into new customers, so its scale is smaller, but the introduction of cloud computing is more in-depth and specific. Therefore, GY_2 has a greater impact on the depth of cloud ecosystem diffusion than on the breadth of cloud ecosystem diffusion.

5. CONTRIBUTION AND IMPLICATION

This study utilizes institutional theory to explore the factors that affect the diffusion of cloud computing in the region, and its results entail the following theoretical contributions and practical implications.

5.1 Theoretical contributions

First, this study supplements previous work on the construction of a cloud ecosystem. The literature mainly focuses on government policies and finds that the government can prompt the diffusion of cloud computing through state subsidies, while this study investigates other means, such as publicity activities organized by the government, that could also promote the development of a cloud ecosystem. In the construction of a cloud ecosystem, core enterprises play an important role. They could promote the diffusion of the cloud ecosystem in an industrial chain, as has been shown in the literature, but we find that they could also promote the diffusion of cloud computing within a region. We speculate that this is because core enterprises can produce atmospheres of mimetic pressure and normative pressure in a region through the demonstration effect, which could stimulate enterprises outside the industrial chain to adopt cloud computing and further promote the diffusion of the cloud ecosystem. The role of the cloud computing supplier is an important one, which has been neglected in previous studies. This study finds that the number of suppliers and the number of activities organized by suppliers in the region have a significant impact on the diffusion of the cloud ecosystem.

Second, this study enriches the research of technological innovation diffusion from a regional perspective. Enterprises in an industrial chain have strong dependency relations and their main business are closely related to

each other, which is conducive to the diffusion of an innovation ecosystem. Therefore, previous research on innovation ecosystems has focused on the industrial chain. However, the cloud ecosystem is different from other innovation ecosystems. On the one hand, it has many service modes, and its different modes have different characteristics. On the other hand, as an infrastructure, it is not limited to a specific industrial chain. Thus, we explore the influencing factors of cloud ecosystem diffusion from a regional perspective. To a certain extent, the core enterprise factors represent a single industrial chain that extends only over a part of the region, while they also represent the impacts caused by different industrial chains. Factors such as the government and suppliers are often ignored by relevant research focusing on the industrial chain. This study brings them into the model to explore the diffusion of the cloud ecosystem in the region and expand the research level.

Third, cloud computing initially only affects single enterprise, but the results of field research show that the diffusion of cloud computing can rise to a higher regional level, and every enterprise in the region is affected by institutional pressures. Compared with other technological innovations, cloud computing has unique characteristics, such as its stronger service capacity and stronger externality, which have been ignored by previous studies of technological diffusion. Considering the influences of externality, core enterprises and suppliers are also included in this research, and the sources of externality are explained by institutional theory. The results confirm that these factors all have significant influences on the diffusion of the cloud ecosystem, which indicates that institutional pressure not only influences single enterprise and single industrial chain but also forms a regional atmosphere. Such an atmosphere may bring pressure on all enterprises in a region and stimulate the diffusion of a cloud ecosystem. Therefore, this study exhibits an innovative use of technological diffusion theory and institutional theory.

5.2 Practical implications

This study provides practical suggestions for the government to formulate policies that can be used to guide the development of the cloud ecosystem. In relation to the diffusion of the cloud ecosystem, one important safeguard measures put forward by the government is to increase state fiscal subsidies. In practice, however, when excessive subsidies go to some enterprises, other enterprises may adopt cloud computing blindly and even seek to gain subsidies by fraud, which will lead to limited subsidies being concentrated in the hands of a few enterprises, making the policy ineffective. However, if an indiscriminate subsidy method is adopted, which provides equal subsidies to all enterprises, the subsidies received by enterprises will be too small, which may lead many enterprises to give up applying subsidies and reduce their willingness to adopt cloud computing. In addition, the government will only rarely carry out other market-oriented measures to promote the diffusion of a cloud ecosystem, because its understanding of new technologies is not thorough. Although this investigation finds that some governments have begun to guide enterprises to adopt cloud computing through market publicity activities, the effect is not satisfactory. Some governments attach too much importance to the number of sessions and participants in publicity activities so that frequent activities are carried out and each activity is overcrowded. These activities increase the burden of the information departments in enterprises and greatly reduce the effectiveness of activities.

The results of this study can lead to recommendations for future government policies and measures. For fiscal subsidy measures, the government could adopt a graded subsidy, giving both large subsidies and general subsidies. For enterprises, even core enterprises, large subsidies are so attractive that they can form coercive pressure that is not based on administrative power. The government could utilize this coercive pressure to set higher standards. For example, only enterprises that implement R&D services, production control services, or even intelligent application services in cloud computing would be granted large subsidies, while other enterprises could receive general subsidies if they implement infrastructure cloud services or platform cloud services. This measure would not only enhance the enthusiasm of enterprises to adopt cloud computing but also

prevent others from blindly adopting cloud computing or seeking to fraudulently receive subsidies. At the same time, this would deepen the use of cloud computing by enterprises. In addition, Breznitz et al. [20] found that venture capital could increase the number of early adopters of cloud services and deepen users' degree of application to cloud computing. We speculate that this may be because there are a series of uncertain risks in the application of new technologies, and the support of external funds can reduce the losses caused by risks. Thus, the government can encourage enterprises to adopt cloud computing by establishing an industry fund.

For non-fiscal subsidy measures, the results indicate that publicity activities can promote the diffusion of a cloud ecosystem. The number of such activities could deepen the application of cloud computing in enterprises, and the number of participants in them could affect the breadth of cloud ecosystem diffusion. Therefore, the government could make a targeted choice of whether to increase the number of activities or the number of participants when organizing publicity activities. If its purpose is to increase the breadth of diffusion of a cloud ecosystem, the number of participants in activities should be increased as much as possible. If the depth needs to be developed, the number of activities for specific enterprises can be increased.

The results of the impact of core enterprises on cloud ecosystem diffusion can provide practical references in two aspects. First, the results indicate that the increases of benchmarking enterprises in the region can prompt enterprises to adopt cloud computing, while the bonus of benchmarking enterprise influences the diffusion of cloud computing only slightly. Therefore, the government should focus on issuing more flexible and appropriate criteria for benchmarking enterprises instead of increasing the number of financial incentives that enterprises have little chance of obtaining. Core enterprises with strong capabilities should also strive to become benchmarking enterprises, which can not only enhance their core competitiveness but also promote the diffusion of a cloud ecosystem.

Second, although the results show that the number of enterprises that have built cloud platforms in the region has little impact on the diffusion of cloud ecosystems, this does not mean that enterprises do not need to establish cloud platforms belonging to their industries. We investigate such enterprises and find that once such platforms are empowered and available, they will obviously affect the application depth of cloud computing within the industrial chain. For example, many enterprises are empowered by the platforms provided by Handu. These enterprises have a comprehensive and in-depth application of cloud computing and can even realize the intelligent application of cloud computing at an early stage. Handu also built its own business ecosystem and enhanced its own competitiveness through its platform. Thus, if core enterprises can obtain financial support from the government and obtain R&D capability, building a cloud platform according to industry characteristics will be a reasonable choice.

ACKNOWLEDGMENTS

This research was supported by the Shandong Province Higher Educational Science and Technology Program of China (No. J17KA066), and the Key Research & Development Program of Shandong Province of China (No. 2019GSF108222).

REFERENCES

- [1] Gill S S, Tuli S, Xu M, et al. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, 8, 100118.
- [2] Khayer A, Talukder S, Bao Y, Hossain N. (2020). Cloud computing adoption and its impact on SMEs' performance for cloud supported operations: a dual-stage analytical approach. *Technol Soc*, 60, 101225.
- [3] Alkhater N, Walters R, Wills G. (2017). An empirical study of factors influencing cloud adoption among private sector organisations. *Telematics & Informatics*, 35(1), 38-54.

- [4] DiMaggio P J, Powell W W (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, 147-160.
- [5] Wuttke D A, Rosenzweig E D, Heese H S. (2019). An empirical analysis of supply chain finance adoption. *Journal of Operations Management*, 65(3), 242-261.
- [6] Kung L A, Cegielski C G, Kung H J. (2015). An integrated environmental perspective on software as a service adoption in manufacturing and retail firms. *Journal of Information Technology*, 30(4), 352-363.
- [7] Kotabe M, Jiang C X, Murray J Y. (2017). Examining the complementary effect of political networking capability with absorptive capacity on the innovative performance of emerging-market firms. *Journal of Management*, 43(4), 1131-1156.
- [8] Martins R, Oliveira T, Thomas M A. (2016). An empirical analysis to assess the determinants of SaaS diffusion in firms. *Computers in Human Behavior*, 62, 19-33.
- [9] Alzadjali K, Elbanna A. (2020). Smart institutional intervention in the adoption of digital infrastructure: the case of government cloud computing in oman. *Information Systems Frontiers*, 22.
- [10] Kharuddin S, Foong S Y, Senik R. (2015). Effects of decision rationality on erp adoption extensiveness and organizational performance. *Journal of Enterprise Information Management*, 28(5), 658-679.
- [11] Melin U, Sarkar Pradip K, Young Leslie W. (2020). To couple or not to couple A case study of institutional legitimacy relating to SaaS applications in two universities. *INFORMATION TECHNOLOGY & PEOPLE*, 33(4), 1149-1173.
- [12] Oliveira T, Martins R, Sarker S, Thomas M, Popovič Aleš. (2019). Understanding saas adoption: the moderating impact of the environment context. *International Journal of Information Management*, 49, 1-12.
- [13] Kushida K, Murray J, Zysman J. (2012). The gathering storm: analyzing the cloud computing ecosystem and implications for public policy. *Communications & Strategies*, 1, 63-85.
- [14] Puthal D, Sahoo B P S, Mishra S, et al. (2015). Cloud computing features, issues, and challenges: a big picture[C]//2015 International Conference on Computational Intelligence and Networks. IEEE, 116-123.
- [15] Ni Yuan. (2019). Core Enterprise Network Capability and Cluster Collaborative Innovation: A Double Moderating Effect Model with Intermediary. *Management Review*, 31(12), 85-99. (in Chinese)
- [16] Iansiti, M., Levien, R. (2004). Strategy as ecology. *Harvard business review*, 82(3), 68-78, 126.
- [17] Trigueros-Preciado S, Pérez-González D, Solana-González P. (2013). Cloud computing in industrial SMEs: identification of the barriers to its adoption and effects of its application. *Electronic Markets*, 23(2), 105-114.
- [18] Chen Xu. (2005). Research on Technology Innovation Diffusion Based on Industrial Cluster. *Chinese Journal of Management*, 3, 333-336. (in Chinese)
- [19] Cui Yi, Shi Lu. (2015). Research on the Diffusion Path of Logistics Innovation in Supply Chain Network. *Journal of Commercial Economics*, 30, 29-31. (in Chinese)
- [20] Breznitz D, Forman C, Wen W. (2018). The Role of Venture Capital in the Formation of a New Technological Ecosystem: Evidence from the Cloud. *MIS Quarterly*, 42(4).

Full Research Paper**A Study on the Characteristics of Regional Social Network for Patent****Technology Transfer of China's "985 Project" universities***Xiaoli Li¹, Zihan Peng¹, Hao Zhou^{2*}*¹School of Management, Wuhan Textile University, China²School of Law and Business, Wuhan Institute of Technology, China

Abstract: Using social network analysis (SNA) method, this study analyzes the regional social network characteristics of patent technology transfer of China's "985 Project" universities. In analyzing university's patent technology transfer network, density, centrality and the number of network edges, etc. are calculated, and the block results are obtained using the Concor method. The study shows: There are obvious differences in outdegree centrality and indegree centrality of each region where "985 Project" universities are located, and the central and western regions are relatively backward; Block 1 and Block 2 are the most active regions of university patent technology transfer activities; The linear fitting coefficient demonstrates that the stronger is constant stronger, and it is particularly important to improve the ability of patent technology transfer in the central and western regions; Betweenness centrality of Liaoning, Tianjin, Fujian, Hubei and Hunan are relatively high to play a role in radiation of technology transfer in the neighboring provinces, etc. All the above results provide directions for improving China's university patent technology transfer.

Keywords: University patent technology transfer, technology transfer, regional network

1. INTRODUCTION

Universities have increasingly been seen as a driver of economic growth (Mowery et al., 2001) ^[1], especially in emerging technology (Etzkowitz et al., 2000; Vallas&Kleinman, 2008) ^[2-3] while Patent is a technical right protected by official certification, which has a series of advantages such as strong authority, complete information and large amount of data ^[4], so it is of great significance to study the related activities of university patent technology to promote economic growth. According to the subject of patent technology implementation, university patent technology transfer can be divided into two types: One is to transfer to derivative enterprises; the other is to transfer to other enterprises and individuals, which specifically includes the transfer of patent application rights, the transfer of patent rights, and the licensing of patented technology. This research mainly examines the transfer of university patent technology to the third party, namely the second type. Based on literature survey, the research of university patent technology transfer focuses on the motivation and type of transfer between universities and enterprises (Chang, 2017) ^[5], or analyzes the patent cooperation behavior based on small sample of universities from the middle and micro levels of patent cooperation behavior ^[6] while there are few studies on the specific characteristics of China university patent technology transfer. In addition, the "985 project" represents the highest level of university scientific research in China, so the research on patent transfer of "985 Project" university can better reflect the implementation efficiency of patent technology in China, and is conducive to exploring the experience of university technology transfer in China. This paper attempts to analyze the specific situation of university patent transfer activities in China from the perspective of "985 project" universities. Although Renmin University of China and Central University for nationalities are also included in the "985 Project", these two universities focus on humanities disciplines and have less patents. Therefore, the patent transfer of China "985 Project" universities involved in this study does not include these two universities.

* Corresponding author. Email: zhouhaoli@163.com (Hao Zhou)

2. RESEARCH DESIGN

2.1 Research method

Social network analysis is a quantitative analysis method developed by sociologists based on mathematical methods and graph theory. It has been developing rapidly since 1990s. Social network similar to a map is a collection of actors and their relationships. Each actor is a single node in the network map. The association between actors forms the connecting line in the network diagram. With the development of social network analysis, social network analysis tools are also constantly introduced and optimized. In particular, Ucinet software developed by Linton Freeman provides a convenient application tool for social network analysis. Ucinet is a powerful social network analysis software. It was originally written by Linton Freeman, an authoritative scholar of social network research at the University of California, Irvine. Later, it was mainly maintained and updated by Steve Borgatti of Boston University and Martin Everett of University of Westminster. Ucinet software can read text files, KrackPlot, Pajek, Negopy, VNA and other files, and can process 32767 network nodes. At the same time, it integrates Netdraw, which includes one-dimensional and two-dimensional data analysis, and mage which is being developed for 3D presentation and analysis, as well as Pajek's Free application program for large-scale network analysis. In recent years, social network research has also been applied in the field of patent measurement. Patent technology transfer network relationships reflects the set of actors and their relationships. Using social network analysis method and related tools can deeply understand the overall network structure and individual network characteristics of patent technology transfer, which is helpful to provide ideas for expanding patent technology transfer channels.

2.2 Data source

Due to the confidentiality of technology transfer contract, it is difficult to obtain the relevant information of patent transfer from universities. However, transfer of patent technology need to be filed or registered with the State Intellectual Property Office, so the relevant information of university patent transfer can be obtained by querying the Patent Affair Database of State Intellectual Property Office, including patent transfer type, transfer object, transfer time, etc., which constitute the main data source of this study. In order to ensure the accuracy of the data, this study retrieved patent transfer from the Patent Affairs Database of State Intellectual Property Office for patent transfer of individual "985 Project" University from 2005 to 2015, including patent right transfer, patent application right transfer and patent licensing respectively. Finally, the original data were aggregated, cleaned and filtered, and 9018 records of patent transfer of "985 Project" University were obtained. The data for this study covers the time period 2005-2015.

2.3 Indicator setting

Network relationship features include the overall network characteristics and individual characteristics of actors. The whole network analysis regards the actor set in the network as a whole, and analyzes the attribute characteristics and rules of the network as a whole. Indicators that measure the overall structural characteristics of social networks generally include indicators such as network centrality, network density, and network cohesion. Centrality indicators are commonly used to measure the influence of individual actors^[7], and can also evaluate superiority or privilege of individual actors in the network, and even prestige^[8]. Freeman (1979) once proposed such indicators as degree centrality and betweenness centrality to describe the centrality of actors in the network^[9]. Wassermann & Faust (1994) proposed that the simplest intuitive definition of the centrality of an actor is that the core member must be the most active and has the most connections with other actors in the network^[10]. The specific indicators are set as follows:

2.3.1 Network centrality

Network centrality is the degree to which a group of actors in a network link up the network around a central point. It depicts the overall centrality of the network and reflects the structural center of the graph (Liu Jun, 2004). The steps to calculate the concentration degree are as follows: first, find the maximum centrality

value (C_{max}) in the network graph, then calculate the difference between the value and the centrality (C_i) of any other node, so as to obtain multiple "differences"; then calculate the sum of the differences; finally divide the sum by the maximum possible value of the sum of the differences^[11]. The network concentration is denoted as C . The higher C is, the higher the degree of centralization of university patent technology transfer network structure is, and the larger the regional gap is; the smaller C is, the lower the centralization degree of university patent technology transfer network structure is, and the patent technology transfer is more balanced among regions. The calculation formula of the network centrality C is as follows in equation (1):

$$C = \frac{\sum_i^n (c_{max} - c_i)}{\max \left[\sum_i^n (c_{max} - c_i) \right]} \quad (1)$$

2.3.2 Network density

Network density reflects the closeness of the relationship between actors in social network. The greater the network density, the closer the relationship between actors. In graph theory, density represents the compactness of each node in a graph. If there are n nodes in the directed network graph, there may be $n(n-1)$ lines at most. The density of a graph is defined as the ratio of the total number of relationships actually in the graph to the maximum possible total number of relationships. The calculation formula is as follows in equation (2), denoted as D , where l represents the actual number of lines in the graph, i and j represent the members in the network, X_{ij} represents the relationship between members i and j , if there is a relationship between them, it is 1, and if there is no relationship, it is 0. The value range of network density is $[0, 1]$. A complete graph is a graph in which all vertices are adjacent, and its density is 1.

$$D = \frac{l}{n(n-1)} = \frac{\sum_{i=1}^n \sum_{j=1}^n X_{ij}}{n(n-1)} \quad (2)$$

2.3.3 Block model

The theory of block model was first put forward by White & Burman and other scholars in 1976. Block model is mainly used to describe the location of network actors, and it is an algebraic analysis of social role description^[12]. The block model divides the actors in the network into different subgroups according to density, distance, or other criteria, usually called blocks, and compares the tightness within the block and the sparseness outside the block. Generally speaking, the analysis of the block model includes two steps: dividing the block and determining the value of the block. Firstly, the actors are assigned to each block of the network by using Concor or hierarchical clustering method. Then, according to some criteria, such as complete fitting, 0-block standard, 1-block standard, α -density index, etc., the values are calculated in each block, so as to understand the overall structure of the network. The α -density metric method is the most commonly used, where α is the critical density value and is often expressed by the average density value of the network, which is also the evaluation criteria used in this study.

2.3.4 Degree centrality

Degree centrality is also known as individual centrality. The concept of degree centrality comes from the concept of "Star" in sociometrics, which indicates the relationship between the actor and other actors in the network. The greater the value of degree centrality, the more connected the actor is to other actors. Degree centrality is an indicator at the individual level. The measurement of degree centrality of an actor is the number of external contacts of the actor, which is the so-called node degree in graph theory. In directed network relations, degree centrality can be divided into outdegree centrality and indegree centrality. The calculation formula of outdegree centrality and indegree centrality is respectively shown in equation (3) and (4). Where a_{ij} represents the number of contacts that actor i points to other actor j ; a_{ji} represents the number of contacts that other actor j points to actor i . In the patent transfer network, indegree centrality and outdegree centrality

respectively measure the important role of regions in the transfer of university patent technology.

$$Outdegree_i = \sum_{j=1}^n a_{ij} \quad (3)$$

$$Indegree_i = \sum_{j=1}^n a_{ji} \quad (4)$$

2.3.5 Betweenness centrality

Betweenness centrality, referred to as intermediary degree, is equivalent to the concept of "bridge" in social metrology. If an actor withdraws from the network relationship, the original sub network of the actor will be disconnected, resulting in other actors in the sub network unable to get contact with each other. If an actor has a high degree of mediation, it means that the actor acts as a middleman to connect many unrelated actors and build a bridge for them. Suppose that the number of shortcuts between actor j and actor k is represented by g_{jk} , where the number of shortcuts between j and k passing through actor i is represented by $g_{jk(i)}$; the ability of actor i to control the interaction between j and k is represented by $b_{jk(i)}$, that is, the probability of i being on the shortcut between j and k , then $b_{jk(i)} = g_{jk(i)}/g_{jk}$. Adding the betweenness centrality of all the actors corresponding to the actor i in the network relationship graph to obtain the betweenness centrality of the actor i , the formula is shown below in equation (5). In the university patent transfer network of this study, betweenness centrality is used to measure the position and dominance of a region in the university patent transfer network. If the betweenness centrality of a certain region is high, it means that this region acts as a bridge between regions in university patent transfer activities [13].

$$Betweenness_i = \sum_j^n \sum_k^n b_{jk}(i), j \neq k \neq i, \text{并且 } j < k \quad (5)$$

3. CONSTRUCTION OF PATENT TECHNOLOGY TRANSFER NETWORK

In this study, the collected patent technology transfer data of China's "985 Project" universities from 2005 to 2015 were classified and counted according to the regions where the transferor and the transferee of the patent technology were located, and an adjacency matrix of 38*38 was constructed, and part of this matrix is as shown in Table 1. The "rows" in the matrix are 33 regions and 5 countries that accept the patent technology of China's "985 Project" universities, and the "columns" in the matrix represent 18 regions where China's "985 Project" universities transferred the patented technology out, and 13 regions or countries added in order to construct the adjacency matrix. The grid value at the intersection of the rows and columns in the adjacency matrix represents the number of patent technologies transferred to the regions in columns from the regions in rows where "985 Project" universities are located; the grid value of the diagonal of the adjacency matrix represents the number of patent technologies implemented within the region where the "985 Project" universities are located. The constructed 38*38 adjacency matrix was input into Ucinet software and the data format was converted. Then NetDraw was used to visualize the regional network relationship of the patent technology transfer for "Project 985" university region as shown in Figure 1. The square nodes in the figure represent regions where the patent technology is transferred out or in. Node size in Figure 1 is set according to the number of patent technology transfer of that region received from "Project 985" universities, that is, the larger the node is, the more patents received from "Project 985" universities. The connection between nodes represents the strength of patent technology transfer between regions, that is, the thicker the connection is, the more frequently the patent technology transfer occurs between regions. The arrows in the line point to the regional transferee of the patent technology. In order to reflect more intuitively the patent technology transfer status of the regions to which the "985 Project" universities belong and other regions, the low intensity of patent technology transfer (i.e. the frequency of patent technology transfer is less than 10 times) is filtered out in

4. THE EMPIRICAL ANALYSIS

4.1 Analysis of the overall network characteristics

Ucinet software was used to calculate the overall characteristic parameters of the regional network of patent technology transfer of "985 Project" universities, including density, number of network edges and network centrality, as shown in Table 2 which shows that except for a few years, the network density and the number of network edges showed an overall rising trend from 2005 to 2015, indicating that the network connection of patent technology transfer was getting closer and closer, and the frequency of patent technology transfer was increasing. Network indegree centrality reflects the concentration of patent technology introduction regions. Table 2 shows that from 2005 to 2015, the regions in which the patent technology is introduced into are relatively dispersed, and they have a trend of centralization from 2010 to 2014, but the distribution concentration of the regions was decreased significantly in 2015, suggesting that some regions firstly introduced university patent technology, and then drove other regions to actively adopt new technology. Therefore, China's introduction of patent technology stepped in a new level in 2015, and this turning point is relatively obvious. However, this conclusion needs to be further verified by future follow-up studies. Network outdegree centrality reflects the concentration of patent technology export regions where the "985 Project" universities are located and this indicator did not show obvious fluctuations from 2005 to 2015, which indicates that the patent technology transfer of each "985 Project" university has a balanced increase in horizontal implementation, and each "985 Project" university attaches more importance to university patent technology transfer activities.

Table 2. Overall network parameters by year

Year	Network density	number of network edges	indegree centrality	outdegree centrality
2005	0.0135	85	3.273%	3.009%
2006	0.0121	51	3.351%	4.045%
2007	0.0185	93	2.514%	3.155%
2008	0.1202	352	2.802%	2.907%
2009	0.2617	820	4.102%	2.690%
2010	0.3435	998	6.099%	4.526%
2011	0.3713	1135	4.621%	4.139%
2012	0.3841	1118	7.205%	4.020%
2013	0.4211	1615	5.104%	4.385%
2014	0.4701	1443	6.727%	2.660%
2015	0.3606	1334	2.982%	3.553%

According to the network management of regions of patent technology transfer from 2005 to 2015, the block results are obtained by using CONCOR method. And the regions involved in the patent technology transfer network of "985 Project" universities in China from 2005 to 2015 can be divided into 6 blocks in total. The first block consists of 7 regions, specifically Jiangsu, Shanghai, Hubei, Beijing, Liaoning, Heilongjiang and Shanxi; The second block consists of four regions: Chongqing, Sichuan, Tianjin and Zhejiang; The third block consists of three regions: The United Kingdom, Jilin and Hunan; The fourth block includes two regions, specifically Gansu and Japan; The fifth block consists of six regions: Guangdong, Qinghai, Shandong, Fujian, Ningxia and the United States. Block 6 consists of 11 regions, specifically Anhui, Hebei, Yunnan, Guangxi, Shanxi, Xinjiang, Hainan, Jiangxi, Inner Mongolia, Henan and Taiwan. Block 7 consists of four regions, specifically Korea, Germany, Guizhou, and Xizang. Block 8 comprises a region, specifically Hong Kong. Meanwhile, the density and interval density of a single block are calculated by CONCOR method, and the overall density of the network is calculated, as shown in Figure 3 (a). Among them, the internal interaction between Block 1 and Block 2 is relatively frequent. Secondly, the transfer out or introduction of patent technology between Block 1 and Block 2 is relatively active. In addition, Block 1 and Block 2 each have a

higher frequency of patent technology transfer out to Block 5. The other blocks have no strong interaction and loose structure. The whole network density α is 2.6947, Considering that the foreign regions involved in the university patent technology transfer network are included, and the amount of university patent technology transferred to foreign countries is very small at present, this part of data reduces the entire network density value, so the obtained α is corrected by rounding, and the value of α is adjusted to 3 as a critical value to get a 0-1 block image matrix shown in figure. 3 (b). The image matrix is calculated as shown in the following formula in equation (6). A highly generalized simplified network diagram is drawn according to the image matrix information, and the inter-block structure is shown in Figure. 3 (c). It can be seen that Block 1 and Block 2 are the most active regions in university patent technology transfer activities. Firstly, Block 1 and Block 2 have a high level of patent technology transfer activity within their own block; In addition, there is a lot of interaction between Block 1 and Block 2, with Block 1 having more patent technology transfer output to Block 6 and Block 5, and Block 2 having more output to Block 5. The activity of Block 1 and Block 2 is directly related to the number of "project 985" universities and the capacity of the patent technology transfer in these two blocks. In addition, these regions are economically developed and have a strong ability to absorb the patent technology of universities. Moreover, the long-term cooperative relationship of patent technology transfer is also conducive to solidify the inter-block and intra-block trading relationship. A simplified diagram of the network provides direction for finding potential areas to target for patent technology transfer.

$$B_{ij} = \begin{cases} 1, & \text{if } a_{ij} > \alpha \\ 0, & \text{else} \end{cases} \quad (6)$$

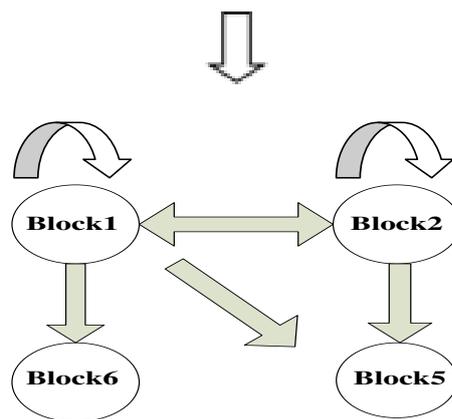
Block	1	2	3	4	5	6	7	8	
1	30.024	11.071	1.667	0.357	14.000	3.013	0.179	0.714	
2	12.464	11.167	1.000	0.000	11.750	2.977	0.438	0.000	
3	2.952	2.417	0.167	0.000	2.389	1.030	0.167	0.000	
4	0.071	0.875	0.000	0.000	0.083	0.048	0.000	0.000	
5	2.929	1.208	0.722	0.000	1.233	0.500	0.208	0.000	
6	0.078	0.023	0.000	0.000	0.030	0.000	0.000	0.000	
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
								Network density=	2.6947

(a)Regional network block model density matrix of "Project 985" university patent technology transfer



Block	1	2	3	4	5	6	7	8
1	1	1	0	0	1	1	0	0
2	1	1	0	0	1	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

(b)Regional network image matrix of university patent technology transfer



(c) Simplified map of the regional network of patent technology transfer of "Project 985" universities

Figure 3. Regional network block model of university patent technology transfer

4.2 Analysis of characteristics of network actors position

Ucinet software is used to respectively calculate indegree centrality and outdegree centrality of the individual "985 Project" university patent technology transfer region, as shown in Table 3. There are obvious differences in both outdegree centrality and indegree centrality among different regions, which proves that there is a serious imbalance in patent transferring behavior in different regions [14]. In terms of regional outdegree centrality, the number of patent technology transfer from Beijing and Shanghai has respectively exceeded 1000. Secondly, Jiangsu, Zhejiang, Guangdong and other regions also rank higher. The difference of this indicator is firstly directly related to the number of "project 985" universities in the output regions, such as Beijing with a total of 6 "985 project" universities, Tsinghua university, Peking University, Beijing University of Aeronautics and Astronautics, Beijing Institute of Technology, Beijing Normal University, China Agricultural University; Shanghai with a total of 4 "985 project" universities, Shanghai Jiaotong university, Tongji University, Fudan University and East China Normal University; Xi 'an with a total of 3 "985 project" universities. Secondly, the regional outdegree centrality is directly related to the patent strength and patent technology transfer ability of "985 project" universities, and regional support, such as Jiangsu, Zhejiang and Guangdong where "985 project" Universities are not concentrated, but Southeast University, Zhejiang University and South China University of Technology in those regions have outstanding patent output and technology trading ability. At the same time, Jiangsu, Zhejiang and Guangdong as economically developed regions attach more importance to technology transfer and industrial innovation, which leads to the higher outdegree centrality. As far as the regional indegree centrality is concerned, this indicator directly reflects the state of patent technology introduction of various regions in China from "985 project" universities. Among them, Jiangsu has the largest number of patent technology introduction with 2029, followed by Beijing, Guangdong, Zhejiang and Shanghai with a relatively large number of patent introductions; while the number of patent technology introduction from "985 project" Universities is relatively small in China's central and western regions and other regions. The central and western regions have relatively weak independent innovation capacity and relatively backward economic development. Therefore, these regions should actively introduce and absorb the advanced technology of other regions to catch up with and surpass in economy. In addition, the scope of patent technology transfer activities in China's "985 Project" universities is still relatively small, as well as the geographical location and degree of technological innovation and other factors lead to the less demand for patents of "985 Project" university from Taiwan, Hong Kong, Macao and foreign countries.

On the whole, the indegree centrality and outdegree centrality of "985 Project" university patent technology transfer reflect that there are obvious differences in the export and introduction of regional patent technology in

general, among which the economically developed regions have outstanding advantages and the number of patent technology export and introduction is the most. But is the gap of the export or introduction of patent technology transfer among regions narrowing, or is it widening? In this study, Ucinet software is used to calculate the annual indegree centrality and outdegree centrality of each region from 2005 to 2015, and this value is taken as the dependent variable of each year variable. The linear fitting coefficient of indegree centrality and outdegree centrality of each region is obtained by using slope function in Excel ^[15], as shown in Table 3. The results show that the export and introduction of patent technology in most regions show an increasing trend, and the fitting coefficient of out of center degree is high in Beijing, Jiangsu, Shanghai, Zhejiang and Chongqing. It can be seen that the export of patent technology of "985 project" universities in these regions grows most rapidly, and the fitting coefficient of outdegree centrality is high in Jiangsu, Beijing, Guangdong, Zhejiang and Shanghai, so it can be seen that the patent technology export from these regions is increasing, while the central and western regions play little role in promoting the transfer of patent technology, and the increasing rate is significantly smaller than that in economically developed regions. It can be seen that in terms of university patent technology transfer in different regions, China shows the situation that the strong regions always keep strong, so it is particularly important to improve the patent technology transfer ability in the central and western regions.

The betweenness centrality of each region from 2005 to 2015 is calculated using Ucinet software as shown in Table 3. Betweenness centrality can be used to measure the intermediary role of a region in the university patent technology transfer network. If betweenness centrality of a region is higher, it indicates that the region plays a role of bridge between regions in the university patent technology transfer network activities, and the radiation effect of geographical location is obvious. The difference of betweenness centrality of regional patent technology transfer is directly related to the ability of regional patent technology transfer, followed by the geographical location advantage. Technology spillover has the characteristics of geographical space, and it is easy to form the technology divergence area in the neighboring provinces or within 800 kilometer ^[16]. Table 3 shows that Zhejiang, Beijing, Shanghai and Tianjin have the highest degree of betweenness centrality, which indicates that these regions have strong technology spillover capacity and play a key role in promoting the whole country's technology transfer capacity. In addition, although Liaoning, Tianjin, Fujian, Hubei and Hunan do not have advantages in outdegree centrality and indegree centrality, the degree of betweenness centrality of these provinces is higher than that of other provinces which shows that these provinces play a greater role in the bridge of regional patent technology transfer, and play a radiation role in the technology transfer of surrounding provinces. In order to promote the ability of patent technology transfer in the central and western regions, these provinces can be used as the engine regions of patent technology transfer.

Region	Outdegree	Indegree	Outdegree coefficient	Indegree coefficient	Betweenness	Region	Outdegree	Indegree	Outdegree coefficient	Indegree coefficient	Betweenness
Beijing	1652	1191	34.92	23.87	37.327	Hebei	-	83	-	1.25	-
Shanghai	1124	710	16.85	10.76	37.073	Jiangxi	-	29	-	0.78	-
Jiangsu	848	2029	20.6	47.3	24.187	Neimenggu	-	21	-	0.42	-
Zhejiang	808	898	13.18	11.5	66.54	Shanxi	-	21	-	0.53	-
Shanxi	709	200	16.23	4.83	6.903	Xinjiang	-	10	-	0.15	-
Guangdong	624	1183	7.31	15.1	25.945	Yunnan	-	30	-	0.33	-
Hubei	513	323	8.57	5.4	12.189	Guangxi	-	53	-	0.66	-
Chongqing	489	270	12.15	6.12	6.694	Henan	-	40	-	0.91	-
Tianjin	406	229	4.6	3.34	33.037	Xizang	-	9	-	0.31	-
Silongjia	401	189	8.1	4.72	16.468	Guizhou	-	6	-	0.009	-
Sichuan	399	300	5.71	4.95	4.53	Ningxia	-	1	-	0.009	-
Hunan	359	273	5.84	4.93	19.775	Qinghai	-	11	-	0.35	-
Shandong	240	330	4.06	5.94	5.837	Hongkong	-	5	-	-0.15	-
Liaoning	189	155	3.75	3	36.197	Taiwan	-	1	-	0.009	-
Jilin	98	72	0.973	0.9	3.278	South Korea	-	1	-	0.018	-
Fujian	91	140	2.36	2.82	26.772	Germany	-	3	-	-0.045	-
Anhui	35	152	-0.2	3.66	0.948	America	-	3	-	0.054	-
Gansu	26	20	0.7	0.68	0.3	Britain	-	2	-	0.091	-
Hainan	-	17	0	0.21	-	Japan	-	1	-	0	-

5. CONCLUSIONS

In this study, Ucinet software is used to calculate the relevant parameters of regional network of "985 Project" university patent technology transfer. In the analysis of the overall network characteristics, we calculate the density, the number of network edges, the degree of network centralization and other overall characteristic parameters, and use the Concor method to calculate the simplified network diagram of the university patent technology transfer block model. In the analysis of the individual characteristics of the network actor position, we calculate outdegree centrality and its linear fitting coefficient for the export regions of "985 Project" university patent technology transfer, and indegree centrality and its linear fitting coefficients for the introduction regions of "985 Project" university patent technology transfer.

The results show that: ①during 2005-2015, the network density and the number of network edges showed an overall upward trend, and the indegree centrality shows that there was a centralization trend in the patent technology introduction region from 2010 to 2014, but the distribution concentration degree of patent technology introduction region decreased obviously in 2015; ②From 2005 to 2015, China's "985 Project" university patent technology transfer network involves a total of six blocks, in which Block 1 and Block 2 are the most active regions of university patent technology transfer activities. Block model results indicate the possible direction of patent technology transfer for each region; ③There are obvious differences in outdegree centrality and indegree centrality of each region, and the central and western regions are relatively backward; ④The linear fitting coefficient shows that for China's university patent technology transfer, the stronger is constant stronger, and it is particularly important to improve the ability of patent technology transfer in the central and western regions; ⑤The betweenness centrality parameter shows that although Liaoning, Tianjin, Fujian, Hubei and Hunan did not have advantages in outdegree centrality and indegree centrality, the betweenness centrality of these provinces was relatively high, which indicates that these provinces played a larger role as a bridge for regional patent technology transfer and played a radiation role in the technology transfer of surrounding provinces, so it is necessary to promote the patent technology transfer in Central and Western regions of China. These provinces can be regarded as the engine areas of patent technology transfer.

ACKNOWLEDGEMENT

This research was supported by the Educational Science Planning of Hubei Province under Grant 2020GA036, and supported by the Education Department of Hubei Province under Grant 20Y083 and DSS20191201.

REFERENCES

- [1] Mowery D, Nelson R, Sampat B, et al. The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh-Dole Act of 1980 [J]. *Research Policy*, 2001, 30(1): 99-119.
- [2] Etzkowitz, H., Webster, A., Gebhardt, C., Terra, B.R.C. The future of the university and the university of the future: evolution of ivory tower to entrepreneurial paradigm[J]. *Research. Policy*, 2000, 29 (2), 313-330.
- [3] Vallas, S.P., Kleinman, D.L. Contradiction, convergence and the knowledge economy: the confluence of academic and commercial biotechnology[J]. *Socio-Economic Review*, 2008, 6(2), 283-311.
- [4] Song Lan, Qin Yunlong, Li Minsi. A study on patent licensing network of the six central provinces under "the Belt and Road" Strategy in China[J]. *Science & Technology Progress and Policy*, 2013,34(14): 37-44. (in Chinese)
- [5] Chang S H. The technology networks and development trends of university-industry collaborative patents[J]. *Technological Forecasting and Social Change*, 2017, 118(5): 107-113.
- [6] Lu Yikai, Shi Liang, Zhang Shanshu. Study on the characteristic and case of university-involved patent collaboration based on the social network analysis[J]. *Journal of Modern Information*, 2015, 35(8): 69-74. (in Chinese)

- [7] Brass D J, Burkhardt M E. Centrality and power in organizations [J]. *Networks and organizations: Structure, form, and action*, 1992, 191(215):198-213.
- [8] He Jiayuan. A study of the social network of high school geography teachers' GIS study participation [D]. Taiwan University, 2012. (in Chinese)
- [9] Freeman L C. Centrality in social networks: conceptual clarification [J]. *Social networks*, 1978, 1(3): 215-239.
- [10] Wasserman S, Faust K. *Social network analysis: Methods and applications* [M]. Cambridge university press, 1994.
- [11] Liu Jun. *Introduction to social network analysis* [M]. Social Science Literature Press, 2004. (in Chinese)
- [12] White H C, Boorman S A, Breiger R L. Social structure from multiple networks. I. Blockmodels of roles and positions [J]. *American Journal of Sociology*, 1976, 81(4): 730-780.
- [13] Liu Jun. *Holistic net analysis handout: UCINET Software practical guide* [M]. Shanghai: Gezi Press, 2009. (in Chinese)
- [14] Wang Yuandi, Chen Fengzhen, Li Min. A Study on patent licensing spatial distribution and flow networks of research institutions in China [J]. *Journal of Intelligence*, 2016, 35(1): 57-63. (in Chinese)
- [15] Zhao Sangmei, Si Hongmei, Du Huadong. Analysis of cross-regional technology transfer based on network model [J]. *R&D Management* 2013, 25(5): 54-61. (in Chinese)
- [16] Fu Miao. Geographical distance and technological spillover effects: a spatial econometric explanation of technological and economic agglomeration phenomena [J]. *China Economic Quarterly*, 2009, 8(4): 1549-1566.(in Chinese)

Full Research Paper

Leveraging Linguistic Signaling to Prompt Feedback in Open Innovation Communities

Suya Hu¹, Di Xu^{2*}, Yan Li³

¹School of Management, Xiamen University, Xiamen, 361005, China

²School of Management, Xiamen University, Xiamen, 361005, China

³School of International Relations, Xiamen University, Xiamen, 361005, China

Abstract: Organizations are increasingly using open innovation communities (OICs) to gain external ideas. The success of OICs in promoting innovation, however, depends not just on posting activity by participants, but, crucially, on whether or not responses are subsequently received. Drawing on signaling theory, our study tries to explore how to leverage linguistic signals expressed in idea descriptions to influence feedback from two key parties: the moderator and peers. Linguistic features are divided into affective signaling (i.e., linguistic style matching, negative emotion, and impoliteness) and informative signaling (i.e., post length and quality). The research model is empirically tested on a large dataset collected from the Huawei community. Results show that both affective and informative signaling are effective in invoking feedback from the moderator. We also find that only negative emotion is positively associated with feedback from peers, while the effects of other signals show different trends. This research provides practical insights into how to maintain the viability of OICs.

Keywords: feedback, signaling theory, ideas, open innovation communities

1. INTRODUCTION

With the paradigm of Open Innovation (OI) ^[1], organizations often have reached outside of their boundaries to elicit ideas from large and diverse crowds. Among various approaches to obtaining external ideas, Open Innovation Communities (OICs) have gained great popularity. Indeed, examples of OICs can be found in different industries such as Lego, Salesforce, and Huawei. Not only the posted ideas are important for innovation, but feedback is also of vital importance ^[2]. Previous studies mainly concentrated on factors such as peer recognition, reciprocity, and community incentives on members' response behavior in OICs context. How the content itself impacts the willingness of participants to expend time and effort to respond is rarely examined. In light of these gaps and concerns, we focus on the antecedents of feedback and examine how these sociolinguistic factors exert influences on feedback behavior from different parties. Our first research question asks: *What kinds of linguistic characteristics embedded in idea posts can affect feedback from the moderators?* Apart from the moderator's feedback, innovation-related conversations among members also play an important role in facilitating the viability of the community. Consequently, the second research question is: *What kinds of linguistic characteristics embedded in idea posts can affect the amount of interactions from peers?*

2. THEORETICAL FOUNDATION AND HYPOTHESES

Signaling theory originates from the research of labor market ^[3]. In the context of OICs, ideators who submit innovation suggestions are signalers. Private information refers to the deliberate strategies that ideators use. Receivers are the moderator or those participants who have access to the information and thereafter potentially offer responses to them. Signals are how these idea providers linguistically communicate, which is carried out via online messages. In the presence of information asymmetries, ideators could choose deliberate linguistic signals to potentially under-inform others in an efficient and effective fashion to obtain a sufficient amount of desired interaction in return. Affective and informative signals are two major types of properties embedded in social media

* Corresponding author. Email: dxu@xmu.edu.cn (Di Xu), 17720170154888@stu.xmu.edu.cn (Suya Hu)

context content ^[4]. We evaluate the effects of three affective signals, namely *linguistic style matching*, *negative emotion*, and *impoliteness*. Similarly, two important informative signals in ideas are *post length* and *quality*.

3. RESEARCH METHODOLOGY

We obtained data from the OIC of Huawei (<https://club.huawei.com/>), which is one of the most important windows for Huawei company to listen to “the voice of the customer” (VOC). The crawler tool Web Scraper inserted in Google Chrome is used for data acquisition. We tracked all information available between March 2018 and March 2020. Our data features 95,668 ideas posted by 70,821 community members, 548,367 responses by peers and 29,544 responses by moderators. We employ two regression models: a logistic model for feedback from the moderator and a NB regression for feedback from peers. We use MLE method to estimate these coefficients and adopt robust standard errors clustered within each member to avoid heteroskedasticity.

4. RESULTS

In the affective signaling dimension, the coefficients of linguistic style matching, negative emotion and impoliteness are 0.348, 0.207 and -7.026, respectively, with statistically significant effects of all these explanatory variables ($p < 0.01$). Thus, hypotheses H1a, H2a and H3a are supported. In the informative signaling dimension, both post length ($\beta = 1.164$, $p < 0.01$) and quality ($\beta = 0.964$, $p < 0.01$) has a significant positive relationship with feedback from the moderator. Moreover, the coefficient value of squared post length is significant and negative ($\beta = -0.196$, $p < 0.01$). Thus, hypotheses H4a and H5a are supported. Negative emotion ($\beta = 0.047$, $p < 0.01$) are positively related to feedback from peers, while linguistic style matching ($\beta = 0.019$, $p > 0.1$) and impoliteness ($\beta = -0.557$, $p > 0.1$) are insignificant, therefore providing support for H2b and H3b, but not for H1b. Both post length ($\beta = -0.089$, $p < 0.01$) and quality ($\beta = -0.284$, $p < 0.01$) has a significant negative relationship with feedback from peers. However, the coefficient of squared post length is significant and positive ($\beta = 0.017$, $p < 0.01$). This result indicates a minor U-shaped relationship between post length and feedback from peers. Therefore, H4b and H5b are not confirmed.

5. CONCLUSION AND DISCUSSIONS

All of these findings are of great significance both in theory and in practice. Theoretically, we investigate a broad array of linguistic signals observed within ideas to better understand what motivates feedback in OIC environment. Moreover, we assess the distinct impacts of these signals on receiving feedback from two parties: the moderator and peers. From a managerial perspective, this study offers insights that can be utilized to better exploit of OICs. Our findings suggest that attention needs to be paid to the linguistic formulation of ideas. Moreover, if receiving feedback from the moderator is the goal of the participant composing a post, the sentiment of the idea, the style of the idea, the level of politeness in the idea, the length of the idea, and the quality of the idea all need to be carefully considered by the ideator. Future research could incorporate different dimensions such as informational or emotional feedback to better explore the mechanism of linguistic features on feedback. Besides, the generalizability of our study might be limited because we examined only a single community. Extending our research to other similar communities could validate and elaborate our findings.

REFERENCES

- [1] Chesbrough H W. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.
- [2] Zhu H, Kock A, Wentker M, et al. (2019). How does Online Interaction Affect Idea Quality? The Effect of Feedback in Firm-Internal Idea Competitions. *Journal of Product Innovation Management*, 36(1):24-40.
- [3] Spence M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3):355-374.
- [4] Chen L, Baird A, Straub D. (2020). A Linguistic Signaling Model of Social Support Exchange in Online Health Communities. *Decision Support Systems*, 130(3):113233.

Full Research Paper**Research on Early Warning of Non-performing Loans of Small and Medium-sized Micro-enterprises Under the Background of COVID-19****-- Taking XX Branch of N Bank as an Example**

Xinying Guo, Cui Guo, Junjie Zhou*, Guoxin Wang, Huanlian Yang*

Business School, Shantou University, Shantou, 515063, China

Abstract: In recent years, in order to solve the problem of small and medium-sized micro enterprises' loan difficulty, the major banks have developed some very relaxed microloan businesses under the guidance of the policy. These businesses help the vigorous development of small and micro enterprises, but also bring the problem of non-performing loan ratio increasing year by year. In 2020, due to the influence of COVID-19, this problem became particularly acute. The sharp increase in the non-performing loan rate posed new challenges to the credit risk management of banks. Do these enterprises have some warning features when they apply for loans before the epidemic? This problem is worthy of further exploration. Therefore, this paper used the LAD method to analyze the approval data of small and medium-sized micro enterprises (SMEs) approved for online loans in XX Branch of N bank in 2019, excavated the pattern characteristics of SMEs with non-performing loans after the epidemic, help banks with similar problems to improve the credit risk assessment mechanism, improve their early warning ability against the epidemic and other force majeure factors, and reduce the systemic financial risk Insurance, maintain financial stability.

Keywords: small and medium-sized micro-enterprises (SMEs), credit risk, logical data analysis(LAD), COVID-19.

1. INTRODUCTION

In recent years, with the continuous optimization of China's economic structure and the continuous promotion of industrial upgrading, SMEs, as the main force of China's national economic and social development, have become one of the new driving forces of economic development^[1]. By 2018, there are 40 million SMEs in China, accounting for 99.7% of the total number of enterprises, contributing 60% of China's GDP, driving the total number of urban employment population to exceed 80% of the domestic employment population, and contributing more than 70% of scientific and technological innovation^[2]. To keep the main body of the market active, China has paid great attention to the development of SMEs and issued a series of preferential policies to promote the solution of financing difficulty and the high financing cost of small and medium-sized micro-enterprises. Besides, in response to the state's support policies for small, medium, and micro-enterprises, many commercial banks have relaxed the threshold of credit for small, medium, and micro enterprises and issued a series of favorable policies. For example, simplify the loan process of SMEs, online application policy, mortgage-free loans, etc. Online application of credit certainly gives SMEs greater flexibility, but for banks, with the increasing credit scale of SMEs, the non-performing rate and overdue rate of credit assets are gradually rising. Especially when it comes to the COVID-19 which attacked human life in 2020, a large number of enterprises stopped production and shut down, the import and export trade was blocked, and the national economy turned into a depression period, and the development of enterprises was also greatly affected. The tertiary industry with the service industry as the main body is severely hit by the epidemic, and it is also the

* Corresponding author: Cui Guo, E-mail: cguo@stu.edu.cn ; Junjie Zhou, E-mail: jjzhou@stu.edu.cn

industry where SMEs are concentrated. According to the survey of Tsinghua Institute of economics and management (Figure 1-1), about 23% of the large enterprises are difficult to support for more than three months, but about 85% of the small, medium, and micro enterprises are unable to support for three months. Compared with large and medium-sized enterprises, small and medium-sized enterprises have weaker business sustainability. Small and micro businesses are facing more novel coronavirus pneumonia and more difficulty in the management of credit risk.

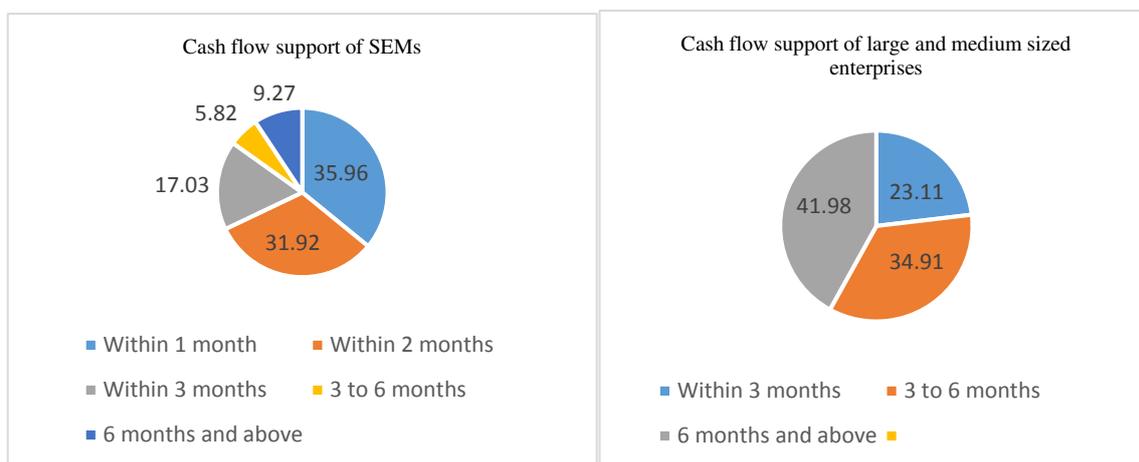


Figure 1-1. Data source: Tsinghua University of economics and management, 1435 samples of small and medium-sized enterprises, 212 samples of large enterprises

In general, the epidemic has undoubtedly brought a big one-time impact on the quality of bank credit assets, and the non-performing loan rate of SMEs has increased. According to the CBRC data, compared with the end of 2019, the non-performing loan ratio of the banking industry increased slightly. By the first quarter of 2020, the non-performing loan ratio of the banking industry will be as high as 2.04%, up 0.06 percentage points from the beginning of the year^[3]. For example, the non-performing loan ratio of the accommodation and catering industry, small and micro enterprises, and personal consumption loans increased by 0.14, 0.12, and 0.13 percentage points respectively. Taking the credit activities of SMEs in XX Branch of N bank in s city as an example, as of June 2020, there are three SMEs with bad repayment behaviors, which are distributed in wholesale and retail, residential decoration, cultural publishing, and other industries. These three enterprises were approved for online credit in 2019, and bad repayment behavior occurred in 2020 (after the outbreak of the COVID-19). To further identify the early warning characteristics of these enterprises in the loan application activities before the outbreak of the epidemic, we take the credit activity data of SMEs and use LAD to explore, in order to get the early warning characteristic pattern. It is expected that the research results will help to improve the bank credit risk assessment mechanism, reduce systemic financial risk and maintain financial stability. Providing valid suggestions is conducive to improving the early warning ability of banks to avoid the risk of force majeure.

2. LITERATURE REVIEW

2.1 Causes of non-performing loans

The related research on non-performing loans began in foreign countries. As non-performing loans are closely related to the operational risk of commercial banks, effective prevention, and control of non-performing loan subjects has become an indispensable part of financial construction. This paper discusses the causes of non-performing loans from the following macro and micro perspectives.

From a macro perspective, macroeconomic fluctuations will lead to changes in non-performing loans. When the economy deteriorates, the non-performing loan ratio will increase^[4]. Specifically, GDP growth rate, credit scale, market, and bank scale have a certain correlation with non-performing loans^[5]. When the GDP growth rate is high, the possibility of non-performing loans will increase. High interest, economic recession, and high inflation rates are also important reasons for the increase in non-performing loan rates. Also, there is a positive correlation between the unemployment rate and non-performing loans^[6]. GDP, consumer price index, one-year benchmark interest rate, and other factors also play an important role in the formation of non-performing loans, and GDP harms the formation of non-performing loans, while consumer price index and one-year benchmark interest rate have a positive impact on the formation of non-performing loans^[7]. The research on non-performing loans with borrowing enterprises as the research subject finds that the government intervention and the degree of Regional Marketization will affect the formation of non-performing loans to varying degrees; the greater the government's intervention on enterprises and the lower the degree of Regional Marketization, the higher the possibility of non-performing loans^[8]. In addition, when a bank enters a new area, due to the lack of local credit management knowledge, market share is also one of the possible reasons for the increase in the probability of non-performing loans^[9]. Secondly, there is a significant negative effect between bank profitability, inflation rate, and non-performing loans, while there is a positive correlation between cost income and non-performing loans^[10].

From a micro perspective, the interest rate of commercial banks is an important factor leading to the formation of non-performing loans. When the interest rate of banks increases, the possibility of non-performing loans will also increase^{[11][12]}. There is a positive and significant relationship between the ownership and scale of banks and non-performing loans^[13]. Bad credit evaluation, bad credit strategy, fraud of banking authorities, competition of other monetary funds and other factors related to banks, as well as enterprise management ability, monetary management ability, bookkeeping, and accounting ability and other factors related to customer activities, will also cause bad loans of commercial banks^[14]. For the bank credit approval personnel, their business ability also affects the formation of non-performing loans to a certain extent^[15]. And for the internal managers of banks, their risk management concept is not deep into the heart, which also leads to the phenomenon of paying more attention to loans than to management, which will accelerate the formation of non-performing loans^[16]. On the whole, these are reflected in the lack of internal management ability of banks, which aggravates the pressure of non-performing loans.

2.2 Research methods on the causes of non-performing loans

Among the different types of operational risks faced by the banking industry, credit risk is one of the most threatening risks to the bank. Credit risk is affected by non-performing loans. Therefore, it is necessary to consider the non-performing loans of the bank when properly determining the credit risk^[17]. The research methods for non-performing loans mainly focus on the case method, game method, and regression method^[17]. As non-performing loans are a part of credit risk, their research methods can also use credit risk research methods.

In terms of credit risk measurement, the New Basel Capital Accord advocates the use of the internal rating method. In this method, four main parameters are mainly considered: default rate, loss under default, default exposure, and duration, and then calculated according to the actual situation to obtain the credit risk assessment results^[18]. Currently, the most widely used models include KMV logistic regression model, KMV logistic mixed model, credit risk additional model, etc^{[19][20][21][22]}. With the development of technology, machine learning has been introduced into the field of credit risk measurement. For example, the Bayesian method, which was proposed by Harold Bierman, was first applied to credit risk measurement in 1970^[23], and then received extensive attention from scholars and played an increasingly important role in technical support^[24]; neural

network algorithm has its unique advantages in dealing with high-dimensional complex data. Since the 1990s, foreign researchers began to use neural network analysis to predict the financial crisis of companies^{[25][26]}. To avoid the problems of neural network structure selection and local few points, Corinna Cortes and Vapnik proposed a support vector machine method based on statistical learning theory in 1995; SVM method can identify the small sample, nonlinear and high-dimensional patterns with abnormally high precision and accuracy, and achieve risk early warning function through periodic fitting and adjustment^[27]. Compared with KMV model, unstructured default distance generated by SVM has a more important application value in credit rating evaluation practice^[28]. However, SVM is difficult to implement for large-scale training samples. At the same time, SVM is also very sensitive to missing data and lack of flexibility; LAD, the full name of which is logical analysis of data, is a supervised learning method based on combinatorics, optimization theory, and Boolean logic. LAD can only be used to analyze binary data with a value of 0-1. Boros (1997)^[29] proposed a data binarization method, which can transform numerical data into 0-1 binary data, which makes it possible for LAD to analyze numerical data and further broaden the application field of LAD. The LAD method has been widely used in cancer diagnosis and prediction^{[30][31]}, risk assessment of patients with heart disease or lung disease^[32], etc. in recent years, the LAD method has also been applied in the field of credit risk rating^[33]. The results show that the correlation and mean absolute error prediction model obtained by the LAD method is always superior to other algorithms (C4.5, support vector machine, logistic regression, and multi-layer perceptron algorithm)^[34].

2.3 Summary

By summarizing the research results of domestic and foreign scholars, it can be found that most scholars summarize the causes of non-performing loans of commercial banks as external factors and internal factors, including macro-economy, government policies, bank management system, etc. Although the research on non-performing loans is more extensive in China, most of the research is from the perspective of commercial banks, few scholars study the characteristics of SMEs with non-performing loans, and few types of research on non-performing loans due to force majeure (such as epidemic). Most of the research is based on the case method, game method, and regression method, which have certain restrictions on the research object. LAD doesn't add constraints to the research object and can generate strong prime and strong spanned patterns. It can determine the credit approval conditions according to the trend of the market economy in different periods, which has high practical significance. Therefore, this paper will study the 30 SMEs in XX Branch of N Bank of s city who received online loan approval before the epidemic (2019), screen the indicators of online credit approval before the epidemic (2019) according to the literature and the actual situation, and get the research indicators in line with this topic. Then, LAD is used to find that the three enterprises with non-performing loan behavior after the epidemic are different from the other 27 enterprises. Based on the results of data analysis and the specific situation of the development period of the market economy, this paper puts forward the improvement scheme to the credit risk assessment system of the bank, improves the early warning ability of its credit business to force majeure events, and reduces the systematic financial risk of the bank; and puts forward suggestions to the small and medium-sized enterprises and relevant policymakers, so as to help them improve their resistance to force majeure the ability to resist factors.

3. METHODOLOGY

3.1 Logical Analysis of Data(LAD)

LAD (Logical Analysis of Data) is a data analysis method based on combinatorics, Boolean function, and optimization, which can solve the problem of binary classification and multivariate classification^[29]. In this chapter, we will introduce the core concepts of LAD. At first, this method was proposed by Peter L. hammer, and then it was developed and supplemented. Nowadays, LAD has become a method of data mining and has

been applied in many fields, such as medicine, commerce, oil exploration, and so on.

In the commercial field, LAD is mainly used in risk assessment. Compared with other research methods, it has five advantages in credit risk rating of development bank

- (1) Since LAD is not based on any statistical assumptions, it has no special requirements for data volume^[35];
- (2) Because the financial system has a certain connection with financial institutions, the complexity of the financial system will have a certain impact on financial institutions. At this time, LAD will allow high-order interaction between variables^[36];
- (3) For different grades, the proportion of the same variable in different credit grades will be different, and LAD allows the total difference^[36];
- (4) The generated pattern rules are not limited by the number and type of variables^[36];
- (5) Based on the MILP model, a variety of rules can be formulated to apply to different financial environments^[37].

Based on the above advantages, the LAD method will be used in this paper. Before expounding on the specific application, the author will elaborate on the related concepts of LAD.

3.1.1 Binarization

LAD can only be used to analyze 0-1 data sets at first. However, with the development of society, most of the data are real values, not simple 0-1 attribute values^[37]. Some programs will assign real-value attributes to the data according to the attributes. Therefore, a binary method is proposed. In this process, we often introduce the data related to the real value attribute of the data as the threshold (or cut point). If the real data value of the value is higher (or lower) than a certain threshold respectively, the attribute of changing the value is defined as 1 (or 0). Namely:

$$Y_i = \begin{cases} 1 & X_i \geq \alpha_i \\ 0 & X_i < \alpha_i \end{cases}$$

As shown in Table 3-1, C+ denotes normal repayment enterprises, C- denotes enterprises with non-performing loans, A and B are observed values under different variables. Since the observations are not binary variables, they need to be binarized. Here we first assume that two thresholds, $\alpha_A=3$, $\alpha_B=2$, The data which is greater than the value of the cut point and then its properties are 1, otherwise, they are 0. According to the Boolean function, we can get the truth Table 3-2.

Table 3-1. Data sets			Table 3-2. True table			Table 3-3. True table			
	A	B		A	B		A	B	B'
C+	3.5	4	C+	1	1	C+	1	1	1
C-	2.5	1.5	C-	0	1	C-	0	1	0

There is not only one cut point, so we should distinguish the positive set from the negative set as much as possible. We can see in Table 3-2 that for the observation value under variable B, the combination of the positive set and the negative set is not completely distinguished. At this time, we can assume a cut point $\alpha_B = 1$, and the resulting binary data is shown in Table 3-3.

3.1.3 Pattern generation

Patterns are the core of LAD. Positive patterns are subsets of covering sets that intersect with positive sets and separate from negative sets^{[29][37]}. Similarly, negative patterns are the same. In this process, the program needs to divide the space of the n-dimensional vector which has been binarized and feature selected, so that each space only contains positive or negative vectors, and then the pattern has been generated. As shown in Figure 3-1, a class a pattern is a spatial area that only covers class a data and does not include class B data, which is represented by a dotted line box. Similarly, the class B pattern is an area that only covers class B data, which is represented by a solid line box.

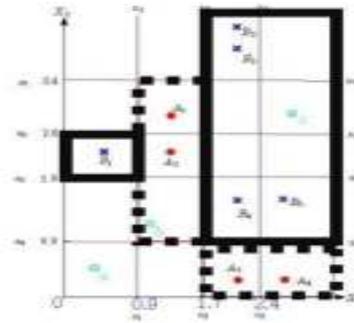


Figure 3-1. Strong pattern

If there are no additional constraints and the size of the space is not considered, any minimum interval or its combination in the space can be used as a pattern. In this pattern, each pattern is required to cover as many points as possible, which is called the strong pattern. (as shown in Figure 3-1)

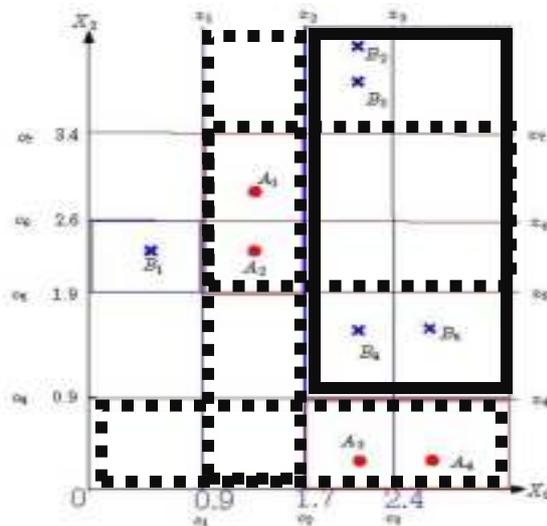


Figure 3-2(a). Strong spanned pattern

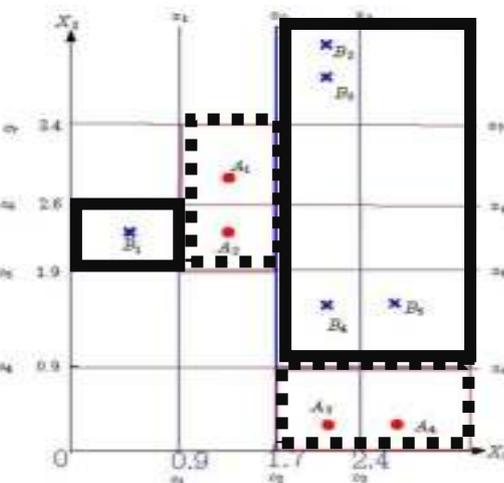


Figure 3-2(b). Strong prime pattern

By dividing the final strong pattern, we can get the strong spanned pattern (SS) and strong prime pattern (SP). As shown in Figure 3-1, we can see that there are some green points A, B, and C, which strictly do not belong to any pattern. When we tighten the condition and require that there should not be any points in the pattern that do not conform to the pattern, that is, we require the solid wireframe in the space vector to cover as many points and as little space as possible, then the pattern is called strong spanned pattern (as shown in Figure 3-2(a)). When we relax the condition and allow those points that can not be judged in the pattern, that is to say, it is here In this case, it is called a strong prime pattern (as shown in Figure 3-2(b)). The division of strong spanned pattern and strong spanned pattern gives us some enlightenment in the credit risk assessment of SMEs.

3.2 Data processing

The data of this study are from N bank in S City, covering the secondary and tertiary industries. The purpose of this study is to explore the characteristics of SMEs with non-performing loans after the epidemic, taking the SMEs with online loans approved in 2019 as the research subject. The sample data includes the registration information, cash flow statement, and loan information audit form of the selected SMEs.

When evaluating the credit risk level of SMEs, we need to refer to a series of indicators to determine their credit rating^[33]. There are 24 variables in the original data table. In the process of screening variables, we use the correlation test, and then according to the test results, we eliminate the variables that are not significant in the

correlation test and finally get 12 variables (Table 3-4).

Table 3-4. Description of variables

Variables	Explanation
HY	Industry of tax-paying enterprise
REGISTER	Registered capital of tax-paying Enterprises
TIME	Loan period of tax-paying Enterprises
LEVEL	Credit rating evaluation of tax-paying enterprises when they apply for loans for the first time
REMAIN	The loan balance of tax-paying Enterprises
PEOPLE	Number of employees in tax-paying Enterprises
FR	The proportion of shares held by legal persons in tax-paying Enterprises
PROFIT	The proportion of profits of tax-paying enterprises in business income (%)
INCOME CHANGE	Year on year change in sales revenue of tax-paying enterprises in recent 12 months (%)
HOLDS	Owner's equity of tax-paying enterprises (yuan)
TAX	The total amount of tax paid by tax-paying enterprises in recent 12 months (yuan)
INCOME	Sales revenue of tax-paying enterprises in recent 12 months (yuan)

According to the data requirements and operation steps of LAD, we first binary the data. In this study, except for the four variables of PEOPLE, HOLDS, TAX, and INCOME, which are assigned according to the industry standard (1 for those higher than the industry standard, and 0 for those lower than the industry standard), other binary threshold selection standards use the data points with obvious segmentation in the positive set and negative set as the cutting points, and the subsequent steps are run by the program instead.

4. RESULT AND ANALYSIS

LAD can determine the early warning characteristics of SMEs that generate non-performing loans according to the approval data of 30 SMEs approved online before the epidemic (2019) and can generate strong prime pattern and strong spanned pattern. In this study, only the strong class has been generated, and the strong prime pattern and the strong spanned pattern will be gradually explored in the future.

In this study, non-performing loan enterprises are regarded as a positive set and assigned with a value of 1; enterprises with good credit are regarded as a negative set and assigned with a value of 0. Through the LAD program, strong classes are obtained. The results are shown in Table 4-1.

Table 4-1. Result

Variables	Class 1	Class 2	Class 3	Class 4	Weight
HY		<3		<3	0.1052
REGISTER		≥2000000	≥2000000	≥2000000	0.1580
TIME		<8		<8	0.1052
LEVEL			<A+		0.0526
REMAIN	≥800000				0.0526
PEOPLE					
FR					
PROFIT	<7.935%		<7.935%	<7.935%	0.1580

Variables	Class 1	Class 2	Class 3	Class 4	Weight
INCOME CHANGE					
HOLDS	=0		=0	=0	0.1580
TAX		=1			0.0526
INCOME	=1		=1	=1	0.1580
Coverage	0.6667	0.6667	0.6667	0.6667	
Accuracy	1.0000	1.0000	1.0000	1.0000	

After LAD program processing, a total of four classes are obtained. In each class, each variable has a different threshold corresponding to it. The specific explanation is as follows:

(1) For class 1, among the three SMEs with non-performing loans, there are two enterprises in line with the class. When the loan balance of an enterprise is more than 800000, its profit accounts for less than 7.935% of the operating income, the owner's equity is lower than the industry standard, and its income is higher than the industry standard of the SMEs, the enterprise is more likely to produce non-performing loans. The higher the loan balance of an enterprise, the less credit interaction between the enterprise and the bank, the lower the bank's understanding of the enterprise and the difficulty in controlling its repayment ability. When we observe the profitability of an enterprise, we can't just look at the business's income. The higher the proportion of profit in business income, the stronger the profitability of an enterprise. When the proportion of profit to income is lower than a certain level, it shows that the profitability of the enterprise is insufficient, so the possibility of non-performing loans will also be greatly increased. The owner's equity refers to the residual equity enjoyed by the owner after deducting the liabilities from the assets of the enterprise. Generally speaking, the higher the owner's equity of an enterprise is, the lower the financial risk is, and the smaller the possibility of non-performing loans is. Therefore, when the owner's equity is lower than the industry standard, the possibility of non-performing loans will be relatively increased. According to the existing national industry standards, when the income of enterprises is higher than the industry standards, it does not meet the standards of small, medium, and micro-enterprises, and tax and other preferential policies are no longer used. Therefore, when the income of enterprises does not meet the industry standards, it will also affect the non-performing loan ratio to a certain extent.

(2) For class 2, among the three SMEs with non-performing loans, there are two enterprises in line with the class. When the enterprise is in the manufacturing or wholesale and retail industry, the registered capital is more than 2000000, the establishment period is less than 8 years, and the total amount of tax in the past 12 months is higher than the industry standard, the small and medium-sized micro-enterprises are more likely to produce non-performing loans. Due to the high demand for raw materials and output for the manufacturing industry and logistics for the wholesale and retail industry, but during the epidemic period, the industries that rely on liquidity are hit to a greater extent. Therefore, the possibility of non-performing loans will increase with the operation of enterprises in this type of industry. The registered capital of an enterprise is not equal to the paid-in capital, so from the perspective of the operation of the enterprise, the paid-in capital is slightly important. When the establishment period of SMEs is short, the bank is less familiar with the enterprise and, so it is inevitable to make mistakes in the approval of the enterprise in the credit approval. Besides, due to the short period, the company's operating conditions are unstable, it increases the risk of non-performing loan. When the total amount of tax paid in the past 12 months is higher than the industry standard, the profit of the enterprise after-tax deduction will decrease, which will increase the repayment pressure, and then affect the possibility of non-performing loans to a certain extent.

(3) For class 3, among the three SMEs with non-performing loans, there are two enterprises in line with the class. When the registered capital of the enterprise is higher than 2000000, the credit rating of the enterprise is below A+, the profit accounts for less than 7.935% of the operating income, the owner's equity is lower than the industry standard, and its income is higher than the industry standard of small, medium and micro enterprises, the enterprise is more likely to produce non-performing loans. The registered capital of an enterprise is not equal to the paid-in capital, so from the perspective of the operation of the enterprise, the paid-in capital is slightly important. Since the data used in this study are small, medium, and micro enterprises applying for credit before the epidemic (2019), their credit rating is not affected by the epidemic. Therefore, after the epidemic, the lower the credit rating of enterprises in loan approval, the higher the possibility of non-performing loans. When the proportion of profit to income is lower than a certain level, the possibility of non-performing loans will also be greatly increased. When the owner's equity is lower than the industry standard, the possibility of non-performing loans will be relatively increased. According to the existing national industry standards, when the income of enterprises is higher than the industry standards, it does not meet the standards of small, medium, and micro-enterprises, and tax and other preferential policies are no longer used. Therefore, when the income of enterprises does not meet the industry standards, it will also affect the non-performing loan ratio to a certain extent.

(4) For class 4, among the three SMEs with non-performing loans, there are two enterprises in line with the class. When the enterprise is in the manufacturing or wholesale and retail industry, the registered capital is more than 2000000, the establishment period is less than 8 years, the proportion of profit in the operating income is less than 7.935%, the owner's equity is lower than the industry standard, and the income is higher than the industry standard, the enterprise has a higher possibility of non-performing loans. Due to the high demand for raw materials and output for the manufacturing industry and logistics for the wholesale and retail industry, but during the epidemic period, the industries that rely on liquidity are hit to a greater extent. Therefore, the possibility of non-performing loans will increase with the operation of enterprises in this type of industry. The registered capital of an enterprise is not equal to the paid-in capital, so from the perspective of the operation of the enterprise, the paid-in capital is slightly important. When the enterprise establishment period is short, the bank is less familiar with the enterprise, so it is inevitable to make mistakes in the approval of the enterprise in the credit approval. Also, the company isn't so steady therefore it may face different ways of difficulties such as socially, financially and so on. When the proportion of profit to income is lower than a certain level, the possibility of non-performing loans will also be greatly increased. When the owner's equity is lower than the industry standard, the possibility of non-performing loans will be relatively increased. According to the existing national industry standards, when the income of enterprises is higher than the industry standards, it does not meet the standards of small, medium, and micro-enterprises, and tax and other preferential policies are no longer used. Therefore, when the income of enterprises does not meet the industry standards, it will also affect the non-performing loan ratio to a certain extent.

To sum up, by analyzing the characteristics of SMEs that generate non-performing loans, it can be concluded that the proportion of owner's equity, profit in operating income, registered capital, and sales income (tax, yuan) in recent 12 months have the highest weight, followed by the industry and establishment period of the enterprise, and finally the credit rating evaluation and the credit rating evaluation of the enterprise in the first loan application. The balance of the loan and the total amount of tax in the past 12 months (yuan). As this study only obtains strong classes, banks should focus on examining the owner's equity, the proportion of profit in operating income, registered capital, and sales income (yuan) in the past 12 months in the process of credit approval for SMEs, without considering the trend of the economic market, and then examine other variables in the classes according to the weight.

5. SUMMARY AND PROSPECT

5.1 Summary

Although domestic scholars have not paid attention to LAD, LAD is rarely applied to the credit evaluation system in the world. However, the study found that, from the perspective of data analysis, it is essentially consistent to judge whether the applicant's credit is good through some indicators and to predict whether the patient has a certain disease through a series of medical examination data. We can expect that the application of the LAD method in credit risk assessment can achieve satisfactory results or even breakthroughs. Although the LAD method is limited to the prediction of class results, in the process of credit approval, financial institutions only need to judge whether the applicant's credit is good or not. The LAD method is just suitable for binary classification. Besides, LAD can selectively generate strong prime pattern and strong spanned pattern, to help financial institutions make "easing" or "tightening" credit approval decisions according to market and policy.

5.2 Suggestions

Since this study only generates strong models, the author puts forward relevant suggestions to banks, SMEs, and policymakers without considering the development of the market economy.

First of all, for banks, appropriate standards should be formulated for the variables appearing in the model, and in the process of credit approval for SMEs, it should focus on examining the owner's equity, the proportion of profit in operating income, registered capital and sales revenue(yuan) in the past 12 months, and then examine other variables appearing in the model according to the weight. As for the industries with high liquidity dependence, they are greatly affected by the force majeure risks such as epidemic situations. Therefore, we should attach great importance to them. During the credit approval, we should strictly review the variables in the pattern, and when appropriate, we can ask for the increase of collateral, to reduce the credit risk of banks.

Secondly, for SMEs, we should pay more attention to profitability. To enhance the profitability of enterprises, on the one hand, it can improve the proportion of owners' equity and profits in operating income, reduce the repayment pressure of enterprises, and then make it easier to pass the bank's credit approval. On the other hand, it can promote the development of enterprises by improving the profitability of enterprises, and store more liquidity, so as to reduce its own financial risk and enhance its awareness of the epidemic situation The ability to resist a force majeure event.

Finally, for policymakers, in order to maintain the vitality of SMEs in China's economy and solve the financing problems of them, we can refer to the relevant variables in the model to make policies and improve their credit passing rate, so as to promote the high-quality development of the economy of China.

5.3 Expectation

Because the data of this study comes from S city and N banks, the conclusion has certain limitations, but it also has a certain reference value for other scholars. Secondly, this study only generated a strong pattern, which did not consider the development of the market economy. Because the market economy is not immutable, the early warning characteristics of different development conditions of the market economy need to be explored. In the future research, the author will broaden the data sources, generate strong prime patterns and strong spanned patterns while generating positive patterns, and generate negative patterns for enterprises with good repayment, combine positive and negative pattern, strong prime pattern and strong spanned pattern, and formulate targeted credit risk assessment rules for specific financial environment, to get more universal results Conclusion: to reduce the bank credit risk, so as to maintain the stability of China's SMEs credit market, to solve the financing problem of small and medium-sized micro-enterprises, and promote the vigorous development of SMEs in China's economic market

ACKNOWLEDGEMENT

This study was supported by the Guangdong Provincial Science and Technology Research Project under Grant 2017A030307026; National Fund Cultivation project of STU under Grant NFC16002 and STU Scientific Research Initiation Grant under Grant STF15003.

REFERENCES

- [1] Xu Hui, Ji Chunli, Li Ji, Zhou Bin, Jin Xin.(2011).Research on risk response mechanism of technology-based SMEs from the perspective of Organizational Immunity [J]. *Management world*, (02): 142-154(in Chinese)
- [2] Zhu Wuxiang, Zhang Ping, Li Pengfei, Wang Ziyang.(2020).The dilemma of small and medium-sized enterprises and the improvement of policy efficiency under the impact of epidemic situation -- Based on the analysis of two national questionnaires [J]. *Management world*, 36 (04): 13-26(in Chinese)
- [3] Wang Gang, Wang Yanwei, Shang Bo Wen.(2020).Impact of the new crown pneumonia epidemic on banking industry and policy optimization suggestions [J]. *China Banking*, (05): 30-32(in Chinese)
- [4] Dimitrios P.Louzis,Angelos T.Vouldis,Vasilios L.Metaxas.(2012).Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios[J]. *Journal of Banking and Finance*,36(4)
- [5] Salas V, saurina J.(2002).Credit risk in two institutional regimes:Spanish commercial and saving banks[J].*Journal of Financial Services Research*,22 (3) : 203~224
- [6] Demirgüçkunt A, Detragiache E.(2016).The Determinants of Banking Crises: Evidence from Industrial and Developing Countries[J]. *Policy Research Working Paper*, 45 (1) : 81~109.
- [7] Liu Yan. (2014).Analysis on the causes and related factors of non performing loans of commercial banks in China [J]. *Systems engineering*, 32 (05): 65-74(in Chinese)
- [8] Tan Jinsong, Jian Yuyin, Chen Ying. (2012).Government intervention and non performing loans: a case study of a state-owned commercial bank from 1988 to 2005 [J]. *Management world*,(07): 29-43 + 187(in Chinese)
- [9] Das, A., & Ghosh, S. (2003). Determinants of credit risk in Indian state-owned banks: An empirical investigation. *Economic Issues-Stoke and Trent*, 12(2), 27-46.
- [10] Mohammad O. Farooq,Mohamed S. Elseoud,Seref Turen,Mohamed Abdulla.(2019).Causes of Non-Performing Loans: the experience of Gulf Cooperation Council countries[J]. *Entrepreneurship and Sustainability Issues*.
- [11] Waweru N M,Kalani V M.(2009).Commercial banking crises in Kenya:Causes and remedies.*African Journal of Accounting,Economic Economics,Finance and banking research*.
- [12] Demirgüçkunt A, Detragiache E. (2016).The Determinants of Banking Crises: Evidence from Industrial and Developing Countries[J]. *Policy Research Working Paper*, 45 (1) : 81~109.
- [13] Warue, B. N. (2013).The effects of bank-specific and macroeconomic factors on non-performing loans in commercial banks in Kenya: A comparative panel data analysis. *Advances in Management & Applied Economics*, 3(2), 135-164.
- [14] A H M Yeaseen Chowdhury, Afifatun Jannah. (2020).Causes of Non-Performing Loan: A Study on Bangladesh Public Sector Banks[J]. *Frontiers in Management Research*,4(1).
- [15] Steave,C.(2012).Theory of Credit Risk and Securitization of Non-performing loans[J]. *Journal of Financial and Quantitative Analysis*, (1): 54~59.
- [16] Zeng Shihong, Liu Yuhui, Duan chenju.(2003).Analysis of the formation mechanism of non-performing loans from the behavior of state-owned commercial banks [J]. *China soft science*, (09): 43-48(in Chinese)
- [17] Zelalem.(2013).Determinants of Non-Performing Loans in Licensed Commercial Banks: Evidence from Ethiopian Banks. *Asian Economic and Financial Review*,5(6), pp.868-882.
- [18] Zhu Xiaozong, Zhang Zongyi, Geng Huadan, Wu Jun.(2006).Empirical comparison and applicability analysis of modern credit risk measurement models [J]. *Journal of management engineering*, (01): 88-93(in Chinese)

- [19] Wu Jianmin, Jia Zhiqing.(2004).Credit risk measurement characteristics of KMV model [J]. Statistics and decision, (11): 122-123(in Chinese)
- [20] Hu Sheng, Lei Huanhuan, Hu Huaqiang.(2018).Research on credit risk measurement of real estate enterprises in China based on logistic model [J]. China soft science, (12): 157-164(in Chinese)
- [21] Lei Chengyao, Wang Xingqi.(2017).Empirical research on credit risk measurement and control of platform like corporate bonds -- Based on improved logistic KMV mixed model [J]. Investment research, 36 (01): 146-159(in Chinese)
- [22] Liu Yingchun.(2012).Research on the application of CreditRisk ~ + model in credit risk measurement of credit portfolio [J]. Practice and understanding of mathematics,42 (09): 45-51(in Chinese)
- [23] Harold Bierman J R, Warren H.(1970).The Credit Granting Decision[J]. Management Science,16(8): 519-532.
- [24] Ding Dongyang, Zhou Lili, Liu Leping.(2013).Review on the application of Bayesian method in credit risk measurement [J]. Mathematical statistics and management,32 (01): 42-56(in Chinese)
- [25] Tom Altman,Paul F. Boulos.(1995).Solving Flow-Constrained Networks: Inverse Problem[J]. Journal of Hydraulic Engineering,121(5).
- [26] Coats, Warren L.(1983).The SDR as a Means of Payment: Reply to Comments by van den Boogaerde, Callier, and Kenen[J]. Staff Papers - International Monetary Fund,30(3).
- [27] Lu Hualan.(2016).Research on credit risk early warning mechanism of small loan in China [J]. Statistics and decision,08:154-157(in Chinese)
- [28] Shen Peilong, Zhou Hao. (2010).Research on credit risk prediction of SMEs based on support vector machine [J]. International financial research, (08): 77-85(in Chinese)
- [29] E. Boros, P. Hammer, T. Ibaraki, A. Kogan.(1997).”Logical analysis of numerical data”, Mathematical Programming,vol. 79, pp. 163–190.
- [30] Axelrod D., T. Bonates, P. L. Hammer, I. Lozina.(2004).From Diagnosis to Therapy via LAD.Invited Lecture at INFORMS Annual Meeting, Denver, CO, October,
- [31] Alexe G, Alexe S, Axelrod DE, Hammer PL, Weissmann D.(2005).Logical analysis of diffuse large B-cell lymphomas. Artif Intell Med.34(3):235-67.
- [32] Martin, S.G., Kronek, L.P., Valeyre, D., Brauner, N., Brillet, P.Y., Nunes, H., Brauner, M.W., Rety,F. (2010).High-resolution computed tomography to differentiate chronic diffuse interstitial lung diseases with predominant ground-glass pattern using logical analysis of data. European Radiology 20(6),1297–1310.
- [33] Isabel Linda Moyo,Aditi Kar Gangopadhyay,Victor Gumbo,Eriyoti Chikodza,Brian Jones.(2020).Application of Logical Analysis of Data (LAD) to Credit Risk Ratings for Banks in Zimbabwe[J].Journal of Risk and Financial Studies.
- [34] Anupama Reddy, Honghui Wang,Hua Yu,Tiberius Bonates, Vimla Gulabani, Joseph Azok, Gerard Hoehn,Peter L Hammer,Alison E Baird,and King C Li.(2008). Logical Analysis of Data (LAD) model for the early diagnosis of acute ischemic stroke. BMC Med Inform Decis Mak.
- [35] Christine Dupuis,Michel Gamache,Jean-François Pagé.(2011). Logical analysis of data for estimating passenger show rates at Air Canada[J]. Journal of Air Transport Management,18(1).
- [36] P. L. Hammer,A. Kogan,M. A.(2011). Lejeune. Reverse-engineering country risk ratings: a combinatorial non-recursive model[J]. Annals of Operations Research,188(1).
- [37] Cui Guo,Hong Seo Ryoo.(2012). Compact MILP models for optimal and Pareto-optimal LAD patterns[J]. Discrete Applied Mathematics,160(16-17).

The Relationship Amongst Big Data Capability, Supply Chain Dynamic Capability, And Dynamic Innovation Capability

Guangqian Peng^{1}, Yang Zhang¹, Ping Gao², Junfeng Yin³*

¹College of Business Administration, Capital University of Economics and Business, China

²School of Environment, Education and Development, The University of Manchester, United Kingdom

³Patent Examination Cooperation (Beijing), Center of the Patent Office, China

1. INTRODUCTION AND RESEARCH FRAMEWORK

With the thriving of digital age, big data is being produced from all industries at an unprecedented rate. This induces organizations to attempt to leverage big data in order to create value. However, very few companies in practice have so far obtained benefits from big data, though with surge of investments in it. Meanwhile, there is still limited studies in academy on what business value can be derived and how to be derived from big data [1-2].

To realize the potential huge value of big data, companies need to develop organizational big data capability to extract relevant information and make sense of it to make decisions. The big data capability of a company is constantly changing over time, including the improvement of analysis technology and market changes. In other words, big data capability is an important dynamic capability of enterprises.

In order to address the critical gaps in the literature and the practice, and to explore the potential value of big data capability, hereby we propose the central research question as: what is the impact of big data capability on companies? To approach the answer to this research question, the sub-research questions are proposed as:

- What are the impacts of big data capability on companies' other dynamic capabilities, specifically, on supply chain dynamic capability?
- What are the impacts of big data capability on companies' other dynamic capabilities, specifically, on dynamic innovation capability?

Grounded on resource-based view, supply chain theory, innovation theory, and expert interview, the hypotheses and research framework are proposed herein (Figure 1).

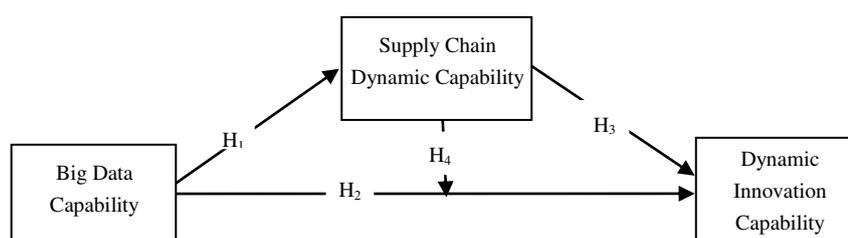


Figure 1. The conceptual framework

The dataset was derived from a data mining of 303 listed hi-tech companies in China. The firms ages range from 3 to over 15 years. Construct validity and reliability were examined. We conducted partial least squares (PLS) modeling to test the proposed model and hypotheses. The overall model explains about 67.5% of the variance of the endogenous constructs, which indicates a satisfactory model fit [3].

* Corresponding author. Email: gqpeng@cueb.edu.cn (Guangqian Peng). Gratitude goes to the National Natural Science Foundation of China (Grant 71472128).

2. MAJOR FINDINGS

This study observed some valuable and interesting findings: (1) The results indicate that, generally speaking, the level of big data capability of the companies in China are still in its infancy. (2) Although so, a company's big data capability already tends to significantly help to improve its supply chain dynamic capability. (3) However, strangely, the big data capability of a company does not necessarily help to improve its dynamic innovation capability currently. On the contrary, the current direct effect of big data capability on dynamic innovation capability tends to be significantly negative. These results are in line with the studies by Shah *et al.* (2012)^[4] and Wamba *et al.* (2015)^[5]. In our opinion, the possibly reason is that big data capability is just initiated recently in most companies, thus, may lead to certain disorders. (4) Meanwhile, a company's supply chain dynamic capability tends to influence its dynamic innovation capability directly and positively. (5) Moreover, supply chain dynamic capability has shown mediating and moderating positive effects on the relationship between big data capability and dynamic innovation capability. (5) Furthermore, owing to the positive mediating and moderating effects of supply chain capability, big data capability tends to have positive total effect on dynamic innovation capability.

3. CONTRIBUTION

This study is amongst the first to explore the potential impact of big data capability on companies' other dynamic capabilities.

First, this study called on scholars and practitioners to distinguish amongst big data, big data analytic capability, and big data capability. Then, based on extant literature and field study, it proposed three dimensions of big data capability, including: resource integration capability, depth analysis capability, and real-time prediction capability. Further, we designed the measurement items of these three dimensions, which can be used to measure and assess the levels of big data capability and its sub-capabilities of a company.

Moreover, grounded on supply chain theory and innovation theory, this study proposed a Big data-supply chain-innovation Model, which is composed of three main parts: big data capability, supply chain dynamic capability and dynamic innovation capability. The model can be used to understand, examine, and assess the impact of big data capability on companies' other dynamic capability. Specifically, this model can be used to understand, examine, and assess how big data capability can help to improve supply chain dynamic capability, and further contribute to higher dynamic innovation capability (indirectly currently).

This study also contributes to the extension of our knowledge on the impacts of company characteristics on the interrelationship of big data capability and companies' other dynamic capabilities. An important finding herein is that company characteristics (firm age, size, and ownership) are likely to affect the levels (magnitude) of supply chain dynamic capability and of dynamic innovation capability, but do not necessarily change the interrelationships between big data capability, supply chain dynamic capability, and dynamic innovation capability.

REFERENCES

- [1] Gupta M, George JF (2016). Toward the development of a big data analytics capability. *Info&Manag* 53(8): 1049-1064.
- [2] Mikalef P, Boura M, Lekakos G, and Krogstie J. (2019). Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment. *Brit J of Manag* 30(2): 272-298.
- [3] Hulland J. (1999). Use of Partial Least Squares (PLS) in Strategic management research: a review of four recent studies. *Strateg Manag J* 20(2): 159.
- [4] Shah S, Horne A, and Capella J. (2012). Good data won't guarantee good decisions. *Harv Bus Rev*, 90(4): 23-25.
- [5] Wamba F F, Akter S, Edwards A, Chopin G, and Gnanzou D. (2015) How 'big data' can make big impact: findings from a systematic review and a longitudinal case study," *Int'l J of Prod Econ*, (165): 234-246.

Short Research Paper

A Meta-analysis of Online Impulse Buying and The Moderating Effect of Economic Development Level

Yang Zhao^{1,2}, Bin Yang², Ning Wang²*

¹Center for Studies of Information Resources, Wuhan University, Wuhan, 430072, China

²School of Information Management, Wuhan University, Wuhan, 430072, China

Abstract: With the prosperity of e-commerce and the increase in per capita disposable income of consumers, online impulse buying has become an epidemic. Various researchers explored what factors influenced consumers' online impulse buying behaviors and the strength of each factor, but the conclusions were usually different. Therefore, this study constructed a comprehensive framework by using a meta-analysis to derive a unified conclusion. The meta-analysis was conducted according to 54 previous empirical studies about online impulse buying. The results show that website security, price, novelty, and negative emotion have no significant impact on online impulse buying. Meanwhile, the economic development level can moderate relationships between website visual appeal, ease of use, price, promotion, pleasure, positive emotion, and online impulse buying.

Keywords: meta-analysis, online impulse buying, economic development level, moderating effect

1. INTRODUCTION

With the prosperity of e-commerce and the increase in per capita disposable income of consumers, online impulse buying has become an epidemic. Triggered by easy access to products, easy buying process (e.g., 1-Click ordering), lack of social pressures, and absence of delivery efforts, impulse buying apparently occurs in about 40% of all online expenditures (Verhagen & Van Dolen, 2011). Online impulse buying can be defined as a sudden and immediate online purchase with no pre-shopping intentions (Verhagen & Van Dolen, 2011). The decision-making is unplanned, spontaneous, unreflective, and dominated by emotions. As insight into consumers' impulse buying behaviors in online environment is vital for e-commerce practitioners, an increasing number of researchers explore the influence of the website quality (Hasim et al., 2018), the product involvement (Bhakat & Muruganatham, 2013; Bukidz & Tielung, 2014), socio-cultural factors (Bhakat & Muruganatham, 2013), and users' emotional states (Bhakat & Muruganatham, 2013) on online impulse buying behaviors.

Despite considerable research on online impulse buying, research results are inconsistent due to the investigation time, the location of the survey, the target population, the number of samples and other force majeure factors. For example, some researchers found that online stores' navigability had a strong positive relationship with online impulse buying (Zou, 2018), while other researchers showed that the influence of e-store's navigation was insignificant (Floh & Madlberger, 2013). Meanwhile, the strength of correlation coefficients obtained in different studies was also inconsistent. These inconsistencies will bring some confusion to further theoretical studies and practical work. Therefore, we attempt to find an explanation for the contradictory results in literature by using a meta-analysis of empirical studies on online impulsive buying. Meta-analysis is a method that statistically analyses a large collection of previous results from empirical studies for the purpose of integrating the findings (Glass, 1976).

* Corresponding author. Email: yangzhao_0813@whu.edu.cn (Yang Zhao)

The online impulse buying behavior was affected by various factors and different scholars classified those factors into diverse categories according to previous research results. In this study, we choose three aspects (i.e., website stimulus, marketing stimulus, and affective stimulus) as the classification category of 13 main factors that affect consumers' online impulse buying. Considering that price plays an important role in consumers' buying behavior, and price sensitivities are different thanks to the imbalanced economic development level. Meanwhile, the development level of information technology and information literacy are closely related to the development of economy. Therefore, we use the economic development level as a moderating variable in the relationship between factors and online impulse buying.

2. LITERATURE REVIEW

2.1 Website stimulus

In an online context, the impulse buying is influenced by not only the product characteristics but also the features of the shopping environment (i.e., the website) (Madhavaram & Laverie, 2004). Results of Zou (2018) showed that website security, navigability, and visual appeal had a strong positive relationship with online impulse buying behavior. In addition, online shopping environments are at a disadvantage when compared with brick-and-mortar stores due to the lack of touch, taste, or try products. Vonkeman et al. (2017) found that increased level of interactivity was likely to diminish a consumers' perception of mediation by creating a sense of presence. On the other hand, the website ease of use refers to the website usability, which describes how easy it is to search information in a site and how easy the site can be navigated. Results of Liu et al. (2013) indicated that perceived website ease of use was a crucial online cue for engendering impulse buying online. Therefore, we consider website security, website navigability, website visual appeal, interactivity, and ease of use as factors in website stimulus that affect consumers' online impulse buying.

2.2 Marketing stimulus

Consumers are constantly exposed to marketing stimuli that promote impulse buying with the popularity of multitudinous e-commerce shopping festivals. Scarcity creates a sense of urgency among buyers which results in increased quantities purchases, shorter searches, and greater satisfaction with products they have bought (Aggarwal et al., 2011). Findings of Wu et al. (2020) showed that both limited-quantity scarcity and limited-time scarcity led to online impulse buying. Price discounts are the most widely employed sales promotion strategies both online and offline, and online shoppers are more price-sensitive because of low search costs and direct price comparisons (Xu & Huang, 2014). Park et al. (2012) confirmed that price attributes on shopping websites played an important role in online impulse buying. Consumers usually engage in impulse buying in order to experience feelings of fun and novelty. Findings of Zou (2018) expressed that novelty has a strong positive influence on online impulse buying behavior. Moreover, in online shopping, a shopper may have no knowledge of certain products or no explicit intention to buy them, but is stimulated by some promotional offers. Conclusions of Hasim et al. (2018) indicated that sales promotions positively impacted online impulse buying behavior. Therefore, we consider scarcity, price, novelty, and promotion as factors in marketing stimulus that affect consumers' online impulse buying.

2.3 Affective stimulus

Shopping in a network environment requires interaction with the website, and the interaction leads to both affective and cognitive reactions (Parboteeah et al., 2009). Affective reactions capture an individual's emotional response when interacting with an environment, which results in subsequent cognitive reactions (e.g., impulse buying intention). For example, Shen and Khalifa (2012) found a significant positive effect of arousal on impulse buying when consumers felt the shopping environment was pleasant. Their research results also showed that pleasure was a significant determinant of impulse buying behavior. Findings of Suhud and Herstanti (2017)

indicated that positive emotion had a direct positive and significant impact on impulse buying. On the other hand, shopping is an experience sought for its ability to alleviate negative emotions and improve moods (Mano, 1999), people in depressed conditions are more likely to shop impulsively for the reason of improving their emotional states. Therefore, we consider arousal, pleasure, positive emotion, and negative emotion as factors in affective stimulus that affect consumers' online impulse buying.

The research model established in this study is shown in Figure 1.

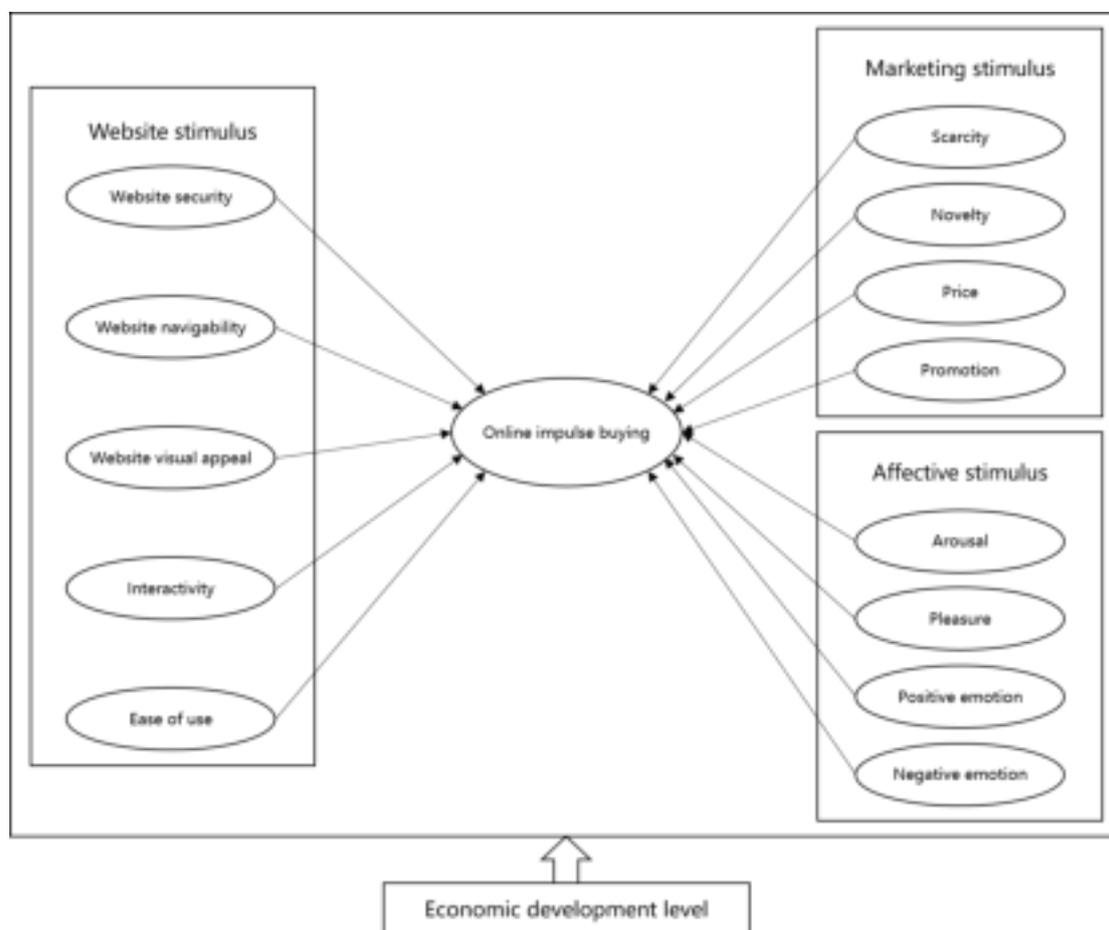


Figure 1. Conceptual model

3. METHODOLOGIES

3.1 Data collection and coding

To retrieve as many articles related to online impulse buying as possible, we used “online impulse buying”, “online impulse purchasing”, “online impulse shopping”, “website stimulus”, “marketing stimulus”, and “affective stimulus” as key words. We searched literatures in some popular databases such as Google Scholar, Web of Science, Science Direct, SpringerLink, CNKI, etc. After the preliminary search, 1,345 initial papers were obtained.

We excluded 1,291 repetitive and inappropriate papers, and finally, 54 articles were included. The criteria of literature selection adopted in this study are: (1) the literature must be an empirical study of online impulse shopping and quantitatively tested relationships between antecedent factors and online impulse buying behavior; (2) the literature must report correlation coefficients or other values (e.g. F-value) that could be converted to correlation coefficients; and (3) the literature must report the sample size.

After carefully reading the articles selected for meta-analysis, we extracted the following information: author name, publication date, publications, investigated countries or regions, sample size, key constructs, and reported

effect sizes. At the same time, we merged the constructs with similar meanings as many constructs involved in these selected articles had different names that expressed similar meanings. Finally, we got a total of 13 relationships. To test the moderating effect of economic development level, all articles included in the meta-analysis were divided into two groups based on the economic development level of the country or region to which the sample belongs. The reference standard we used to divide is the data published by the World Bank ("Classifying countries by income," 2019).

3.2 Statistical analysis

We adopted R 3.6.0 to conduct the meta-analysis. It is important to choose the fixed-effect model or the random-effect model in a meta-analysis. When there is a single effect in the hypothesis sample, the fixed-effect model is selected, and when there are different effects in the hypothesis sample, the random-effect model is selected (Borenstein et al., 2007). In this study, we chose the random-effect model due to the differences in the samples. In addition, we used Q-test to examine the heterogeneity of the distribution of effect sizes, and calculated the Z-score to find potential moderating effects. Meanwhile, the publication bias of each relationship was examined by fail-safe N. If the fail-safe N is larger than "5*K + 10" (K is the number of studies), it can be argued that there is no publication bias (Rosenthal, 1979).

4. RESULTS

4.1 Correlation analysis

The effect sizes of each relationship in the model are combined, and the results are shown in Table 1.

95% CI (confidence intervals) of relationships between website security, price, novelty, negative emotion, and online impulse buying all include 0, and their combined effect sizes are not significant. Our findings point out that the website stimulus factors have a significantly positive relationship with online impulse buying except the website security, and the interactivity has the weakest relationship as its combined effect size is only 0.17. The scarcity and promotion both have a significantly positive relationship with online impulse buying in the marketing stimulus factors. The affective stimulus factors are significantly and positively related to online impulse buying except the negative emotion.

From the perspective of publication bias, values of fail-safe N of all factors are greater than "5*K + 10", which means that all factors in this study have no publication bias.

Table 1. The results of correlation analysis

Categories	Factors	Q-value	Combined effect size	95% CI	N _{fs.05}
Website stimulus	Website security	311.40***	0.32	[-0.12; 0.65]	250
	Website navigability	15.12**	0.36***	[0.23; 0.47]	126
	Website visual appeal	170.72***	0.29***	[0.15; 0.41]	815
	Interactivity	48.60***	0.17***	[0.07; 0.27]	138
	Ease of use	232.43***	0.22*	[0.02; 0.40]	402
Marketing stimulus	Scarcity	9.26**	0.32***	[0.21; 0.42]	29
	Price	1587.28***	0.42	[-0.39; 0.86]	610
	Novelty	111.39***	0.38	[-0.06; 0.70]	135
	Promotion	597.29***	0.37***	[0.23; 0.51]	3383
Affective stimulus	Arousal	106.03***	0.40***	[0.29; 0.50]	1347
	Pleasure	279.96***	0.41***	[0.31; 0.50]	3904
	Positive emotion	238.27***	0.41***	[0.25; 0.54]	891
	Negative emotion	828.47***	0.21	[-0.35; 0.66]	967

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.2 Moderator analysis

Due to the limited number of selected articles, only one developed country's data had been collected in the relationships between website security, interactivity, scarcity, novelty, negative emotion, and online impulse buying. Therefore, only 8 pairs of relationships can conduct the moderator analysis, and the results are shown in Table 2.

Most 95% confidence intervals exclude 0 except the relationships between ease of use and online impulse buying in the developed subgroup, price and online impulse buying in both groups, and positive emotion and online impulse buying in the developed subgroup. On the other hand, the results of the moderator analysis demonstrate that the economic development level moderates 6 pairs of relationships as Z-score of the website navigability and arousal are not significant. Specifically, the combined effect size of website visual appeal on online impulse buying in developed countries or regions is higher than that in developing countries or regions, which means that the website visual appeal is a vital predictor for online impulse buying in developed countries or regions. Likewise, the combined effect size of promotion on online impulse buying in developed countries or regions is higher than that in developing countries or regions, which means that online consumers in developed countries or regions are more likely to be influenced by sales promotion to make impulse consumptions. The combined effect sizes of other factors are opposite, values in developing countries or regions are higher than those in developed countries or regions. It is worth mentioning that the combined effect size of price on online impulse buying in developed countries or regions is very little, which indicates that consumers in developed countries or regions are more likely to ignore price factors when make online impulse buying.

Table 2. The results of moderator analysis

Categories	Factors	Economic development level	Combined effect size	95% CI	Z-score
Website stimulus	Website navigability	Developed	0.35	[0.26; 0.43]	0.13
		Developing	0.36	[0.08; 0.59]	
	Website visual appeal	Developed	0.50	[0.33; 0.63]	-8.62 ^{***}
		Developing	0.23	[0.07; 0.38]	
	Ease of use	Developed	0.10	[-0.08; 0.26]	4.63 ^{***}
		Developing	0.27	[0.02; 0.48]	
Marketing stimulus	Price	Developed	-0.04	[-0.47; 0.41]	18.53 ^{***}
		Developing	0.64	[-0.43; 0.96]	
	Promotion	Developed	0.43	[0.06; 0.69]	-3.07 ^{**}
		Developing	0.36	[0.19; 0.50]	
Affective stimulus	Arousal	Developed	0.42	[0.18; 0.61]	0.61
		Developing	0.39	[0.27; 0.51]	
	Pleasure	Developed	0.33	[0.01; 0.58]	3.85 ^{***}
		Developing	0.44	[0.34; 0.53]	
	Positive emotion	Developed	0.22	[-0.14; 0.53]	5.41 ^{***}
		Developing	0.43	[0.24; 0.58]	

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. DISCUSSION

This study examined the impact of 13 factors (i.e., website security, website navigability, website visual appeal, interactivity, ease of use, scarcity, price, novelty, promotion, arousal, pleasure, positive emotion, and negative emotion) on consumers' online impulse buying, and explored the moderating effect of economic development level by using a meta-analysis. The results of this study provide both theoretical and practical constructions for e-commerce platforms and merchants.

First, this study proposed a comprehensive framework by integrating 54 previous empirical studies and divided 13 selected factors into website stimulus, marketing stimulus and affective stimulus, which helped scholars understand consumers' online impulse buying behaviors from multiple dimensions.

Second, this study chose economic development level as the moderator of relationships between factors and online impulse buying, and verified that economic development level had a significant moderating effect on relationships between website visual appeal, ease of use, price, promotion, pleasure, positive emotion, and online impulse buying. The result indicates that when researchers explore the impact of website visual appeal, ease of use, price, promotion, pleasure, and positive emotion on online impulse buying in the future, they should pay full attention to the moderating effect of economic development level.

Third, managers and developers of e-commerce platforms should take website navigability, website visual appeal, interactivity, ease of use, scarcity, promotion, arousal, pleasure, and positive emotion seriously as these factors are significantly and positively related to online impulse buying in the research results.

Finally, results of the moderator analysis can give an inspiration for cross-border trade practitioners. For cross-border e-commerce practitioners, when the target customers are people in developed countries or regions, website visual appeal and promotion are the main attractions to consumers. In contrast, when the target customers are people in developing countries or regions, ease of use, price, pleasure, and positive emotion more appeal to consumers. That is, in developing countries or regions, to provide users with a pleasant buying experience, more efforts should be put into website operation simplification and product pricing strategy.

6. CONCLUSION

Based on the meta-analysis, this study utilizes a quantitative review method to integrate the consistent or inconsistent conclusions of previous studies, and reaches a unified view. The results of this study have specific development significance for both researchers and online buying service providers. Despite the implication of this study, there are still some limitations.

First, the number of relevant studies that can be used for meta-analysis of online impulse buying is limited. In addition, although we have included the most investigated factors of online impulse buying in the meta-analysis, there are other factors that may also have a significant effect. Finally, the classification of the sample into developed and developing countries or regions based on the economic development level may be too crude. The moderating effect of 5 relationships were not examined due to the limited number of samples in developed countries or regions.

ACKNOWLEDGEMENT

This study was supported by the National Key R&D Program of China (2019YFC010167) and "User Information Behavior Research Team under Big Data and Intelligent Environment" by the academic team of young scholars of Humanities and Social Sciences of Wuhan University.

REFERENCES

- [1] Aggarwal, Praveen, Jun, Sung Youl, & Huh, Jong Ho. (2011). Scarcity Messages. *Journal of Advertising*, 40(3), 19-30. doi:10.2753/joa0091-3367400302
- [2] Bhakat, Ravi Shankar, & Muruganantham, G. (2013). A Review of Impulse Buying Behavior. *International Journal of Marketing Studies*, 5(3). doi:10.5539/ijms.v5n3p149
- [3] Borenstein, Michael, Hedges, Larry, & Rothstein, Hannah. (2007). *Meta-Analysis Fixed effect vs. random effects* (pp. 1-162).
- [4] Bukidz, Rocky San Seaver, & Tielung, Maria V. J. (2014). THE EFFECT OF PRODUCT INVOLVEMENT AND SOCIAL INFLUENCE ON ONLINE IMPULSE BUYING OF IBA STUDENTS. *Jurnal EMBA*, 2(3), 1019-1027.
- [5] Classifying countries by income. (2019). Retrieved from <https://datatopics.worldbank.org/world-development-indicators/stories/the-classification-of-countries-by-income.html>
- [6] Floh, Arne, & Madlberger, Maria. (2013). The role of atmospheric cues in online impulse-buying behavior. *Electronic Commerce Research and Applications*, 12(6), 425-439. doi:10.1016/j.elerap.2013.06.001
- [7] GLASS, GENE V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3-8.
- [8] Hasim, Muhammad Asyraf, Shamsudin, Mohd Farid, Ali, Aeshah Mohd, et al. (2018). The Relationship between Sales Promotions and Online Impulse Buying in Malaysia. *Opcion*, 34(16), 295-308.
- [9] Liu, Yong, Li, Hongxiu, & Hu, Feng. (2013). Website attributes in urging online impulse purchase: An empirical investigation on consumer perceptions. *Decision Support Systems*, 55(3), 829-837. doi:10.1016/j.dss.2013.04.001
- [10] Madhavaram, Sreedhar Rao, & Laverie, Debra A. (2004). Exploring Impulse Purchasing on the Internet. *Advances in consumer research*, 31, 59-66.
- [11] MANO, HAIM. (1999). The Influence of Pre-Existing Negative Affect on Store Purchase Intentions. *journal of retailing*, 75(2), 149-172.
- [12] Parboteeah, D. Veena, Valacich, Joseph S., & Wells, John D. (2009). The Influence of Website Characteristics on a Consumer's Urge to Buy Impulsively. *Information Systems Research*, 20(1), 60-78. doi:10.1287/isre.1070.0157
- [13] Park, Eun Joo, Kim, Eun Young, Funches, Venessa Martin, et al. (2012). Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research*, 65(11), 1583-1589. doi:10.1016/j.jbusres.2011.02.043
- [14] Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641.
- [15] Shen, Kathy Ning, & Khalifa, Mohamed. (2012). System design effects on online impulse buying. *Internet Research*, 22(4), 396-425. doi:10.1108/10662241211250962
- [16] Suhud, Usep, & Herstanti, Ghassani. (2017). Investigating the Impulse Buying of Young Online Shoppers. *Advanced Science Letters*, 23(1), 660-664. doi:10.1166/asl.2017.7290
- [17] Verhagen, Tibert, & van Dolen, Willemijn. (2011). The influence of online store beliefs on consumer online impulse buying: A model and empirical application. *Information & Management*, 48(8), 320-327. doi:10.1016/j.im.2011.08.001
- [18] Vonkeman, Charlotte, Verhagen, Tibert, & van Dolen, Willemijn. (2017). Role of local presence in online impulse buying. *Information & Management*, 54(8), 1038-1048. doi:10.1016/j.im.2017.02.008
- [19] Wu, Yi, Xin, Liwei, Li, Dahui, et al. (2020). How does scarcity promotion lead to impulse purchase in the online market? A field experiment. *Information & Management*. doi:10.1016/j.im.2020.103283
- [20] Xu, Yin, & Huang, Jin-Song. (2014). Effects of Price Discounts and Bonus Packs on Online Impulse Buying. *Social Behavior and Personality: an international journal*, 42(8), 1293-1302. doi:10.2224/sbp.2014.42.8.1293
- [21] Zou, Tingting. (2018). Online Impulse Buying Behavior Amongst Undergraduate Students in Tianjin, the People's Republic of China. *ABAC Journal*, 38(2), 94-113.

Short Research Paper

The Content Influence Mechanism of the Behavior of Poverty Alleviation Crowdfunding Users: An Empirical Study of A Chinese Crowdfunding Platform

Xiaoping Liu^{1*}, Shuang Chen¹, Yi Yang¹

¹School of Economics and Management, Chongqing University of Posts and Telecommunications, China

Abstract: Poverty alleviation crowdfunding is an "E-commerce + Agriculture + Crowdfunding" innovative mode, which promotes the targeted poverty alleviation and rural economic development in China. Previous researches mainly focus on non-content elements (e.g., the number of comments and praise) while publications are rarely related to content elements. We suggest that content elements (e.g., text of products) also have an important role in helping users to make different decisions (i.e., purchasing or donating). We select antipoverty programs in JD as a sample and adopt the linear regression model to analyze the influence of text type on the behavior of crowdfunding users. Specifically, the experimental results indicate that the emotion text (ET) make users tend to donate while the product text (PT) make users tend to purchase. This study will help project sponsors to achieve better effects on poverty alleviation by adjusting the ratio of ET to PT.

Keywords: user behavior, text feature, poverty alleviation, crowdfunding

1. INTRODUCTION

Crowdfunding refers to a fundraising channel through which the fundraiser can integrate resources of the mass public via the Internet to provide fund assistance for certain activities performed by other organizations or individuals^[1]. Compared to traditional fundraising approaches, crowdfunding has fewer restrictions and higher financing efficiency^[2]. crowdfunding can be categorized into four types: reward-based, equity-based, loan-based and donation-based^[3]. Poverty alleviation crowdfunding includes the characteristics of two crowdfunding models: reward-type and donation-type. Once a project is successfully implemented, the reward-type of investors will get products, while the donation-type will not^{[4],[5]}. Reward-type of crowdfunding offers an investment opportunity associated with a consumption experience^[6] where consumers can choose the amount they give to a project which allows individuals to reveal their valuation about a certain product.

Previous researches mainly focus on the reward-based model^[7] and donation-based model of crowdfunding^[8], and little attention has been paid to. The combination of two crowdfunding model where there are explicit project or non-project gains for the project backers. The development in Web 2.0 has propelled the growth of crowdfunding platforms, which integrate the requests for reward-based or donation-based under one domain and make it available to everyone who is interested in Poverty alleviation crowdfunding. There are several online crowdfunding sites that host crowdfunding projects combine charitable and reward-based such as Kickstarter, Ulule, Taobao and JD. Therefore, we review one e-commerce platform, named "JD" to be the research object. Specifically, we seek to answer two research questions: (1) Do crowdfunding investors prefer purchase or donation? (2) What are the factors that influence crowdfunding users to make donation or consumption decisions?

Crowdfunding's fundamental purpose is to ensure the successful implementation of the project. The success of crowdfunding projects does not only depend on the financial performance but also the number of

* Corresponding author. Email: liuxiaoping@cqupt.edu.cn(Xiaoping Liu)

project backers. However, few research has analyzed the mechanisms of project backer's number to crowdfunding projects' success. Scholars have conducted a series of research in raise funds, they^{[9]-[11]} tried to explore determinants of crowdfunding performance and delineate the influence of non-content characteristics on the financial performance of crowdfunding projects. These works help to understand the non-content elements related to financing performance, but they often overlook the influences of the content characteristics. The role of the text is important as project creators could to persuade the potential backers with a convincing story, and push project backers to make crowdfunding decision. Research on the content characteristics of the crowdfunding projects is limited^{[12]-[14]}. In our exploration, we try to demonstrate how different types of item description appeals in a textual message influence the project backer's decisions.

2. MATERIALS AND METHODS

In existing research, influencing factors of user's crowdfunding decision were usually categorized into two types: non-content factors or content factors. The non-content factors refer to the elements that irrelevant to the description information on the project interface, (e.g., social capital of fundraisers, project duration and funding goals). In contrast, content factors refer to the elements that description information of project, which can be divided into emotion text (ET) and product text (PT), these two types of content factors play different roles in crowdfunding decision of project backers.

ET refers to the description information full of personal feelings, which describes personal circumstances and shows the identity of special groups, (e.g., poor farmers) and uses a positive or negative narrative tone. We identify language indicative of positive or passive psychological capital: consisting of failure, risk and achievement, as an important emotional signal in poverty alleviation crowdfunding. The personal background description of the project creators is an important basis for project backers to make donation decisions. Personal background intends to constructs project creator's "weak" identity through the description of their social identity, occupation and economic status, so as to constructs its "weak" identity in the society and stimulates the compassion of the donor and obtain a better fund-raising effect.

We found that women related words (e.g., daughter and mother) appear in the emotion text of crowdfunding projects will inspire the trust and empathy of crowdfunding users to make donation decisions^[8]. The emotional text of the project creators can improve the accuracy of crowdfunding project performance prediction^[15], and the background information can produce good charitable results^[16]. In total, the longer ET, the more empathic users will be aroused, and users will be inclined make donation decisions. Accordingly, we propose two hypotheses as shown in Figure 1:

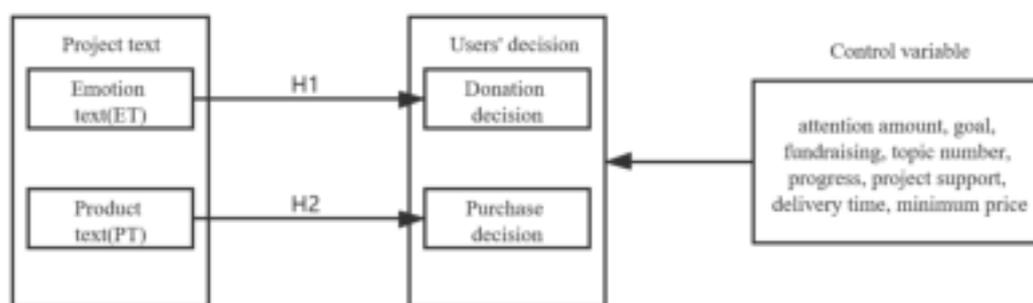


Figure 1. Theoretical framework

H1: The presence of emotion text (ET) in a Poverty alleviation crowdfunding project is positively related to

the project backer likelihood of making a donation decision.

PT refers to the overall description of the product parameters, including the detailed characteristics, brand, functions and place of origin, which is more objective than the ET. The description of PT, including product parameters needed to be identified by project backers, such as appearance, taste and quality etc.

For project backers, product text is an important way to understand products and make consumption decisions. The study found that the use of specific and accurate description in crowdfunding projects can make project backers more aware of the project, enable project backers to make consumption decisions^[17]. Project backers possess greater ability and motivation to make careful evaluations on crowdfunding projects, when there is objective evidence about the quality of the product, the specific benefits and costs of the product^[13]. Project backers need to rely on a large number of product texts for the comparison and selection of crowdfunding products. The presentation of these product texts is a "signal" sent by the project sponsor. With the help of product texts, Project backers can have a better understanding of products, which has an important impact on user crowdfunding decision-making. Therefore, the longer the PT, the more product information will be provided, and crowdfunding users will be inclined make purchase decisions. Thus, we propose the following:

H2: The presence of product text (PT) in a Poverty alleviation crowdfunding project is positively related to the project backer likelihood of making a purchase decision.

Prior studies on crowdfunding have indicated that non-content factors are important predictors of success of crowdfunding projects. Thus, we use the presence of some non-content factors as control variables in this analysis.

We recorded the project data, that include 14 information, apart from the textual requests, the data also contain additional project information such as history of posts, comments, and his/her account age, of all the projects on this platform From January 2015 to October 2019. Using the crawlers Python converted to a comma separated values file format and removing the projects with missing information. Finally, we obtain detailed data of 583 projects with poverty alleviation.

The proportion of different types of text in the total text reflects the attention of project creators to different attributes of poverty alleviation crowdfunding projects, which further affects the decision-making between user consumption and donation. Therefore, we make detailed statistics on the number of words of emotion text (ET) and product text (PT) to support subsequent experiments.

3. RESULTS

3.1 Variable descriptions and measurements

3.1.1 Dependent variable

We explore the degree of user's inclination between donation decision-making and purchase decision-making. In line with this research, we also operationalize crowdfunding user's decision as a continuous variable measuring the number of committed to the project by investors called funds raised. The increase or decrease of the number of crowdfunding users can directly reflect the user's inclination of decision-making. Therefore, we use the proportion of donators FD% (free donation), and the percentage of purchasers PD% (purchase donation) as the index to measure the user's inclination of decision-making. If FD% increases, it indicates that the crowdfunding users tend to make donation decisions, otherwise, it indicates that the crowdfunding users tend to make purchase decisions ($PD\% + FD\% = 1$).

3.1.2 Explanatory variables

For the sake of describing the level of detail in the project text and eliminating the effects of other factors on the decision of crowdfunding users, we apply ET % as a measure index of how much emotional text occupies the full length, and the PT% is take as a measure index of how much product text occupies the full length. When

the ET% of crowdfunding project is more than 50%, it is considered that the description of the project is dominated by the emotional text. When the ET% of crowdfunding project is less than 50%, it is considered that the description of the project is dominated by the product text (PT) ($PT\% + ET\% = 1$).

3.2 Data analysis

3.2.1 Summary statistics of variables

The descriptive statistics and the summary of the variables in this paper are shown in Table 1. The mean of the funding goal is ¥21759.55, the funding goal has a maximum value span ranging from ¥38 to ¥1,000,000, indicating that the fund expectation of project sponsors varies greatly. Concerning the length of project text, the mean length of emotional text is 713.7 and the mean length of product text is 584.78. From Table 2, the means of donators and purchasers are 130.4 and 586.19, respectively. we notice that the number of consumers is significantly more than the number of donors. The standard deviation of donators and purchasers are much higher and more unbalance.

Table 1. Summary statistics of variables

	Min	Max	Mean	Std. Dev	Obs.
DA	0	8620	130.04	700.875	583
PC	4	76577	586.19	3278.439	583
<i>lgFD%</i>	0	0.9850303	0.1033138	0.180438125	583
<i>lgPD%</i>	0.015	1	0.8966862	0.180438125	583
GL	38	1000000	21759.55	61332.767	583
FR	1340	1058069	69032.42	134078.614	583
PG	18	16727	540.4974	1266.34721	583
PT	30	2490	713.7	358.887	583
ET	0	3652	584.78	382.544	583
PT%	0.019	1	0.5598272	0.184931768	583
ET%	0	0.9809765	0.4401728	0.184931768	583
AT	5	4000	163.84	337.602	583
PS	0	974	10.78	71.019	583
TP	0	953	31.46	59.471	583
DT	1	299	14.17	20.471	583
LP	0	798	47.2	65.664	583

DA: Donator, PC: Purchaser, GL: Goal, FR: Fund-raising, PG: Progress, AT: Attention amount, PS: Project support, TP: Topic number, DT: Delivery time, LP: Lowest price

Our funds raised independent variable followed a right skewed, gamma distribution, which presents analytical challenges. For variables with non-zero values we used a natural log transformation to correct for this skewness. This transformation allows us to interpret variables transformed using the natural log and those using the inverse hyperbolic sine transformation in the same way. Positive psychological capital, funds raised, created, Facebook friends, and backed were all transformed using the inverse hyperbolic sine transformation.

For the purpose of improving the regression fit, and make dependent variables present normal distribution, eliminate the heteroscedasticity problem, and do not change the relative relationship between variables, we log-transform dependent variable with the equation $\lg(\text{original value})$. in the later model, the variables FD and PD are log converted, *lgFD%* and *lgPD%* are used as dependent variables. We delete the variables of original value is 0, and ensure that all the variables are valid.

3.2.2 Variable correlation analysis

Interactions are provided in Table 2 below. provides the correlation between the variables. It can be seen that the correlation ($\gamma = -1.000$, $P = 0.00 < 0.01$) is between $lgFD\%$ and $lgPD\%$, it is a relation of completely negative correlation. Because the product text (PT) and emotion text (ET) constitute the overall text information of the project. It is normal to observe a high correlation ($\gamma = 0.292$, $P = 0.000 < 0.001$) between $ET\%$ and $lgFD\%$. From Table 2, between $PT\%$ and $lgPD\%$ have the high correlation $\gamma = 0.32$, $P = 0.001 < 0.05$. Meanwhile, we believe that the goal ($\gamma = -0.04$, $P = 0.34 > 0.05$), fundraising progress ($\gamma = 0.064$, $P = 0.121 > 0.05$), attention amount ($\gamma = -0.079$, $P = 0.057 > 0.05$), support form project creator ($\gamma = 0.025$, $P = 0.548 > 0.05$), topic number ($\gamma = -0.059$, $P = 0.158$), expected return delivery time ($\gamma = 0.029$, $P = 0.485 > 0.05$) and minimum price ($\gamma = 0.007$, $P = 0.86 > 0.05$) are unrelated to the $ET\%$.

Table 2. Variable correlation matrix

	Variables	1	2	3	4	5	6	7	8	9	10
1	$lgFD\%$	1									
2	$lgPD\%$	-.593** 0	1								
3	$ET\%$.292** 0	-.132** 0.001	1							
4	$PT\%$	-.292** 0	.132** 0.001	-1.000** 0	1						
5	GL	-.113** 0.008	0.057 0.166	-0.04 0.34	0.04 0.34	1					
6	FR	.093* 0.03	-.286** 0	0.064 0.121	-0.064 0.121	-0.063 0.131	1				
7	AT	-.240** 0	0.068 0.101	-0.079 0.057	0.079 0.057	.142** 0.001	.199** 0	1			
8	ST	.182** 0	-.334** 0	0.025 0.548	-0.025 0.548	-0.024 0.568	.135** 0.001	-0.001 0.979	1		
9	TP	-.222** 0	-0.007 0.861	-0.059 0.158	0.059 0.158	0.017 0.69	.548** 0	.380** 0	0.054 0.192	1	
10	DT	.113** 0.008	-.088* 0.035	-0.029 0.485	0.029 0.485	.194** 0	-0.06 0.145	0.025 0.554	-0.007 0.869	0.036 0.389	1
11	LP	.286** 0	-.114** 0.006	-0.007 0.86	0.007 0.86	-0.024 0.568	-0.063 0.126	-.109** 0.008	-0.051 0.215	-.084* 0.042	.161** 0

GL: Goal, FR: Fund-raising, AT: Attention amount, ST: Supporter, TP: Topic number, DT: Delivery time, LP: Lowest price

All correlations with an absolute value greater than (0.05) are significant at $p < 0.05$ and an absolute value greater than (0.07) at $p < 0.01$.

3.3 Empirical Model

According to the qualitative analysis above, we construct the econometric model, we use a Hierarchical regression model with robust standard errors. To better appreciate the impact of content factors and non-content factors, we run three different models, that is Model 1, Model 2 and Model 3. Model 1 is the controls only model. We incrementally add the non-content factors to analyze the impact of the crowdfunding users of donation or purchase decision, expressed as formula (1) and formula (2). Model 2 highlights the scenario where we study the emotional text (ET) on the crowdfunding users of donation decision, expressed as formula (3). Model 3 considers

the product text (PT) on the crowdfunding users of donation decision, expressed as formula (4).

Model lists look like this:

- Model 1 (Controls only model)

$$lgFD\% = \alpha_0 + \alpha_1 \text{Fundraising} + \alpha_2 \text{Progress} + \alpha_3 \text{Attention} + \alpha_4 \text{Support Projects} + \alpha_5 \text{Project Topics} + \alpha_6 \text{DeliveryTime} + \alpha_7 \text{Lowest Price} + \varepsilon \quad (1)$$

$$lgPD\% = \alpha_0 + \alpha_1 \text{Progress} + \alpha_2 \text{Support Projects} + \alpha_3 \text{Delivery Time} + \alpha_4 \text{Lowest Price} + \varepsilon \quad (2)$$

- Model 2 (ET on the crowdfunding user's donation decision)

$$lgFD\% = \alpha_0 + \alpha_1 \text{Fundraising} + \alpha_2 \text{Progress} + \alpha_3 \text{Attention} + \alpha_4 \text{Support Projects} + \alpha_5 \text{Project Topics} + \alpha_6 \text{DeliveryTime} + \alpha_7 \text{Lowest Price} + \alpha_8 \text{ET}\% + \varepsilon \quad (3)$$

- Model 3 (PT on the crowdfunding user's donation decision)

$$lgPD\% = \alpha_0 + \alpha_1 \text{Progress} + \alpha_2 \text{Support Projects} + \alpha_3 \text{Delivery Time} + \alpha_4 \text{Lowest Price} + \alpha_5 \text{PT}\% + \varepsilon \quad (4)$$

3.4 Estimation Results

Table 3. Estimation Results

Variables	Model1		Model2	Model3
	lgFD%	lgPD%	lgFD%	lgPD%
FR	-0.082* 0.032		-0.081* 0.028	
PG	0.291** 0.000	-0.258** 0.000	0.261** 0.000	0.251** 0.000
AT	-0.141** 0.001		-0.126* 0.001	
PS	0.172** 0.000	-0.307** 0.000	0.170** 0.000	-0.305** 0.000
TP	-0.317** 0.000		-0.288 0.000	
DT	0.115* 0.003	-0.084* 0.027	0.114* 0.002	-0.087* 0.021
LP	0.248** 0.000	-0.132** 0.001	0.250** 0.000	-0.132** 0.000
ET%			0.244** 0.000	
PT%				0.112* 0.003
R ²	0.257	0.199	0.315	0.211
Adjusted- R ²	0.247	0.193	0.305	0.204
D-W	1.900	2.039	2.059	2.071
F	26.441	35.845	30.737	30.914
VIF			1.021	1.005

FR: Fund-raising, PG: Progress, AT: Attention amount, PS: Project support, TP: Topic number, LP: Lowest price, DT: Delivery time *p<.05;

**p<.01.

Table 3 provides the estimation results for all three models. Model 1 studies the impact of control variables only. Model 2 considers the emotional text (ET) on the influence of crowdfunding user's decision of a donation. Model 3 considers the product text (PT) on the influence of crowdfunding user's decision of purchase. Table 3 reports the coefficients of the hierarchy regression and some of the fit statistics (R^2 , Adjusted- R^2 , and D-W). We specify two forms of R^2 in our research. The values of VIF is less than 5, which not strong enough to evoke the problem of multi-collinearity. The D-W values of the three models are all around 2. It is proved that there is no sequence correlation.

It can be seen from the regression results that the overall fitting effect of model 1 (1) is better ($R^2 = 0.257$). Then, add the explanatory variable ET% to the model 2, the linear regression effect of ET% is significant ($\beta = 0.244$, $P = 0.000 < 0.001$), and R^2 is increased, the overall effect of the model is improved ($R^2 = 0.315$). It is proved that the emotional support text can significantly affect the crowd funding users who are more inclined to make donation decisions. The significance of emotional text implies that H1 is supported. H1 illustrates the importance of emotional text appeals for persuading crowdfunding users to donate.

Similarly, the overall fitting effect of model 1 (2) is good ($R^2 = 0.199$). When the explanatory variable PT% is added to the model, model 3 is obtained. At this time, the linear regression effect of explanatory variable PT% is significant ($\beta = 0.112$, $P=0.003<0.05$), and R^2 increases, the overall effect of the model is improved ($R^2=0.211$), which proves that product description text can significantly affect crowdfunding users' preference to make consumption decisions. We find support for H2, and it indicates that the presence of product text in a crowdfunding project has significant impact on the success of product purchase.

4. CONCLUSIONS

Our research illustrates the importance of textual factors in determining the success of a user's decision in an online platform. Based on the content categorization, we further expound on the emotional text, and product text appeals amongst the content features. We show content categories (ET vs. PT), have a significant impact on the likelihood of the user's crowdfunding decision. Specifically, we find that presence of emotion text increases the probability of user's donate intention, and product text can stimulate the users' to make a purchase intention decision, it positively impacts the likelihood of decision success. Our research will be beneficial to all those project creators who post their requests on crowdfunding platforms.

Our study is the first to investigate content text as an important influence factor and explore its role in the entrepreneurial fundraising process. For scholars, our study advances understanding of the determinants of successful crowdfunding campaigns - introducing content text as an important theoretical lens for understanding crowdfunding user's decision. Our study also underscores the importance of considering the interactions of two crowdfunding model, versus studying one model in isolation. For creators, our research suggests that creators would benefit from proactively signaling emotion content and product content when raising funds through crowdfunding. We hope that these findings and their associated implications lead to further academic inquiry regarding the role of content factor in poverty alleviation crowdfunding.

ACKNOWLEDGEMENT

This research was supported by the Chongqing Social Science Planning Project under Grant 2018YBGL065 and 2019YBGL065, Chongqing Municipal Education Commission Humanities and Social Sciences Research Project under Grant 18SKGH041.

REFERENCES

- [1] Lin Y, Boh WF, Goh KH. (2004). How different are crowd funders? Examining archetypes of crowd funders and their choice of projects. *Social Science Electronic Publishing*
- [2] Allison TH, Davis BC, Short JC, Webb JW. (2015). Crowdfunding in a prosocial micro-lending environment: examining the role of intrinsic versus extrinsic cues. *Entrepreneurship: Theory and Practice*, 39(1): 53-73
- [3] Beaulieu T, Sarker SA. (2015). Conceptual framework for understanding crowdfunding. *Visual arts research*, 37(1): 1-31
- [4] Gleasure R, Feller J. (2016). A Rift in the ground: theorizing the evolution of anchor values in crowdfunding communities through the oculus rift case study. *Journal of the Association of Information Systems*, 17(10): 708-736
- [5] Gleasure R, Feller J. (2016). Emerging technologies and the democratisation of financial services: a metatriangulation of crowdfunding research. *Information and Organization*, 26(4): 101–115
- [6] Schwienbacher, A. (2015). Entrepreneurial risk-taking in crowdfunding campaigns. *Small Business Economics*.
- [7] H. Zheng, D. Li, J. Wu. (2014). The role of multidimensional social capital in crowdfunding: A comparative study in China and US. *Information & Management*, 51 (4): 488–496
- [8] Majumdar A, Bose I. (2018). My Words for Your Pizza: An Analysis of Persuasive Narratives in Online Crowdfunding. *Information & Management*, 55(6): 781-794.
- [9] Mollick E R. (2014). The Dynamics of Crowdfunding: An Exploratory Study. *Social Science Electronic Publishing*, 29(1): 1-16
- [10] André, Kévin, Bureau S, Gautier A. (2017). Beyond the Opposition Between Altruism and Self-interest: Reciprocal Giving in Reward-Based Crowdfunding. *Journal of Business Ethics*, 146(2): 313-332
- [11] Lukkariinen A, Teich J E, Wallenius H. (2016). Success drivers of online equity crowdfunding campaigns. *Decision Support Systems*, 87: 26-38
- [12] Allison T H, Mckenny A F, Short J C. (2013). The effect of entrepreneurial rhetoric on microlending investment: An examination of the warm-glow effect. *Journal of Business Venturing*, 28(6): 690-707
- [13] Allison T H, Davis B C, Webb J W, et al. (2017). Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. *Journal of Business Venturing*, 32(6): 707-725
- [14] Parhankangas A, Renko M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing*, 32(2): 215-236.
- [15] Mitra T, Gilbert E. (2014). The language that gets people to give: phrases that predict success on kickstarter. *Acm Conference on Computer Supported Cooperative Work & Social Computing*. ACM.
- [16] Grau S L, Folse J A G. (2007). Cause Related Marketing: The Influence of Donation Proximity and Message Framing Cues on the Less Involved Consumer. *Journal of Advertising*, 36(04): 19-33
- [17] Parhankangas A, Renko M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing*, 32(2): 215-236

Full Research Paper**Research on Continuous Use of B2B Platform in Chinese Intelligent Engineering Companies Based on the Theory of Resource Complementary***Zhichao Zhang**, *Hongbo Jiang*, *Ting Shao*

School of Economics and Management, Xiamen University of Technology, Xiamen, 361024, China

Abstract: What factors affect the continuous use of B2B platforms by intelligent engineering companies is an important issue. Based on the theory of resource complementarity, a model of influencing factors for intelligent engineering companies to continue using the B2B platform is constructed, and four influencing factors including the complementary resources given by the platform, the complementary resources given by the company, interaction quality, and exploration and exploitation capability are analyzed. Using SmartPLS 3.0 to analyze 217 survey data, the results show that the complementary resources given by the platform and companies have positive impact on the interaction quality, and the complementary resources given by the platform have positive effect on the exploration and exploitation capabilities. Interaction quality has positive impact on the company's continuous use intention of the B2B platform. The complementary resources given by the company have no positive impact on the exploration and exploitation capabilities, and the exploration and exploitation capabilities have no positive impact on the company's continuous use intention of the B2B platform. Finally, some suggestions are proposed to increase the company's continuous use intention of B2B platform.

Keywords: Intelligent engineering company, B2B platform, theory of resource complementarity, continuous use intention.

1. INTRODUCTION

In recent years, China's B2B transaction service revenue has shown a year-on-year growth trend. In 2019, the transaction size of China's B2B market reached 25.94 trillion yuan, a year-on-year increase of 21.39% [1]. With the rapid development of B2B platforms, industries such as agriculture, automotive, and intelligent engineering industries have actively joined various B2B platforms to develop their own B2B e-commerce. As an emerging field, the intelligent engineering company is developing rapidly, but has problems such as difficulty in obtaining information [2] and poor service quality [3]. The advantages of the B2B platform that are conducive to the company's digitalization [4], integrate the relationship between all parties in the company operation [5], and improve the quality of management decision-making [5]. So it can to help companies integrate information and improve service quality [6, 7]. The B2B platform improves corporate performance by saving costs [8] and reducing the risk of credit and capital in the transaction process [9]. The B2B platform and the intelligent engineering company have formed a complementary relationship of resources through cooperation, generating synergy effects and promoting each other's development. Resource complementarity has significant positive impact on the cooperative innovation performance of high-tech companies [10]; organizations can gain resource value through complementary effects, thereby improving cooperation performance [11]; resource complementarity enables individuals to break through their own capabilities and enhance the overall level of cooperation [12]. The existing literature mainly studies how resource complementarity can improve the performance of company, but has not used the theory in the research between companies and B2B platforms. This is the innovation of this research.

At present, scholars mainly apply the Technology Acceptance Model (TAM) [13], Theory of Rational Action (TRA) [14], Theory of Planned Behavior (TPB) [15], and Diffusion of Innovations (DI) [16] and other theories

* Corresponding author. Email: hbjiang@xmut.edu.cn(Hongbo Jiang)

(models) study the continuous use intention of individuals. In the study of the continuous use intention of the companies, Wu Chaoyan and Huang Lei (2015) applied the value adoption model (VAM) and the expectation confirmation model (ECM) theory to analyze the intention of retail companies' continuous use of mobile social media, and found that both perceived value and satisfaction helped to increase the continuous use intention of retail companies [17]. Pei Qian and Hu Yanhua (2020) found through overall analysis and industry comparison that feedback to logistics companies on the use of government service items will increase logistics companies' intention to continue using government service items [18]. The theory of resource complementarity refers to the fact that a single resource cannot exert its maximum utility. Only when it is combined with other resources to produce complementary effects can it create greater value and bring about resource appreciation [19]. The theory of resource complementarity has been used to explore the relationship between ERP and e-commerce [20] and company performance [21], but rarely used to analyze companies' continuous use of B2B e-commerce platforms. B2B platforms and companies maximize the utility of resources by complementing each other's resources, which not only promotes the development of each other, but also brings higher benefits to themselves. Therefore, this article applies the theory of resource complementarity to explore the factors that affect the intention of intelligent engineering companies' continuous use of B2B e-commerce platforms. The research's objectives are:

1. Based on the theory of resource complementarity, explore the factors influencing companies' continuous use of B2B platforms;
2. Analyze the four influencing factors that affect the companies' continuous use of the B2B platform, that is, the relationship between the complementary resources given by the platform, the complementary resources given by the companies, interaction quality, and exploration and exploitation capability;
3. Provide countermeasures and suggestions for improving the intelligent engineering companies' continuous use of B2B platforms ;
4. Expand the application scenarios of resource complementarity theory.

The research route of this article is as follows: The first part is introduction, which introduces the research background, the problems to be studied, the goals, and the significance of the research. The second part is literature review, which reviewing the important contributions in this field. The third part is research hypothesis and model. Based on previous research results, it proposes the factors affecting the intelligent engineering companies' continuous use of B2B e-commerce platforms, and constructs a model of the factors. The fourth part is data collection and analysis. Finally are research conclusions and limitations.

2. LIERATURE REVIEW

2.1 The Theory of Resource Complementarity

The theory of resource complementarity was proposed in the 1990s. It means that a single resource cannot exert its maximum utility. Only when it is combined with other resources to produce complementary effects, can it create greater value and realize resource appreciation [19]. Complementary resources are not a single resource, but a collection formed by the aggregation of many resources, which are the value added by the existence of other resources [22]. In recent years, the research directions mainly include the following:

The first is to choose a more suitable partner by studying the complementation of heterogeneous resources. Liu Keyin (2016) found the three key factors that determine the complementary matching of resources in cooperation-resource heterogeneity, resource relevance, and various transaction costs of acquiring and applying resources. Based on this, they proposed the Euclid cooperation distance evaluation method and the partner attraction method. That improves the scientificity and accuracy of the company's selection of partners [23]. For companies facing a clear task, it is very convincing to analyze their resource needs and use this information to evaluate the performance results of alternative partners [24]. In a survey, it was found that the higher the matching between tasks and resources, the greater the probability of contact between partners. When companies require

partners to have different resources, contact will drop. The higher the matching of resources and capabilities with key companies and tasks, the partnership will be more stable and firm [25].

The second is to promote cooperation through the theory of resource complementarity. Complementary resources and complementary capabilities between companies have positive impact on the cooperative relationship [26]. Dussauge (2000) believes that greater resource complementarity can lead to less knowledge overlap, and less knowledge overlap is one of the starting points for companies to adopt cooperation. Different alliances can interact to develop complementary resources and credit [27]. Complementary resources can improve the credit and performance of the alliance and promote better cooperation between companies [28]. By matching satisfactory partners and strengthening the cooperative relationship between companies, the alliance can develop better and obtain higher economic benefits.

The third is that the theory of resource complementarity can improve alliance performance. The performance of the joint venture is driven by the ability of the joint venture partners to create synergies by adding complementary resources. The complementary resources of the two partners have both independent and joint impact on the performance of the joint venture [29]. It shows that the synergy effect of the complementary resources of the partners on the performance of the joint venture is substantial. Under the premise that the partners' behavior is less uncertain, and there is strong local system development and moderate government support, resource complementarity can improve alliance performance [30].

Resource complementarity can help companies choose better partners, promote cooperation, and improve alliance performance. However, there are still few studies on the theory of resource complementarity for continuous use. Complementation of resources is to generate synergy through the complementary resources provided by the two parties. In the cooperative relationship between the company and the platform, the company provides users to increase the flow of the platform, and the platform provides the company with a safer and more reliable trading place and customers information help companies conduct transactions quickly and efficiently. In 2020, the top domestic intelligent engineering industry and finance platform, Yigongcheng, has accumulated nearly 1.3 billion yuan in matching transaction projects. Yigongcheng platform and intelligent engineering companies achieve a win-win situation through complementarity of resources, which can improve efficiency and maintain a long-term and stable cooperative relationship. This reflects the research significance of this article. Through this kind of resource complementary cooperation relationship between the company and the platform, we can discover the factors that affect the continuous use of the platform by the company, and provide suggestions to improve the intention of the company to continue using the platform.

2.2 Continuous Use Intention

For a long time, companies have paid attention to the factors that influence the continuous use intention of users. The research on the continuous use intention is mainly divided into two categories: the continuous use intention of individuals and the continuous use intention of companies.

Regarding the individual's continuous use intention, improving the user's perceived ease of use can increase user's satisfaction, and then enhance the user's continuous use intention [31]. Gbongli et al. (2019) investigated 539 virtual currency users by extending the original technology acceptance model and using the structural equation-artificial neural network method, confirming that perceived usefulness can affect continuous use [32]. Consumers' perceived ease of use of bike-sharing will affect consumers' behavioral intentions, which in turn affects consumers' continuous use behavior [33]. Qi et al. (2020) based on the theory of planned behavior (TPB) and found that consumers only decide whether to use plastic surgery apps after considering factors such as physical satisfaction and perceived convenience, which will also affect the length of time consumers use apps [34].

For companies' continuous use intention, perceived usefulness, perceived ease of use, satisfaction, etc.

have significant impact on retail companies and logistics companies' intention of continue using mobile social media and Internet of Things Technologies [21, 35]. Factors such as the government's preferential policies, service quality, and technological innovation services will affect the intention of companies to continue using government service projects [18].

Most of the existing literature on the continuous use intention research is aimed at individual users, and there are few studies about the companies. Research on the continuous use of B2B platforms by companies can help B2B platforms increase usage of platform services, the number of users, and improve operating benefit. This article discusses the factors affecting the continuous use of B2B platforms by intelligent engineering companies based on the theory of resource complementarity, and provides some suggestions for enhancing the companies' use intention.

3. RESEARCH DESIGN

3.1 Research model

In the process of using the B2B platform, intelligent engineering companies have formed complementary resources, including the resources given by the platform and the resources given by the company. Complementary resources will affect the interaction quality between companies and platforms and their exploration and exploitation capabilities, and will increase the intention of intelligent engineering companies to continue using the B2B platform. The research model is constructed as shown in Figure 1.

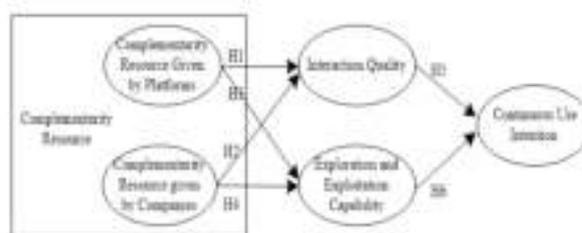


Figure 1. Research model.

3.2 Research hypothesis

3.2.1 Complementary resources and interaction quality

In this study, complementary resources include two parts: the resources given by the platform and the resources given by companies. Interaction quality refers to the communication quality and conversation level between intelligent engineering companies and B2B platforms [36]. More resource complementarity can bring less overlap of knowledge, and can more easily absorb and transform each other's resources. Frequent exchanges between the two parties promote the improvement of the interaction quality [23]. Therefore, this research proposes the following hypotheses:

H1: The resources given by the platform have significant positive impact on the quality of interaction.

H2: The resources given by the company have significant positive impact on the quality of interaction.

3.2.2 Complementary resources and exploration and exploitation capabilities.

Exploration capabilities include things captured by terms such as search, change, risk-taking, experiment, flexibility, discovery, and innovation. Exploitation capabilities include refinement, production, efficiency, selection, implementation, execution, etc. [37]. Complementary resources between companies can ensure the effective development of cooperative utilization and cooperative exploration activities [38]. Due to lack of resources, companies have always competed fiercely for resources such as capital, technology, and management. For managers, pursuing both exploration and exploitation capabilities at the same time will be a challenge. The key to solving this problem is to seek external resource supply, achieve cooperation, and use the complementary resources to improve exploration and exploitation capabilities [37, 39]. After acquiring resources, the capabilities of companies and platforms in the application of existing knowledge and the development of innovative activities will be greatly improved, that is, complementary resources will promote the exploration and development of capabilities. Therefore, this research proposes the following hypotheses:

H3: The resources given by the platform have significant positive impact on the exploration and exploitation capabilities.

H4: The resources given by companies have significant positive impact on the exploration and exploitation capabilities.

3.2.3 Interaction quality, exploration and exploitation capabilities and continuous use intention.

In the existing research, scholars have done lots of research about interaction quality on satisfaction, and satisfaction is one of the key factors affecting continuous use. Bhattacharjee (2001) believes that perceived usefulness and satisfaction will affect users' intention to continuous use of information systems [40]. Different types of interactivity can positively affect users' immersion and satisfaction. Immersion and satisfaction are regarded as the important factors that affect continuous use of information systems [41]. The three dimensions of interactivity have varying degrees of influence on immersion and satisfaction, and immersion and satisfaction have positive effect on continuous use [42]. Therefore, this research proposes the following hypotheses:

H5: Interaction quality has significant positive impact on continuous use intention.

The interaction between exploration and development capabilities will have a positive impact on a company's sales growth rate [43]. Exploitation capabilities provide a basis for the development of exploration capabilities, and exploration capabilities have positive impact on product innovation and market performance [44]. The strength of the exploration and exploitation capability will directly affect the sales growth rate, product innovation and market performance of the company. If the exploration and exploitation capability is stronger, the company can get more returns, then the company is more willing to continue use the B2B platform continuously. Therefore, this research proposes the following hypotheses:

H6: Exploration and exploitation capabilities have significant positive impact on the continuous use intention.

3.3 Questionnaire design

The questionnaire is divided into two parts: the basic information of the company and the measurement items. The basic information of the company includes the nature of the company, the time it was established, the scale of the company, and the time of entering the platform. The measurement items are shown in Table 1. There are 5 dimensions, including resources given by the platform, resources given by the company, interaction quality, exploration and exploitation capabilities, and continuous use intention. Experts are invited to review the translated content for all items. Using Likert 7-level scale, 1 means "completely disagree", 4 means "neutral", and 7 means "completely agree".

Table 1. Dimensional measurement

Dimension	Number	Item	Reference
complementary resource given by the platform(CRP)	CRP ₁	The online payment function provided by the platform is safe and fast.	Harrison (2001)
	CRP ₂	The potential market demand within the platform is large, the market prospect is broad, and the market resources are abundant.	
	CRP ₃	The cooperative customers in the platform are loyal and credible.	
	CRP ₄	The platform has powerful interactive functions, and it is convenient for companies to communicate on the platform.	
	CRP ₅	The platform provides resources that the company does not have, and makes a significant contribution to the company.	
complementary resource given by the company(CRC)	CRC ₁	The service quality of the company is great, and platform orders can be met in time.	Harrison (2001)
	CRC ₂	The products or services output by the company to the platform are of high quality.	
	CRC ₃	The brand influence of companies on the platform is strong.	
	CRC ₄	The presence of companies can expand the market share and influence of the platform.	
	CRC ₅	The resources provided by companies that are not available on the platform themselves are a significant contribution to the platform.	
interaction quality (IQ)	IQ ₁	Companies and platforms can instantly share market information or customer demand information with each other.	Dussauge (2000)
	IQ ₂	Companies and platforms can share information about technological	

		changes in important products and services.	
	<i>IQ₃</i>	The interaction between the company and the platform is seamless and the communication is constructive.	
	<i>IQ₄</i>	The connection between the company and the platform is very close.	
exploration and exploitation capability (EEC)	<i>EEC₁</i>	After entering the platform, the company's ability to develop new markets has become stronger.	March (1991), Gupta (2006)
	<i>EEC₂</i>	After entering the platform, companies have become more capable of reducing costs.	
	<i>EEC₃</i>	After entering the platform, companies are more capable of improving product flexibility and personalization.	
	<i>EEC₄</i>	After companies settled on the platform, the platform's operational capabilities and scope have been improved.	
	<i>EEC₅</i>	After companies settled on the platform, the platform has strengthened its ability to improve product and service quality.	
	<i>EEC₆</i>	The business that the platform gives to the company is taking up an increasing proportion of the overall business of the company.	
continuous use intention (CUI)	<i>CUI₁</i>	The intention is to continue using the platform instead of discontinuing it.	Bhattacharjee (2001)
	<i>CUI₂</i>	Intend to continue to use this platform instead of using any other channels or platforms.	
	<i>CUI₃</i>	Recommend other companies to use this platform.	
	<i>CUI₄</i>	If possible, plan to stop using the platform. (reverse coding)	

4. DATA COLLECTION AND ANALYSIS

4.1 Data collection

This research takes company registered on the intelligent engineering B2B platform-Yigongcheng (<https://www.yigongcheng.com/>) as the survey objects, conducts surveys through Internet, and distributes the questionnaires by email, WeChat, QQ and other social software. The survey started in December 2018 and lasted for one month. A total of 300 questionnaires were distributed and 217 valid questionnaires were returned. The effective response rate of the questionnaire was 71.6%.

In terms of the nature of companies, the percentages of private companies, joint ventures, sole proprietorships and state-owned companies in the survey were 38.2%, 34.6%, 20.3%, and 6.9% respectively. According to the proportions of various types of companies in my country, the surveyed companies among them, private companies and joint ventures account for a relatively large proportion, while state-owned companies account for a small proportion, indicating that the structure of the sample conform to the actual situation. From the perspective of the establishment time of companies, most companies are established in the two intervals of 3 years or less and 5-10 years, accounting for 67.8% of the total. From the perspective of company scale structure, 30.4% have 100 or less employees, 28.1% have 101-300 employees, and 34.6% of the companies have 301-500 employees. From the perspective of the time of companies entering the B2B platform, the largest part is within 1-3 years, accounting for 32.7%. In this period, e-commerce was developing rapidly, attracting a large number of companies to settle on the B2B platform. The mean value of each dimension in this study tends to the middle value of 5, and there are no extreme values and missing values. Generally speaking, the data of the questionnaire basically obeys the normal distribution, and subsequent research and analysis can be carried out.

4.2 Reliability and validity analysis

4.2.1 Reliability test

The reliability of the data is tested by the Cronbach's α value and the CITC value of the total correlation coefficient of the corrected term. First, check Cronbach's α value. If it is higher than 0.8, the reliability is high; if it is between 0.7 and 0.8, it is good; if it is between 0.6 and 0.7, it is acceptable; if it is less than 0.6, it is not good. Secondly, the CITC value is generally greater than 0.5. In addition, observe the Cronbach's α value of the item and the Cronbach's α value after the item is deleted. If the Cronbach's α value of the item is less than the Cronbach's α value after the item is deleted, then the item should be deleted. The Cronbach's α value of CRP, CRC, IQ, EEC and CUI are all greater than 0.6, indicating that the credibility is acceptable. The Cronbach's α

value after the item is deleted of IQ_2 and EEC_6 is greater than the Cronbach's α value, and the CITC value of CUI_2 and CUI_6 is less than 0.5, so the item IQ_2 , EEC_6 , CUI_2 and CUI_4 must be deleted. Table 2 shows the reliability test of deleting items IQ_2 , EEC_6 , CUI_2 and CUI_4 .

Table 2. Reliability test after deleting item IQ_2 , EEC_6 , CUI_2 , CUI_4

Dimension	Number	CITC	Cronbach's α after deleting the item	Cronbach's α
complementary resource given by the platform(CRP)	CRP_1	.796	.868	0.900
	CRP_2	.767	.874	
	CRP_3	.782	.871	
	CRP_4	.659	.897	
	CRP_5	.755	.877	
complementary resource given by the company(CRC)	CRC_1	.733	.857	0.884
	CRC_2	.678	.870	
	CRC_3	.770	.848	
	CRC_4	.647	.876	
	CRC_5	.782	.845	
interaction quality (IQ)	IQ_1	.793	.757	0.863
	IQ_3	.735	.812	
	IQ_4	.694	.850	
exploration and exploitation capability (EEC)	EEC_1	.787	.855	0.891
	EEC_2	.678	.881	
	EEC_3	.736	.869	
	EEC_4	.710	.873	
	EEC_5	.776	.858	
continuous use intention (CUI)	CUI_1	.569		0.724
	CUI_3	.569		

4.2.2 Validity test.

In this study, statistical analysis software SPSS26.0 was used to test the validity of the sample data. The KMO statistic was 0.736, the value of Sig was less than 0.001, and the Bartlett sphere test was significant, indicating that the data is suitable for factor analysis. After rotating and deleting the factors of items less than 0.5, the factor load matrix after rotating shows that the range of factor load is 0.614-0.830. Extracting 5 factors, the total explained variance is 68.961%, and the load of each index on the corresponding factor is greater than the cross load on the other factors, indicating that the questions have good validity.

4.2.3 Convergent validity and discriminative validity.

The test criteria for convergent validity are combined reliability (CR) and average variance extracted (AVE). When the CR value and the AVE value are greater than 0.7 and 0.5, It shows that the data has good convergent validity. The results of the convergent validity test are shown in Table 3. The CR values of the sample data are all greater than 0.7, and the AVE values are all greater than 0.5, indicating that each dimension has good convergent validity.

Table 3. Convergent validity test

Dimension	Number	CR	AVE
complementary resource given by the platform(CRP)	CRP_1	0.871	0.575
	CRP_2		
	CRP_3		
	CRP_4		
	CRP_5		
complementary resource given by the company(CRC)	CRC_1	0.873	0.578
	CRC_2		
	CRC_3		
	CRC_4		
	CRC_5		
interaction quality (IQ)	IQ_1	0.750	0.504
	IQ_3		
	IQ_4		
exploration and exploitation capability (EEC)	EEC_1	0.874	0.539
	EEC_2		
	EEC_3		
	EEC_4		

Dimension	Number	CR	AVE
continuous use intention (CUI)	<i>EEC₅</i>	0.802	0.576
	<i>EEC₆</i>		
	<i>CUI₁</i>		
	<i>CUI₂</i>		
	<i>CUI₃</i>		

The discriminative validity is shown in Table 4. The correlation coefficient between each dimension and other dimensions is significantly smaller than the arithmetic square root of its AVE value, indicating that the scale has good discriminant validity.

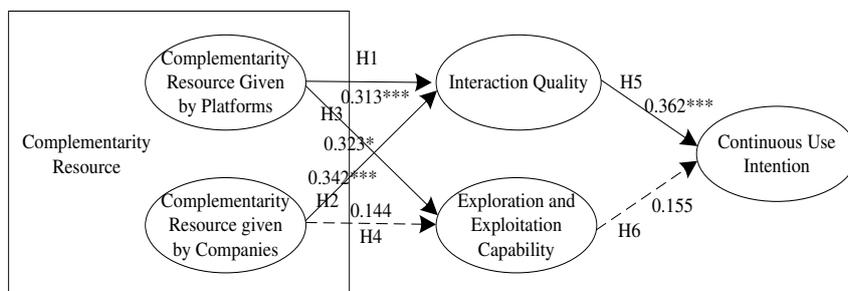
Table 4. Discriminant validity

	CRP	CRC	IQ	EEC	CUI
CRP	0.758				
CRC	0.612	0.760			
IQ	0.502	0.464	0.709		
EEC	0.411	0.343	0.536	0.734	
CUI	0.369	0.329	0.372	0.356	0.759

Note: The bold numbers on the diagonal are the arithmetic square roots of AVE

4.3 The structural model

Using SmartPLS 3.0 to test the hypotheses, the results are shown in Figure 2. H1, H2, H3, and H5 are supported, and H4, H5 are rejected. The hypotheses testing results are shown in Table 5.



Note: ***: P<0.001, **: P<0.01, *: P<0.05

Figure 2. The structural model testing results

Table 5. Hypotheses testing results

Hypothesis	Causal path	Path coefficient	P	Result
H1	CRP→IQ	0.313	0.000	Supported
H2	CRC→IQ	0.342	0.000	Supported
H3	CRP→EEC	0.323	0.012	Supported
H4	CRC→EEC	0.144	0.245	Not supported
H5	IQ→CUI	0.362	0.000	Supported
H6	EEC→CUI	0.155	0.126	Not supported

5. CONCLUSIONS

This study draws on research results in relevant fields in China and abroad, and studies the factors that influence the continuous use of B2B platforms by intelligent engineering companies through empirical and theoretical aspects. Based on the theory of resource complementarity, a research model of factors affecting the continued use of B2B platforms by intelligent engineering companies was constructed and the model was verified. Conclusions are as following.

First, the resources given by the platform have a positive impact on the interaction quality and the exploration and exploitation capability. This is consistent with the view of Harrison (2001), March (1991) and others. The platform provides companies with secure payment venues, complete information of the market and customers, and products and services that can increase the company’s trust in the platform. The increase in trust

can make the communication between the platform and the company more constructive, and the information sharing will be more accurate and complete. Complementary resources such as complete information and high-quality services obtained by companies can help companies improve their ability to change, improve product flexibility, and open new markets, which has significant impact on their exploration and development capabilities.

Second, the resources given by companies have positive impact on the quality of interaction, but they have no positive impact on the exploration and exploitation capabilities. It has positive impact on the interaction quality because the high-quality products provided by the company, corporate brand influence and other resources can help the platform expand its market share and influence, and connect the platform and the company more tightly, which is similar to Lavie (2006) and Gupta (2006)'s conclusion. The main reason why the resources given by companies do not have positive impact on exploration and exploitation capabilities is that the intelligent engineering industry is an emerging industry, and the platform has not yet formed an effective system to use this new type of complementary resources provided by the intelligent engineering industry. Complementary resources cannot be used to their maximum effect, and ultimately failed to improve the platform's exploration and exploitation capabilities.

Third, the interaction quality has positive effect on the company's continuous use intention, but exploration and exploitation capabilities have no positive impact on the company's continuous use intention. The good interactive communication between the platform and the company can help the company obtain market information, customer demand information, and important product and service technological change information, which is conducive to improving the decision-making quality of managers, improving the overall rapid response ability of the company, and helping the company improve market competition. This enhances the company's intention to continuous use of the platform. The interaction quality positively affects the continuous use intention, which is consistent with the research conclusions of Lu (2010), Chunmei Gan (2015) and others. The exploration and exploitation capability has no effect on the company's intention to continuous use of the platform. If a platform is limited to its exploration and exploitation capabilities, companies will continue to explore new platforms.

6. RECOMMENDATIONS

First, the platform should guarantee the stability of the security payment venues, the complete information of markets and customers, product services and other resources, and improve it on this basis. For example: Maintain the platform payment environment regularly to prevent others from using vulnerabilities to damage the interests of the platform or companies. A secure payment environment is crucial for the platform and companies. Develop a complete information collection process to ensure the integrity and credibility of the information. The internal information can be collected by interviewing relevant personnel of the company and issuing questionnaires. For external information, it can be collected through product information, service concepts, marketing strategies, etc. released by the company on the official website. Develop an information report and publish it on the platform, which is conducive to quickly understanding supplier company information. Develop different service strategies for different companies, provide platform financial credit services, and stimulate online transactions of small and medium companies. Develop margin strategies for purchasers to ensure the fairness of transactions. Develop strategies to increase platform exposure for suppliers and increase the opportunities for suppliers to obtain orders.

Second, the supplier companies that settle on the platform should focus on improving the quality of their products or services, respond in a timely manner to buyers' questions about products, and deal with orders in a timely and effective manner. Attach importance to the buyer's experience and evaluation after the transaction is

completed, provides timely feedback and resolves unsatisfactory orders, establishes a good reputation, and reaches a friendly cooperative relationship with buyers. The purchaser company should pay the deposit or security deposit for the transaction on time, and complete the payment of the balance payment in time after the transaction is completed to improve the corporate reputation. Good cooperation between purchaser companies and supplier companies will increase their trust in the platform, believing that the platform is a convenient, fast and safe trading place. At the same time, it will improve the market influence and competitiveness of the platform.

Third, the platform and companies should strengthen communication and exchanges to realize information sharing. In terms of data information, platforms and companies must share non-confidential data and information with each other as much as possible. The frankness of both parties can enhance the quality of communication. The platform should set up a separate question answering webpage to help newly settled companies or companies that are not familiar with the platform operation process to adapt to the platform's trading mode more quickly. Companies should also have specialized personnel to regularly provide the platform with the latest information on corporate products and services to ensure the accuracy and effectiveness of corporate information on the platform. Promote the formation of good communication between the platform and the company can strengthen the company's continuous use intention.

7. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

This article is based on the theory of resource complementarity to study the companies' intention to continue to use, while the existing literature mainly applies the technology acceptance model and the theory of rational behavior to study continuous use intention. In future research, it may use other theories to investigate in-depth research on continuous use intention. For the research objects, it is mainly aimed at companies in the intelligent engineering industry. There is lack of research on the intention of companies in the fields of medicine, machinery, and transportation to continuous use of B2B platforms. In further research, the theory of resource complementarity can be applied to analyze and explore companies in other fields. This research confirms that complementary resources can improve the interaction quality, and the interaction quality has a significant impact on the company's continuous use intention. The performance impact of complementary resources on the platform and the company has not yet been studied and analyzed, so the next research will focus on whether the complementary resources have a significant impact on the performance of platforms and companies.

ACKNOWLEDGMENTS

This research was supported by the Social Science Fund of Fujian Province of China (FJ2020B023), the Humanities and Social Sciences Youth Project of Education Ministry of China (18YJC630140), Major Project of Industry-university-research Innovation Fund of Chinese Universities at New Generation Information Technology Innovation of China (2019ITA01011), and Major Project of Social Science Research Base of Fujian Province, China (FJ2019JDZ051) and Xiamen University of Technology's 2020 Graduate Science and Technology Innovation Program Project (YKJCX2020137)

REFERENCES

- [1] IiMedia. (2020). Research Report on China's company procurement industry from 2019 to 2020. <https://report.iimedia.cn/repo7-0/39021.html> (in Chinese)
- [2] Z.G. Hu, P.Y. Wang. (1997). Intelligent engineering and its applications in power engineering in china. IFAC Proceedings Volumes, 30 (17): 651-656.
- [3] Qi Di. (2016). The current problems in the intelligent building industry. Theoretical Research In Urban Construction,

- (32): 82-83. (in Chinese)
- [4] Thomas Ritter, Carsten Lund Pedersen. (2020). Digitization capability and the digitalization of business models in business-to-business firms: Past, present, and future. *Industrial Marketing Management*, 86: 180-190.
- [5] Zonghai Xie. (2020). Research on the application of B2B e-commerce in company procurement supply chain management. *China Business & Trade*, (16): 102-103. (in Chinese)
- [6] Chao Yang. (2020). Research on the architecture of supply information module based on B2B e-commerce platform. *China Market*, (14): 180-181. (in Chinese)
- [7] Woo ka shing, Christine Ennew. (2004). Business-to-business relationship quality An IMP interaction-based conceptualization and measurement. *European Journal of Marketing*, 38 (9): 1252-1271.
- [8] Suxiu Xu, Meng Cheng, George Q. Huang. (2015). Efficient intermodal transportation auctions for B2B e-commerce logistics with transaction costs. *Transportation Research Part B: Methodological*, 80: 322-337.
- [9] Zhou Zhou. (2009). Analysis of the status quo of China's B2B e-commerce, *Management Observer*, (11): 230-231. (in Chinese)
- [10] Hedong Liu, Wenxiao Chen. (2020). The black-box deconstruction of resource complementation and behavior synergy to enhance cooperation performance: an empirical analysis of high-tech companies. *Studies in Science of Science*, 38(10): 1847-1857. (in Chinese)
- [11] Paul Milgrom, John Roberts. (1990). The economics of modern manufacturing: Technology, strategy, and organization. *American Economic Review*, 80(3): 511-528.
- [12] Ard-Pieter de Man, Geert Duysters. (2005). Collaboration and innovation: A review of the effects of mergers, acquisitions and alliances on innovation. *Technovation*, 25(12): 1377-1387.
- [13] Cho Heetae, Chi Christina, Chiu Weisheng. (2020). Understanding sustained use of health and fitness apps: Incorporating the technology acceptance model with the investment model. *Technology in Society*, 63: 101429.
- [14] Enrique Bigne, Luisa Andreu, Blanca Hernandez, Carla Ruiz. (2019). The impact of social media and offline influences on consumer behaviour. An analysis of the low-cost airline industry. *Current Issues in Tourism*, 21 (9): 1014-1032.
- [15] Dongxiao Gu, Sophia Shanova, Xuejie Yang, Ada Hakimova. (2020). Research on Continuous Information Search in Online Health Community. *Information Science*, (11): 92-97. (in Chinese)
- [16] Chunhui Tan, Yao Li. (2020). Research on the Influencing Factors of Users' Willingness to Continue Using in Virtual Academic Community. *Researches In Library Science*, (20): 28-38. (in Chinese)
- [17] Chaoyan Wu, Lei Huang. (2015). Retail companies' willingness to continue using mobile social media. *China Business and Market*, (6): 89-95. (in Chinese)
- [18] Qian Pei, Yanhua Hu. (2020). Investigation and Analysis of Shenzhen Logistics companies' Use of Government Company Services. *Logistics Technology*, 39 (9): 34-39+48. (in Chinese)
- [19] Paul Milgrom, John Roberts. (1995). Complementarities and fit Strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics*, 19 (2): 179-208.
- [20] Pei-Fang Hsu. (2013). Integrating ERP and e-business Resource complementarity in business value creation. *Decision Support Systems*, 56: 334-347.
- [21] Zhiang (John) Lin, Haibin Yang, Bindu Arya. (2009). Alliance Partners and Firm Performance: Resource Complementarity and Status Association. *Strategic Management Journal*, 30(9): 921-940.
- [22] Teece David, Gary Pisano. (1994). The Dynamic Capabilities of Firms: an Introduction. *Industrial and Corporate Change*, 3(3): 537-556.
- [23] Keyin Liu, Jialin Tang. (2016). A Method of Selecting Company Cooperative Innovation Partners Based on Complementary Matching of Heterogeneous Resources. *Science and Technology Management Research*, (21): 145-150+56. (in Chinese)
- [24] Giuseppe Soda, Marco Furlotti. (2017). Bringing Tasks Back In: An Organizational Theory of Resource

- Complementarity and Partner Selection. *Journal of Management*, 43(2): 348-375.
- [25] Marco Furlotti, Giuseppe Soda. (2018).Fit for the Task: Complementarity, Asymmetry, and Partner Selection in Alliances. *Organization Science*, 29(5): 837-854.
- [26] Paul W. Beamish, John C. Banks. (1987).Equity Joint Ventures and Theory of the Multinational Company. *Journal of International Business Studies*, 18(2): 1-16.
- [27] Pierre Dussauge, Bernard Garrette, Will Mitchell. (2000).Learning from Competing Partners Outcomes and Durations of Scale and Link Alliances in Europe, “North America and Asia” Strategic. *Management Journal*, 21(2): 99-126.
- [28] Matthew J. Robson, Constantine S. Katsikeas, Bodo B. Schlegelmilch, Barbara Pramböck. (2019).Alliance Capabilities, Interpartner Attributes, and Performance Outcomes in International Strategic Alliances. *Journal of World Business*, 54(2): 137-153.
- [29] Chang-Bum Choi, Paul W. Beamish. (2013).Resource complementarity and international joint venture performance in Korea. *Asia Pacific Journal of Management*, 30(2): 561-576.
- [30] Feifei Jiang, Xu Jiang. (2019).The Contingent Value of Resource Complementarity for Alliance Performance: Evidence From Chinese Manufacturing Firms. *IEEE Transactions on Engineering Management*, 66(3): 354-367.
- [31] Carolyn H. Dawson, Jamie B. Mackrill, Rebecca Cain. (2017).Assessing user acceptance towards automated and conventional sink use for hand decontamination using the technology acceptance model. *Ergonomics*, 60(12): 1621-1633.
- [32] Komlan Gbongli, Yongan Xu, Komi Mawugbe Amedjonekou. (2019).Extended Technology Acceptance Model to Predict Mobile-Based Money Acceptance and Sustainability: A Multi-Analytical Structural Equation Modeling and Neural Network Approach. *Sustainability*, 11(13).
- [33] Chuanhong Chen, Xueyan Li. (2018).Research on the Influencing Factors of Citizens’ Willingness to Use Bicycles. *Chinese Journal of Management*, 15(11): 1601-1610. (in Chinese)
- [34] Yixuan Qi, Jin Pan. (2020).Research on the effect of body satisfaction on plastic surgery app users’ willingness to use. *China Market*, (16): 130-131. (in Chinese)
- [35] Lin Dong, Yue Hu. (2020).Research on the Willingness of Logistics companies to Continue Using the Internet of Things Technology. *Logistics Engineering and Management*, 42(10): 57-60. (in Chinese)
- [36] Jeffrey S. Harrison, Michael A. Hitt, Robert E. Hoskisson, R. Duane Ireland. (2001).Resource complementarity in business combinations: Extending the logic to organizational alliances. *Journal of Management*, 27(6): 679-690.
- [37] James G. March. (1991).Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1): 71-87.
- [38] Dovev Lavie, Lori Rosenkopf. (2006).Balancing Exploration and Exploitation Alliance Formation. *The Academy of Management Journal*, 49(4): 797-818.
- [39] Anil Gupta, Ken G. Smith, Christina E. Shalley. (2006).The Interplay Between Exploration and Exploitation. *Academy of Management Journal*, 49(4): 693-706.
- [40] Bhattacharjee Anol. (2001).Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly*, 25(3): 351-370.
- [41] Hsi-Peng Lu, Judy Chuan-Chuan Lin, Kuo-Lun Hsiao, Li-Ting Cheng. (2010).Information sharing behaviour on blogs in Taiwan: Effects of interactivities and gender differences. *Journal of Information Science*, 36(3): 401-416.
- [42] Chunmei Gan, Weijun Wang. (2015).The willingness to continue using academic blogs: the influence of interactivity, immersion and satisfaction. *Information Science*, 33(3): 70-74+94. (in Chinese)
- [43] Zi-Lin He, Poh-Kam Wong. (2004).Exploration vs. Exploitation: An Empirical Test of the Ambidexterity Hypothesis. *Organization Science*, 15(4): 481-494.
- [44] Goksel Yalcinkaya, Roger J. Calantone, David A. Griffith. (2007).An Examination of Exploration and Exploitation Capabilities: Implications for Product Innovation and Market Performance. *Journal of International Marketing*, 15(4): 63-93.

Full Research Paper

Research on the Development of Regional E-commerce Based on the Perspective of Industrial Conjunction: A Case Study of Hubei Province

Xiangbing Chen^{1,2}, Haiying You³*

¹Evergrande School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

²Hubei Industrial Policy and Management Research Center, Wuhan, 430065, China

³Finance Department, Wuhan University of Science and Technology, Wuhan, 430081, China

Abstract: In order to explore the relationship between regional e-commerce and other regional industries, this paper uses the input-output table of Hubei Province in 2012. By using the input-output analysis method, the consumption coefficient, distribution coefficient, influence coefficient and sensitivity coefficient of Hubei Province are obtained, and the relationship between regional industries and regional e-commerce development is discussed. It is considered that the level of regional e-commerce in Hubei Province is not high, but it has great development potential.

Keywords: input-output analysis, industry conjunction, regional e-commerce

1. INTRODUCTION

With the rapid development of China's Internet economy, China's e-commerce industry and its related industries occupy an important position in China's national economy. E-commerce economy plays an important role in promoting China's employment, economic development and industrial transformation.

This paper will take the development of regional e-commerce industry as the research object and take Hubei Province as an example. From the perspective of industrial association, this paper explores the development characteristics of regional e-commerce and the key factors of industrial collaborative development.

2. THEORETICAL BASIS

In the infancy of software development, designers held functionality (what the system does to the world) as the primary goal of software development.

Wassily Leontief put forward the industrial conjunction theory to study the economic and technological association between intermediate inputs and intermediate outputs among industries. Industry conjunction theory is also called input-output theory, which mainly researches the broad, complicated and consanguineous technological and economical relations existed in social and economical activities^[1].

In the study of the relationship between regional e-commerce and related industries, there are currently some articles using input-output analysis method. Duan Lufeng and Tang Wenwen theoretically analyzed the causes of the development of rural e-commerce from the perspective of industrial linkage. They believe that agriculture is the foundation and driving force for the development of the secondary and tertiary industries, and other industries are highly dependent on agriculture. The improvement of agricultural productivity can promote urban and rural development and regional development, which is one of the reasons for the development of rural e-commerce^[2]. Zhang Wanshu and Yu Wantang combined with industry correlation analysis and system dynamics analysis, selected e-commerce transaction scale, express delivery business volume and express delivery service efficiency as key variables, and concluded that the forward related industry of express delivery industry is e-commerce, and they interact with each other and are closely related^[3].

* Corresponding author. Email: xbchen@wust.edu.cn(Xiangbing Chen)

This paper will use the input-output analysis method, based on the perspective of industrial correlation to carry out an empirical analysis of China's regional e-commerce level. To explore the industrial factors that cause the differences, taking Hubei Province as an example, this paper links the development of e-commerce with basic industries, and further explores the impact of these industries on the development of regional e-commerce.

3. INPUT-OUTPUT MODEL CONSTRUCTION

The input-output model can be used to systematically quantify the interdependence of various sectors in a complex economic system. It can not only reflect the intermediate input and demand of various industries, but also analyze the relationship between related industries and the chain reaction of industries. The input-output analysis method is a quantitative analysis method used to study the economic system such as national economy, regional economy, sectoral economy, company or enterprise economic unit, and is used to explore the interdependence of input and output among various departments [4].

According to the input-output theory, the input-output model constructed in this paper is shown in the Table 1.

This paper considers that information technology and business service are the support and carrier of e-commerce. Therefore, this paper will select these two industry indicators to merge, as a representative indicator of the e-commerce industry, and extend the periphery of the e-commerce industry to IT industry and business industry.

The first step is to introduce the direct consumption coefficient into the equation and establish the input-output model. This paper will use it to analyze the backward direct correlation effect of Hubei e-commerce on other economic sectors. The formula can be expressed as shown in equation (1).

$$A = Xq^{-1} \quad (1)$$

The direct consumption coefficient can be expressed as shown in equation (2).

$$a_{ij} = \frac{x_{ij}}{q_j} \quad (i, j = 1, 2, \dots, n) \quad (2)$$

The degree of economic connection between intermediate input sector and a certain sector is expressed by the direct consumption coefficient. A represents the direct consumption matrix, which reflects the economic relations among departments and products in the input-output table. In the input-output table, x_{ij} represents the product value of industry sector j consuming industry sector i , and q_j represents the total input of industry sector j .

The second step is to use the complete consumption coefficient to explore the backward complete correlation effect between e-commerce and other economic sectors in Hubei Province. Its essence is to explore the indirect relationship between the two departments. The complete consumption coefficient is obtained by adding the direct consumption coefficient to the indirect consumption coefficient.

The complete consumption coefficient can be expressed as equation (3).

$$b_{ij} = a_{ij} + \sum_{k=1}^n b_{ik}a_{kj} \quad (i, j = 1, 2, \dots, n) \quad (3)$$

The formula indicates that the indirect consumption coefficient is the product consisting of k kinds of intermediate products, namely the total indirect consumption of product j to product i .

The matrix can be expressed as equation (4).

$$B = (I - A)^{-1} - I \quad (4)$$

Table 1. Input output basic model

Input \ Output		Intermediate Use				Final Use				Inflow	Other	Total of output
		Sector 1	Sector 2	...	Sector n	Total of intermediate use	Final consumption on expenditure	Total investment	Total of final use			
Intermediate input	Sector 1	x_{11}	x_{12}	...	x_{1n}	$\sum x_{1i}$		f_1			q_1	
	Sector 2	x_{21}	x_{22}	...	x_{2n}	$\sum x_{2i}$		f_2			q_2	
	
	Sector n	x_{n1}	x_{n2}	...	x_{nn}	$\sum x_{ni}$		f_n			q_n	
Total of input		$\sum x_{i1}$	$\sum x_{i2}$...	$\sum x_{in}$	$\sum \sum x_{ij}$		$\sum f_i$			$\sum q_i$	
Added value	Labor compensation	l_1	l_2	...	l_n	$\sum l_i$						
	Net production tax	m_1	m_2	...	m_n	$\sum m_i$						
	Depreciation of fixed assets	r_1	r_2	...	r_n	$\sum r_i$						
	Operating surplus	e_1	e_2	...	e_n	$\sum e_i$						
	Total added value	y_1	y_2	...	y_n	$\sum y_i$						
Total of input		q_1	q_2	...	q_n	$\sum q_i$						

The third step is to use the distribution coefficient to calculate and analyze the proportion of e-commerce in various industries in Hubei Province.

The distribution coefficient is expressed as equation (5).

$$h_{ij} = \frac{x_{ij}}{q_i} \quad (i, j = 1, 2, 3, \dots, n) \tag{5}$$

q_i is the total output of sector i . The distribution coefficient can be used to represent the proportion of the product q_i input in the production of industry sector j .

The fourth step is to use the sensitivity coefficient to explore the influence degree of e-commerce by different industries, and can be used to analyze the backward response degree of Hubei e-commerce to different economic sectors.

The sensitivity coefficient is expressed as equation (6).

$$\delta_i = \frac{\sum_{j=1}^n b_{ij}}{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n b_{ij}} \quad (i = 1, 2, \dots, n) \tag{6}$$

According to the formula of sensitivity coefficient, the actual meaning of its value is defined in the following Table 2.

Table 2. Meaning of sensitivity coefficient

Sensitivity coefficient	Influence from other industries
>1	Above average
=1	Equal to the average
<1	Below average

Finally, the influence coefficient is used to explore the impact of e-commerce on different industrial sectors in Hubei Province. It can be used to analyze the demand of different economic sectors in Hubei Province. The degree of the influence of one industry on other industries is shown by the influence coefficient.

The influence coefficient is expressed as equation (7).

$$\lambda_j = \frac{\sum_{i=1}^n b_{ij}}{\frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n b_{ij}} \quad (7)$$

The actual significance of influence coefficient is shown in the following Table 3.

Table 3. Meaning of influence coefficient

Influence coefficient	Influence on other industries
>1	Above average
=1	Equal to the average
<1	Below average

4. INDUSTRY CONJUNCTION ANALYSIS OF E-COMMERCE INDUSTRY

4.1 Data and index selection.

Since the national input-output table is published every five years, the regional input-output table is released later than the national one, and the regional input-output table for 2017 has not been published. Therefore, this paper selects the input-output table of Hubei Province in 2012 for analysis, which can make the research results more convincing.

The selection of indicators will be explained from the following perspectives.

- **Research Methods.** Based on the perspective of industry association, this paper uses the input-output method to analyze the development level of regional e-commerce, and selects the basic industry sector as the alternative variable of e-commerce.
- **Industrial structure.** China's input table includes 42 sectors, covering agriculture, industry and services. Among them, the service industry is closely related to the development of China's e-commerce, and the foundation of e-commerce is inseparable from the development of information technology, which belongs to the category of the tertiary industry.
- **Select representative e-commerce industries.** Referring to Li Chenggang (2020) and Liu Jun et. (2018), This paper selects information technology service industry (ITS) and leasing business service industry (LBS) as the substitution variables of e-commerce in the input-output table.
- **Treatment methods.** ITS specific includes computer service industry, software industry, telecommunication and other information transmission service industry. LBS include rental industry, business service industry and tourism. Since the scope of ITS or LBS is larger than that of the e-commerce industry and the e-commerce industry belongs to ITS and LBS, this paper takes the overlapping part of the input-output

results of the ITS and LBS to analysis.

This paper explains the reasons for selecting the index of input-output analysis from the above four perspectives. Based on these, this paper will select ITS and LBS as the alternative variables of e-commerce in the input-output table. Through the input-output analysis of these two industries, the characteristics and trends of regional e-commerce development are explored.

4.2 The correlation analysis of electronic commerce industry.

4.2.1 Calculation of the direct consumption coefficient

The direct consumption coefficient is calculated by dividing the intermediate input in the input-output table by the total input^[5]. Using Equation (2), the direct consumption matrix A can be obtained, and the direct consumption coefficients of ITS and LBS can be obtained.

Through analysis and comparison, the top 10 related industries with direct consumption coefficients of ITS and LBS in Hubei Province are shown in Table 4.

Table 4. Top 10 direct consumption coefficient related industries of ITS and LBS

Rank	Direct consumption coefficient			
	ITS		LBS	
1	Electrical machinery and equipment industry	0.1255	Paper making and printing industry, cultural, educational and sporting goods industry	0.1095
2	Information transmission, software and information technology services industry	0.0878	Transportation, storage and post industry	0.0787
3	Telecommunications equipment, computers and other electronic equipment industry	0.0499	Accommodation and catering industry	0.0571
4	Power and heat production and supply industry	0.0262	Telecommunications equipment, computers and other electronic equipment industry	0.0515
5	Leasing and business service industries	0.0245	The financial industry	0.0313
6	Wholesale and retail	0.0188	Wholesale and retail	0.0301
7	Real estate	0.0166	Electrical machinery and equipment industry	0.0277
8	Paper making and printing industry, cultural, educational and sporting goods industry	0.0152	Petroleum, coking products and nuclear fuel processing products	0.0259
9	Accommodation and catering industry	0.0148	Leasing and business service industries	0.0220
10	The construction industry	0.0123	Real estate	0.0211

4.2.2 Calculation of the complete consumption coefficient

The complete consumption coefficient is obtained by adding the direct and indirect consumption coefficients. Based on the direct consumption coefficient, the unit matrix I is constructed. Using Equation (4), the complete consumption coefficient matrix can be obtained to obtain the complete consumption coefficient of ITS and LBS.

Through analysis and comparison, the top 10 related industries with complete consumption coefficients of ITS and LBS in Hubei Province are shown in Table 5.

Table 5. Top 10 complete consumption coefficient related industries of ITS and LBS

Rank	Complete consumption coefficient			
	ITS		LBS	
1	Electrical machinery and equipment industry	0.1622	Paper making and printing industry, cultural, educational and sporting goods industry	0.1801
2	Information transmission, software and information technology services industry	0.1004	Transportation, storage and post industry	0.1159
3	Power and heat production and supply industry	0.0828	Power and heat production and supply industry	0.0713
4	Telecommunications equipment, computers and other electronic equipment industry	0.0767	Telecommunications equipment, computers and other electronic equipment industry	0.0711
5	Metal smelting and calendering industry	0.0517	Accommodation and catering industry	0.0703
6	Paper making and printing industry, cultural, educational and sporting goods industry	0.0473	The financial industry	0.0602
7	Wholesale and retail	0.0411	Petroleum, coking products and nuclear fuel processing products	0.0563
8	Leasing and business service industries	0.0399	Wholesale and retail	0.0553
9	Chemical industry	0.0366	Chemical industry	0.0514
10	Transportation, storage and post industry	0.0309	Food and tobacco industry	0.0467

According to the comparison of Table 4 and Table 5, it can be seen that the ranking of the same industry under different coefficient matrices changes, which is caused by the change of their indirect consumption coefficient. Moreover, the total consumption coefficient is larger than the corresponding direct consumption coefficient, which can more accurately reflect the relationship between industries. Therefore, this paper will focus on the analysis of the complete consumption coefficient.

Table 6. Calculation results of ITS and LBS consumption coefficients

	ITS	LBS
Mean	0.0234	0.0291
Maximum	0.1622	0.1801
Minimum	0	0
Standard deviation	0.0322	0.0354
Sum of mean and standard deviation	0.0556	0.0645

Combined with Table 5 and Table 6, the main consumption industries of ITS and LBS are shown in Table 7.

Table 7. Major consumer industries of ITS and LBS

Major consuming industries (total consumption coefficient is greater than or equal to the sum of mean value and standard deviation)	
ITS	LBS
Electrical machinery and equipment industry	Paper making and printing industry, cultural, educational and sporting goods industry
Information transmission, software and information technology services industry	Transportation, storage and post industry
Power and heat production and supply industry	Power and heat production and supply industry
Telecommunications equipment, computers and other electronic equipment industry	Telecommunications equipment, computers and other electronic equipment industry
	Accommodation and catering industry

It can be seen from Table 7 that the largest consumption of ITS is electricity related industries and ITS own consumption. The use of information transmission, software and information technology services is the consumption of electronic equipment and electricity energy, which is the technical aspect of e-commerce. It can be seen from Table 4.4 that the industries that are closely related to consumers' lives are the most consumed by LBS. This reflects the industrial structure of e-commerce.

The service object of electronic commerce is consumer. It is a web-based business service. Based on the analysis of the main consumption industries of ITS and LBS, it can be concluded that the development of e-commerce in Hubei is closely related to ITS, transportation, warehousing, postal service and general equipment manufacturing.

4.2.3 Calculation of the distribution coefficient

Distribution coefficient refers to the proportion of a certain industrial sector used in other industries, and can reflect the influence of different industries on the industry^[6]. By using Equation (5), the distribution coefficients of ITS and LBS can be obtained.

Through analysis and comparison, the top ten industries with distribution coefficients of ITS and LBS in Hubei Province are ranked in Table 8.

Table 8. Top 10 distribution consumption coefficient related industries of ITS and LBS

Rank	Distribution consumption coefficient			
	ITS		LBS	
1	Wholesale and retail	0.1343	The construction industry	0.2658
2	Information transmission, software and information technology services industry	0.0878	Wholesale and retail	0.2360
3	The financial industry	0.0289	Transportation equipment industry	0.0613
4	Transportation, storage and post industry	0.0258	Food and tobacco industry	0.0540
5	Leasing and business service industries	0.0182	Scientific research and technical services	0.0486
6	Scientific research and technical services	0.0112	Chemical industry	0.0414
7	Real estate	0.0088	The financial industry	0.0262
8	Chemical industry	0.0055	Leasing and business service industries	0.0220
9	Power and heat production and supply industry	0.0053	The textile industry	0.0220
10	Food and tobacco industry	0.0050	Nonmetallic minerals and other mining products industry	0.0209

The mean distribution coefficient of ITS is 0.009, the standard deviation is 0.0245, and the sum of the mean and standard deviation is 0.0335.

The mean value of the distribution coefficient of LBS is 0.0233, the standard deviation is 0.0537, and the sum of the mean value and standard deviation is 0.077.

Combined with Table 8, the main consumption industries of ITS and LBS are obtained through comparative analysis, as shown in Table 9.

Based on the primary distribution industry and the second distribution industry of ITS and LBS, it can be concluded that the development of e-commerce in Hubei is closely related to ITS, Transportation, storage and post industry, and wholesale and retail industry.

Table 9. The main consumption industries of ITS and LBS

	ITS	LBS
Primary distribution industry (The distribution coefficient is greater than or equal to the sum of the mean value and standard deviation)	Wholesale and retail Information transmission, software and information technology services industry	The construction industry Wholesale and retail
Secondary distribution industry (The distribution coefficient is greater than or equal to the mean, less than or equal to the sum of the mean and standard deviation)	The financial industry Transportation, storage and post industry Leasing and business service industries Scientific research and technical services	Transportation equipment industry Food and tobacco industry Scientific research and technical services Chemical industry The financial industry

4.2.4 Calculation of sensitivity coefficient and influence coefficient

Using Equation (6) and Equation (7), the sensitivity coefficient and influence coefficient of ITS and LBS can be obtained as shown in Table 10.

Table 10. Sensitivity coefficient and influence coefficient of ITS and LBS

Sensitivity coefficient	ITS	0.6445
	LBS	1.0507
Influence coefficient	ITS	0.8546
	LBS	0.9575

As can be seen from Table 10, the sensitivity coefficient and influence coefficient of ITS are lower than those of LBS, indicating that the development level of ITS in Hubei Province is still lower than that of LBS. The sensitivity coefficient and influence coefficient of ITS are less than 1, below the average level, indicating that the development level of ITS in Hubei Province is low. The induction coefficient of LBS is greater than 1, and the influence coefficient is less than 1, indicating that LBS in Hubei Province is affected by other industries more than the average level, but its influence on other industries is lower than the average level, indicating that LBS in Hubei Province has a weak driving effect on other industries, but it has a great development potential.

5. CONCLUSIONS

First of all, in terms of consumption, the development of e-commerce in Hubei Province needs the support of good traffic and transportation conditions and information service conditions, and the development of e-commerce industry cannot be separated from the support of electronic equipment manufacturing industry. The reason is that the main consumption industries of A and B are in sales, postal and electronic equipment manufacturing, indicating that to vigorously develop e-commerce in Hubei Province, A large number of transportation services, information products and services, and electronic equipment manufacturing products need to be consumed. Therefore, if Hubei province wants to improve the level of e-commerce, it should plan and develop scientifically according to the rank of total consumption coefficient.

Secondly, in terms of distribution, the input distribution of e-commerce in other industries in Hubei Province mainly focuses on wholesale and retail, transportation, storage and transportation, ITS. It can be seen that the development direction of e-commerce in Hubei Province is effective, and it is necessary to increase the investment in industries that are more closely related to e-commerce consumption, such as wholesale and retail, transportation, storage and transportation, ITS, etc., and reduce the investment in industries that are less closely related to e-commerce consumption, such as construction industry and chemical industry.

Finally, from the analysis of the sensitivity coefficient and influence coefficient, it can be seen that the level of e-commerce in Hubei province is not high and has a great potential for development. Therefore, it is necessary to vigorously develop science and technology service industry and business industry, and make it become the pillar industry of the tertiary industry, so as to improve the development level of e-commerce in Hubei Province.

ACKNOWLEDGEMENT

This research was supported by philosophy and social science project of Hubei Education Department under Grant 20D022.

REFERENCES

- [1] Tiago Camarinha Lopes, Henrique Dantas Neder. (2017). Sraffa, Leontief, Lange: The Political Economy of Input-output Economics. *EconomiA*, 18(2): 192-211
- [2] Duan Lufeng, Tang Wenwen. (2016). Theory, Externality and Policy Choice of Agricultural E-commerce Development. *Jiangsu Agricultural Sciences*, 44 (4): 4-7(in Chinese)
- [3] Zhang Wanshu, Yu Wantang. (2016). Empirical Analysis on the Industrial Relationship Between Express Delivery Industry and E-commerce. *China Management Informatization*, 19(3): 165-167(in Chinese)
- [4] Adam Rose. (1995). Input-output Economics and Computable General Equilibrium Models. *Structural Change and Economic Dynamics*, 6(3): 295-304
- [5] Hu ronghua, Guan Yuhua. (2001). The Application of Excel in Input-output Analysis -- A Series of Lectures on Excel Application (3). *Jiangsu Statistics*, (12): 25-27(in Chinese)
- [6] Yang Guizhong. (2016). Application of Excel in Case Teaching of Input-output Method. *Science and Technology Horizon*, (20): 125-126(in Chinese)
- [7] Li Chenggang.(2020). Empirical Research on the Promotion of Real Economy by Big Data Development -- Based on Provincial Panel Data from 2013 to 2018. *Guizhou Social Sciences*, (1): 132-141(in Chinese)
- [8] Liu Jun, Wang Xiufeng. (2018). Research on the Impact of Big Data Industry on the National Economy of Southeast Guizhou. *China's Collective Economy*, (16): 17-19(in Chinese)
- [9] Folorunso Olusegun, Awe Oludare Gaoriel, Sharma Sushil K, Jeff Zhang. (2006). Factors Affecting the Adoption of E-commerce: A Study in Nigeria. *Journal of Applied Sciences*, 6(10): 2224-2230

-
- [10] George Saridakis, Yanqing Lai, Anne-Marie Mohammed, Jared M. Hansen. (2018). Industry Characteristics, Stages of E-commerce Communications, and Entrepreneurs and SMEs Revenue Growth. *Technological Forecasting and Social Change*, 128(1): 55-66
- [11] Jinshui Zhang. (2011). The Analytical Solution of Balanced Growth of Non-linear Dynamic Multi-sector Economic Model. *Economic Modelling*, 28:410-421
- [12] Elco E. Koks, Lorenzo Carrera, Olaf Jonkeren, Jeroen C.J.H.Aerts. (2016). Regional Disaster Impact Analysis: Comparing Input-output and Computable General Equilibrium Models. *Natural Hazards and Earth System Sciences*, 16(8):1911-1924

Short Research Paper

An Empirical Research on Online Negative Comments Control Based on Consumer Satisfaction

Haiying You¹, Xiangbing Chen^{2,3*}

¹Fiance Department, Wuhan University of Science and Technology, Wuhan, 430081, China

²Evergrande School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

³Hubei Industrial Policy and Management Research Center, Wuhan, 430065, China

Abstract: This paper studies how to achieve consumer satisfaction toward different types of online negative comment through service recovery, and constructs online negative comments evaluation model based on the concepts of service recovery and customer satisfaction. Through the scenario simulation, this paper makes different service remedies to the online negative comment, and compares its impact on consumer satisfaction. Empirical analysis shows that both spiritual and material remedies have positive impact on consumer satisfaction, and the material remedies are more effective than spiritual remedies for the same causes of the online negative comment.

Keywords: online negative comment, consumer satisfaction, service recovery

1. INTRODUCTION

The convenience of information communication and non-contact environment make online shopping become an important form of shopping. This change is particularly evident in the first half of 2020. Consumers trend to publish comment on their own shopping experience and consumer experience on the relevant web page. At the same time, consumers are also used to learning about the product information through internet before making a purchase decision. These consumer habits will make online comments have great influence: positive online comments can not only improve the image of the enterprise, also can promote the consumption; negative online comments can not only brings to the enterprise loss of reputation, also can affect sales, even will bring enterprise bankruptcy risk when the enterprise cannot handled properly. In the perspective of customer satisfaction, this paper constructs the online negative online comment control model, and tires to help the enterprises to do the negative control through survey method, and empirical analysis.

2. THE RELATIONSHIP BETWEEN NETWORK NEGATIVE COMMENT AND CUSTOMER SATISFACTION

Negative Internet Word of Mouth (IWOM) is a kind of customer's behavior when consumer is not satisfied with the corporation, the main form of which is the online negative comment^[1]. Online negative comment is a process of that the consumers publish their own dissatisfaction or unpleasant consumer experience in the form of text, pictures, video, and other form on the online platform^[2].

Factors that affect the negative evaluation of consumers can be analyzed from the three perspectives: products, services and third-party logistics. The factor of product is including product quality problems, the question that product does not matching the actual, product cargo damage, and so on. The factor of service is including service attitude, service quality, and so on. The factor of third-party logistics is including the delivery time, the quality of distribution services, logistics cargo damage risk, and so on.

* Corresponding author. Email: xbchen@wust.edu.cn(Xiangbing Chen)

Negative online comment is a form of IWOM. Customer oriented enterprises are faced with how to improve customer satisfaction and reduce customer dissatisfaction. This is because customer satisfaction is related to word-of-mouth, transaction costs, and customer loyalty^[3]. And the correlation under the condition of the customers are very satisfied will be particularly notable, While customers are not satisfied can lead to negative word of mouth, no longer to patronize and complain to the third party such as the adverse consequences.

Customer dissatisfaction will lead to the emergence of negative comments^[4]. If the negative comment of the consumer is improperly handled, the negative comment will give the enterprises a very serious crisis^[5]. When there have been some online negative comments, rebuilding consumer satisfaction has become one of the goals of the enterprise. The enterprise that takes effective service recovery can not only rebuild the consumer satisfaction again, but also increase consumer repeat purchase probability and positive word of mouth, reduce consumer conversion behavior and bring high profits for the enterprise.

According to Oliver's expectation-gap model, consumer satisfaction comes from a comparison of the customer's expected performance of a product or service with the actual performance perceived by the customer^[6]. When the actual perception is greater than the expected, the consumer will be satisfied. According to the service recovery evaluation model with fairness theory, the measurement of consumer satisfaction mainly lies in the evaluation of consumers' actual perception of service recovery^[7]. The fairness of consumer perception could include three parts: fairness of distribution, fairness of procedure and fairness of interaction^[8]. The three dimensions of the fairness of consumer perception are used as the standard to measure the satisfaction of consumers in this paper. Fairness of distribution refers to the fairness perception of service recovery results. Procedural fairness refers to the fairness perception of the service recovery process. Interaction fairness refers to the fairness perception of corporation's attitude towards service recovery.

3. RESEARCH MODEL AND HYPOTHESIS

3.1 Research model.

It must understand the reasons of the online negative comment. And then, the corporation can target to make service recovery, and form reconstruction of consumer satisfaction. This paper studies how to achieve consumer satisfaction toward different types of online negative comment through service recovery. According to the above analysis, this paper builds a theoretical model, as shown in Figure 1.

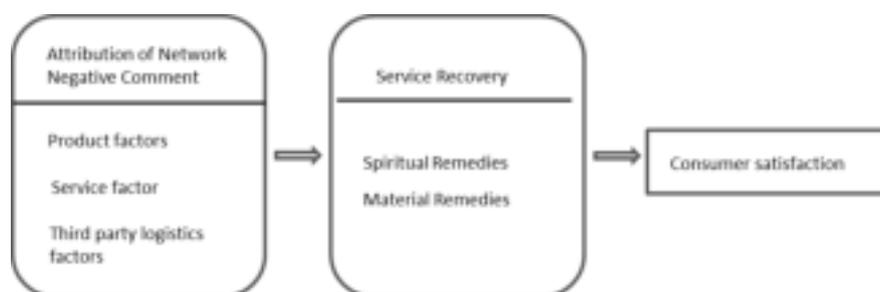


Figure 1. The theory model of online negative comment control based on customer satisfaction

3.2 Hypothesis.

Service recovery is an effort by firms to avoid negative impacts by mitigating consumer dissatisfaction at the spiritual level through communication, exchange, apology, and interpretation, as well as at the material level economic compensation, giving some small gifts, discounts, etc. to make up for consumer economic losses^[9]. Both types of remedies can partly eliminate consumer dissatisfaction, and may have a positive impact on

consumer satisfaction. Thus, this paper raises the following hypothesis:

H1: Service recovery has a positive effect on consumer satisfaction;

H11: Spiritual remedies have a positive effect on consumer satisfaction;

H12: Material remedies have a positive effect on consumer satisfaction.

Online consumers may encounter some service failures. The service failures may include poor product quality, poor service attitude, slow logistics and other issues. When the benefit of consumers is damaged, consumer will inevitably produce discontent mood and consumers may have a negative evaluation of the enterprise, so as to bring harm to the enterprise^[10]. Corporations should take effective service remedies to compensate for negligence and eliminate grievances. Because the product problems is a kind of visible damage to consumers, material compensation can make up for the loss; because the service factor may be more psychological harm to the consumer, apology and explanation can a more effective way to eliminate consumer dissatisfaction; logistics problems come from the third party logistics enterprises operation, material remediation than the spirit for the consumer satisfaction with a more significant impact because online enterprises and logistics enterprises are cooperative relations. Therefore, it states the following assumptions:

H2: Online negative comment appraisal of product factors, material recovery than the spiritual remedy has a more significant impact to consumer satisfaction;

H3: Online negative comment appraisal of service factors, spiritual remedy than the material remedy has a more significant impact to consumer satisfaction;

H4: Online negative appraisal of the third party logistics, material recovery than the spirit of remedy has a more significant impact to consumer satisfaction.

4. EMPIRICAL ANALYSIS

4.1 Research design.

This paper uses the form of a questionnaire to study the online negative comment control based on consumer satisfaction through service recovery. The questionnaire includes two parts: one is the basic information of the respondents, including gender, age, education, income, etc.; another is the scenario simulation of consumer satisfaction under service recovery. In the three scenarios of Product factors, service factors and third-party logistics factors, the enterprises are assumed to adopt mental remedies and material remedies respectively. Through the distributive fairness, procedural fairness and interactive fairness, the questionnaire measures the satisfaction of the respondent with respect to the simulated service remedy. Refer to Likert 5 sub-scale, the number 1 to 5 in the questionnaire represent completely disagree, do not agree, generally agree, agree, fully agree. Specific questionnaires are as following:

Table 1. Questionnaire items of service remedies and consumer satisfaction

Measurements		Numbering	Item	
Scene One	Spiritual Remedies (SR _{SF})	Spiritual Remedies1 (SR1 _{SF})	Q1	The customer service staff is kind to me
		Spiritual Remedies2 (SR2 _{SF})	Q2	Business performance makes me feel very warm
		Spiritual Remedies3 (SR3 _{SF})	Q3	Business pays attention to me a lot
	Consumer Satisfaction after Spiritual Remedies (SRS _{SF})	Result of Spiritual Remedies 1 (FSRR)	Q4	I am satisfied with the results of the online shop service remedies
		Process of Spiritual Remedies 1 (FSRP)	Q5	I am satisfied with the process of the online shop service remedies
		Attitude of Spiritual	Q6	I am satisfied with the attitude of the

		Remedies1 (FSRA)		online shop service remedies
	Material Remedies (SR _{MF})	Material Remedies1 (SR _{1MF})	Q7	I deserve the compensation of the customer service staff for me
		Material Remedies2 (SR _{2MF})	Q8	The compensation of the online shop for me makes up for my losses
		Material Remedies3 (SR _{3MF})	Q9	The compensation of the online shop for me is more than that I expected
	Consumer Satisfaction after Material Remedies (SRS _{MF})	Result of Material Remedies1 (FMRR)	Q10	I am satisfied with the results of the online shop service remedies
		Process of Material Remedies1 (FMRP)	Q11	I am satisfied with the process of the online shop service remedies
		Attitude of Material Remedies1 (FMRA)	Q12	I am satisfied with the attitude of the online shop service remedies
Scene Two	Spiritual Remedies (SR _S)	Spiritual Remedies1 (SR _{1S})	Q13	The customer service staff is kind to me
		Spiritual Remedies2 (SR _{2S})	Q14	Business performance makes me feel very warm
		Spiritual Remedies3 (SR _{3S})	Q15	Business pays attention to me a lot
	Consumer Satisfaction after Spiritual Remedies (SRS _S)	Result of Spiritual Remedies2 (SSRR)	Q16	I am satisfied with the result of the online shop service remedies
		Process of Spiritual Remedies2 (SSRP)	Q17	I am satisfied with the process of the online shop service remedies
		Attitude of Spiritual Remedies2 (SSRA)	Q18	I am satisfied with the attitude of the online shop service remedies
	Material Remedies (SR _M)	Material Remedies1 (SR _{1M})	Q19	I deserve the compensation of the customer service staff for me
		Material Remedies2 (SR _{2M})	Q20	The compensation of the online shop for me makes up for my losses
		Material Remedies3 (SR _{3M})	Q21	The compensation of the online shop for me is more than that I expected
	Consumer Satisfaction after Material Remedies (SRS _M)	Result of Material Remedies2 (SMRR)	Q22	I am satisfied with the results of the online shop service remedies
		Process of Material Remedies2 (SMRP)	Q23	I am satisfied with the process of the online shop service remedies
		Attitude of Material Remedies2 (SMRA)	Q24	I am satisfied with the attitude of the online shop service remedies
Scene Three	Spiritual Remedies (SR _{TS})	Spiritual Remedies1 (SR _{1TS})	Q25	The customer service staff is kind to me
		Spiritual Remedies2 (SR _{2TS})	Q26	Business performance makes me feel very warm
		Spiritual Remedies3 (SR _{3TS})	Q27	Business pays attention to me a lot
	Consumer Satisfaction after Spiritual Remedies (SRS _{TS})	Result of Spiritual Remedies3 (TSRR)	Q28	I am satisfied with the results of the online shop service remedies
		Result of Spiritual Remedies3 (TSRP)	Q29	I am satisfied with the process of the online shop service remedies

		Result of Spiritual Remedies3 (TSRA)	Q30	I am satisfied with the attitude of the online shop service remedies
Material Remedies (SR _{TM})		Material Remedie1 (SR1 _{TM})	Q31	I deserve the compensation of he customer service staff for me
		Material Remedie2 (SR2 _{TM})	Q32	The compensation of the online shop for me makes up for my losses
		Material Remedie3 (SR3 _{TM})	Q33	The compensation of the online shop for me is more than that I expected
Consumer Satisfaction after Material Remedic (SRS _{TM})		Result of Material Remedies3 (TMRR)	Q34	I am satisfied with the results of the online shop service remedies
		Process of Material Remedies3 (TMRP)	Q35	I am satisfied with the process of the online shop service remedies
		Attitude of Material Remedie3 (TMRA)	Q36	I am satisfied with the attitude of the online shop service remedies

4.2 Data analysis and hypothesis testing.

The questionnaires were distributed 170 copies and 150 were actually recovered. 22 questionnaires were removed, 128 valid questionnaires were validated, and the recovery rate was 85%. Among them, the proportion of men and women were 35.11% and 64.1%, mainly concentrated in the age of 25 years of age, 79.7% of undergraduate education, the monthly disposable income of respondents focused on 501-1000 yuan, the proportion of shopping more than three times in six months is 83.6%.

4.2.1 Regression analysis

In the regression analysis, this paper takes the three items of service recovery as the independent variables (SR1-SR3) and the service satisfaction (SRS) as the dependent variable. The regression equation for consumer satisfaction with service recovery and service recovery is in the form of:

$$SRS = \beta_0 + \beta_1 SR1 + \beta_2 SR2 + \beta_3 SR3 \quad (1)$$

The following is a regression analysis of service recovery and consumer satisfaction, as shown in Table 2 and Table 3.

Table 2. Regression analysis of spiritual recovery and consumer satisfaction

model	Scene one			Scene two			Scene three		
	Non - normalized coefficient B	t	Sig.	Non - normalized coefficient B	t	Sig.	Non - normalized coefficient B	t	Sig.
(constant)	.754	2.933	.004	1.107	5.1038	.000	1.072	5.1046	.000
Spiritual Remedies 1	-.020	-.230	.818	.076	1.084	.281	.010	.128	.898
Spiritual Remedies 2	.381***	3.621	.000	.229**	2.204	.029	.297***	2.839	.005
Spiritual Remedies 3	.377***	4.220	.000	.341***	3.710	.000	.341***	3.582	.000
Square R	.445			.465			.501		
F	34.909			37.851			43.521		

Remarks: *** indicates a significant level of 1%; ** indicates a significant level of 5%; * indicates a significant level of 10%.

Table 3. Regression analysis of material recovery and consumer satisfaction

model	Scene one			Scene two			Scene three		
	Non - normalized coefficient	t	Sig.	Non - normalized coefficient	t	Sig.	Non - normalized coefficient	t	Sig.
	B			B			B		
(constant)	.701	2.906	.004	.820	3.772	.000	.984	4.318	.000
Material remedies1	.275***	4.747	.000	.246***	3.991	.000	.023	.398	.691
Material remedies2	.314***	5.524	.000	.243***	3.177	.002	.406***	4.917	.000
Material remedies3	.269***	5.032	.000	.328***	5.683	.000	.333***	4.949	.000
Square R	.584			.593			.590		
F	60.505			62.635			61.909		

Remarks: *** indicates a significant level of 1%; ** indicates a significant level of 5%; * indicates a significant level of 10%.

According to the results of regression analysis, it can be seen that spiritual and material remedies have a positive effect on consumer satisfaction. It can be seen from the significant level of regression coefficient that spiritual recovery 1 has less effect on consumer satisfaction, while material recovery 1 has less effect on consumer satisfaction in scenario 3. The significance level of all other dependent variables was less than 0.05, which indicated that H1, H11 and H12 were established.

4.2.2 Paired sample T test

In this paper, the three dimensions of perceived fairness are taken as the measurement of consumer satisfaction after service recovery of online negative comment. Customer satisfaction after service recovery (SAS) can be divided into three categories: remedial results satisfied (RR), remedial process satisfaction (RP) and remedial attitude satisfaction (RA). Reference paired sample T test, the functional relationship between them is:

$$SRS=f(RR,RP,RA) \quad (2)$$

It can be seen from Table 4 that the average of spiritual remedies is less than the mean value of material remedies, and the two-tailed probability significance values (2 tailed) were 0.000, indicating that there is significant difference between the spiritual remedies and material remedies. The mean difference of each variable is negative, that shows that the effect of material compensation measures on the same online negative comment is better than that of spiritual compensation measures. Therefore, it can be inferred that H2, H4 hold and H3 does not hold.

Table 4. Paired samples T test

		Scene one		Scene two		Scene three	
		mean	Sig.	mean	Sig.	mean	Sig.
scene 1	Mental Remedy Result / material Recovery Result	3.08/3.82	.000	3.00/3.63	.000	2.95/3.41	.000
scene 2	Mental recovery process / material recovery process	3.03/3.77	.000	3.05/3.73	.000	3.03/3.42	.000
scene 3	Mental remedy attitude / material remedy attitude	3.42/3.80	.000	3.02/3.69	.000	3.12/3.51	.000

5. CONCLUSION AND SUGGESTION

5.1 Conclusion

This paper confirms the positive relationship between service recovery and consumer satisfaction through regression analysis. That is, service recovery has a positive effect on customer satisfaction after service recovery. Table 5 summarizes the test results for the hypothesis. In addition, the paired sample T test shows that material remedies have a more significant impact on consumer satisfaction than spiritual remedies in different scenarios, indicates that the effectiveness of material recovery is bigger than that of mental recovery. Therefore, enterprises in the service remedy should pay more attention to the material compensation to consumers to make up for consumer economic losses, and maximize customer satisfaction. The results of hypothesis test based on empirical analysis are shown in Table 5.

Table 5. Hypothesis test conclusion

Hypothesis	Content of Assumptions	Verification conclusion
H1	Service recovery has a positive effect on consumer satisfaction	confirmed
H11	Spiritual recovery has a positive effect on consumer satisfaction	confirmed
H12	Material remedies have a positive effect on consumer satisfaction	confirmed
H2	Online negative comment appraisal of product factors, material recovery than the spiritual remedy has a more significant impact to consumer satisfaction	confirmed
H3	Online negative comment appraisal of service factors, spiritual remedy than the material remedy has a more significant impact to consumer satisfaction	disconfirmed
H4	Online negative appraisal of the third party logistics, material recovery than the spirit of remedy has a more significant impact to consumer satisfaction	confirmed

The conclusion of empirical analysis is that material remedies always have a more significant effect than mental remedies on the same online negative comment. For H3 is not validated, it is attributed to the virtual and intangible nature of the internet. Because most of online services are homogeneous services, consumers may increasingly feel the enthusiasm of the business more difficultly. Therefore, the spiritual remedies have been unable to make up for the loss of consumer, and consumer will be more concerned about the material compensation.

5.2 Suggestion

According to the results of empirical research, the effect of material recovery is better than the effect of spiritual recovery. But it does not mean that corporation can ignore spiritual remedy. When the online negative comment appears, corporation should first resolve the dissatisfaction of consumers through the spiritual remedy. Corporation should be from the customer point of view, to give a sincere apology, should not in order to evade responsibility, cover up their mistakes, argue with customers and even shirk their responsibilities. For the sake of customers and maximize access to their trust and understanding, corporation should respect customers, so as to better communicate with consumers, to solve their problems, to meet customer needs. In addition, company can skillfully use a variety of situational factors, and take the initiative to communicate with customers to customers targeted remediation. Finally, companies can provide consumers with material compensation, (including gifts, returns, replacement, discounts, free, etc.) to make up for consumer losses from dissatisfaction product or service.

Therefore, corporation need to fully examine their own shortcomings, and to improve them, such as

strengthening the internal staff training, improving service levels, so as to reduce service errors, improve service quality. Company has to identify negative reasons for consumer online comment: the reason is due to product itself, the company should do some product price changes and the implementation of strict product quality control; the reason is due to service process, the company should train service personnel, improve the quality of staff, and strengthen the supervision of service to avoid the recurrence of the same problem; the reason is due to the third-part logistics corporation, company can negotiate with the logistics enterprises on logistics services, or replace the logistics service providers and choose a better partner.

ACKNOWLEDGEMENT

This research was supported by philosophy and social science project of Hubei Education Department under Grant 20D022.

REFERENCES

- [1] Verhagen T, Nauta A, Feldberg F.(2013).Negative Online Word-of-mouth: Behavioral Indicator or Emotional Release?. *Computers in Human Behavior*, 29(2): 1430-1440
- [2] Yang Xueqin. (2011). A Comparative Study of WOM and IWOM. *Journal of Southwest Agricultural University (Social Science Edition)*, 9(6):44-46(in Chinese)
- [3] IM Wetzer, M. Zeelenberg, R. Pieters.(2007).Never Eat in That Restaurant, I Did!: Exploring Why People Engage in Negative Word-of-mouth Communication. *Psychology & Marketing*, 24(8): 661-680
- [4] Litvin S. W, Goldsmith R. E, Pan B.(2008).Electronic Word of Mouth in Hospitality and Tourism Management. *Tourism Management*, 29(3): 458-468
- [5] Nie Ling.(2013). A Study on the Impact of Negative IWOM on Consumers' Brand Transfer. Ms D Thesis. Hangzhou, China: University of Zhejiang(in Chinese)
- [6] Oliver Richard L.(1980).A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, 17(4): 460-469
- [7] Boshoff C..(2005).A Reassessment and Refinement of RECOVSAT: An Instrument to Measure Satisfaction with Transaction-specific Service Recovery. *Managing Service Quality*, 15(5): 410-425
- [8] Chia-Chi Chang.(2008).Choice, Perceived Control and Customer Satisfaction: The Psychology of Online Service Recovery. *Journal of Cyber Psychology &Behavior*, 11(3): 321-328
- [9] Othman Z, Zahari M. S. M, Radzi S. M..(2013).Customer Behavioral Intention: Influence of Service Delivery Failures and Service Recovery in Malay Restaurants. *Procedia-Social and Behavioral Sciences*, 105(1): 115-121
- [10] Kim J. H, Jang S. C..(2014).The Fading Affect Bias: Examining Changes in Affect and Behavioral Intentions in Restaurant Service Failures and Recoveries. *International Journal of Hospitality Management*, 40: 109-119

Full Research Paper**A Study on the Characteristics of the Publishers of Rumor-Refuting****Information about COVID-19***Xiaping Xin¹, Yan Tu¹, Tianmei Wang^{1*}*¹School of Information, Central University of Finance and Economics, Beijing, 100081, China

Abstract: The uncertainty of public health emergencies leads to the diffusion of rumors on social media, and the extensive forwarding of rumor-refuting information has become a way of rumor management. Determining the characteristics of rumor-refuting information publishers that are conducive to rumor information being forwarded has become a concern. This article crawls 3,035 rumor-refuting information data about the COVID-19 event. Based on social capital theory, from the perspective of rumor-refuting information publishers, the characteristic variables of social capital of publishers are selected from the three dimensions of structure, relationship, and cognition to construct the model of influencing factors of information-forwarding behavior of rumor-refuting information, and the negative binomial regression model is used for data analysis and robustness test. Results show that the number of followers, disclosure of geographical location information, gaining membership, and obtaining Weibo professional certification will increase the number of forwarding of the rumor-refuting information, and whether or not having a personal introduction will not significantly affect the forwarding of rumor-refuting information. At the same time, the number of Weibo and fans will reduce the retransmission of the rumor-refuting information. Research results of this paper can help the platform make targeted choices and encourage users to release rumor-refuting information, expand the scope of rumor-refuting information dissemination, and effectively achieve rumor governance.

Keywords: rumor management, COVID-19, social capital, negative binomial regression

1. INTRODUCTION

At the end of 2019, COVID-19 swept the world and posed a huge threat to the social economy because of its spread and destruction. At the same time, it brought panic to the people. Negative emotions and rumors caused depression, generated excessive stress reaction, and induced people to suffer from "emotional plague." These problems then lead to rumors and to a flood of negative emotions, thereby creating a vicious circle.

Social media has become a platform for promoting interaction between people and reflecting social values. It has also become a virtual place where people can share personal opinions, experiences, and insights. However, the unknown public health emergencies and the spread of panic often lead to rumors on social media platforms [1]. This confusing information has brought added pressure on the prevention and treatment of the already arduous public health emergencies; if the action to deal with rumors lags behind, it may have a profound impact on public order [2].

As the second major public health incident of pneumonia in China since the beginning of the 21st century, the COVID-19 incident has caused pronounced damage to the social economy and a huge threat to people's lives and safety. Uncertainty and panic have become the breeding ground for rumors, and social media has led to the rapid spread of rumors. Since the outbreak of the COVID-19 incident, the Baidu index of "rumors" has risen sharply. Rumors, such as that about drinking and taking Shuanghuanglian oral liquid which is believed to effectively prevent an individual from catching the virus, lead people to snap up goods and consequently increase the risk of infection. The spread of rumors on public health emergencies will not only cause panic but

* Corresponding author. Email: wangtianmei@cufe.edu.cn (Tianmei Wang)

may even delay the timing of treatment among patients, consequently increasing the seriousness of the damage caused by the incident. In 2019 alone, the Weibo platform effectively handled 77,742 pieces of false information and 470 new rumor cases. Looking at the headlines at the peak of rumors, a number of rising rumors in 2017, 2018, and 2019 are found to be related to public health emergencies. Therefore, the management of online rumors has become a serious challenge to the management of public health emergencies in China.

The Letter of Commitment to Jointly Resist Online Rumors signed by the Cyberspace Administration of China mentions that we should take the initiative to identify, resolutely curb, effectively control, and work together to dispel rumors. As Lai ^[3] said, online rumors are akin to viruses in the network environment, and the online rumor governance system corresponds to the immune system of the network environment. Therefore, a good rumor refutation and rumor governance system is the key step to prevent the spread and even the emergence of rumors.

The most effective method of controlling rumors is to publish information to refute rumors ^[2]. In the face of this severe challenge, authoritative organizations and professionals actively publish information on social media to refute rumors, which are forwarded by the majority of netizens. However, the research on the forwarding behavior of rumor-refuting information remains rare. Although Wang ^[4] uses machine learning methods to identify users who may publish rumor-refuting information through user characteristics, the focus of this study lies on the characteristics and attributes of rumor-refuting users. In fact, the publishers of rumor-refuting information on social media have additional complex social attributes and social capital. Identifying which characteristic factors will affect the forwarding behavior of rumor-refuting information is worthy of in-depth exploration, which is conducive to the selection of social platforms and will encourage relevant users to improve the breadth of rumor-refuting information forwarding.

Existing studies believe that social capital is a major component of social media and has an important impact on the forwarding of information in the media ^[5]. Social media can affect users' perception, emotion, and behavior ^[6]. The sustainable development of social media is inseparable from the active participation and interaction of users, including search, browsing, retweeting, comments, attention, making friends, feedback, and other behaviors. The relationship network structure formed by these behaviors has a profound impact on user behavior. In social media network relationships, users obtain the attention of others and social capital through free knowledge sharing behavior, comment behavior, and like behavior, among others. Therefore, from the perspective of social capital and the characteristics of the publishers of rumor-refuting information, this study examines the factors affecting the amount of information forwarding. Specifically, based on social capital theory, starting from the characteristics of the publishers of rumor-refuting information and taking the COVID-19 event as an example, this research constructs a model of the factors affecting the forwarding of rumor-refuting information and examines the characteristics of rumor-refuting that affect the amount of forwarding. This study aims to obtain the data of refuting rumors from the Sina Weibo platform and conduct empirical research to provide some reference for the platform to expand the scope of the dissemination of rumors and effectively realize the governance of rumors.

2. LITERATURE REVIEW AND MODEL HYPOTHESIS

2.1 Literature review

Rumor is a kind of news that spreads among specific people; it is generally accepted and different from facts, and has the characteristics of sudden, rapid spread, and wide radiation ^[7]. The destructive and unknown characteristics of public health emergencies accelerate the growth of rumors ^[8] and infiltrate the media network faster than real information ^[9]. For the rumor management of public health emergencies, the existing research is carried out from three angles: rumor dissemination process governance, rumor dissemination subject and object

governance, and rumor governance mechanism.

In the study of the spread of rumors, researchers draw lessons from the idea of the spread of infectious diseases. Some scholars divide the spreading process of rumors into the stages of breeding, spreading, and refuting rumors^[3] and latent outbreak period, mutation period, and extinction period^[11], which are basically in line with the law of growth and disappearance in the spread of infectious diseases. Many researchers carry out mathematical modeling research on the process of rumor spread. Daley and Kendal^[12] put forward the classical D-K model, drawing lessons from the idea of infectious disease transmission; in this model, people are classified as ignorant, disseminator, and suppressor, in which the disseminator will transform the ignorant into a disseminator, and the suppressor will transform the disseminator into a suppressor. Maki and Thomso^[13] further improve the D-K model and put forward the M-T model. The core of the model is that rumors spread through direct contact with the communicator, and the disseminator will be transformed into a suppressor when the disseminator touches other communicators in time. Also drawing lessons from the idea of infectious disease transmission, Dodds^[14] puts forward the SIR model, which includes rumor-susceptible people, rumor-infected people, and rumor-immune people. The researchers also proposed variants of various SIR, SIS, and DK models^[15,16]. As far as the object of rumor governance is concerned, for general rumors, dealing with false texts is necessary, and for online false information with profound destructiveness, we can also investigate the corresponding legal liability fields while dealing with false texts^[17]. Official information can restrain the spread of rumors and group behavior when official information somewhat spreads^[18]. For rumor disseminators, first of all, government departments should severely crack down on the rumor disseminator field^[19]. In addition, the quality of netizens and the government's behavior of refuting rumors are important factors affecting the interruption of rumors^[11]. In terms of rumor control and governance mechanism, researchers advocate all-round ecological governance. Dou^[20] points out that rumor management should be carried out from the subject-information-environment as a whole. Liu^[21] points out that rumor management should be carried out in such aspects as perfecting relevant laws and regulations, establishing an information certification center, enhancing citizens' moral awareness, strengthening humanistic quality education, and improving citizens' awareness of responsibility. Li^[22] conducts ecological management of rumors from the perspective of risk analysis from the three dimensions of subject-process-environment.

However, most of the existing studies start from the rumor itself, ignoring the perspective of refuting rumors. First of all, in the aspect of refuting rumors, researchers pay attention to the importance of releasing and refuting rumors, and researchers believe that the release of rumor-refuting information is a powerful response to rumors. Although publishing information to refute rumors is much less frequent than publishing rumors^[13], timely and accurate release of rumor information can effectively maintain network security and public order^[23]. Yuan^[11] also believes that the behavior of refuting rumors is an important factor affecting the interruption of rumor transmission. Second, the researchers also believe that the main approach to eliminate rumors is to refute rumors. Shuhud^[24] proposes six steps to deal with rumor information: filter, subject classification, identification, notification, verification, and rebuttal. Paek^[2] also points out that rumor-refuting information is the most effective way for the government to deal with rumors rather than simply make a counterstatement or complete denial of rumor information. Finally, the researchers put forward the way to refute rumors. Wen^[25] believes that the best way to stop rumors is for influential users to come out and refute rumors in the early stages of the spread of rumors. Wang^[4] classifies rumor refuters through feature engineering and identifies those who are rumor-refuting users. However, the research focuses on the characteristic attributes of rumor-refuting users, without considering social attributes and social capital. Therefore, most of the current studies focus on the rumor itself and ignore the perspective of rumor refutation. Some scholars start their research from the perspective of rumor refutation, but they still do not consider the attributes of social media and social capital.

When the information of refuting rumors is widely spread, it will reduce the harm of rumors. Therefore, from the perspective of refuting rumors, combined with social capital theory, this study will examine the influence of the characteristics of the publishers on the amount of rumor-refuting information forwarding and provide a clearer understanding of the forwarding mechanism of rumor-refuting information. The goal is for the platform to formulate rumor-refuting strategies to provide a certain theoretical basis and practical reference, thus promoting the spread of rumor-refuting information.

2.2 Research models and hypotheses

Social media is the product of the Internet era. Users form a social network with information dissemination as the core through mutual attention, comment, like, retweet, browsing, and other behaviors. The social capital formed by the social network will affect the behavior of users in social activities. Social capital has been widely studied, which helps understand social relations. As the driving force of social behavior in human society, social capital theory is widely used in the analysis of users' shopping behavior, business operation, social media use, and user knowledge payment behavior [3,14,26,27].

The publishers of rumor-refuting information accumulate social capital in the network through the interaction between social media, such as paying attention to, giving likes, publishing articles, to name a few. This social capital will inevitably affect the forwarding behavior of other users for the rumor-refuting information. Therefore, this paper applies social capital theory to investigate the effect of the social capital of users who publish rumor-refuting information on the amount of the rumor-refuting information forwarding in the case of public health emergencies.

Social capital is rooted in the structure of lasting social networks and the relationship between people who know and recognize each other. Bourdieu [28] first makes a systematic description of social capital; taking "connection" as the starting point, he proposes that social capital is "a collection of actual or undiscovered resources related to the relationship network of mutual understanding." For Coleman [29], social capital is a variety of resources to promote the realization of individual goals. Then, "social capital theory" proposed by Putnam [6] extends social capital to the macro social network structure to describe the reasons for individual behavior in the social network and defines social capital as the characteristics of social organizations, including social networks, social norms, and social trust. Considering that social capital is an invisible force combined with social interaction and relationship, shared values and interests, and common belief and trust, this paper uses the commonly used three-level classification of social capital, namely, structural level, relational level, and cognitive level fields [11].

In essence, the forwarding of rumor-refuting information belongs to a specific user behavior model. Therefore, the rest of this paper will draw lessons from the theory of social capital to construct a model of influencing factors of rumor-refuting information forwarding from the perspective of social capital of the publishers of rumor-refuting information from three levels: structural, relational, and cognitive dimensions.

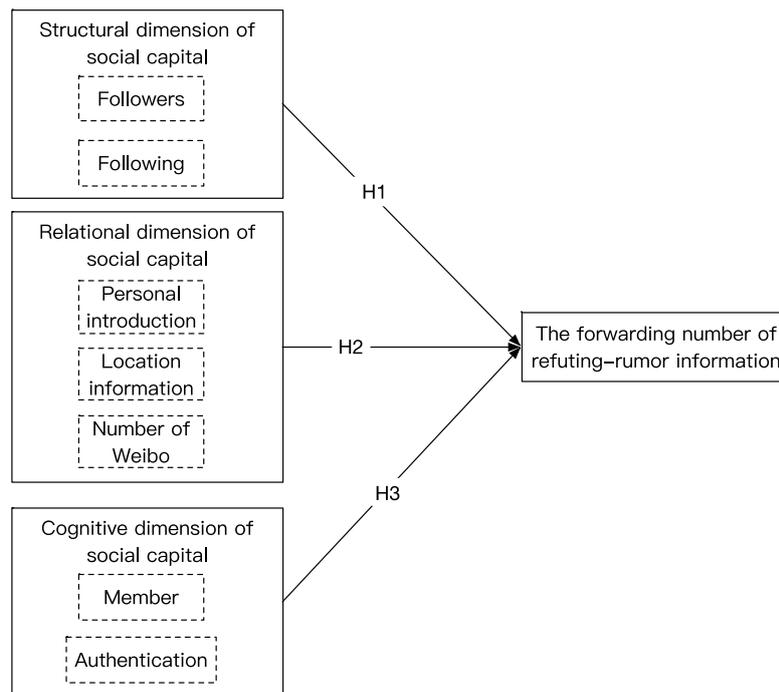


Figure 1. Model of factors affecting the forwarding number of refuting-rumor information

The structural dimension of social capital refers to the trend of establishing interpersonal relationships formed by members and the intensity of members' efforts to connect with other people in the network ^[11]. It is the relationship between individuals formed through connections and reflects the structural characteristics of the network ^[15]. These connections are made possible through interaction with others to maintain contact and expand interpersonal relationships, thus affecting the willingness of users to participate in activities ^[30]. Bi ^[31] believes that on Weibo platform, social capital can be formed by fan relationships and attention relationships. The number of fans and followers has a positive relationship with users' social capital and has a positive impact on the attention of their published content. In addition, social structural capital can enhance users' subjective well-being of using social media and has a positive impact on users' question-and-answer and shopping behaviors ^[7,32].

Sina Weibo users can have a certain number of followers and fans because of their interest and influence. From the perspective of social capital, these users and their followers constitute the social capital of the user's structural hierarchy. Therefore, combined with the number of Weibo users and the number of followers (fans), this study puts forward the following assumptions:

H1a: The number of followers of rumor-refuting microblog publishers will have a positive impact on the forwarding amount of rumor-refuting information.

H1b: The number of followings of rumor-refuting microblog publishers will have a positive impact on the forwarding amount of rumor-refuting information.

Relationship dimension social capital is the capital rooted in the continuous development of interpersonal relationships ^[15], which refers to the trust, mutual benefit, and win-win situation generated by the interpersonal relationship in the network. Individuals build relationship trust by participating in interpersonal relationships. Relationship trust refers to the accessibility of other people's access to informative comments and the availability of other people's emotional support. It is an important relationship resource and represents a relationship attribute. The generation of trust begins when users show their true information. Trust has a positive impact on the identity of interpersonal relationships. It can largely enhance knowledge acquisition behavior and

positively promote users' purchase behavior and willingness to pay. Reciprocity refers to the behavior caused by the relationship of interests in the relational community [12,26].

Sina Weibo users can choose whether to disclose their information, and the platform will record the number of historical Weibo posts by the user, which constitute the social capital of the user's relationship dimension. Therefore, combined with the openness of user information and the number of Weibo, this research puts forward the following assumptions.

H2a: For rumor-refuting microblog publishers, displaying personal introduction will have a positive impact on the forwarding amount of rumor-refuting information.

H2b: For rumor-refuting microblog publishers, displaying personal geographic information will have a positive impact on the forwarding amount of rumor-refuting information.

H2c: The number of Weibo posted by rumor-refuting microblog publishers will have a positive impact on the forwarding amount of rumor-refuting information.

Cognitive dimension social capital refers to the resources that provide common explanation and collective vision in the collective, including topic and vision consistencies, describing the common interests, values, and expression language of relationship members. Cognitive social capital has a positive effect on users' product selection behavior and information exchange behavior [33].

Weibo has membership functions, and Weibo users who gain membership have the same vision and experience. In addition, Weibo will authenticate professionals or celebrities, which constitute the cognitive dimension capital of the user. Based on the above analysis, the following assumptions are put forward:

H3a: For rumor-refuting microblog publishers, obtaining membership will have a positive impact on the forwarding amount of rumor-refuting information.

H3b: For rumor-refuting microblog publishers, gaining Weibo professional certification will have a positive impact on the forwarding amount of rumor-refuting information.

3. DATA DESCRIPTION

Sina Weibo has more than 440 million monthly active users, making it the largest and most active Weibo platform in China. This research crawls 3,035 rumor-refuting information and corresponding user data released in the "COVID-19" incident on the platform from January 22, 2020 to February 22, 2020.

The dependent variable is the forwarding number of rumor-refuting information, and the independent variable is based on the social capital of the publishers of refuting rumors. The number of fans, followers, Weibo posted; whether to disclose the profile, geographic location information; whether to be a Weibo member; and whether to obtain Weibo professional certification are selected from the structural, relational, and cognitive dimensions. Among them, the number of followers (fans), followings, and Weibo posts are discrete non-negative integers, and the other independent variables are 0–1 variable. The corresponding symbols for the data are as follows:

Table 1. Corresponding table of the personal information of the subject of the publication of rumor-refuting information

Symbol		Meaning
Dependent variable	repostsnum	Number of information forwarded
Structural dimension independent variable	followers	Number of followers (fans) of Publishers
	followings	Number of publishers focused on others
Relational dimensional independent variable	weibonum	Number of Weibo posted by publisher
	hasintroduce	Whether the publisher opens his/her personal introduction. 1 means yes, 0 means no.
	hasplace	Whether the publisher opens the geolocation information. 1 means yes, 0 means no.

Cognitive dimension	ismember	Whether the publisher is a Weibo member. 1 means yes, 0 means no.
independent variable	hasverified	Whether the publisher has obtained Weibo professional certification. 1 means yes, 0 means no.

The data are described statistically, and the specific description results are as follows:

Table 2. Information description statistics of the subject of the publication of rumor-refuting information

Variable	Mean	Standard deviation	Maximum	Minimum	Skewness
repostsnum	8.071e+01	1.995e+03	9.970e+04	1	43.69
followers	8.848e+02	1.229e+03	2.000e+04	1	5.81
followings	1.159e+06	5.331e+06	1.593e+04	1	15.73
weibonum	2.555e+04	3.674e+04	7.562 e+05	1	5.90
hasintroduce	9.600e-01	1.900e-01	1	0	-4.99
hasplace	5.800e-01	4.900e-01	1	0	-0.32
ismember	6.900e-01	4.600e-01	1	0	-0.80
hasverified	7.400e-01	4.400e-01	1	0	-1.11

The deviation values of the three non-binomial independent variables followers, followings, and weibonum are 5.81, 15.73, and 5.90, respectively, all greater than 5, indicating that their distribution is not normal. Considering the subsequent regression model, normalizing the distribution of this variable is necessary. Here, logarithmic transformation is used to logarithmize the three variables, that is, the variables are logarithmic to form new variables, and the corresponding new variables are described as follows.

Table 3. Logarithmic description statistics of information of the subject of the publication of rumor-refuting information

Variable	Mean	Standard deviation	Skewness
log(followers)	6.240e+00	1.080e+00	-0.47
log(followings)	1.038e+01	3.510e+00	-0.24
log(weibonum)	9.160e+00	1.800e+00	-1.14

The skewness value of the logarithmic variable accords with the normal distribution, and the original variable is replaced and used as the follow-up regression model.

Finally, through the kappa test of the eigenvalue of the independent variable matrix, the condition number k is 10.74, which is much less than 100. This finding indicates that no multicollinearity problem exists in the model and can be analyzed by regression.

4. EMPIRICAL ANALYSIS

4.1 Negative binomial regression analysis

Considering that the dependent variable is a discrete non-negative integer and has a residual with non-normal distribution, the regression model chooses the negative binomial regression model for estimation. The following is a brief description of the basic concepts of the negative binomial regression model.

Negative binomial regression is similar to conventional multiple regression, except that the dependent variable y follows the observed value of negative binomial distribution. Thus, the possible value of y is a non-negative integer. It is a generalization of Poisson regression model, in which the mean and variance of variables cannot be equal. The negative binomial distribution in negative binomial regression is composed of Poisson distribution plus a gamma noise variable with a mean value of 1 and a scale parameter (scala parameter)

of v. Therefore, the negative binomial distribution is expressed as follows:

$$P_r(Y = y_i | u_i = \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + u_i}\right)^{\alpha^{-1}} \left(\frac{u_i}{\alpha^{-1} + u_i}\right)^{y_i}, \tag{1}$$

$$u_i = t_i u, \quad \alpha = \frac{1}{v}, \quad u \text{ is the probability of } y \text{ occurrence at each exposure (described by time here).}$$

In negative binomial regression, the average value of y is determined by exposure time t and k regression variables (x values) as described in formula 2.

$$u_i = \exp(\ln(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}). \tag{2}$$

β is the parameter to be estimated.

Furthermore, the negative binomial regression model can be described as equation 3.

$$P_r(Y = y_i | u_i = \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{1}{\alpha^{-1} + 1}\right)^{\alpha^{-1}} \left(\frac{\alpha u_i}{1 + \alpha u_i}\right)^{y_i}. \tag{3}$$

Finally, the parameters can be obtained by the maximum likelihood estimation method.

The generalized linear model glm in MASS package in R language integrates the negative binomial regression function. This article uses R for programming. The version number of R is 3.6.2. The version number of the Master package is 7.3–51.4.

The regression results are as follows:

Table 4. Results of negative binomial regression model

Variable		Coefficient	Standard error	Significance level
Constant term	(Intercept)	1.302	0.233	2.330e-08***
Structural dimension dependent variable	Followers	-0.333	0.030	<2e-16***
	Followings	0.471	0.013	<2e-16***
Relational dimension dependent variable	Weibonum	-0.309	0.021	<2e-16***
	Hasintroduce	-0.196	0.164	0.234
	Hasplace	1.555	0.074	<2e-16***
Cognitive dimension dependent variable	Ismember	0.243	0.070	0.001***
	Hasverified	0.734	0.092	1.210e-15***

Signif.codes:0‘***’0.001‘**’0.01‘*’0.05‘.’0.1‘ ’1

The rest of the model parameters are as follows:

Table 5. Other parameters of negative binomial regression model

Parameters	Parameter value
Null deviance	8,049.0 on 3,034 degrees of freedom
Residual deviance	3,575.6 on 3,027 degrees of freedom
AIC	21,362
2 x log-likelihood	-21,344.199

4.2 Robustness test

To test the robustness of the model, this study divides the data and uses the divided data for regression fitting. Specifically, the original data are sorted according to the user id (uid). In part, user id indicates the event that a user creates an account. The smaller the id, the sooner the account is created. Therefore, by classifying users by the size of id and analyzing the model results of classified user data, we can further explore the

robustness of the model.

A total of 3,035 pieces of data are found in the original data, and the first 1,517 pieces of data sorted from small to large are selected for negative binomial regression. The regression results correspond to the coefficient parameters of the model as shown in the table below.

Table6. Results of negative binomial regression model after screening data

Variable		Coefficient	Standard error	Significance level
Constant term	(Intercept)	0.418	0.393	0.287
Structural dimension dependent variable	followers	-0.341	0.048	7.610e-13***
	followings	0.509	0.018	<2e-16***
Relational dimension independent variable	weibonum	-0.263	0.033	8.610e-16***
	hasintroduce	-0.114	0.246	0.644
	hasplace	1.897	0.104	<2e-16***
Cognitive dimension dependent variable	ismember	0.271	0.098	0.006**
	hasverified	0.441	0.127	0.001***

Signif.codes:0'***'0.001'***'0.01'*'0.05'.'0.1' '1

By taking the sorted "forerunner" data for negative binomial regression, except for the decrease in the significance level of "ismember" in this variable (here is also significant), the significance level and parameter symbols of other parameters have not changed. Therefore, the change in the data has no huge impact on the results. The model is robust.

4.3 Result analysis

For the structural dimension of social capital from the level of significance, the number of Weibo publishers' followings has a positive impact on the number of retweets, and the result is significant, assuming that H1b is established. Although the number of followers' (fans) parameter is significant, its symbol is negative. This finding indicates that the number of fans of the publisher has a negative impact on the forwarding of rumor-refuting information, assuming that H1a is not true. From the significance level, the number of Weibo publishers' followings has a positive impact on the number of reposts of refuting rumors, and its coefficient is 0.471, indicating that for each additional following of the user, the user publishes 0.471 more retweets of rumor-refuting information. Additionally, the meaning of the other coefficients is similar, which shows that when the activity and influence of users are enhanced, they will have closer ties with other users and more influence on the forwarding behavior of others. This finding is consistent with the description of Bi^[31]. At the same time, the number of followers of Weibo publishers and the number of Weibo posts significantly reduced the number of Weibo retweets, with a regression coefficient of -0.33309. This finding is inconsistent with the previous hypothesis that the number of followers represents the social cost of the user's structural dimension. This result may be related to the proliferation of the Weibo platform, and the excessive number of fans may be due to the water army rather than the influence gained by the user through the network relationship. No significant positive or even negative correlation exists between the number of fans and the forwarding number of refuting-rumor information.

For the relational dimension of social capital, the number of Weibo and whether open personal introduction are negative and the introduction parameters are insignificant, indicating that H2a and H2c are not valid. Whether the disclosure of geographical location information is significant, indicates that H2b is established. The regression coefficient of Weibo variable is -0.30897, indicating that a negative correlation exists between the number of Weibo and the forwarding number of refuting-rumor information. The higher the number of Weibo,

the more it can reflect the social capital of the user relationship dimension. However, at the same time, combined with the entertainment of the media platform, the high number of Weibo may reflect that the user has a wide range of interests and distractions but cannot be extremely professional and can reduce others' willingness to forward refuting rumors. The variable of whether or not to display profiles is insignificant. It is speculated to be related to the fact that most users display profiles, which can also be seen from the descriptive statistics. Thus, no significant relationship exists between an open personal introduction and the number of Weibo retweets. The disclosure of geographic location information by Weibo publishers has a positive impact on their forwarding volume, indicating that Weibo publishers' trust in others will positively affect the forwarding behavior of others. The results show that certain information disclosure on Weibo platform can improve the social capital of users and promote the operation of the information dissemination network. In addition, the coefficient of whether the geographic location information variable is disclosed or not is remarkably at 1.555, which has the greatest positive impact on the dependent variable; this finding is consistent with the fact that trust is an important social capital and can promote user behavior in Wu [2] and Zhou [26].

For the cognitive dimension of social capital, whether it is a member and whether it has received professional certification are significant and its symbol is positive, indicating the assumption that H3a and H3b are established. Opening members and obtaining Weibo professional certification can increase the retransmission of Weibo, which is similar to the research results in Wu [2]. This finding indicates that users will consider the professional background of the posting subject and the same intention to use the platform when forwarding refuting rumors. Professionals on the Weibo platform are also indicated to be more recognized, and people easily resonate with those having the same willingness to use the platform.

5. CONCLUSION AND DISCUSSION

The problem of rumors of public health emergencies has become a major challenge to digital governance in China. To deal with the rumors in public health emergencies, social media platforms, government agencies, and netizens actively carry out measures to refute rumors. Therefore, based on the theory of social capital, this study constructs a model of factors affecting the forwarding of rumors from three levels: structural, relational, and cognitive dimensions. Taking the COVID-19 event as an example, we obtain the data of the Sina Weibo platform, investigate the characteristics of the publishers that affect users' forwarding and dispelling rumors, and use the negative binomial regression to model. On this basis, we carry out a robustness test.

The results show that structural, relational, and cognitive dimensions of social capital all have an impact on the forwarding behavior of rumor-refuting information. Specifically, factors such as the number of followers of publishers, disclosure of geographic location information, opening members, and obtaining Weibo professional certification will increase the forwarding number of the rumor-refuting information, and whether or not open personal introduction will not significantly affect the forwarding of the rumor-refuting information. At the same time, the number of Weibo and fans will reduce the retransmission of the rumor-refuting information.

The current study aims to enrich and expand the theoretical research results of public health emergency rumor governance and social capital by exploring the behavior of refuting rumor information in public health emergency. At the same time, it provides certain empirical and theoretical bases for the national Internet digital governance.

Based on the above data analysis results, this study puts forward the following suggestions.

First, expanding the scope of information to refute rumors is a direct method of effectively achieving the effect of refuting rumors. Media platforms should pay attention to users' social capital, and social media platforms must further control the water army on the platform. The research results overturn the hypothesis that the number of fans represents the social cost of the structural dimension of users. Too many fans do not

represent the influence that users gain through network relationships. This phenomenon may be related to the proliferation of the water army on the Weibo platform. Therefore, purifying the platform environment and optimizing the fan recommendation mechanism are both necessary.

Second, users must be pertinently encouraged to conduct behaviors such as information disclosure, the degree of trust among users must be enhanced, various new functions for users must be customized, the social capital of users' relationship dimension must be improved, and the retransmission of information to refute rumors must be further promoted to enhance the status of user relations.

Finally, the platform should develop a reasonable authentication mechanism to improve the professionalism of platform users, give platform members and professional certification qualifications, and improve users' cognitive dimension of social capital.

In a word, expanding the spread scope of rumor information is the direct approach to effectively achieve the effect of rumor information. The media platform should pay attention to the social capital of platform users, expand the spread of rumor information, and curb the Internet rumor industry chain.

Of course, as an exploratory study, this research has the following shortcomings.

First, although Sina Weibo is a representative social media, this study only selects the Sina Weibo platform as the starting point. It lacks horizontal research between platforms. Notably, the factors affecting the forwarding behavior of rumor-refuting information on other platforms may be different.

Second, this research does not consider the dynamic nature of the forwarding of rumor-refuting information and does not explore from the time series. Nonetheless, the behavior of forwarding and rumor-refuting information itself is a continuous process, and investigating the influencing factors of rumor-refuting information forwarding in the dynamic process is a worthy endeavor.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 72061147005, 72072194, and the National Social Science Foundation of China under Grant 16BXW045.

REFERENCES

- [1] Gao Y, Wang S, Padmanabhan A, Yin J, Cao G (2018) Mapping spatiotemporal patterns of events using social media: a case study of influenza trends. *International Journal of Geographical Information Science* 32: 425-449
- [2] Paek HJ, Hove T (2019) Mediating and moderating roles of trust in government in effective risk rumor management: A test case of radiation-contaminated seafood in South Korea. *Risk Analysis* 39: 2653-2667
- [3] Lai S, Tang X (2017) Research on Internet rumor Management based on immune system Theory. *Modern intelligence* 30-34
- [4] Wang S, Li Z, Wang Y, Zhang Q (2019) Machine learning methods to predict social media disaster rumor refuters. *International journal of environmental research and public health* 16: 1452
- [5] Recuero R, Araujo R, Zago G (2011) How does social capital affect retweets? In: *Proceedings of the International AAAI Conference on Web and Social Media*
- [6] Hwang H, Kim KO (2015) Social media as a tool for social movements: The effect of social media use and social capital on intention to participate in social movements. *International Journal of Consumer Studies* 39: 478-488
- [7] Pan X (2002) Internet rumors and news truth. Master's degree thesis East China normal University
- [8] Gao Y, Wang S, Padmanabhan A, Yin J, Cao G (2018) Mapping spatiotemporal patterns of events using social media: a case study of influenza trends. *International Journal of Geographical Information Science* 32: 425-449
- [9] Vosoughi S, Zhou H, Roy D (2016) Enhanced twitter sentiment classification using contextual information. *arXiv preprint arXiv:160505195*

- [10] Hong Y, Jia L (2019) Research on Rumor Collaborative Governance Mechanism of Public Emergencies from the Perspective of Actor Network. *Modern intelligence* 14
- [11] Daley DJ, Kendall DG (1965) Stochastic rumours. *IMA Journal of Applied Mathematics* 1: 42-55
- [12] Maki DP, Thompson M (1973) *Mathematical models and applications: with emphasis on the social life, and management sciences.*
- [13] Dodds PS, Watts DJ (2004) Universal behavior in a generalized model of contagion. *Physical review letters* 92: 218701
- [14] Linhe Z, Ling L (2020) Dynamic Analysis of time-delay rumor Propagation Model based on dispel rumor Mechanism. *Journal of physics* 69: 020501
- [15] Qianqian Q, Hua H, Yanan L, Chengfeng J, Yuanyuan M (2019) S2IR rumor spreading Model based on the structural characteristics of Social Network. *Complex system and complexity science* 16: 48-59
- [16] Hua L (2019) The Government Mechanism of Internet rumor Governance: legal boundaries and Power constraints. *Financial and economic law* 9
- [17] Yang Y, Fan Z, Yang J, Jin Y (2018) A Multi-layer Network rumor Communication Model based on official Information Control. *Computer application research* 35: 1294-1297
- [18] Ozturk P, Li H, Sakamoto Y (2015) Combating rumor spread on social media: The effectiveness of refutation and warning. In: 2015 48th Hawaii International Conference on System Sciences. IEEE, p. 2406-2414
- [19] Dou Y, Li H, Zhang P, Wang B (2018) A study on the Ecological Countermeasures of Wechat's political rumors. *Modern intelligence* 6
- [20] Liu X, Li P, Zhou M, Li C (2018) Research on Harms and Governance of Internet Rumors. In: 2nd International Conference on Economics and Management, Education, Humanities and Social Sciences (EMEHSS 2018). Atlantis Press, p. 74-78
- [21] Li H, Hong L (2019) Research on risk Analysis and Prevention Strategy of Network rumors for political Security. *Modern intelligence* 20
- [22] Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital, and the organizational advantage. *Academy of management review* 23: 242-266
- [23] Shuhud MIM, Alwi NHM, Halim AHA (2017) Six Critical Processes to Refute Social-Media-Rumor. *Advanced Science Letters* 23: 4929-4933
- [24] Wen S, Jiang J, Xiang Y, Yu S, Zhou W, Jia W (2014) To shut them up or to clarify: Restraining the spread of rumors in online social networks. *IEEE Transactions on Parallel and Distributed Systems* 25: 3306-3316
- [25] Putnam RD (2000) Bowling alone: America's declining social capital. In: *Culture and politics*. Springer
- [26] Birasnav M, Chaudhary R, Scillitoe J (2019) Integration of social capital and organizational learning theories to improve operational performance. *Global Journal of Flexible Systems Management* 20: 141-155
- [27] Bourdieu P (1980) Le capital social: notes provisoires. *Actes de la recherche en sciences sociales* 31: 2-3
- [28] Coleman JS (1988) Social capital in the creation of human capital. *American journal of sociology* 94: S95-S120
- [29] Hajli N, Sims J, Zadeh AH, Richard M-O (2017) A social commerce investigation of the role of trust in a social networking site on purchase intentions. *Journal of Business Research* 71: 133-141
- [30] Bi Y (2015) *Social Capital Measurement of Weibo users*. Guangzhou: South China University of Technology
- [31] Lin K-Y, Lu H-P (2011) Intention to continue using Facebook fan pages from the perspective of social capital theory. *Cyberpsychology, Behavior, and Social Networking* 14: 565-570
- [32] Chiu C-M, Hsu M-H, Wang ET (2006) Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision support systems* 42: 1872-1888
- [33] Paek HJ, Hove T (2019) Mediating and moderating roles of trust in government in effective risk rumor management: A test case of radiation-contaminated seafood in South Korea. *Risk Analysis* 39: 2653-2667

Short Research Paper**Public Negative Emotions Regulation During the COVID-19 Emergency:****GIR's Content Features and Lingual Form Do Matter***Xie Yu¹, Xuefeng Li¹, Wei Zhang^{1*}, Mei Zhang², Yanchun Zhu³*

¹School of Information, Central University of Finance and Economics, Beijing, 100081, China, ²School of Sociology and Psychology, Central University of Finance and Economics, Beijing, 100081, China, ³Business School, Beijing Normal University, Beijing, 100875, China

Abstract: Emergencies and their associated negative emotions have a great effect on public health. As a key part of emergency management, government information release (GIR) not only meets the public's health information seeking but also helps to eliminate the breeding and spreading of negative social emotions. From the two aspects of content features and lingual forms, a regression model was built to explore the mechanism of GIR on the regulation of public negative emotions by adopting the theoretical methods of content analysis and emotion calculation. During the emergency outbreak, if the government can timely release information on the incident and respond to the public using rational language, public negative emotions can be alleviated. During the emergency peak, the government should release the event progress, resolution, and disposal information to improve the recognition of public and eliminate negative emotions. According to different stages of emergencies, the government should timely and reasonably utilize the attitude tendency, content type, and lingual form of GIR to effectively regulate the public negative emotions.

Keywords: Public Health, Negative emotions, Government information release, Rhetorical strategies

1. INTRODUCTION

It is well known that emergencies and their associated social emotions, especially the collision of various negative emotions, have a great effect on public health^[1,2]. During emergencies, the positive psychological suggestion will make positive emotions generate, to avoid the influence caused by negative emotion, which is of great help in controlling disease and restoring healthy life^[3,4].

As a key part of emergency management, the government information release (GIR) not only meets the public's health behavior, but also helps to eliminate the public negative emotions, avoid secondary crises, and stabilize the social order^[5]. In recent years, governments at all levels have formed a relatively complete communication system in terms of the information disclosure of emergencies^[6]. However, in practice, such chronic diseases as the lack of dominant agenda setting, the information release delay, the misjudgment of public concern, and the loss of negative emotion guidance are still serious^[6,7]. Based on scientifically analyzing the rules of negative emotion infection-evolution, how to explore GIR strategies to effectively adjust and channel social emotions derived from emergencies has become a new topic to improve government emergency management level^[8].

*Address for correspondence: Dr. Wei Zhang, E-mail: weizhang@cufe.edu.cn.

At present, the research on the GIR strategy during emergencies is still in its infancy. The existing studies focus on the perspectives of public opinion and analyze the distribution characteristics of positive and negative emotions at each stage of emergencies. Most of the proposed release strategies are descriptive and speculative, with weak operability^[8,9]. In terms of specific release strategy formulation, such as release timing, content composition, language style selection, and emotional guidance, no specific quantitative feasible plan is provided, which cannot provide decision support for information release strategy in emergency management and effective regulation of negative social emotions^[10,11].

Therefore, this paper tries to explore the mechanism of GIR on the regulation of public negative emotions based on the content type, attitude orientation, lingual form to accurately grasp the distribution of the public emotions in emergencies and to effectively regulate the social negative emotions.

The contributions of this paper: (1) deconstructed the content type and attitude tendency of GIR and measuring the polarity of emotion by using content analysis, and text mining; (2) conducted empirical researches on the influence of content features and lingual forms of GIR on public emotions through the construction of regression model.

The rest of the paper is organized as follows. First, we review previous literature on GIR. Second, we introduce our theoretical foundation and the analysis procedure of the influence of GIR content features. Next, we state data collection procedures, data analysis, and results. The effect of GIR lingual forms on regulating public negative emotion is conducted and discussed. Finally, implications, as well as opportunities for future studies, are discussed.

2. RELATED WORKS

2.1. Content components of the GIR

Scholars have studied the content components of GIR using theoretical methods of journalism, public management, administration, and communication, etc. They believe that the cause, nature, degree of harm, and disposal measures of the incident are the main components of information disclosure content^{[12][13]}. Most scholars at home and abroad think that the content of the government's emergency information disclosure is closely related to the development stage of the incident, and it varies at the three stages of the early warning period, the outbreak period, and the post-restoration period, each with its emphasis. However, in practice, it still needs to be enriched and improved on how to optimize the content components of GIR information disclosure^[14].

2.2. Disclosure strategy of the GIR

Scholars mostly are based on the case analysis and content analysis methods to summarize the existing problems in the GIR disclosure on emergencies, and also to propose improvement measures^[15-17]. Existing research shows that the current GIR disclosure focuses on eliminating the strangeness of the people and reducing the intrusion of the social order by the incident, but pays little attention to the adjustment of negative emotions^[15,18,19], resulting in the defects such as slow response, lack of guidance of topic setting, and insufficient guiding role for emotions, etc^[20], so the relevant countermeasures have been proposed; however, most of them are descriptive with low operability^[9,16].

2.3. Rhetorical features of the GIR

The rhetorical strategy is an important means to improve the persuasiveness of information posted by the GIR and effectively express their positions and opinions^[21-23].

Based on Aristotle's rhetoric theory of three persuasive audience appeal, Lim analyzed the specific manifestations of the lack of emotional appeal, lack of rhetorical personality, and messiness of rational appeal in the response of local governments to crisis events, and proposed that the use of multiple rhetoric methods should be emphasized when dealing with emergencies^[21]; Arendt et al. believed that charismatic political language and pro-political behavior can help the government to alleviate the short-term crisis^[22].

In summary, the scholars have summarized the problems in GIR based on the micro-blogs data and verified the effectiveness of the strategy through simulation. These results provide theoretical references for this study, but there are still some shortcomings: (1) They focus on the information dissemination of emergencies, construct the content of released information, and clarify the form of information content, but rarely involving the mechanism of action, and paying less attention to the content type and components of GIR, attitudes, and the influence of lingual forms on social-emotional communication; (2) The improvement measures and policy suggestions are mostly descriptive and speculative discussion.

3. RESEARCH ON THE INFLUENCE OF THE GIR CONTENT FEATURES ON THE public'

EMOTIONS

3.1. Data collection and processing

Considering the event types, timeliness, attention, and social impact of public emergencies, we selected recent seven typical emergencies as research cases. The GIR of the above incidents was selected for the trial survey. On this basis, the number of published micro-blogs and their forwards, comments, and likes were used as indicators of activity, and the GIR with more active performance were selected as samples. It collected a total of 413 pieces of GIR and 203,854 comments.

Based on the trial survey, inspired by reference^[15], this paper divides the content features of the GIR samples into the following two aspects:

- **Attitudes** There are three main types of attitudes: Positive attitudes; neutral attitudes; negative attitudes.
- **Content types** It includes 6 types, namely event progress, popular science; rumor-removal; resolution and disposal; summary and reflection; others.

When classifying the emotions of the GIR comment data in this study, we adopted the classification method of seven emotions, that is, anger, disgust, fear, joy, love, sadness, and surprise. The manual coding was used to classify the types of GIR samples. "NLPIR Chinese word segmentation system" was applied to perform GIR recognition and emotion computing. The evaluation of emotions should take into account the GIR comment data.

Through the kappa calculator, the reliability test was performed on the variables of the GIR content type, to obtain the reliability value of about 86.8%. This indicates higher reliability. For emotion analysis of comments, 1,000 comments were randomly selected for manual labeling, and then compared with the emotional analysis results of the NLPIR, to obtain the reliability value of the information expression form for about 75.2%, indicating that the platform is credible for emotional analysis on GIR to a greater degree.

3.2. Descriptive statistical analysis

Through screening all 15,200 comments, a total of 165 met the requirements. Emotion analysis was then performed on these comments. In terms of the number of likes, there were 446,076 comments from public, of which the comments with anger had the highest proportion of up to 38.04%, and those with surprise and joy were the lowest, to be 0.97% and 0.10% respectively. Therefore, this paper selects anger, sadness, disgust, fear, and love as research objects, excluding surprises and joy.

Public emotion have different characteristics for GIR in various phases, which indicates that GIR information influences public emotions in emergencies (see Fig 1.). It can be seen that the number of public in anger was relatively high during the outbreak phase of an emergency; as the government continues to release information, the proportion of public in anger constantly declines; sadness peaks during the climax phase; the proportion of disgust is relatively high in the flat phase; the proportion of public negative emotions declines in the recession phase, and the proportion of love reaches a peak.

Therefore, in the outbreak and climax phases, the GIR should play a good role in guiding public opinion and better assuming the government's social responsibility, and efforts should also be done to emotional channeling in the flat phase, but during the recession phase, the frequency of information disclosure should be reduced accordingly to allow emergencies to slowly withdraw from the public eye.

3.3. The influence of GIR's attitudes on public emotions

To study the relationship between public emotions and attitudes, an independent sample means the test was used to statistically analyze three qualitative variables: positive, neutral, and negative. A regression model of GIR attitude tendency and netizen emotions was established

$$attitude_i = a_1 \ln(fear) + a_2 \ln(anger) + a_3 \ln(sadness) + a_4 \ln(disgust) + \varepsilon \tag{1}$$

Here, $attitude_i$ indicates three tendencies of attitudes: positive, negative, and neutral; fear, anger, sadness, and disgust respectively indicate four types of negative emotions.

The results of regression analysis validate: (1) A positive attitude to posting information has a significant effect on the "fear" emotion of public ($P < 0.01$), and the mean number of "fear" emotion (0.767) using a positive attitude in posting information was significantly lower than that (1.412) in other types of attitudes, indicating that positive attitudes can relieve the fear emotions of public. Meanwhile, positive attitudes in posting information are significant for the public disgust emotions ($P < 0.05$), and the mean number of disgust (2.057) generated by using positive attitudes was significantly lower than that (2.771) in other types of attitudes, indicating that using positive attitudes in posting information can reduce the public disgust emotions. (2) The neutral attitude in posting information shows significance for public fear emotions ($P < 0.01$), and the mean number of fear emotions (1.526) using positive attitudes was significantly higher than that (0.684) caused by other types of attitudes, indicating that using a neutral attitude can stimulate public fear emotions. (3) The use of negative attitudes in the information disclosure can reduce the fear and sadness of public, while significantly stimulating their anger. For example, in the Wuxi viaduct

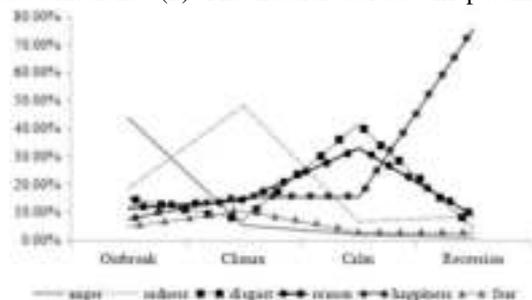


Fig 1. Emotional changes in various periods.

collapse incident, the GIR, Wuxi's Release was questioned for the delayed disclosure of the emergency information, while it posted short comments to criticize online rumors, which triggered a wider range of controversy. This shows that the GIR should talk to public with an equal attitude when facing public opinions, rather than criticize public with a negative attitude.

3.4. The influence of GIR content types on public emotions

There are 6 content types of GIR information: event progress, popular science, rumor removal, resolution and disposal, summary and reflection, and others. Through statistical analysis of the 165 pieces of information released, it can be seen that the GIR information content is mainly the event progress, and the resolution and disposal; the information on the event progress accounted for the vast majority, up to 41.2%, followed by information on incident resolution and disposal, accounting for 25.5%.

The independent sample means the test was used to study the relationship between the content type of information and the public emotions. The regression model is constructed as:

$$c_type_j = a_1 \ln(fear) + a_2 \ln(anger) + a_3 \ln(sadness) + a_4 \ln(disgust) + \varepsilon \quad (2)$$

The results of the regression coefficients and their significance show that: (1) The disclosure of event progress information by the GIR shows significance for public fear emotions ($P < 0.01$), and the mean number of "fear" emotions (1.529) generated by disclosing event progress was significantly higher than that by the release of other types of information (1.038). This shows that the disclosure of event progress by the GIR can easily stimulate the fear of public. (2) The disclosure of popular science information by the GIR is significant for "anger" and "sadness" emotions ($P < 0.01$). From the mean numbers of these two emotions, the disclosure of popular science information by the GIR can easily dispel public anger and sadness. (3) The disclosure of rumor removal information by the GIR has a significant effect on public "anger" ($P < 0.1$) and "sadness" emotions ($P < 0.05$). From the mean of these two emotions, the disclosure of rumor removal information by the GIR can easily dispel public anger and sadness. (4) The disclosure of resolution and disposal information by the GIR is significant for public sadness emotions ($P < 0.05$), and the mean number of sadness emotions (3.909) generated by disclosing such information was significantly higher than that by the release of other types of information (2.75). This shows that the disclosure of resolution and disposal information by the GIR can easily stimulate the sadness of public. (5) The disclosure of summary and reflection information by the GIR shows significance for public negative emotions such as fear, anger, sadness, and disgust. From the mean numbers of these four emotions, the disclosure of summary and reflection information by the GIR can easily dispel these emotions of the public. (6) The disclosure of other information by the GIR has no significant influence on negative emotions.

4. RESEARCH ON THE INFLUENCE OF GIR LINGUAL FORMS ON public' EMOTIONS

4.1. Theoretical analysis and main hypothesis

Rhetoric is the art of persuasion for the audience, allowing them to form some kind of judgment, and recognize, approve and adopt the opinions or take some action. Aristotle's rhetorical theory divides lingual forms into three types, namely, ethos, pathos, logos. Wang cited the three elements of propaganda in journalism: truth, emotion, and reason as theoretical references, and divided the lingual form of GIR into three major categories: character appeal (i.e., concise in text, with a clear attitude and a firm tone), emotional appeal (i.e., mainly uses

interrogative and exclamatory sentences and makes good use of implicit expressions with the diverse texts and soft tone), and logic appeal (i.e., personalized positioning, strict wording, and neutral tone, with statements mostly, and updated and profound content)^[17]. Wang's classification method of the micro-blogs is based on the theory of three persuasive audience appeals, and it is more suitable for the classification of the GIR lingual form. Thus, in this paper, the GIR lingual form was classified in terms of character appeal, emotional appeal, and logic appeal.

Studies have shown that in case of an emergency, a "venting" response and dishonest manner are likely to cause public outrage^[23], and a positive response to public concerns and a frank and humble attitude can win the favor of public and resolve the crisis of public opinion. Hovland's persuasion model considers the objectivity and credibility of the disseminator as the basic conditions for persuasion. Among them, credibility depends mainly on the qualifications and reliability of experts^[21]. This is similar to Aristotle's reliance on the "character" element to increase the credibility of the persuasive and to improve the persuasive effect. Both of them emphasize the credibility of the information disseminator to achieve the purpose of persuading the audience. Therefore, the following hypotheses have been made.

H1: The character appeal-based GIR disclosure has a dispelling effect on negative emotions.

In terms of different negative emotions, H1 can be clearly defined as the following hypotheses:

H1a: The character appeal-based GIR disclosure has a dispelling effect on fear.

H1b: The character appeal-based GIR disclosure has a dispelling effect on anger.

H1c: The character appeal-based GIR disclosure has a dispelling effect on sadness.

H1d: The character appeal-based GIR disclosure has a dispelling effect on disgust.

With the development of the incident, the GIR disclosure is adjusted timely in a humane and daily language style. This will not only show the government's confidence and ability to handle the incident to the public but also minimize public panic^[16]. The key to relief of the public's "hate" is to alleviate the repression of institutions on them in terms of information, power, and discourse^[17], and rely on metaphorical rhetoric, symbolic and mythical interpretive policy discourse systems in information disclosure. Thus, it can better achieve the purpose of maintaining and realizing the public interest, and actively lead the public's feelings to positive emotion, thereby forming a harmonious atmosphere of public opinion^[23].

Therefore, the following hypotheses have been made:

H2: In emergencies, the emotional appeal-based GIR disclosure is prone to sadness, but alleviates other negative emotions of the public, thereby improving the public recognition of government work.

In terms of emotion classification, H2 can be clearly defined as:

H2a: The emotional appeal-based GIR disclosure has a dispelling effect on fear.

H2b: The emotional appeal-based GIR disclosure has a dispelling effect on anger.

H2c: The emotional appeal-based GIR disclosure is prone to sadness.

H2d: The emotional appeal-based GIR disclosure has a dispelling effect on disgust.

H2e: The emotional appeal-based GIR disclosure is prone to love emotions.

Studies have shown that accurate numbers in authoritative information released by government departments are more effective than vague numbers in reducing public anxiety^[17-20]. Accurate numbers can also increase the credibility of information disseminators. Zhou et al. found that the structure of the speech according to the steps

of the current situation of the incident, the cause of the incident outbreak, and the measures taken by the government is relatively clear, allowing the audience to understand the content of government leaders’ speech clearly^[11]. Thus, the following hypotheses have been made:

H3: The logic appeal-based GIR disclosure has a dispelling effect on negative emotions.

In terms of emotion classification, H3 can be clearly defined as:

H3a: The logic appeal-based GIR disclosure has a dispelling effect on fear.

H3b: The logic appeal-based GIR disclosure has a dispelling effect on anger.

H3c: The logic appeal-based GIR disclosure has a dispelling effect on sadness.

H3d: The logic appeal-based GIR disclosure has a dispelling effect on disgust.

4.2. Research design

In this paper, a multiple regression model was adopted to study the influence of the GIR lingual form on public emotions. It usually includes the following steps:

(1) Variable selection

The lingual form of GIR information disclosure was used as independent variables, the public emotion ratio was used as the dependent variable, and GIR content features were used as control variables^[10] (shown in Fig. 2).

(2) Model construction

The regression model was analyzed using the multivariate regression analysis method. The basic regression model established was as follows:

$$Emotion = \beta_0 + \beta_1 Ethos + \beta_2 Pathos + \beta_3 Logos + \beta_4 phase + \beta_5 video + \beta_6 picture + \beta_7 link + \beta_8 topic + \beta_9 original + \beta_{10} len + \beta_{11} content + \beta_{12} attitude + \beta_{13} follower + \beta_{14} forward + \beta_{15} comment + \beta_{16} like + \varepsilon \quad (2)$$

Here, *Emotion*, as the dependent variable, indicates the emotional characteristics of public comments on the GIR, including five types of emotions: love, anger, sadness, fear, and disgust; the number of public emotions was calculated by adding the number of comments and likes; *Ethos*, *Pathos*, *Logos* are the lingual forms of the GIR, which are character appeal, emotional appeal, and logic appeal, the classification method is based on the method given by Wang et al.^[17]; *phase* is the stage of GIR information disclosure, which is divided into outbreak phase, climax phase, flat phase, and recession phase^[34]; *video*, *picture*, and *link* indicate whether there are related videos, pictures, external links in the GIR disclosure; *topic* indicates whether it is included in the information; *original* refers to whether the information is original; *len* refers to the length of the micro-blog words; *content* refers to the content type, including event progress, popular science, rumor removal, resolution and disposal, summary and reflection, and others; *attitude* means the attitude tendency, which is divided into positive, neutral, and negative; *follower* refers to the number of GIR fans; *forward* and *comment* are the number of GIR messages forwarded and commented by public; *like* is the number of likes of the GIR information.

According to the classification method given by Wang et al. ^[17], the GIR is classified in lingual forms (as shown in Table1).

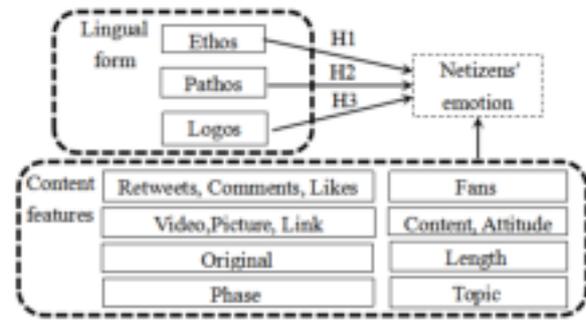


Fig 2. Conceptual model

Table 1. Classification basis of lingual form.

Lingual form	Feature	Sentence structure	Intention
Ethos	The text is concise, the attitude is clear, and the tone is firm	Commonly used imperative sentences, parallelism, and contrast	Explain the reason and show the attitude of the government
Pathos	Variety of text, soft tone	Commonly used interrogative sentences and exclamation sentences	Emotional counseling
Logos	Strictly worded and neutral	Commonly used declarative sentences	Convey the facts

4.3. Empirical analysis

According to the regression analysis results of Model 1, R^2 was adjusted to 0.402, indicating the influence of the GIR information disclosure on public loves. The coefficient of the emotional appeal-based lingual form was 0.213, which indicates that this lingual form can increase the love emotion of public, while the coefficients of the character appeal-based and logic appeal-based lingual forms were -0.361 and -0.401, indicating that these two lingual forms will reduce the proportion of public love emotion. In addition, the coefficient of the original was 0.107, and that of the video was 0.002, indicating that the original micro-blog disclosure by the GIR, attached with video can improve the recognition of public (as shown in Table 2).

For Model 2, R^2 was adjusted to 0.357, indicating the influence of the GIR information disclosure on public fear. The coefficients of the lingual forms based on character appeal, emotional appeal, and logic appeal were -0.259, -0.326, and -0.129 respectively, which indicates that these three forms can all reduce the fear of public, and the emotional appeal-based lingual form has the greatest influence. Meanwhile, the coefficients of the pictures and videos in the GIR were positive, and passed the significance test, indicating that the GIR attached with pictures and videos are more likely to generate fear.

For Model 4, R^2 was adjusted to 0.668, indicating the influence of the GIR information disclosure on public sadness. The coefficient of emotional appeal-based lingual form was 0.097, which indicates that this form can easily lead to the sadness emotion of public. Also, the pictures and original information in the GIR can easily cause the public to have sad emotions.

For Model 5, R^2 was adjusted to 0.549, indicating the influence of the GIR information disclosure on public which can dispel the disgust emotion of public. What's more, the GIR information with external links and more

Table 2. Regression results of language forms on the public emotions.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
	Happiness	Fear	Anger	Sadness	Disgust
Con	1.335**	0.063***	3.759**	-1.74**	0.461*
Outbreak	-0.267	-0.113	0.126*	0.254*	0.114*
Climax	-0.101	-0.104	0.035	0.156	0.026
Calm	-0.067	-0.043	-0.03	-0.069	0.07
Video	0.002*	0.052**	0.008	-0.005	-0.081
Picture	-0.005	0.044*	0.041*	0.063*	0.036
Link	0.027	0.017	0.039	-0.058	0.09***
Topic	0.076	-0.092	-0.153**	0.023*	0.031
Original	0.107**	0.007	0.001	0.023**	-0.033*
Length	-0.007	-0.009	0.118	0.055	0.071*
Fans	0.043	0.012	-0.335	-0.017	-0.162
Retweets	0.094	0.026	0.036	0.017	0.038
Comments	0.498*	0.574***	0.483**	0.656***	0.691***
Content	—	—	—	—	—

Attitude	——	——	——	——	——
Ethos	-0.361**	-0.259**	-0.035*	0.095	-0.054
Pathos	0.213***	-0.326**	-0.119**	0.097***	-0.179*
Logos	-0.401**	-0.129***	-0.102*	0.089	-0.15*
R2	0.464	0.424	0.532	0.703	0.595
Adj R2	0.402	0.357	0.477	0.668	0.549
N	165	165	165	165	165

* p < 0.1, ** p < 0.05, *** p < 0.01.

length will also cause public to have disgust. The original video will have a dispel effect on disgust.

According to the results of regression analysis, the authors concluded the influence of the GIR lingual form on the emotions of public and the verification of their hypotheses, as shown in Table 3

Table 3. Regression results of lingual forms on the public emotions.

Lingual Form	Happiness	Fear	Anger	Sadness	Disgust
Ethos	Negative	Negative (H1a support)	Negative (H1b support)	(H1c unsupported)	(H1d unsupported)
Pathos	Positive (H2e support)	Negative (H2a support)	Negative (H2b support)	Positive (H2c support)	Negative (H2d support)
Logos	Negative	Negative (H3a support)	Negative (H3b support)	(H3c unsupported)	Negative (H3d support)

5. CONCLUSION

By using content analysis and regression analysis, this paper analyzed the content features of GIR and the influence mechanism of lingual forms on public emotions and made a case analysis based on the COVID-19 emergency in Wuhan. The results show that: (1) The event progress information issued by GIR tends to generate negative emotions such as anger, sadness, and disgust; Refuting rumors can reduce public sadness, but it is also positively correlated with "disgust", indicating that refuting rumors may increase public disgust. (2) The positive attitude released by GIR can stimulate public "likes" and relieve their "fears" and "dislikes". And the neutral information can increase public "fear" emotion. While the negative information can stimulate "anger" and relieve "fear" and "sadness". (3) The use of character appeal-based lingual form has a dispelling effect on negative emotions like "anger" and "fear", that is, it can dispel public negative emotions and improve the proportion of rational emotions. While The emotional appeal-based form can help eliminate negative emotions such as fear, anger, and disgust, and increase the proportion of "sadness" and "preference". The use of the logic appeal-based lingual form has a good dispelling effect on public negative emotions such as "fear", "anger" and "disgust", could promote public to be rational, but it will reduce the goodwill of public to government agencies.

ACKNOWLEDGEMENT

This work was supported by Beijing Natural Science Foundation (9182016, 9194031), National Natural Science Foundation of China (72061147005, 71874215, 71571191) and MOE (Ministry of Education in China) Project of Humanities and Social Sciences (15YJJCZH081, 17YJAZH120, 19YJJCZH253), Fundamental Research Funds for the Central Universities (SKZZY2015021) and Program for Innovation Research in Central University of Finance and Economics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- [1] Li J, Stephens, KK, Zhu Y, Murthy D. (2019). Using social media to call for help in Hurricane Harvey: Bonding emotion, culture, and community relationships. *International Journal of Disaster Risk Reduction*. 2019; 38:101212.
- [2] McMahon SA, Ho LS, Brown H, Miller L, Ansumana R, Kennedy CE. Healthcare providers on the frontlines: a qualitative investigation of the social and emotional impact of delivering health services during Sierra Leone's Ebola epidemic. *Health Policy and Planning*.2016;31(9):1232-1239.
- [3] Smith EN, Romero C, Donovan B, et al. Emotion theories and adolescent well-being: Results of an online intervention. *Emotion*. 2018; 18(6):781.
- [4] Jamshidian Y, Kiani A, Dargahi S. Relationship of cognitive emotion regulation and meaning in life with health anxiety among emergency nurses. *Health in Emergencies and Disasters*.2018;3(4):199-206.
- [5] Liu Y, Xu W, Qiao H, et al. Key scientific problems in public emergency management. *Management Review*.2016; 28(8): 3-5.
- [6] Rodríguez Bolívar MP, Alcaide Muñoz L, López Hernández AM. Scientometric study of the progress and development of e-government research during the period 2000-2012. *Information Technology for Development*.2016; 22(1): 36-74.
- [7] Li J, Dong Y.Ensuring effective news coverage in disasters: a study of news coverage capability in China. *Global Business and Organizational Excellence*.2016;35(4):55-66.
- [8] Liu YJ, Chen SJ, et al. The analysis and policy recommendations on the spread of network public opinion in major production safety accidents—A case study of the 8.12 Tianjin port explosion.*Management Review*.2016; 28(3):221-229
- [9] Yu L, Li L, Tang L, et al. A multi-agent-based online opinion dissemination model for China's crisis information release policy during hazardous chemical leakage emergencies into rivers. *Online Information Review*.2017;41(4): 537-557.
- [10] Zhang W, Wang M, Zhu YC. Does government information release really matter in regulating contagion-evolution of negative emotion during public emergencies? From the perspective of cognitive big data analytics. *International Journal of Information Management*.2020; 50:498-514.
- [11] Zhou L, Cai L, Ye Y. Online emotional expression in response to an emergency: A sentiment analysis of public discourse on micro-blogs in response to a heavy rainfall in Wuhan, China. *China Media Research*.2019; 15(1):52-66.
- [12] Choi J, Lee S. Managing a crisis: A framing analysis of press releases dealing with the Fukushima nuclear power station crisis. *Public Relations Review*.2017; 43(5):1016-1024.
- [13] Henderson AC, Bromberg DE. Performance information use in local government: Monitoring relationships with emergency medical services agencies. *Public Performance & Management Review*.2015;39(1):58-82.
- [14] Kim Y. Understanding publics' perception and behaviors in crisis communication: Effects of crisis news framing and publics' acquisition, selection, and transmission of information in crisis situations. *Journal of Public Relations Research*.2016;28(1):35-50.
- [15] Zhou L, Li X, Huang J. Research on information release and its influence during unexpected events-in the views of government micro-blog. *Journalism Bimonthly*.2015; 2:144-152
- [16] Zhang, Y, Wang JC. (2015). On the problems and countermeasures of government information release mechanism in emergency—based on the case study of 2015 Shanghai bund stampede incident .*Journal of Intelligence*.2015;34(5):111-117.
- [17] Wang Q, Hu, S. Constructing media framework for government micro-blog's information publication under emergent events. *Modern Communication (Journal of Communication University of China)*.2016;38(11): 71-75.

- [18] Dai X, Zhang L, Dai WH, Zhou H. Research on emotional mapping of social networks. *Management Review*. 2016;28(8):79-86.
- [19] Li B. Research on the current social emotion situation and the steering of discourse space. *Jinan Journal (Philosophy & Social Science Edition)*. 2016;38(6):90-98
- [20] Chen SY, Huang C, Chen Q, Yang LR, Xu XL. Information-releasing strategy of local government microblog group during public emergency--a case study of Tianjin port explosion. *Journal of Intelligence*. 2016; 35(12):28-33.
- [21] Lim JS. The effectiveness of refutation with logic vs. indignation in restoring the credibility of and trust in a government organization: A heuristic-systematic model of crisis communication processing. *Journal of Contingencies and Crisis Management*. 2019; 27(2):157-167.
- [22] Arendt C, LaFleche M, Limperopulos MA. A qualitative meta-analysis of apologia, image repair, and crisis communication: Implications for theory and practice. *Public Relations Review*. 2017;43(3):517-526.
- [23] McNeill A, Harris PR, Briggs P. Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic. *Frontiers in public health*. 2016;4: 26.

Full Research Paper

From Liking to Following: Research on the Antecedents of Middle-aged and Elderly Douyin Users' Fans

*Jin Yan, Ningyue Cao, Changyu Wang**

School of Business, Jiangnan University, Wuxi, 214122, China

Abstract: Less is known about the determinants of middle-aged and elderly users' number of fans on the short-video platform. To help middle-aged and elderly users understand how to increase fans on the short-video platform, this paper thus draws on social penetration theory and information foraging theory to analyze the data collecting from the personal homepage of middle-aged and elderly users (including the information about users and videos, the number of likes and the number of fans) on Douyin platform which is the Chinese version of TikTok. The results show that: (1) The number of likes received by middle-aged and elderly users is positively correlated with the number of fans; (2) Getting verification from the platform, the number of bio words, the number of videos, and the professionalism of videos have no significant direct relations to the number of fans; (3) Getting verification from the platform, the number of bio words and the professionalism of videos can strengthen the relationship between the number of likes and the number of fans, while the number of videos has no significant moderating effect on the above relationship.

Keywords: middle-aged and elderly users, short-video social media platform, Douyin, fans, likes

1. INTRODUCTION

Since population aging is an issue of concern to all nations of the world in the past, present, and future, helping senior citizens live a healthy and happy life has become a critical social issue^[1]. Social media platforms play a critical role in improving seniors' social lives because they can help elderly people to interact with others and actively participate in society^[2]. For example, on the largest Chinese mobile short-video social media platform Douyin (<https://www.douyin.com/>), people can create simple short videos and use self-generated videos to share interesting something with others, such as new products and nice places; they can also watch others' videos similarly. Though psychological aging theory^[3] suggested that older adults have limited cognitive abilities, such as interacting with online surfaces, middle-aged and elderly users could use short-video social medial apps flexibly due to the ease of them. According to the 2019 China Internet Audiovisual Development Research Report, the utilization rate of short videos by users aged 50 and above has risen to 66.7%. Thus, mobile short-video social media platforms become an important channel for middle-aged and elderly users to participate in society.

It should be noted that the core resource of social media is "traffic". Many studies about short-video reveal the importance of fans for users. For example, reference^[4] shows that fans play a key role in disseminating information in the information age; reference^[5] shows that many sports fans can watch Olympic events online through short videos, taking the opportunity to drive the sports economy. Therefore, to help middle-aged and elderly users better keep up with the trend of the Internet era and participate more deeply in society, it is necessary not only to help them watch short videos but also to help them develop a group of users into their fans on short-video platforms, thereby helping them carry out economic activities while achieving social participation. Previous studies showed that the popularity of the user's posts affects the attention of other users^{[5], [6]}. According to these studies, we may infer that sharing interesting short videos to attract the users and get

* Changyu Wang is the corresponding author and can be contacted at: changyu@jiangnan.edu.cn.

likes are the basic activities the middle-aged or elderly users need to do if they want to get a large number of fans on short-video platforms. However, on short-video platforms most users just view short videos quickly to get fun, even middle-aged or elderly users get many likes through sharing videos, whether these likes can transform into fans is uncertain. Therefore, it is also important for platform managers and middle-aged or elderly users to understand how to improve the transformation of the likes into fans.

Most researches on the number of fans gained by users mainly focus on traditional social media platform situations. For example, reference ^[5] found that the number of Twitters published and the number of other users' likes affect users' attention to the US Olympic governing bodies; reference ^[6] showed the popularity of the user's videos affects the attention of other users. However, less is known about the factors that can promote the link of likes to fans. The research on short-video platforms focuses on exploring the influencing factors of general users using short-video platform^s ^{[1], [7], [8]}, little is known about how to develop a group of fans, especially for middle-aged or elderly video sharer. In the current information system field, scholars have begun to take the special group of middle-aged and elderly users as the research object and carry out related research. However, in the current literature, most studies focus on elderly users' motivations of social media use ^[9] and behaviors after social media use ^{[10]-[12]}, less attention is paid to the following questions: **What factors would influence the number of fans of middle-aged and elderly users on short-video platforms? What factors can strengthen the transformation of the number of likes into the number of fans?**

This study thus tries to take Douyin platform as an example to explore the relationships among the information about middle-aged and elderly users themselves and their videos provided on the platform, the number of likes, and the number of fans. The study may derive the following main contributions. First of all, this paper enriches previous short-video social media research and improves our understanding of the use of social media by middle-aged and elderly users. Secondly, it enriches the deciding factors for other users to become fans of a focal user through the information about users themselves and their videos provided by middle-aged and elderly users on the Douyin platform. Finally, these results can also help middle-aged and elderly users become "middle-aged and elderly Internet celebrities" with more fans, so as to develop a "fan economy".

2. THEORETICAL MODEL AND HYPOTHESES DEVELOPMENT

Taking Douyin platform as an example, this study deduced the determinants of the number of fans of the focal middle-aged or elderly user from the perspectives of the number of likes and self-disclosure information on the user's personal homepage and built a theoretical model framework based on social penetration theory ^[13] and information foraging theory ^[14]. According to social penetration theory ^[15], the self should consist of several layers from peripheral layers (e.g., demographic information) to central layers (e.g., personal values). More importantly, human relationships develop through revealing the self, and this process primarily occurs through self-disclosure. However, information searchers rely on the rule of maximizing self-interest, so it is hard for them to develop final relationships only due to users' self-disclosure. Information foraging theory ^[16] reveals that the relationship between information providers and information searchers in a given environment. Based on information foraging theory ^[14], searching information on the web is similar to searching food for predators. Thus, information foragers often minimize the costs and maximize the speed of finding information. As a mobile short-video social media platform, Douyin provides much online information, where users can acquire knowledge. To better understand the antecedents of fans, we believe that the above theories can be also used in the given environment and provide insight for exploring middle-aged and elderly users' decisions of following users.

2.1 The number of likes

The number of likes refers to the total number of likes for videos a middle-aged or elderly user received from other users after they enter Douyin platform. Following their favorite organizations or users in social media can be seen as a way to express users' interests^[17]. Therefore, the number of likes represents the current likeability of other users to the focal user's videos. According to the information foraging theory^[14], users attempt to maximize the rate of obtaining valuable information and minimize the cost of finding and understanding the gathered information. Thus, they may measure the value of following the focal middle-aged or elderly user to get valuable content in the future based on information cues about existed videos on Douyin platform. In other words, if a user likes the videos provided by the focal middle-aged or elderly user now, he/she may perceive that the focal user would provide valuable videos in the future, thus he/she would follow the focal user. Moreover, researchers can observe how many people have liked the video after watching the videos through TikTok which is the international version of Douyin^[6]. If other users observe that the middle-aged or elderly user has more likes, they would believe that the content of the videos published by the user has a higher likeability, that is, the more likely they would follow the user to watch their videos in the future. Thus,

H1: The number of likes a middle-aged or elderly user receives is positively related to the number of fans. That is, the more the user gets likes, the more the number of fans.

2.2 The users' self-disclosure information on personal homepage

Besides the number of likes about the specific videos, middle-aged or elderly users can disclose information about both themselves and videos on their personal homepages through Douyin platform.

2.2.1 User's own information disclosure

On the top of the Douyin user's personal homepage, there is the user's personal information such as the number of bio words, and verification. The number of bio words refers to the number of words of a personalized signature displayed on the personal homepage by the middle-aged and elderly users themselves. Usually, the bio is presented in the form of text. When users edit their own bio, the more words they make, the more comprehensive self-disclosure other users will feel. In addition, providing bio with more words may also allow other users to perceive the user's effort to provide content services. The verification is a certification that middle-aged and elderly users apply to the official. After meeting the requirements of the platform, the user can apply and obtain the verification. Usually, the verification marks include Associate Chief Physician of Pediatric Pediatrics, Actor, Midwife, etc. Users' verifications. Thus, to a certain extent, it reflects user-related professional information.

Based on the above arguments, in the context of Douyin platform, users would use the number of bio words, and the verification to induce the degree of self-disclosure about the focal middle-aged or elderly user's positive attitudes and professions. According to the social penetration theory, users will decide whether they are willing to establish a relationship with the focal user according to the degree of his/her self-disclosure^[13]. Thus, we proposed that,

H2a-H3a: the number of bio words (H2a) and the verification (H3a) are positively correlated with the number of fans.

According to information foraging theory^[14], information searchers try to maximize the speed of information search. Thus, users on Douyin platform may use the focal middle-aged or elderly user's self-disclosure information (e.g., the number of bio words, and the verification) to confirm their judgments through their liking videos, such as they believe that focal middle-aged or elderly user with more self-disclosure information will provide higher quality video. In other words, the focal middle-aged or elderly user's own information disclosure can inspire other users to change from just liking the user's video to following the user. Accordingly, this article proposes the following hypotheses:

H2b-H3b: the number of bio words (H2b) and the verification (H3b) can strengthen the positive relationship between the number of likes and the number of fans.

2.2.2 User's video information disclosure

On the bottom of the homepage, there is a brief description of the videos added by the focal middle-aged or elderly user, including the number of videos and the professionalism of videos. The number of videos refers to the total amount of published videos after the focal middle-aged or elderly user entering Douyin platform. The more videos a user posts, the more people they may attract. The number of videos can also stand for the strong desire and ability to display videos. The professionalism of videos refers to whether most videos a focal user posted consistent with his/her identity and the contents of all videos are professional and belong to one topic. For example, a middle-aged or elderly user has a user name "Sister Xiaohong's Stories of DeliveryRoom(小红姐的产房故事)", most videos this user shared is knowledge related to midwifery, and other users would believe that the videos published by the user is related to the user's identity and the type of videos is considered as professional. Thus, the professionalism of videos can disclose a user's profession.

In Douyin platform, users would use the number of videos and the professionalism of videos to deduce information disclosure about the focal user's desire to post more videos and to attract others and ability to create professional videos which would encourage them to follow the focal user. Social penetration theory^[13] also believes that the degree of information self-disclosure can affect the degree of relationship establishment. Therefore, we proposed that,

H4a-H5a: the number of videos (H4a) and the professionalism of videos (H5a) are positively correlated with the number of fans.

According to information foraging theory^[14], users would combine all relevant pieces of information to determine whether it is worth following a user. Thus, users on Douyin platform may use the focal middle-aged or elderly user's self-disclosure information (e.g., the number of videos and the professionalism of videos) to confirm their judgments through their liking videos. In other words, the focal middle-aged or elderly user's video information disclosure can inspire other users to change from just like the user's video to follow the user. Accordingly, this article proposes following hypotheses:

H4b-H5b: the number of videos (H4b) and the professionalism of videos (H5b) can strengthen the positive relationship between the number of likes and the number of fans.

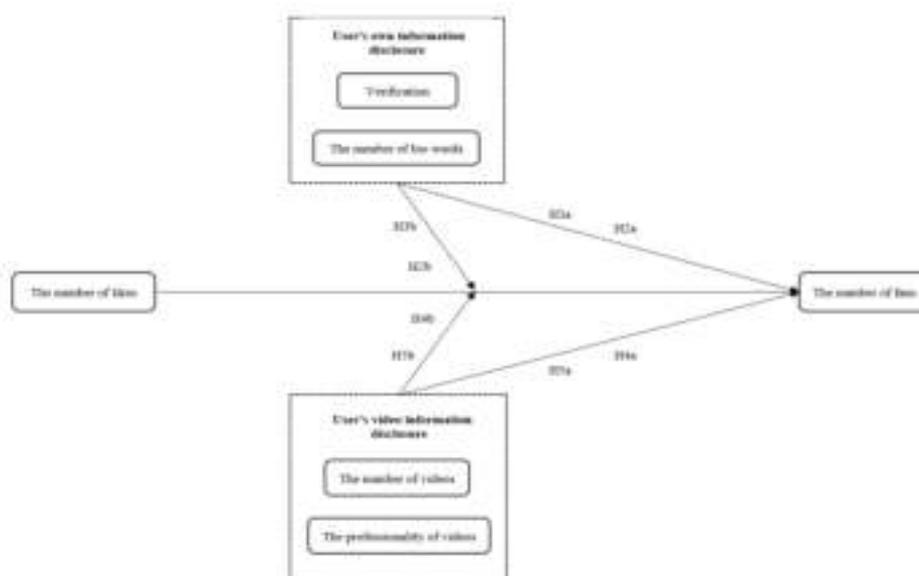


Figure 1. Research model

The research model of this paper is shown in Figure 1. The research object is the number of fans of middle-aged and elderly Douyin users. We also control some variables (such as the user's gender, age, city level, education, tenure of sharing videos, shop entrance, and the number of top videos).

3. METHODOLOGY

3.1 Data collection

The relevant data of this research is obtained from the Douyin platform in China. The APP was launched in 2016. Its main purpose is to help users record their daily life, express themselves and convey personal ideas. Through the platform, more and more middle-aged and elderly users can create and share their short-form videos to gain attention. This study mainly explores the factors affecting the number of fans of middle-aged and elderly users. We selected the sample according to the following requirements: 1) Middle-aged and elderly users are 40 years old or above; 2) Middle-aged or elderly users have a certain degree of activity (the number of fans or the number of likes received by the user reaches 50,000). Based on the above requirements, this study screened the samples. Due to the limited data of middle-aged and elderly users, we first used the keywords (e.g., 爷爷(grandfather; father's father); 奶奶(grandmother, father's mother); 姥姥(grandmother; mother's mother); 姥爷(grandfather; mother's father); 爸(father); 叔(uncle; father's younger brother); 伯(uncle; father's elder brother); 妈(mother); 姨(aunt; mother's sister); 姑(aunt; father's sister); 婶(aunt; wife of one's father's younger brother); 姐(sister); 哥(brother)) that most middle-aged and elderly users used as the user name to search the sample list and 165 middle-aged and elderly users who are 40 years old or above were selected. Then, we screened the personal information displayed on users' homepages to justify the degree of users' activity and only 118 middle-aged and elderly users who have a certain degree of activity were selected as the research sample.

We collected variables for each user at two time periods. The number of likes, the number of bio words, the verification, the number of videos, the professionalism of videos, age, gender, city level, education background, tenure of sharing videos, shop entrance, and the number of top videos were collected at phase 1 (from July 1, 2020, to July 23, 2020); after three months, the number of users' fans were collected at phase 2 (from October 1, 2020, to October 13, 2020).

3.2 Variable measurement

The data obtained from Douyin platform mainly includes numbers and content. The number of likes, the number of bio words, the number of videos, and the number of fans are available on the user's homepage and can be directly collected. The data relating to the content is judged and calculated through text reading and converted into numerical variables by authors. According to the definition of professionalism of videos, authors read each user's personal homepage and videos posted and judged the professionalism of videos each user posted. If the contents of videos are consistent with the user's identity and are professional and unified, the value of the professionalism of videos for the user equals 1, otherwise, it is 0. The verification can be determined by whether the user has platform certification (if the user has it the value is 1 or is 0).

Considering that other characteristics may affect the outcome variables, we included age, gender, educational background, city, tenure of sharing videos, shop entrance, and the number of top videos as control variables. First, the user's age and the number of top videos can be directly obtained from the user's homepage, which can be directly represented by numbers; Second, the user's gender, education, the city, and shop entrance can be directly obtained through the user's homepage (the gender of users is 1 for male, otherwise it is 0; the user's disclosure of education information is 1, otherwise it is 0; the user's city level, using numbers 1-5 to represent first-tier cities to fifth-tier cities; the shop entrance can be determined by whether the user has the channel to store, if the user has it the value is 1 or is 0); Third, the user's tenure of sharing videos refers to the number of days it has been since the user first posted a video.

3.3 Variable results

This study uses ordinary least squares regression for hypothesis testing. While the data show the evidence of overdispersion, we scaled variables to z-standardized values [$z(x) = (x-x)/sd(x)$] before regression analysis. In order to test the moderating effects, this article firstly centered all the independent variables, then multiplied the number of likes with other main variables, calculated the values of the interaction terms (for example, the number of likes * the number of bio words). Independent variables, the interaction terms, and control variables were included in a regression model. Table 1 (Model 1) shows the results of the estimation and significance test. Figure 2 shows the interactions between the users' information disclosure and the number of likes.

Table 1. Estimations results

Variables	Model 1	Hypotheses
Main Variables		
The number of likes	0.494*	H1 Supported
The number of bio words	-0.083	H2a Not supported
Verification	0.075	H3a Not supported
The number of videos	-0.047	H4a Not supported
Professionalism of videos	-0.061	H5a Not supported
Interaction variables		
The number of likes*the number of bio words	0.360*	H2b Supported
The number of likes* verification	0.168*	H3b Supported
The number of likes* the number of videos	0.049	H4b Not supported
The number of likes* professionalism of videos	0.173*	H5b Supported
Control variables		
Age	-0.027	
Gender	-0.041	
Educational background	-0.007	
City	0.066	
Tenure of sharing videos	-0.061	
Shop entrance	0.073	
The number of top videos	0.017	
Notes: * significant at 5%		

According to regression analysis results, we found that the number of likes ($\beta=0.494$, $p<0.05$) is significantly positively correlated with the number of fans, so H1 is supported. The number of bio words ($\beta=-0.083$, $p>0.05$) and verification ($\beta=0.075$, $p>0.05$) are not significantly related to the number of fans. Thus, H2a and H3a are not supported. The number of videos ($\beta=-0.047$, $p>0.05$), and the professionalism of videos ($\beta=-0.061$, $p>0.05$) are not significantly correlated with the number of fans, so there is no evidence supporting H4a, and H5a. Age ($\beta=-0.027$, $p>0.05$), gender ($\beta=-0.041$, $p>0.05$), education ($\beta=-0.007$, $p>0.05$), city level ($\beta=0.066$, $p>0.05$), the first dynamic time ($\beta=-0.061$, $p>0.05$), shop entrance ($\beta=0.073$, $p>0.05$) and the number of top videos ($\beta=0.017$, $p>0.05$) has no significant relationship with the number of fans.

According to moderating effects results as shown in Table 1 and Figure 2, we found that for middle-aged and elderly users with a larger number of bio words, the number of likes is more positively related to the number of fans ($\beta=0.360$, $p<0.05$), which indicates that the number of bio words can strengthen the positive relationship

between the number of likes and the number of fans. Thus, H2b is supported. In addition, for middle-aged and elderly users with verification, the number of likes is positively related to the number of fans ($\beta=0.168, p<0.05$), indicating that the verification can strengthen the positive relationship between the number of likes and the number of fans. Thus, H3b is supported. The number of videos has no significant moderating effect on the relationship between the number of likes and the number of fans ($\beta=0.049, p> 0.05$). Therefore, H4b is not supported. The positive relation of the number of likes to the number of fans is stronger when the contents of videos are more professional ($\beta=0.173, p<0.05$), which supports H5b.

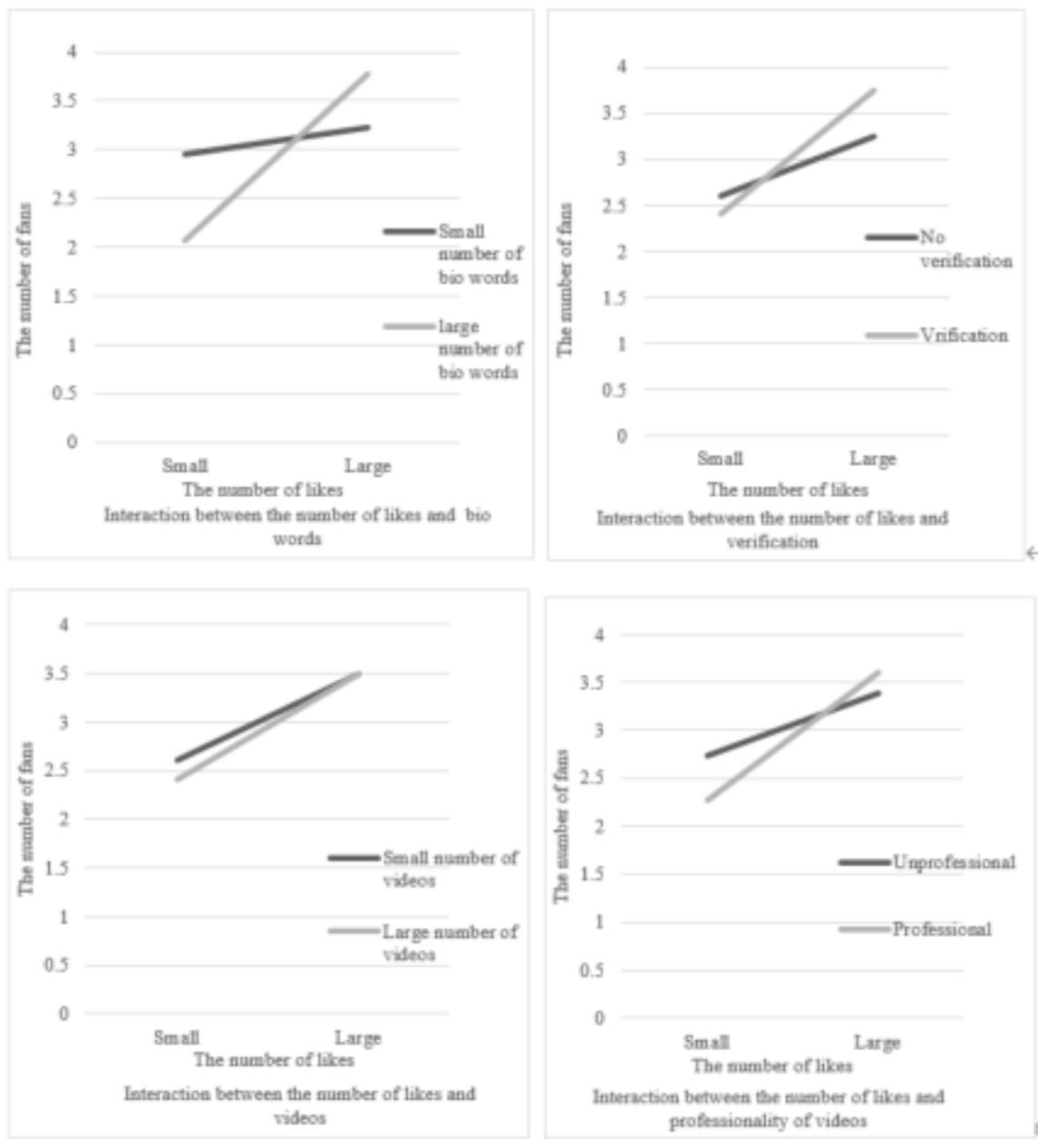


Figure 2. The interactions between the users' information disclosure and the number of likes

4. CONCLUSION AND DISCUSSION

4.1 Main findings

Nowadays, more and more scholars focused on research in social media scenarios. However, for

middle-aged and elderly groups, little empirical research is on the antecedents of the number of fans. Therefore, we took Douyin as an example and answered the following question based on social penetration theory and information foraging theory: on short-video social media platforms, what factors would influence the number of fans of middle-aged and elderly users? what factors can strengthen the transformation of the number of likes into the number of fans? Empirical results show that the number of likes received by middle-aged and elderly users was positively related to the number of fans they developed. That is to say, the more likes the user received, the more fans they will get. Besides, compared with users who don't have verification, likes are more likely to transform into fans for users who have verification. Also, the number of bio words and the professionalism of videos can strengthen the positive relationship between the number of likes and the number of fans. Detailed discussions about our empirical findings are shown as follows.

The number of likes received by middle-aged and elderly users is positively related to the number of fans. The result is consistent with a simple decision-making rule that “people generally recognize those who preferred by the public”^[18], which indicates that focal users with more likes will stimulate other users to follow them, then get more fans. This provides evidence that other users are concerned about the number of likes the focal user received for videos on Douyin.

The number of bio words, as one of the user's own information disclosure, is not significantly related to the number of fans. A possible explanation is that middle-aged and elderly people have already formed their own social circle in real life, only judging the number of bio words is not obvious for building new connections. Therefore, the number of bio words would not increase the number of fans on Douyin. Though the number of bio words is not significantly related to the number of fans, it is proved to strengthen the positive effect of the number of likes on the number of fans as we hypothesized. Even reference^[19] believes that people can re-select and shape their identity in virtual cyberspace, middle-aged and elderly users can re-shape their identity through detailed bio. Users on Douyin would confirm that the number of likes can stand for the number of fans when middle-aged and elderly users provide a more detailed bio, thus they would follow middle-aged and elderly users with more likes.

The verification, as another user's own information disclosure, does not influence other users' decision-making on following the middle-aged and elderly users. One possible explanation is that although focal users have a verification, other users may not need related knowledge, for example, users with the “midwife” verification are unlikely to directly gain the attention of the youth and the old. However, the verification can strengthen the relationship between the number of likes and the number of fans as we hypothesized. Users with verification will be appropriately promoted by Douyin to increase the popularity of their videos, for example, their user name will be more likely to be in the front when being searched, and videos of focal users appear more frequently, resulting in curiosity and interests of other users, then will strengthen the relationship between the number of likes and the number of fans.

The number of videos, as a user's video information disclosure, neither affects the number of fans of the users directly nor strengthens the relationship between the number of likes and the number of fans, which contrary to the research about fans in Twitter^[5]. One possible explanation is that the quantitative index does not represent the quality of the videos, so it does not influence the number of fans. For example, someone may share several moments every day, resulting in the tiredness of other users, rather than freshness at first. Though it is easy to operate this APP, it still takes a lot of time and effort to produce high-quality videos, such as adding popular background music, text, special effect, and so on. Therefore, the number of videos doesn't influence the number of fans directly or indirectly.

The professionalism of videos, as another information disclosure of user's videos, is not significantly related to the number of fans. A possible explanation is that the amount of users on Douyin is huge, and their purpose of

using it is also different, so the professionalism of videos does not directly affect the number of fans. However, it can indirectly affect the number of fans by strengthening the relationship between the number of likes and the number of fans. The possible explanation is that Douyin App is mostly designed for young people, but it allows middle-aged and elderly people to enhance the opportunities to “touch the net” and participate in society. While middle-aged and elderly users appear, it may trigger other users to feel a sense of freshness. Especially when the content of videos is professional, the positive relationship between the number of likes and the number of fans is higher.

4.2 Theoretical contributions

Our findings would present several theoretical contributions to social media literature. First of all, the existing social media research on the number of fans can be supplemented by introducing Douyin platform as a new type of social media platform. This study regards the way that the middle-aged and elderly people gain more fans on Douyin as a “fan economy” and explores antecedents of the number of fans. In addition, though a few studies focused on short-video platforms [20], [5]-[8], little is about the middle-aged and elderly users [10]-[12]. This study would enrich the insufficiency of the previous literature on short-video social media such as Douyin and advance our understandings of middle-aged and elderly users online.

Secondly, this study enriches the deciding factors for other users to become fans of a focal user through the self-disclosure information about users themselves and their videos provided by middle-aged and elderly users on the Douyin platform. While most previous studies focused on that the traits of the user’s videos affect the popularity of users’ videos [5], [6], this study can supplement previous research through self-disclosure information of both the focal users themselves and their videos. Moreover, we have found that on Douyin, the self-disclosure information has no influence on the number of fans, which contraries to previous studies about social penetration theory [15].

Finally, through examining the interaction effects of users’ self-disclosure information and the number of likes, this study enriches the literature about the impact of self-disclosure information and the number of likes on the number of fans. Different from what has been found in previous studies [19], [20], this study firstly explores the deciding factors of fans from the perspective of self-disclosure information. In the context of Douyin platform, the number of bio words, the professionalism of videos, and the verification can increase the positive effect of the number of likes.

4.3 Practical implications

These findings can provide focal users and platform managers with some practical implications. Firstly, the number of bio words can strengthen the positive effect of the number of likes in this study. The result shows that, if the middle-aged and elderly users want to acquire more fans, they need to edit more bio words as possible as they can, so that this information could arouse users’ behavior of following focal users. For platform operators, they should give full play to their regulatory responsibilities to ensure the authenticity of focal users’ information.

Secondly, verification can also strengthen the positive effect of the number of likes on the number of fans. This result shows that middle-aged and elderly users should continuously increase the degree of self-disclosure by applying for platform verification. For platform operators, they should provide the chance of applying the verification for focal users, as long as focal users meet hard requirements.

Thirdly, the professionalism of videos can strengthen the positive effect of the number of likes on the number of fans. This result shows that middle-aged and elderly users should create more professional videos by expertise or tools of editing their works. It’s noted that they should emphasize the quality of videos rather than the quantity. For platform operators, they should play the role of administrators to monitor this platform and resist the publication of vulgar videos.

Finally, platform operators can use the research results to transform the middle-aged and elderly people who have already settled on the platform into an “elderly internet celebrity” with a large number of fans, so as to develop an “Internet celebrity economy”. The success of the internet celebrity economy mainly depends on the fans. Platform operators also utilize the fans to the realization of Internet economy so that they can achieve a win-win between the platforms and users.

4.4 Limitations and future studies

Though this research has some implications, we still have some limitations and plan to solve these problems in the future. First, it is not certain whether the model based on Douyin platform is applicable to other social media platforms and needs to be verified in future. Second, improvements can be made in the data collecting process. For example, we can use Python to crawl data to avoid the small sample size. Thirdly, whether these findings hold true for young Douyin users will be tested. Finally, we can use mixed-methods to verify our findings.

ACKNOWLEDGEMENT

This work was supported in part by National Natural Science Foundation of China under Grants 71901106 and the Ministry of Education of Humanities and Social Science under Grant 18YJC630169.

REFERENCES

- [1] Talaei-Khoei A., Daniel J. (2018). How younger elderly realize usefulness of cognitive training video games to maintain their independent living. *International Journal of Information Management*, 42: 1-12.
- [2] Pan S.. (2017). Examining the effects of social networks formed in a senior-oriented online community on older participants' subjective well-being in China. *Chinese Journal of Communication*, 11: 135-154.
- [3] Kurniawan S. (2008). Older people and mobile phones: a multi-method investigation. *International Journal of Human - Computer Studies*, 66(12): 889-901.
- [4] Jensen K B, Rothenbuhler E W, Pooley J D, Craig R T. (2016). The international encyclopedia of communication theory and philosophy. *References Reviews*, 31(6): 17-18.
- [5] Bo L, Scott O, Dittmore S W. (2018). Twitter and olympics: exploring factors which impact fans following american olympic governing bodies. *International Journal of Sports Marketing and Sponsorship*, 19(4): 370-383
- [6] Zhu C, Xu X, Zhang W, Chen J, Evans R. (2020). How health communication via TikTok makes a difference: a content analysis of TikTok accounts run by Chinese provincial health committees. *International Journal of Environmental Research and Public Health*, 17(1): 192-204.
- [7] Wang Y. (2020). Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of TikTok (Douyin). *Computers in Human Behavior*, 110: 1-13.
- [8] Chen Y. (2020). The status quo, characteristics and application strategies of TikTok in public hospitals. *Jiangsu Health Service Management*, 192(08):123-125.
- [9] SCH Corrêa, Soares J L, Christino J, Gosling M, Goncalves C A. (2020). The influence of YouTubers on followers' use intention. *Journal of Research in Interactive Marketing*, 14(2): 173-194.
- [10] Todd W. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, 16(1): 85-102.
- [11] Shang L, Zhou J, Zuo M. (2020). Understanding older adults' intention to share health information on social media: the role of health belief and information processing. *Internet Research*, 31(1): 100-122.
- [12] Teerakapibal S, Melanthiou Y. (2019). The new helping the old: social media as a facilitator for variety seeking in food choices of the grey population. *British Food Journal*, 122(1): 272-290.

- [13] Kim J, Song H. (2016). Celebrity's self-disclosure on Twitter and parasocial relationships. *Computers in Human Behavior*, 62: 570-577.
- [14] Pirolli P, Stuart C. (1999). Information foraging. *Psychological Review*, 106(4): 643–675
- [15] Altman I, Taylor D A. (1973). Social penetration: The development of interpersonal relationships. *Journal of Social Psychology*, 75(1): 79-90.
- [16] Petter S, Straub D W, Rai A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 31(4): 623-656.
- [17] Watanabe N, Yan G, Soebbing B P. (2015). Major league baseball and Twitter usage: the economics of social media use. *Journal of Sport Management*, 29(6): 619-632.
- [18] Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality & Social Psychology*, 39(5): 752-766.
- [19] Jeng W, Desautels S, He D, et al. (2017). Information Exchange on an Academic Social Networking Site: A Multi-discipline Comparison on ResearchGate Q&A. *Journal of the Association for Information Science and Technology*, 68(3): 638-652.
- [20] Zhao J, Wang J. (2020). Health advertising on short-video social media: a study on user attitudes based on the extended technology acceptance model. *International Journal of Environmental Research and Public Health*, 17(5): 1501-1521

Full Research Paper

An Analysis of the Characteristics and Evolution of China's Pension Policy Based on Text Mining

Liang Yang¹, Meng Zhao¹, Miyan Liao¹, Liangqiang Li^{1*}, Xiaolin Li¹

¹Business and Tourism School, Sichuan Agricultural University, China

Abstract: With the economic and social development, China has changed from the initial demographic dividend to an aging population, and governments at all levels have introduced a series of policies to cope with the negative effects caused by the aging population. This paper collects a total of 167 pension policy documents from 2010 to 2020, and uses text mining techniques to systematically study the characteristics and evolutionary trends of pension policies in terms of issuing agencies and contents. The conclusions drawn in this paper are as follows: First, the Ministry of Civil Affairs and other departments play a major role in the formulation of pension policies. Second, the pension policies formulated by the government during the study cycle are biased toward public service aspects, and the content of medical care, investment, and pensions is mainly indirectly focused on a cyclical basis. The findings of this paper can help the government to check the gaps and improve the existing policy system, and the research methodology can help the government to make more scientific decisions.

Keywords: Pension policy, Policy characteristics, Text mining

1. INTRODUCTION

With the continuous development of China's economy and society, and the continuous improvement of medical and health conditions, the issue of population aging is becoming an increasingly prominent issue in social governance. There are three main factors influencing population aging: declining fertility rate, longer life expectancy, and a large number of people entering old age. According to the data released by the National Bureau of Statistics, the number of elderly people aged 60 and above in China reached 254 million at the end of 2019, accounting for 18.1% of the total population, and the number of elderly people aged 65 and above reached 176 million, accounting for 12.6% of the total population, representing a significant increase compared to 7% in 2000. Along with the growth of urban and rural residents' income level, residents' willingness to have children is getting lower and lower, and the trend of population aging in China may get further accelerated.

The rapid growth of population aging may bring a series of social problems in the background. Firstly, the growth of the number of elderly population puts forward higher requirements for local services in medical care, maintenance, welfare, transportation, etc. Secondly, the decrease in the proportion of young and middle-aged population makes the number of labor force decrease, and the aging of the group presents more difficulties for social development^[1]. The increase of population aging makes the burden of society and family increase, and how to solve the elderly problem has become a high concern of the society. Meeting the demand for elderly services is the core and key link of the national strategy of actively corresponding to the aging population, as well as an important part of social governance^[2]. In addition, due to the weak awareness of family planning and uneven government investment, the problem of rural retirement is becoming increasingly prominent, and there is an extreme imbalance between urban and rural retirement resources^[3].

Government departments have played a key leading role in actively responding to the social aging problem. As early as 1980, the Ministry of Justice issued the "Ministry of Justice Notary Lawyer Department on the Issuance of Unclaimed Pension Certificates" to focus on pension-related matters. Since the 18th National

* Corresponding author. Email: lilq@sicau.edu.cn

Congress, national ministries and commissions have issued successive policies from different aspects such as finance, medical care and scientific innovation to focus on the development of the elderly and the construction of the pension system. 2019, the General Office of the State Council issued "Opinions on Promoting the Development of Pension Services" systematically proposed 28 measures to systematically solve pension problems, and the pension service policy has achieved a breakthrough in its entirety. For a long time, government departments have mostly formulated domain-related policies from their own perspectives, which eventually led to a segmented and fragmented political system^[4]. Although a series of policies have been introduced by various ministries to solve the pension problems, the characteristics presented by the existing pension policies, the deficiencies of the existing policies, and the characteristics of the evolutionary trends of the policies over time are still half understood. Based on this, this paper uses text mining techniques to dig into the characteristics of existing pension policies and their dynamic evolution patterns, explore the shortcomings of existing pension policies, and provide suggestions for government policy formulation.

2. PENSION POLICY RESEARCH

Rapid changes in economic conditions and demographic characteristics require that policy design be closely aligned with reality and adjusted in a timely manner in order to advance the development of senior care. The Chinese government foresees an old-age problem arising from the one-child policy in 2020, but is overly optimistic about the outcome and solutions, believing that the problem will be solved through economic development, improved living conditions, and increased social welfare and security^[3]. The reality is that China's social pension problem has far exceeded the government's expectations. In recent years, the Chinese government has introduced a series of policies and established a universal social security for the elderly and a social care system for the elderly^[5].

Government documents dealing with elder care can be considered as belonging to a continuously evolving public policy narrative, and existing studies have explored existing policies in terms of their characteristics^[6]. Zhai et al. argue that the aging of China's population now shows new characteristics such as a significant slowdown in the rate of aging, an increase in the size of empty nesters living alone, and an increase in the literacy level of the elderly, and in this context, the existing elderly care policies need to be updated to promote the development of the elderly care industry^[7]. Wang studied the evolution path from the perspective of financial security of elderly services, and found that the financial security responsibility of the existing elderly services showed the characteristics of "less welfare - lack of financial security responsibility - return of financial security responsibility - multiple government, market, society and family Governance"^[8]. Yi et al. systematically analyzed the current status of social pensions in China and the existing dilemmas, and they argued that the key factors hindering the sustainable development of social pensions are reflected in two aspects: first, there is no policy to clarify the nature of services and support allocation of relevant resources: second, social pension policies are formulated in a stylized manner rather than based on actual interactive processes^[9].

Although many studies have used pension policies as an object to actively suggest suggestions for governmental decision-making, most previous studies are qualitative in nature and few have used quantitative studies to explore the characteristics of pension policies and their evolutionary features. Based on this, this paper intends to use text mining to explore the characteristics of existing pension policies and the analysis of policy evolution, and the findings can help the government make scientific decisions.

3. STUDY DESIGN

3.1 Data source

In order to ensure the accuracy, relevance and representativeness of the policy text data, this study took the

period from January 2010 to January 2021 as the time interval, and selected the relevant pension policy texts issued by the State Council, the Ministry of Civil Affairs, the Ministry of Finance, the National Health and Welfare Commission, the National Health Insurance Administration, the Ministry of Human Resources and Social Security, and the Ministry of Housing and Urban-Rural Development as the data base. When collecting the data, we followed the principles of authority, openness and relevance, and searched the legal database of Peking University with the title of "pension". After screening the policy texts one by one according to the policy content, excluding the policy texts, approvals, activity notices and other policy texts that have become invalid or partially invalid, we finally sorted out a total of 167 valid policy documents, and the specific adoption situation is shown in Table 1.

Table 1. Raw data results and adoption

Policy Category	Number of Post	Number of Adoption
Administrative Laws and Regulations	17	16
Judicial Interpretation	1	0
Regulation	791	149
The Party's Rules and Regulations	2	2
Industry Regulation	8	0

3.2 Research methods and analytical tools

For policy text research, this paper wants to adopt bibliometric and text mining methods.

(1) Bibliometrics

Bibliometrics takes literature and intelligence as the object of study, and uses mathematics, statistics, intelligence and other disciplines to investigate its structural characteristics, quantitative characteristics, and the laws of change, and to further investigate certain structures and characteristics of science and technology^[10]. Generally, the results of quantitative analysis of literature using bibliometric methods are clear and lucid, so in this paper, we use bibliometric methods to analyze the external characteristics of current policies on new coronary pneumonia, and describe and evaluate the existing policies through two levels: chronological order and the subject of issuance.

In conducting the analysis of issuing subjects, the visualization is presented through social network analysis. A social network is a network structure composed of multiple actors, and social network analysis is used to show the connections among actors in the form of a social network diagram^[11]. In this paper, we study the collaboration between departments by counting the joint and independent postings of departments and importing the data into Gephi software for visualization.

(2) Text mining

Text mining, or text data mining, refers to the pattern of mining valuable information from a large collection of unstructured texts and can be considered as an extension of data mining or knowledge discovery from structured databases^{[12] [13]}. Text mining techniques enable the discovery of potential connections, patterns, etc., and the extraction of valuable information from a large amount of unstructured text. In the context of this paper, text mining is used to perform cluster analysis, classify different policies, and explore the characteristics of existing policies in terms of content structure. The specific process of text mining in this paper is shown in Figure 1.

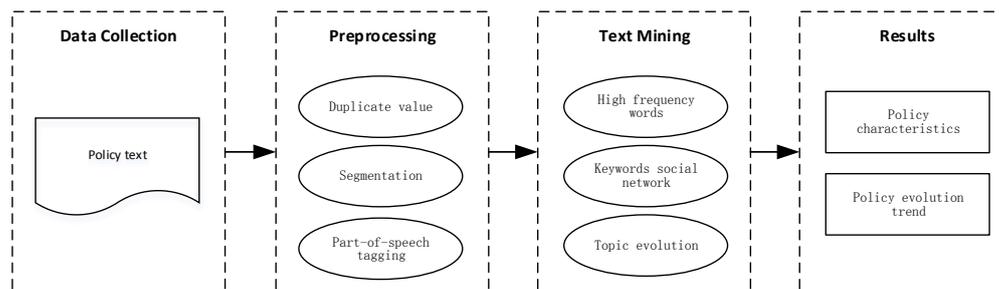


Figure 1. Text mining flowchart

3.3 Keyword extraction

Keywords are words that can maximize the information on the topic of the text, and keyword extraction can summarize the text information briefly [14]. In the research environment of this paper, keyword extraction can grasp the topic of the policy text to the maximum extent quickly and provide a basis for content analysis.

In this paper, the TF-IDF algorithm is used for keyword extraction, and the first part TF (Term Frequency) indicates the frequency of words appearing in a text, which is calculated as (1).

$$TF = \frac{tf}{doc_length} \quad (1)$$

The second part, IDF (Inverse Document Frequency), is used to measure the general importance of words and is calculated as (2).

$$IDF = \log\left(\frac{N}{df}\right) + 1 \quad (2)$$

where, tf is the number of keyword appearances, doc_length is the length of the text, N is the total number of text, df is the number of text containing keywords. The final TF-IDF is the product of TF and IDF.

3.4 Latent dirichlet allocation

The Latent Dirichlet distribution, abbreviated as LDA, is a generative probabilistic model for discrete data collections such as text corpora, which assigns topics to documents and generates topic distributions over the words of a given text collection [15]. LDA contains three main layers of structure: words, topics, and documents, as shown in Figure 2.

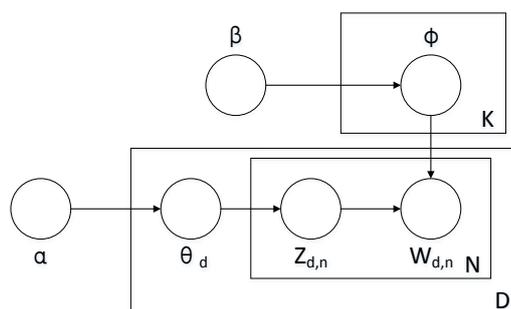


Figure 2. LDA structure

Next we briefly describe the process of LDA generation of documents. First, a document d_i is selected according to the prior probability $p(d_i)$; second, the topic distribution θ_i of document d_i is generated by sampling from the Dirichlet distribution α , and the topic distribution is generated by the Dirichlet distribution with hyperparameter α ; third, the topic $z_{i,j}$ of the j th word of document d_i is generated by sampling from the

polynomial distribution θ_i of the topic; fourth, the topic $z_{i,j}$ of the Dirichlet distribution β to generate the word distribution $\phi_{z_{i,j}}$ corresponding to the topic $z_{i,j}$, the word distribution $\phi_{z_{i,j}}$ is generated by the Dirichlet distribution with parameter β ; finally, the word $\omega_{i,j}$ is sampled from the polynomial distribution $\phi_{z_{i,j}}$ of words to finally generate the word $\omega_{i,j}$.

4 QUANTITATIVE ANALYSIS OF PENSION POLICIES

4.1 Number of posts

The last ten years of pension policy releases are shown in Figure 3. According to the table, it is found that the number of pension policies peaked in 2014, the second highest number of pension policies in 2019, and the lowest number of pension policies in 2011.

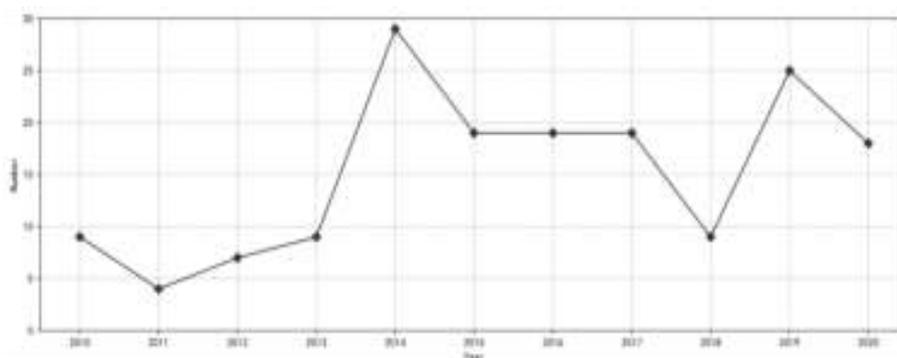


Figure 3. Pension policy year distribution

4.2 Institutional characteristics

Based on the collected 167 policy texts, this paper made statistics on the number of policy issuing institutions. A total of 59 departments have formulated and promulgated policies regarding the pension issue. The top 10 institutions with issued policies are shown in Table 2. It can be found that the Ministry of Civil Affairs solves the pension problem by formulating a variety of relief and welfare policies. It plays the role of the founder in maintaining social stability, improving the happiness of the elderly and promoting economic development. In addition, the Ministry of Human Resources and Social Security, the Ministry of Finance, the State Council and other departments have further improved pension policies from different perspectives in order to meet the needs of diversified and multi-level pension services for the elderly.

Table 2. Number of institutional policy releases

Institution	Number	Institution	Number
Ministry of Civil Affairs (民政部)	78	Insurance Regulatory Commission (保监会)	11
Ministry of Human Resources and Social Security (人力资源与社会保障部)	49	National Health and Wellness Commission (国家卫生健康委)	10
Ministry of Finance (财政部)	45	National Health and Family Planning Commission (国家卫生计生委)	8
State Council (国务院)	18	Ministry of Commerce (商务部)	6
National Development and Reform Commission (国家发改委)	18	Ministry of Housing and Urban-Rural Development (住房和城乡建设部)	5

We use Gephi software to describe the social network between sectors. The joint issuance of pension policy departments is shown in Figure 4. The size of the network nodes reflects the number of documents issued by the

departments, and the thickness of the connecting lines between the nodes reflects the closeness of the departments. According to the network diagram, we can find that the Ministry of Civil Affairs has the largest node in the network diagram, and at the same time, the number of nodes connected to it is the most numerous and complex, which indicates that the Ministry of Civil Affairs has issued the largest number of policies in the research cycle of this paper, and at the same time, the policy cooperation with other departments is also the most frequent. However, in terms of the complexity of the relationships of the nodes, the Ministry of Finance, the Ministry of Land and Resources, the Ministry of Housing and Urban-Rural Development, and the Ministry of Human Resources and Social Security jointly issued relatively more policies, which played an important role in the solution of the pension problem. Finally, the National Bureau of Statistics, the National People's Committee, and the National Open Bank have issued relatively few documents, but they have actively cooperated with other departments and jointly issued policies, which have played an active role in promoting the development of China's pensions.

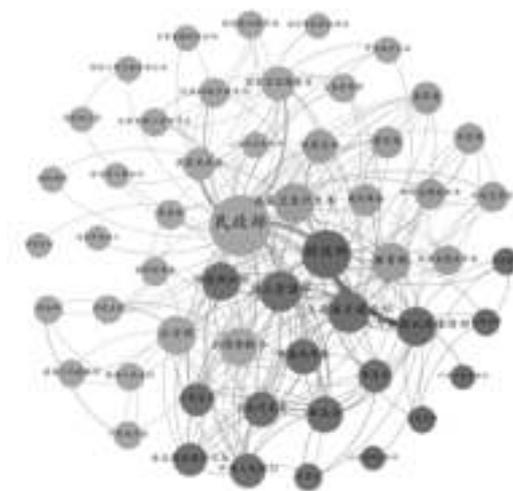


Figure 4. Departments jointly issued network

5 PENSION POLICY CONTENT FEATURES

5.1 High frequency word analysis

In this paper, we use the jieba library of Python to carry out the word segmentation work, and use the deactivated word list of HIT and the deactivated word bank of Machine Intelligence Laboratory of Sichuan University to get a total of 1897 deactivated words. A total of 167 policy texts were divided into words using the precise mode of jieba library, and according to the research purpose of this paper, only verbs as well as nouns were retained, and a total of 132,160 words and 6,456 dissimilar words were obtained. According to the formula of high and low frequency words proposed by Qinglan Sun: $D = \sqrt{n}$. (D indicating the critical value of high and low frequency words, n indicating the total number of dissimilar words) [16], the threshold of high and low frequency words in this paper was calculated as 80. Combined with the actual situation, the top 60 high-frequency words are finally selected, as shown in Table 3.

High-frequency words can reflect the focus of a government's policy over time. It can be seen from the high-frequency words that "pension" appears the most frequently, which corresponds to the theme of this study. "Government", "business unit", "the Ministry of Civil Affairs", "agency", "personnel", "enterprise", "department", "organization" and other subjects related to the high frequency words reflects the Chinese government attaches great importance to the elderly pension problem, governments at all levels not only play a good role itself, and promote enterprise pension services better, which is beneficial to improve the people's feeling, happiness and security. The two high-frequency words "senior citizens" and "urban and rural residents" reflect the targets of the

pension policy. The four high-frequency keywords of "society", "region", "community" and "country" reflect the geographical scope that the pension policies formulated by government departments at all levels are adapted to. High-frequency words such as "service", "development", "management", "work", "construction", "support", "establishment", "system", "guarantee", "reform" and "policy" reflect the action arrangements made by government departments to promote the government and enterprises and institutions to improve the quality of old-age service and promote the healthy development of old-age service. "Endowment insurance", "investment", "join", "funds", "offer", "payment", "product", "information", "pension" and other high frequency words reflect the government to ensure that the elderly pension life, in terms of money, products, information services such as content to make detailed regulations, safeguard the rights and interests of the elderly and the life.

Table 3. High frequency words (top 60)

Number	Word	Frequency	Number	Word	Frequency	Number	Word	Frequency
1	Pension Policy	5690	21	Payment	640	41	Unit	434
2	Service	3339	22	System	626	42	Household	433
3	Institution	2745	23	Pilot	620	43	Way	427
4	Endowment Insurance	1548	24	Guarantee	596	44	Treatment	413
5	Elderly	1423	25	Reform	588	45	Service Facility	404
6	Development	1185	26	Department	544	46	Boost	403
7	Management	1178	27	Project	544	47	Business	390
8	Work	1155	28	Region	531	48	Life Insurance	381
9	Construction	1055	29	Policy	530	49	Complete	375
10	Investment	990	30	Government	523	50	Public Institution	372
11	Society	879	31	Situation	523	51	Transfer	366
12	Personnel	856	32	Product	523	52	Condition	364
13	Insured	799	33	Information	501	53	Service Industry	360
14	Community	791	34	Capital	498	54	Urban-Rural Residents	347
15	Enterprise	766	35	Pension	490	55	Ministry of Civil Affairs	341
16	Support	736	36	Standard	481	56	Make	338
17	Correlation	723	37	Organization	472	57	Accomplish	337
18	Fund	711	38	Handle	449	58	Office	328
19	Provide	707	39	Implement	441	59	Country	327
20	Establish	645	40	Encourage	436	60	Mechanism	323

5.2 Keywords social network

In this paper, we use the TF-IDF method to calculate the word weights in 167 policy texts, take the top 60 keywords in descending order of weights, and generate a keyword co-occurrence matrix, filter the nodes with co-occurrence times lower than 60, and finally use Gephi software to generate a keyword social network graph, and the results are shown in Figure 5, which includes a total of 44 nodes and 1252 edges.

In the network graph of this study, the stronger the intermediary centrality of the nodes, the stronger their role. The results show that the intermediary centrality of "work" is above 90, and it has the largest intermediary centrality, and it plays the most important role in the whole network graph. This work mainly refers to the work activities related to the elderly, which corresponds to the theme of this study. "related", "people", "establishment", "elderly", "development", "institutions", "services", "organization", "management", "policy", "construction" and other nodes have a large degree of intermediary centrality, indicating that in order to improve the quality of elderly care and elderly services, relevant institutions and organizations strengthen management and improve services according to elderly care policies are important elements throughout most of the elderly care. The nodes "human resources", "construction" and "construction" have a high intermediary center. The keywords "human resources", "pension insurance", "project", "home", "ten points" and "ten points" are the most important elements of the policy.

Keywords such as "ten points" are at the margins of the whole study, with a small number of interconnections, and these are not commonly concerned in elderly care policies.

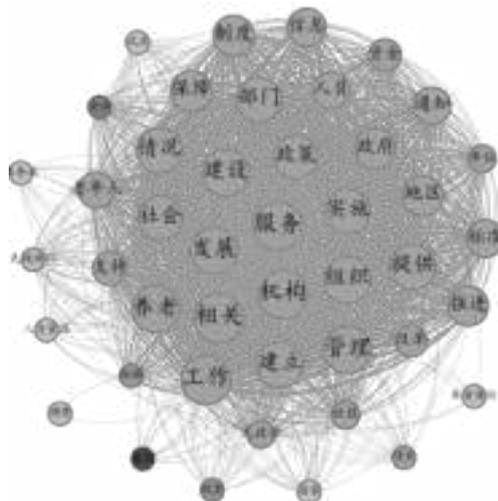


Figure 5. Keywords social network

5.3 Policy evolution analysis

In this paper, we use the topic probability model to cluster the policy text content, and according to Figure 6, the coherence score is highest when the number of topics is 6. Therefore, the policy text is divided into six categories, and 10 topic words are taken under each category, and the results of clustering are shown in Table 4^[17]. The first policy category was named as retirement policy based on the key words of funding, community, subsidy, and central finance. The second policy category was named as insurance-type policies based on key words such as pension insurance, participation, contribution, treatment, and institutions. The third policy category is named as investment category policy according to key words such as investment, management, product, fund, and asset. The fourth policy category is named as epidemic prevention and control policy according to key words such as prevention and control, epidemic, area, staff, disinfection, and isolation. Based on the key words of service, institution, development, elderly, and service facilities, the fifth category of policies was named as public service category policies. Based on the key words of life insurance, support, indicator, medical care, and life insurance, the sixth category of policies was named as medical care category policies.

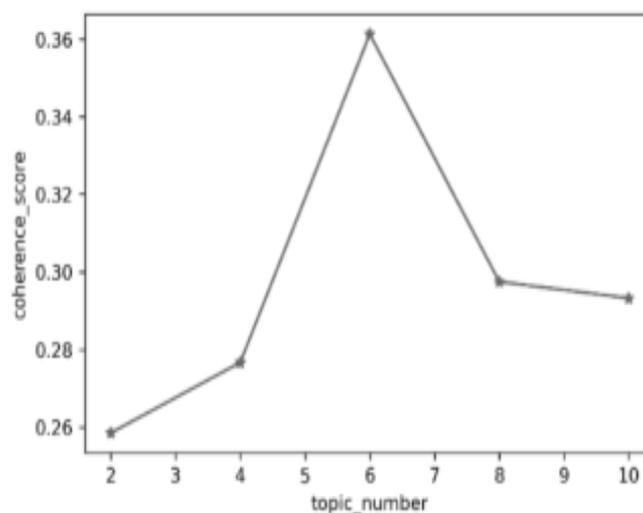


Figure 6. Coherence score

Table 4. Policy topic classification

First-level Topic	Retirement	Probability	Insurance	Probability	Investment	Probability	Epidemic prevention and control	Probability	Public Service	Probability	Medical	Probability
Second-level Topic	Capital	0.027	Endowment Insurance	0.045	Investment	0.047	Prevention and Control	0.016	Service	0.047	Life Insurance	0.017
	Community	0.025	Insured	0.027	Management	0.045	Epidemic	0.014	Institution	0.025	Support	0.009
	Subsidies	0.025	Payment	0.021	Product	0.031	Region	0.013	Development	0.017	Indicator	0.008
	Military	0.014	Treatment	0.014	Funds	0.025	Staff	0.011	Elderly	0.017	Development	0.008
	Central Finance	0.011	Public Institution	0.012	Assets	0.017	Disinfection	0.009	Society	0.012	Type	0.007
	Retirement	0.01	Transfer	0.012	Business	0.011	Isolation	0.008	Construction	0.012	Medical	0.007
	Investment	0.009	Urban-Rural Residents	0.011	Escrow	0.009	Pilot work	0.007	Reformation	0.007	Premium	0.007
	Retirement	0.009	Personal Account	0.009	Account	0.008	Region	0.007	Government	0.006	Group	0.007
	Pension	0.009	Pickup	0.009	Combination	0.007	Civil Affairs	0.007	Service Facilities	0.006	Personal Insurance	0.006
	Ministry of Finance	0.008	Social Insurance	0.008	Risk	0.007	Pneumonia	0.006	Service Industry	0.006	Product Design	0.006

Based on the results of the clustering of the thematic probability model, the frequency of occurrence of policy contents in different categories in that year is counted on a yearly basis, and the results are normalized in order to compare the results between different categories in the time dimension, and the normalization formula is shown as (3).

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where, x_{min} denotes the minimum value and x_{max} denotes the maximum value.

The policies promulgated by the government often reflect the focus of the whole society at one stage. Based on the evolution of the focus of the policy content, we can find out the patterns, shortcomings and future trends of government policy making, and the evolution trend of the pension policy from 2010 to 2020 is shown in Figure 7.

First of all, from an overall perspective, the number of public service policy contents has been maintained at a high level for most of the time, and as the number of elderly people continues to grow, the number of elderly institutions, elderly activity places, and elderly service nursing staff is stretched to the limit, making it difficult for the needs of the elderly to be met. In 2011, the General Office of the State Council issued the Notice on the Issuance of the Construction Plan for Social Aged Care Service System (2011-2015), which systematically provided guidance for the socialization of public services for the aged as well as the industrialization of a phase of development.

Secondly, the number of epidemic prevention and control type policies increased sharply in 2020 due to the impact of the New Crown Pneumonia epidemic, and government policies from this time onwards focused on restoring elderly services in terms of epidemic services to ensure the life and health safety of the elderly and to

better serve their needs.

Finally, some of the thematic contents, although the overall number is small, the government policy will focus on the topic within a certain period. For insurance policies, some policies have been formulated for targeted solutions in response to the current low level of protection for the elderly and the temporary uncertainty of the treatment mechanism. 2010 and 2014 saw certain increases in related policies, and in 2018, the Ministry of Human Resources and Social Security and the Ministry of Finance formulated the "Guidance on the Establishment of a Normal Adjustment Mechanism for the Treatment of Basic Pension Insurance and Basic Pension for Urban and Rural Residents", the policy is closely focused on the establishment of a guaranteed urban and rural residents' pension insurance as well as pension mechanism. For retirement policies, according to the trend of policy evolution, government policies will mention retirement-related matters for the elderly within 2-3 years. For investment-type policies, in order to strengthen the management of assets for the elderly, the government has indirectly developed some initiatives to address such issues, with a high number of such policies in 2013, 2015, and 2018, with the central government giving full play to its role as a demonstration driver for budgetary investment. For medical policies, the current limited medical resources and independent service system are difficult to meet the medical needs of the elderly, and the government rarely formulates relevant policies directly; instead, comprehensive policies will be mentioned from the medical perspective, and such policy measures still need to be improved.

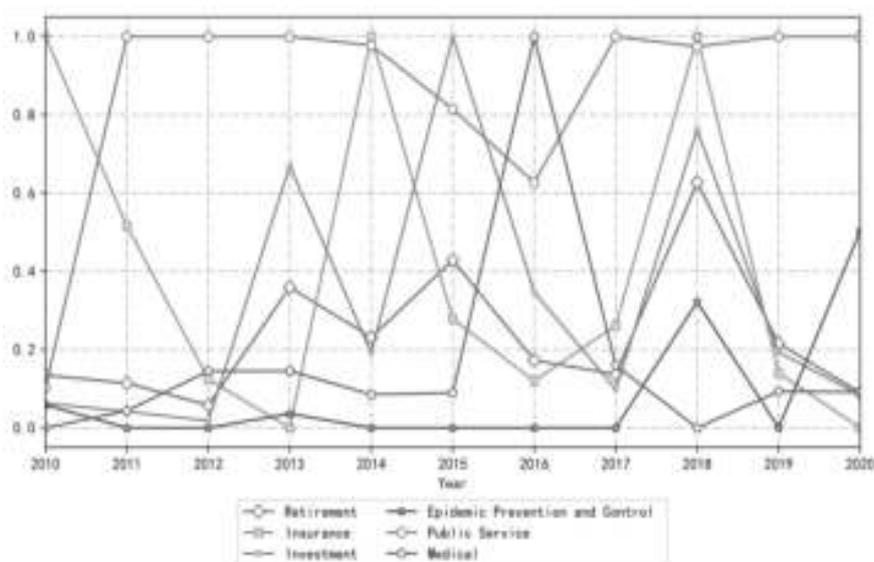


Figure 7. Policy evolution trends

6 CONCLUSIONS

In the face of the rapidly growing trend of aging, the government should implement an active aging strategy and formulate an aging policy that matches economic and social development. This paper collects senior care policy texts for a total of 11 years from 2010 to 2020, and uses text mining techniques to study the characteristics of policy texts and their evolutionary features in this stage, in order to suggest suggestions for the government's next stage of policy formulation. The main findings of this paper are as follows.

First, in terms of policy issuing institutions, the number of pension policies issued by national-level departments such as the Ministry of Civil Affairs, the Ministry of Resources and Social Security, and the Ministry of Finance is high. The Ministry of Civil Affairs, the Ministry of Finance, and the Ministry of Housing and Construction are located in the center of the social network, and these departments often jointly participate in the

formulation of the elderly policy to enhance the scientific as well as perfection of the policy from their own perspective. Secondly, from the perspective of policy characteristics, the existing policies focus on urban and rural elderly people from the perspectives of community, medical institutions, and elderly care institutions to pay full attention to pension insurance, elderly care work, infrastructure construction, and old-age security. Finally, from the perspective of policy evolution, the government has been focusing on making policies from the perspective of public services for the elderly, and indirectly making policies for retirement, investment, insurance and medical care to solve the contradiction between supply and demand of elderly services, in addition, influenced by the epidemic, the government has been focusing on epidemic prevention and control for the elderly since 2020.

This paper explores the content of existing pension policy texts through text mining methods and explores the evolution of existing pension policies from the time dimension. This study is valuable in identifying the characteristics of existing policies and the laws of policy evolution, which can help to grasp the current policy structure, improve the existing policies' shortcomings, and ultimately improve the science of policy formulation.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China Western Project 20XZZ011.

REFERENCES

- [1] Strunk B C, Ginsburg P B, Banker M I. (2006). The Effect Of Population Aging On Future Hospital Demand. *Health Affairs*, 25(3): 141-149.
- [2] Chen G, Zhao X Y, Suo H Y. (2021). Opportunities and Challenges for the High-Quality Development of Elderly Services in the 14th Five-Year Plan Period. *Administration Reform*:1-10. <https://doi.org/10.14150/j.cnki.1674-7453.20210115.001>. (in Chinese)
- [3] Song Y. (2014). Losing an only child: the one-child policy and elderly care in China. *Reproductive Health Matters*, 22(43): 113-124.
- [4] Liu R C, Chen Q J. (2016). Path Dependence in Policy Changes: Examination on the Issue of the Pension Equality in China. *Academic Journal of Zhongzhou*, (11):56-61. (in Chinese)
- [5] Zhang Y. (2013). Meeting the ageing challenge: China's social care policy for the elderly. *China: development and governance*, 2013: 343-349.
- [6] Jacobsen F F. (2015). Understanding public elderly care policy in Norway: A narrative analysis of governmental White papers. *Journal of aging studies*, 34: 199-205.
- [7] Zhai Z W, Chen J J, Li L. (2016). Aging in China: General Trends, New Characteristics and Corresponding Policies. *Journal of Shandong University(Philosophy and Social Sciences)*, (03):27-35. (in Chinese)
- [8] Wang X J. (2021). Evolutionary Features and Prospects of China's Financial Security Policy for Elderly Services under Population Aging. *Economy and Management*, 35(01):13-19. (in Chinese)
- [9] Yi J, Lu D, Deng Y. (2016). The future of social elderly care in China: from the perspective of service-oriented government. *Journal of Service Science and Management*, 9(03): 211.
- [10] Hicks D, Wouters P, Waltman L, et al. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature News*, 2015, 520(7548): 429.
- [11] Marin A, Wellman B. (2011). Social network analysis: An introduction. *The SAGE handbook of social network analysis*, 11:25.
- [12] Vishal G, Lehal G S. (2009). A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence*, 1(1):60-76.
- [13] Tan A H. (1999). Text mining: The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*. sn, 8:65-70.

- [14] Wu H C, Luk R W P, Wong K F, et al. (2008). Interpreting TF-IDF term weights as making relevance decisions. *ACM Transactions on Information Systems*, 26(3):55-59.
- [15] Blei D M, Ng A, Jordan M I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993-1022.
- [16] Sun Q L. (1992). Boundaries of high-frequency and low-frequency words and word frequency estimation methods. *Information Science*, (02):28-32. (in Chinese)
- [17] O'Callaghan D, Greene D, Carthy J, et al. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, 42(13):5645-5657.

Short Research Paper**Digital Technology-driven Business Model Innovations:****A Bibliometric Analysis**

*Jiangping Wan^{1*1}, Mengyan Liu¹, ChunxiaoYang², Hexin Zhang³*

¹School of Business Administration, South China University of Technology, Guangzhou, 510640,China

²Guangdong twenty commercial chain co., LTD., Guangzhou, 510610,China

³The board of directors office of Guangdong province company, China national tobacco corporation, Guangzhou, 510610,China

Abstract: With the advent of the data age, digital technology has been widely used in business model innovation. To understand the current research situation in the field of digital technology-driven business model innovation and reveal the knowledge structure, research hotspots, and development trends in this research field, this paper adopts statistical analysis, co-citation analysis, cluster analysis and other methods to carry out bibliometric analysis and knowledge mapping on the relevant literature included in the Web of Science database. The research results show that customer relationship management, digital economy and financial service system, sustainable development and digital service innovation, and the competition and cooperation mechanism of enterprises are hot topics in this field. Moreover, digital platform, firm performance, and value creation are the main research directions in the future.

Keywords: digital technology, business model innovation, bibliometrics analysis

1. INTRODUCTION

The business model is a hot topic in the Internet boom. A good business model is essential to the success of an organization, which can explain the operation mechanism of an enterprise. The ability of enterprises to develop innovative business models is of great commercial significance, but there are many challenges to improve innovation ability^[1]. Digital technology has promoted the development of the industry and become a booster of business model innovation^[2]. Enterprises need to use digital technology to explore more capabilities and establish a new business model to create long-term value.

At present, bibliometric analysis has played an essential role in showing the research status of disciplines or fields. Therefore, it is of considerable significance for future research to fully and accurately reveal the overall knowledge structure and frontier dynamics in the field of business model innovation driven by this digital technology.

Above all, this paper takes the relevant literature in the web of science database as the research object, reveals the research content and knowledge structure in this field, and explores the research hotspot and development trend. This paper can help scholars understand the research status of digital technology-driven business model innovation, and provide some reference for the in-depth exploration in this field.

2. DATA SOURCES AND RESEARCH METHODS

2.1 Data sources

To ensure the accuracy and reliability of the retrieval results, this paper selects the Web of Science Core Collection as the data source. Using "Digital technologies" and "business model innovations" as keywords to carry out theme retrieval, the period is 2014 to 2020, and the data collection time is March 12, 2020. We obtain

* Corresponding author. Email: csjpw@scut.edu.cn(Jiangping Wan) , liumengyan1998@163.com (Mengyan Liu)

a total of 401 literature records after the retrieval results were deduplicated and corrected. Every literature record includes the title, abstract, keywords, author and institution information of the author, the year and journal of publication, and all citation information.

2.2 Research methods

This paper uses the bibliometric method and visual knowledge mapping to study the relevant literature about digital technology promoting business model innovation. (1) Statistical analysis method: With the statistical analysis method, this paper studies the number of published literature, citation number, national distribution and subject category distribution. (2) Literature co-citation analysis: it means the number of times that two documents are cited simultaneously by one or more documents. Highly cited documents represent the knowledge base of the research field. (3) Keyword clustering analysis: it is to cluster keywords according to their co-occurrence intensity and classify keywords with substantial similarity into one category, to explore research hotspots in this field. (4) Keyword emergence analysis: it can reflect the dynamic changes in a research field and identify research frontiers and trends by identifying keywords that rapidly increase in a certain period.

3. STATISTICAL ANALYSIS

3.1 The overall growth trend analysis

The trend change of the number of published literature and cited literature is an important indicator to measure the research progress of a specific discipline or field. It is of considerable significance to evaluate the stage of this field and predict the future development trend. Table 1 shows the annual number of publications and citations in the research field of digital technologies promoting business model innovation from 2014 to 2020. Since the retrieval date is March 12, 2020, the number of papers in this year is not complete. Figure 1 shows the fitting function results of the number of annual publications and citations in the six years from 2014 to 2019, which directly reflects the overall growth trend of this research field.

Table 1. The number of annual published literature and cited literature

Year	2014	2015	2016	2017	2018	2019	2020
Published literature	19	39	42	78	87	125	11
Cited literature	8	33	136	255	501	1135	187

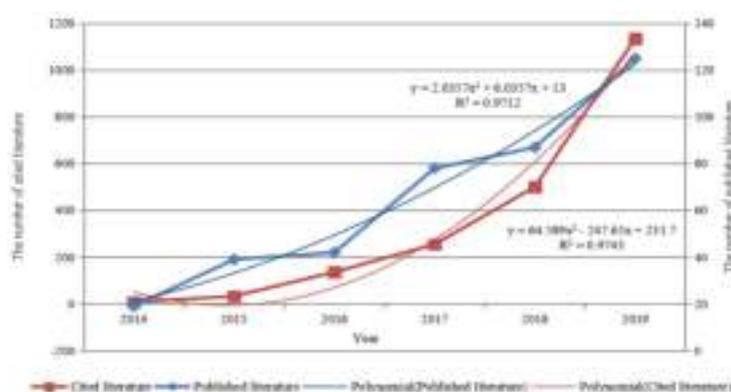


Figure 1. The fitting function results of annual published literature and cited literature

On the whole, the number of published literature each year shows an increasing trend, which can be roughly divided into two stages: the first stage is the foundation stage (before 2016). The number of literature published in this stage increases slowly, which indicates that digital technology has not widely applied to business model innovation. The second stage is the rapid development stage (2017 to date), with 47 literature published from 2017 to 2019, an increase of nearly 100% compared with the 23 literature published from 2014

to 2016 in the first stage, showing an explosive growth trend. The polynomial curve R^2 fitted is 0.9712, which shows that digital technology is increasingly widely used in business model innovation. The number of cited literature per year increased rapidly from 8 times in 2014 to 1135 times in 2019. The fitted polynomial curve R^2 of the cited literature every year is 0.9743. Digital technology has become a hot spot and research direction of promoting business model innovation, which has attracted extensive attention from scholars. This indicates that with the rise of artificial intelligence, the Internet, and digital economy, digital technology plays an increasingly important role in business model innovation.

3.2 The country analysis

To understand the research status and progress in the field of digital technology promoting business model innovation around the world, this part analyzes the national spatial characteristics of relevant research. Table 2 shows the betweenness centrality of the top ten countries in the total amount of literature issued. The betweenness centrality is an indicator of the importance of nodes in the network, and it measures the cooperation between countries. The higher the betweenness centrality, the closer the cooperation with other countries. Figure 2 shows the changing trend of annual publication volume in the top ten countries in publishing frequency.

Table 2. The betweenness centrality of the top ten countries

Country	USA	ITALY	ENGLAND	GERMANY	SPAIN	AUSTRALIA	RUSSIA	FINLAND	SWEDEN	CHINA
Centrality	0.12	0.32	0.24	0.31	0.03	0.08	0.00	0.06	0.01	0.05

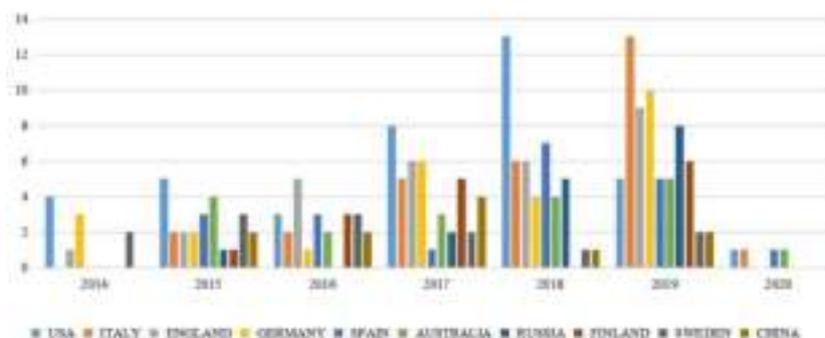


Figure 2. The annual publication volume in the top ten countries

In the field of digital technology promoting business model innovation, the most significant number and contribution is the USA, followed by Italy, England, Germany, Spain, and other countries. The total amount of published literature shows an increasing trend. However, the number is still small, which indicates that the research on digital technology to promote business model innovation is still in its infancy. The number of published literature in the USA showed an increasing trend from 2014 to 2018 but decreased in 2019, in other countries increased rapidly this year, indicating that other countries attached importance to research in this field. As far as China is concerned, the number of published documents ranks 10th, indicating that the degree of attention needs to be further improved. China's betweenness centrality is low (0.05), suggesting that China's international cooperation in this field is not close, which suggests that Chinese scholars need to strengthen foreign exchanges and further increase international academic cooperation.

3.3 The category analysis

The distribution of research literature categories can help scholars understand the core content of the research field of digital technology promoting business model innovation, which provides more effective

reference sources for further research in this field. The Citespace can identify burst words of the research category, which can understand the characteristics of temporal distribution and dynamic change and better reflect the research frontier and development trend in the knowledge domain.

Table 3. The research High-Frequency Categories

Frequency	Centrality	Year	Category
96	0.09	2014	Management
87	0.19	2014	Computer Science
80	0.03	2014	Business
75	0.39	2014	Engineering
61	0.07	2014	Computer Science, Information Systems
31	0.08	2014	Information Science & Library Science
27	0.01	2014	Engineering, Electrical& Electronic
27	0.01	2014	Computer Science, Theory & Methods
22	0.02	2015	Economics
22	0.00	2014	Public Administration

Table 4. Top 5 Categories with the Strongest Citation Bursts

Subject Categories	Year	Strength	Begin	End	2014--2020
Operation Research&Management Science	2014	2.2109	2015	2016	
Computer Science, Theory & Methods	2014	3.3447	2015	2016	
Computer Science,Artificial Intelligence	2014	1.3922	2015	2017	
Education,Scientific Disciplines	2014	1.3745	2016	2017	
Economics	2014	2.6116	2018	2020	

Table 3 is illustrated that the research category is not limited to the field of management but also involves Computer Science, Business, Engineering, Information Systems, Library Science, and other disciplines. It shows that the research on digital technology in the field of promoting business model innovation presents a diversified and cross-development trend. From Table 4 that " Computer Science, Artificial Intelligence " and " Economics " are the research categories with the most prolonged period (2015-2017), and " Computer Science, Theory & Methods " are the research categories with the most considerable burst strength. Before 2017, the main research fields were Operation Research, Management Science, and Computer Science. In the past three years, the research field has mainly focused on economics.

4. RESEARCH TREND ANALYSIS

4.1 The research foundation

Co-citation analysis of literature means that other papers cite two or more literature at the same time, then the two literature form a co-citation relationship. The more frequently cited literature can represent the research foundation in this field. Table 5 shows the top 8 references cited in the research field of digital technology-driven business model innovation. Reading them can identify a research basis.

The research basis mainly includes two aspects: (1) The Concept and Application of Business Model: Based on the multi-disciplinary perspective, the concept of business model should be more precise, and the business model can conceptualize as an interdependent activity system. Zott C et al. put forward a method to promote the study of business model, expanding the application scenarios of business model^[3]. (2) The Rise of Digital Business Model: Digital technology is gradually applying to the business strategy, organization

operation, products, and services of enterprises, which fundamentally promotes the innovation of the business model. Bharadwaj A et al. point out that cross-industry and cross-department enterprises should rethink the role of IT strategy and innovate the digital business model promptly^[4].

Table 5. The high frequency cited literature

Cited Frequency	Centrality	First author(year)	Literature title
40	0.12	Zott, C(2011)	The Business Model: Recent Developments and Future Research.
34	0.05	Teece, D. J (2010)	Business Model, Business Strategy, and Innovation.
32	0.17	Bharadwaj, A(2013)	Digital business strategy: toward a next generation of insights
22	0.10	Yoo, Y(2012)	Organizing for Innovation in the Digitized World
19	0.11	Yoo, Y(2010)	Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research
17	0.15	Gawer, A (2013).	Industry Platforms and Ecosystem Innovation
17	0.16	Chesbrough, H.(2010)	Business Model Innovation: Opportunities and Barriers
16	0.09	Zott, C(2010)	Business Model Design: An Activity System Perspective

4.2 The research hotspots

Keywords are highly abstracted and summarized from the literature topics. It is very representative to examine the hot research topics in a research field through high-frequency keywords, which can effectively determine the hot research area. Use CiteSpace, set the time slice to 1 year, extract the keywords of Top50, and use Pathfinder algorithm to cut. By combining keywords with the same meaning and removing meaningless high-frequency words, this part obtains the final high-frequency keywords. Next, we use LLR to name the cluster, where $N = 201$, $E = 404$. Modular Q and Silhouette are the indicators to measure the clustering effect. Modular Q is usually higher than 0.3, and the Silhouette is higher than 0.5. The high-frequency keyword clustering graph obtained in this paper has the Modularity $Q = 0.7668$, and the Silhouette = 0.7158, which shows that the clustering result is reliable and significant. Figure 4 is illustrated the clustering map of high-frequency keywords and Table 6 is illustrated details of citation network clustering. Besides, the Silhouette under each cluster is higher than 0.7, which shows that the clustering result is perfect.

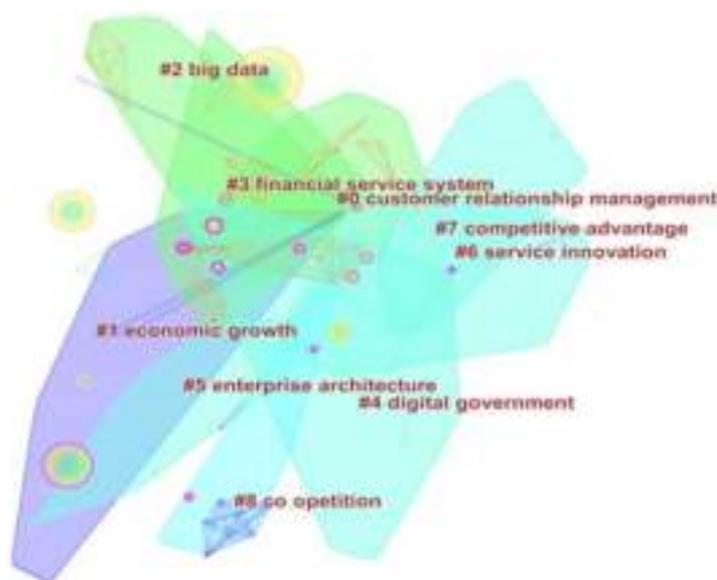


Figure 4. High-frequency keyword clustering mapping

Table 6. The details of citation network clustering

Cluster label	Size	Silhouette	Mean	Top Terms(LLR)
customer relationship management (#0)	28	0.775	2017	customer relationship management;high-speed broadband; big data
economic growth (#1)	19	0.881	2014	economic growth; technology dominance theory; telecommunication service
big data (#2)	18	0.832	2017	big data; precision medicine; supply chain
financial service system (#3)	18	0.915	2017	financial service system; digital transformation; research priority
digital government (#4)	17	0.724	2016	digital government; digital capability ; user acceptance
enterprise architecture (#5)	17	0.706	2016	enterprise architecture; digital economy ; decision making
service innovation (#6)	14	0.790	2016	service innovation ; business model innovation; digital transformation
competitive advantage (#7)	14	0.848	2016	competitive advantage ; firm performance ; capturing value
coopetition (#8)	12	0.913	2015	coopetition ; case study ; coopetition-based business model

According to the high-frequency keyword clustering mapping and reading the representative literature under each cluster, the research focus of digital technology-driven business model innovation mainly focuses on four aspects: customer relationship management (#0); digital economy and financial service system (#1,#2,#3); sustainable development and digital service innovation (#4,# 6); the competition and cooperation mechanism of enterprises (#5,#7,#8).

(1) Customer relationship management: This topic focuses on digital technology that has changed the operation mode of customer relationships in the traditional business model. The role of the Internet of Things and Big Data has affected investment in customer relationship management in the modern service industry. To prevent market share from being influenced by traditional competitors and disruptive business models, the customer relationship management industry has adopted digital innovation-driven business model. Bogers M et al. discuss how enterprises should organize a more open business model when digital technology changes the business model into a consumer-centered one^[5].

(2) Digital economy and financial service system: Digital technology has widely used in finance, finance, banking, and other fields. The digital economy era has brought new opportunities for business model innovation. The digital transformation of the financial service industry challenges the market position of mainstream financial institutions and achieves disruptive innovation. Breidbach CF et al. discuss the significance of the digital economy model in establishing a new value chain in the banking industry^[6]. Liebenau JM et al. point out the challenges related to the digital transformation of the financial service system to guide future research directions^[7].

(3) Sustainable development and digital service innovation: Business model innovation usually fails to fully include the dimension of sustainability, while digital technology accelerates the sustainable development of enterprises. Business model innovation-oriented to sustainable development has become a research hotspot. Pappas IO et al. propose a conceptual model to provide a broader basis for the digital transformation and sustainable development of enterprises^[8]. Besides, service providers are actively improving their competitiveness through a digital service-oriented business model. Based on case studies, Frank AG et al. point out how manufacturing enterprises can overcome challenges when expanding product focus with digital services^[9].

(4)The competition and cooperation mechanism of enterprises: Digital platforms promote new forms of competition and cooperation mechanisms. Digital technology gives enterprises the ability to compete in the

digital world, changes the rules of competition and cooperation, and promotes the innovation of business model^[10]. In the digital world, cooperation among competing enterprises has also become a phenomenon. Øiestad S et al. point out that the book publishing industry is currently undergoing a digital transformation, which challenges the existing business model and urges enterprises to re-examine their product mix and core competitiveness^[11].

4.3 The research frontiers

The research frontier is the most potential research topic or field in scientific research. Different from the research hotspot, the research frontier usually shows a group of emergent dynamic concepts and potential research problems. Using keyword burst analysis can understand the temporal distribution and dynamic change characteristics of burst words. Table 7 shows the top 10 keywords with the most vigorous citation bursts.

Table 7. Top 10 keywords with the strongest citation bursts

Keywords	Year	Strength	Begin	End	2014--2020
diffusion	2014	1.7762	2014	2016	
product	2014	2.1697	2014	2015	
competitive advantage	2014	1.4577	2015	2016	
absorptive capacity	2014	1.6159	2015	2017	
digital	2014	1.6682	2016	2017	
open innovation	2014	2.0874	2016	2017	
ecosystem	2014	1.8185	2017	2018	
digital platform	2014	1.7349	2017	2018	
firm performance	2014	2.2006	2017	2018	
value creation	2014	1.6547	2018	2020	

Table 7 is illustrated that the research field of digital technology promoting business model innovation presents diversified characteristics. The three keywords that have emerged in the past four years are "digital platform," "firm performance," and "value creation," which indicate that these research topics are relatively active and may become the frontier topics of this research field in the future. Using digital technology to build a digital platform provides technical support for enterprises and further promotes the innovation of business model. Future research topics will focus on the application of digital technology to the business model and where the outstanding performance and precious value brought to enterprises are reflected.

5. RESEARCH CONCLUSION AND PROSPECT

This paper uses bibliometrics to summarize the external characteristics and knowledge structure of digital technology-driven business model innovation, besides reveals the research hotspot and research frontier. The results of the statistical analysis show that the number of published literature and cited literature each year shows an increasing trend, which shows that the influence is increasing year by year. The publishing countries mainly focus on the USA, Italy, England, Germany, and other countries. The research category involves management, computer science, business, engineering, information system, and other fields, showing a diversified and cross-development trend.

The research foundation mainly includes two aspects: the concept and application of the business model and the rise of the digital business model. In terms of research hotspot and research frontier, research hotspot focuses on four research themes: customer relationship management, digital economy and financial service system, sustainable development and digital service innovation, enterprise competition and cooperation

mechanism. The research frontier includes digital platforms, enterprise performance, and value creation.

Although the research theories and methods of the business model are relatively mature, due to the rise of digital technology in recent years, most of the research on the application of digital technology in business model innovation stays at the stage of theory and assumption. Furthermore, we should actively strengthen technical research and practical application in this field in the future. For example, in building a digital platform, how to make a breakthrough in digital technology to create more value for enterprises and customers is the crucial point to be solved.

ACKNOWLEDGEMEN

This research was supported by Guangdong Tobacco Monopoly Bureau (Company) science and technology projects (Contract No.201944000200036).

REFERENCES

- [1] Chesbrough, H. (2010). Business model innovation: opportunities and barriers. *Long range planning*, 43(2-3):354-363.
- [2] Mihadjo L W W, Alamsjah F. (2019). Mediating Role of Co-creation Strategy on the Relationship between Business Model Innovation and Corporate Reputation: A Case Study on Indonesian Telecommunication Firms. *Journal of Technical Education and Training*, 11(4) :67-76.
- [3] Zott C, Amit R, Massa L. (2011). The business model: recent developments and future research. *Journal of Management*, 37(4):1019-1042.
- [4] Bharadwaj A, El Sawy, O. A, Pavlou, et al. (2013). Digital business strategy: toward a next generation of insights. *MIS quarterly*, 37(2):471-482.
- [5] Yoo Y, Boland Jr, R. J, et al. (2012). Organizing for innovation in the digitized world. *Organization Science*, 23(5):1398-1408.
- [6] Breidbach CF, Keating BW, Lim, C. (2020). Fintech: research directions to explore the digital transformation of financial service systems. *Journal of Service Theory and Practice*. 30(1):79-102
- [7] Liebenau JM, Elaluf-Calderwood SM, Bonina CM. (2014). Modularity and network integration: Emergent business model in banking. In 2014 47th Hawaii International Conference on System Sciences,1183-1192.
- [8] Pappas I O, Mikalef P, Giannakos M N, et al. (2018) Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *Information Systems and e-Business Management*, 16(3) : 479-491.
- [9] Frank A G, Mendes GH, Ayala NF, et al. (2019). Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. *Technological Forecasting and Social Change* 141:341-351
- [10] Padayachee R, Matthee M, van der Merwe A.(2017) Disruptive technologies and IT decision making in an agile business environment. In 2017 IEEE AFRICON, 843-848.
- [11] Øiestad S, Bugge MM.(2014).Digitization of publishing: Exploration based on existing business model. *Technological Forecasting and Social Change*, 83: 54-65.

Full Research Paper

Enterprise Technology Innovation and Customer Concentration under the Uncertainty of Market Environment

Wendi Wang¹, Wenbin Qu^{1*}, Xiaofeng Wang²

¹School of Economics and Management, China University of Geosciences, Wuhan, 430078, China

²Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan, 430078, China

Abstract: Digital technology has brought new technological changes to economic development and has become one of the important factors affecting the uncertainty of the market environment. It also provides a new direction for enterprises' technological innovation. In the past two years, the spread of COVID-19 has further aggravated the uncertainty of the market environment. In this context, this article uses China's Shanghai and Shenzhen A-share listed companies as a sample to empirically test the impact of corporate technological innovation on customer concentration and the impact of market environment uncertainty on the relationship between the two. Then further explored whether the effect of market environment uncertainty has differences on enterprises with different property rights. The study found that enterprise technological innovation has a significant positive effect on customer concentration, the product market competition has a positive adjustment effect on both. In contrast, environmental uncertainty has a negative adjustment effect on both. The uncertainty of the market environment has a different impact on enterprises with different property rights.

Keywords: product market competition, environmental uncertainty, technological innovation, customer concentration, property rights.

1. INTRODUCTION

With the advent of the digital economy era, emerging technologies such as big data, cloud computing, blockchain, artificial intelligence, and 5G have brought a series of changes to the modern economy and society. Digital technology is being integrated into the production, sales, service, and other processes, and at the same time enriches the mode of enterprise technological innovation. China is also facing new economic transformation requirements, and the technological innovation by enterprises is an essential manifestation of enterprises' transformation. The market environment's uncertainty has become a crucial factor affecting enterprises' investment in technological innovation and restricting their development. In a highly uncertain market environment, the customer preferences, technological development, and competitive situation in the industrial environment where the enterprise is located have significantly changed, and the market has a shorter waiting time^[1]. Enterprises face a high uncertainty and ambiguity level because they do not have complete confidence and accurate predictions about new technologies, new products, and market reactions^[2].

On the one hand, in order to achieve long-term growth, enterprises need to increase investment in technological innovation, which would be of great help for differentiating products, building a moat around business, and gaining enough market share. As The Accenture Technology Outlook 2020 report points out, people will increase expectations of technology application as digital advances become normal. Thus, if the enterprises continue to create new technology-driven products and services traditionally, there will be a rising tension between the customers and the companies. Therefore, only by technological and product innovation can enterprises get

* Corresponding author. Email: quwenbin@cug.edu.cn (Wenbin Qu)

their own place in the digital economy's increasingly competitive market.

On the other hand, customers are the service objects of business activities and the basis for realizing business results. Customer service, which is the core competitiveness of enterprises, is also facing digital transformation. Enterprises need to achieve higher quality and smarter customer service through technological innovation to ensure a stable source of customers and achieve sustainable business operations^[3]. Technological changes have changed people's way of production and life and have also improved people's ability to understand and transform the world. Nowadays, customers and consumers pay more and more attention to the subsidiary value of products brought by services and even regard the quality of services as an essential part of measuring product value and corporate value. Therefore, in addition to attracting customers and consumers by enhancing the product's value, companies also need to focus on innovation in customer service. Through technological innovation to achieve online, platform-based, intelligent, and transparent customer service, so as to enhance the customer's good experience of the company's products.

Previous studies mostly analyzed the impact of customer concentration on corporate innovation, mainly from the perspective of corporate innovation in a passive state. However, in today's digital economy era, technological change has become an important factor affecting the market environment's uncertainty. Companies increasingly choose to actively carry out technological innovations to stabilize existing customers and expand customer sources to ensure that they operate in a highly uncertain market environment. Moreover, with the addition of force majeure factors represented by COVID-19 in recent years, the market environment's uncertainty has reached an unpredictable height, and the stability of major corporate customers has become an important guarantee for most companies to continue their operations. Therefore, this article studies how companies can influence customer concentration through technological innovation in an uncertain market environment, further exploring how the market environment's uncertainty impacts the relationship between technological innovation and customer concentration for companies with different property rights. It is expected to provide ideas for preventing the loss of essential customers under the high uncertainty of the market environment, and further enrich the relevant content of product market competition, environmental uncertainty, technological innovation, and have specific innovation and research value.

2. THEORETICAL ANALYSIS AND RESEARCH HYPOTHESIS

2.1 The impact of technological innovation on customer concentration

Excellent customer value depends on efficient operations, close customer relationships and leading products, which in turn often depend on capabilities such as new product development, customer relationship management, and supply chain management. Among them, the research and development of new products is a way for enterprises to carry out technological innovation. It can create and enhance customer value by improving product innovation, improved product quality and faster research and development cycles. The technological innovation activities of enterprises are gradually regarded as a special promise that enterprises give to customers in order to achieve and maintain long-term cooperation^{[4]-[5]}. The main purpose of enterprises for technological innovation is to improve product quality and competitiveness to meet customers' diversified needs, thereby improving customer service satisfaction, which has the characteristics of value, uniqueness, and risk. Porter's five forces model points out that the market competition faced by enterprises comes from five aspects, including the competitive ability to existing competitors in the same industry, the ability of potential competitors to enter, the substitution ability of substitutes, the bargaining power of suppliers and the bargaining power of buyers, among which the customer concentration mainly responds to the bargaining power of buyers. For one thing, enterprises continue to reduce their own products' substitutability through technological innovation, thereby reducing buyers' bargaining power and increasing the cost of finding new suppliers for customers, who are increasingly dependent on enterprises and

thus more willing to establish long-term and stable relationships. Inderst and Wei^[6] found by constructing a bargaining model that customers with more robust asset size will attract more attention from firms, and companies will take the initiative to adopt innovations such as improved products to enhance the irreplaceability of their own products, so as to gain an advantageous position when negotiating with critical customers. And for another, companies improve existing products through R&D innovation in order to reduce product material and labor costs and achieve low-cost advantages for their own products. By strengthening R&D investment in new products, they convert R&D ideas into research results to meet customers' needs for new and diversified products and attract major customers with cost advantages and innovative advantages. Zhang Feng^[7] found that technological innovation can also attract more new customers by creating new customer value. At the same time, it can better maintain the relationship with existing customers and help companies gain closer contact with customers. It is conducive to creativity and demand for product innovation and improves customer acceptance of new products. Simultaneously, the company's technological innovation can be spread among consumers and customer groups through mass media, interpersonal communication, etc.^[8]. Companies provide customers with better products and services through technological innovation to increase customers' trust and dependence on the company. Therefore, the following hypothesis is proposed in this paper:

H1: There is a significant positive relationship between technological innovation and customer concentration.

2.2 The impact of market environment uncertainty on the relationship between technological innovation and customer concentration

2.2.1 The moderating effect of product market competition

To survive and develop in a fiercely competitive market, companies need to continuously create and provide new customer value, establishing leading innovative advantages. This is an important guarantee for companies to win customers and obtain sustainable income. The ability and level of technological innovation of an enterprise determine whether it can launch new products or services faster and in higher quality, thereby establishing an innovative advantage in the market. The theory of effective competition shows that when the market is in a state of effective competition, under the uncertainty of market structure changes, companies will choose to increase technological innovation under the pressure of high product price sensitivity and the loss of external resources, thereby improving production technology and product quality^[9]. Aghion et al.^[10] put forward the "escape effect" and "Schumpeter effect." As the degree of competition in the product market increases, companies will choose to increase innovation in order to escape the current competitive situation, which is manifested as the "escape effect." Current product market competition is conducive to the enhancement of technological innovation. However, when the degree of product market competition reaches a certain threshold, the enterprise will gradually reduce the marginal value of innovation revenue brought by technological innovation while taking innovation risks, which is manifested as the "Schumpeter effect." At this time, the enterprise will choose to reduce R&D and innovation investment. But the protection of enterprise innovation results can suppress the "Schumpeter effect." When an enterprise intends to carry out technological innovation activities, the future income of new products can be more reasonably estimated, which reduces the risk of innovation to a certain extent^[11]. Under the situation of increasing competition in the product market, the life cycle of products will become shorter and shorter, and the speed of product upgrading will accelerate. If the investment in innovation and R&D is not accelerated to improve their core competitiveness, then companies will face elimination^{[12]-[13]}; This competition is becoming the norm in today's era of digital economic transformation. The degree of product market competition reflects the company's market position in the industry. Companies with a strong market position tend to have poorer product substitutability and less competitive pressure. Thus less attention is paid to strengthening technological innovation to improve customer concentration. And in a market with low product market competition, corporate managers are more likely to exhibit "agent" behavior that avoids risks, preserves existing positions, and prioritizes personal

interests^[14]. Managers are often not motivated to make highly uncertain long-term investments, such as technological innovation R&D investments. However, when the product market competition becomes fiercer, to occupy a particular market share and take the initiative in negotiations with customers, companies will further increase R&D investment and reduce the substitutability of products and customers' bargaining power. At the same time, product market competition can alleviate the degree of information asymmetry between corporate management and shareholders, investors and customers, thus shareholders and investors will have a better understanding of management. The reputation mechanism of the manager's market will play a better role, and the management is more motivated to reduce production costs and seize the market initiative by increasing R&D and innovation, and then win greater customer resources. Based on this, the following hypothesis is proposed in this paper:

H2: Product market competition will promote the positive effect of technological innovation on customer concentration; the more intense the product market competition, the stronger the positive impact of technological innovation on customer concentration.

2.2.2 The moderating effect of environmental uncertainty

The uncertainty of the environment faced by enterprises comes from the dynamics and unpredictability of multiple aspects such as consumers, competitors, collaborators, and government policies, while customers' choice of enterprises is often based on a combination of factors such as product quality, product price, corporate reputation, and service. The higher the environmental uncertainty a company faces, the higher its business risk, strategic risk, and decision risk. The more likely it is that its products and image will be negatively affected, which will harm customer concentration. As environmental uncertainty increases, it becomes more difficult for decision-makers to predict the future of the business. On the one hand, higher environmental uncertainty can exacerbate the degree of information asymmetry. Customers cannot obtain valid cognition and value perception of the enterprise's R&D work, making the subsequent promotion of the new products developed by the enterprise more difficult and reduces the acceptability of the R&D results to customers and consumers. Simultaneously, information asymmetry will also result in investors not being able to make accurate assessments and the company's measurements. Auditors will be more inclined to issue non-standard audit reports under high environmental uncertainty, which will further increase the difficulty of financing for enterprises; and which will not be conducive to the survival of business and reputation image of enterprises and will eventually lead to a decrease in the trust of major customers in enterprises. On the other hand, environmental uncertainty will increase the uncertainty of product material supply. The delivery quality and delivery time of the product material supplier is uncertain; this will lead to the untimely supply of products to miss the time to win the market, while the uncertainty of supply will also bring additional costs of repeated inspection of materials, which will ultimately harm the timely fulfillment of contracts with customers and reduce corporate reputation. Finally, environmental uncertainty also makes the business risk rise; enterprises need to continually update the budget to respond to the changing external environment. Short-term strategic planning and frequent budget plan changes will harm long-term development, increasing enterprises' business risk. As a result, customers often receive unfavorable feedback about the company's low financial position and reduced cash flow earnings, leading to significant customer loss. Joseph^[15] pointed out that companies will choose to invest limited resources in business activities to improve their response speed and ability to respond to the dynamic environment when environmental uncertainty is high. Tung^[16] believes that companies tend to formulate short-term strategic plans. Shen Huihui^[17] found that environmental uncertainty would increase the degree of volatility in firms' surplus and that firm management would be reluctant to take on the additional risk associated with technological innovation. Therefore, the following hypothesis is proposed in this paper:

H3: Environmental uncertainty weakens the positive effect of technological innovation on customer

concentration, and the higher the environmental uncertainty, the weaker the contribution of technological innovation to customer concentration.

2.3 The impact of property rights on the regulation

According to the system theory, in the economic transition environment, system factors significantly influence business operations development^[18]. China's economic development is currently in a transitional period, the government's role in resource allocation still cannot be ignored. At the same time, enterprises need a lot of financial support for technological innovation, so they often need to carry out external financing to obtain financial aid. Compared to non-SOEs, SOEs can rely on government and political connections to allocate resources due to their government-enterprise solid linkages. Thus SOEs often enjoy a certain amount of soft budget constraints in financing. In a highly uncertain market environment, SOEs can still obtain sufficient financial support from banks and other financial institutions or even relief from the government, which will signal investors and customers that corporate risks are under control in times of high economic volatility economic downturn. However, the financing cost of non-SOEs will be much higher than that of SOEs in the face of bank credit policies subject to government intervention. This financing constraint will be more serious when the market environment is highly uncertain, which will send a signal to investors, customers, and consumers that the company's economy is in the doldrums and its survival risk is great.

There is also heterogeneity in the strength of the impact of product market competition and environmental uncertainty on firms with different property rights, which leads to varying effects of market environment uncertainty on the relationship between technological innovation and customer concentration. It can obtain customers, investors, and other resources at lower transaction costs in business for state-owned enterprises, often showing stronger operational capabilities and occupying a strong position in the market; thus, they are less dependent on significant customers^[19]. The environmental uncertainty comes from the uncertainty of macro policy environment and micro market environment and the uncertainty of any force majeure factors. The variability and unpredictability of these many external factors exert many uncontrollable environmental coping pressures on SOEs, so the impact of environmental uncertainty on technological innovation and customer concentration of SOEs will be stronger than the competition in product markets with advantages. For non-state enterprises, their main business goal is to obtain greater economic benefits and maintain their survival and development^[20]. The incentive mechanism of non-state enterprises are linked to the economic benefits of the enterprise, which will motivate the management of the enterprise to increase R&D investment to improve the innovation performance of the enterprise, so the enterprises will pay more attention to the role they play in the product market and adjust the level of R&D investment for the degree of market competition in order to improve their market position and product pricing ability, while the product price is often one of the primary factors for customers to consider the supplied object. Therefore, the measurement of the level of technological innovation by non-state-owned enterprises and the customers' choices of private enterprises are often more affected by the fierce competition in the product market. Based on this, the following hypotheses are proposed in this paper:

H4: The contribution of technological innovation to customer concentration is more pronounced for SOEs under uncertain market conditions than for non-SOEs.

H4a: Compared with the positive adjustment effect of product market competition, environmental uncertainty weakens the positive impact of SOEs technological innovation on customer concentration is more obvious.

H4b: Compared with the negative adjustment effect of environmental uncertainty, product market competition has a more apparent positive impact on customer concentration by promoting technological innovation of non-SOEs.

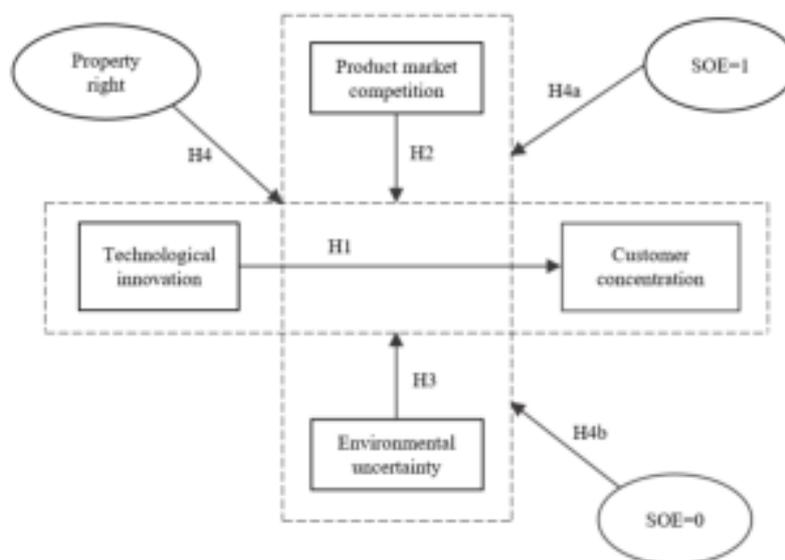


Figure 1. Research Framework

3. RESEARCH DESIGN

3.1 Variable description

(1) Customer concentration (*Customer*): The share of the top five sales of an enterprise can reflect the degree of concentration at the customer level to a certain extent. The logarithmic value of the sum of the purchase amount of the top five customers of a listed company as a proportion of the operating revenue is used to measure the customer concentration, and the larger the value is, the higher the customer concentration is.

(2) Technological innovation investment (*RD*): Technology R&D innovation investment usually includes the investment of tangible and intangible assets, and intangible assets such as technology and knowledge are also based on the support of tangible resources (such as capital, personnel, equipment, etc.), so this paper uses the logarithm of the ratio of R&D expenditure to operating revenue to measure the level of technology innovation investment.

(3) Product market competition (*FEPCM*): In this paper, the opposite of the industry-adjusted Lerner index (*EPCM*) is used to measure the degree of product market competition, which reflects the product market competition among different enterprises in the same industry. A larger value of *FEPCM* means that the weaker the pricing power of the enterprise in the industry it is in, the more inferior the competitive position it is in the market and the greater degree of product market competition it faces, and vice versa. The calculation formula is shown below:

$$FEPCM = -EPCM = \sum_{i=1}^{N_{jt}} \frac{S_{ijt} \frac{OI_{it} - OC_{it} - SE_{it} - MF_{it}}{OI_{it}}}{S_{jt}} - \frac{OI_{it} - OC_{it} - SE_{it} - MF_{it}}{OI_{it}} \quad (1)$$

Where S_{ijt} denotes the operating revenue of enterprise i in industry j in year t , S_{jt} denotes the total operating revenue of industry j in year t , N_{jt} denotes the total number of enterprises in industry j in year t , OI_{it} , OC_{it} , SE_{it} , MF_{it} denote the operating revenue, operating cost, selling expense, and administrative expense of enterprise i in year t , respectively.

(4) Environmental uncertainty (*EU*): The measurement of environmental uncertainty in this paper draws on the method of Shen Huihui^[13] and uses the volatility of corporate sales revenue to measure. First, the least-squares method (*OLS*) is used to estimate the company's abnormal sales revenue in the past five years. The specific Model is as follows:

$$Sale = \varphi_0 + \varphi_1 Year + \varepsilon \quad (2)$$

Where *Sale* is the sales revenue, *Year* is the annual variable. The observation is the current year, *Year* takes 5; if it is the past year, it takes 4, and so on, when the observation is the past fourth year, *Year* takes 1; ε is the residual of the Model (2), which is the requested abnormal sales revenue; Next, calculate the standard deviation of the abnormal sales revenue of the enterprise in the past five years, and divide it by the average sales revenue of the past five years to get unadjusted environmental uncertainty, and the median of the non-industry-adjusted environmental uncertainty of all firms in the same industry in the same year is the industry environmental uncertainty. The larger the value of the *EU*, the higher the environmental uncertainty faced by the firm.

(5) The nature of property rights (*SOE*): The nature of property rights is a dummy variable. When the enterprise is a state-owned enterprise, the value of *SOE* is 1; otherwise, the value is 0.

(6) Control variables set: reference to relevant previous studies of customer concentration, and select enterprise asset size (*Size*), financial leverage (*Lev*), companies listed age (*Age*), return on assets (*ROA*), and book capitalization ratio (*BM*) constitute the set of control variables in this article; at the same time, this article also controls the fixed effects of industry (*Ind*) and year (*Year*).

3.2 Model construction

The model construction of this article is mainly composed of four parts:

In order to test the relationship between technological innovation and customer concentration, verify *H1*. and establish Model (3):

$$Cus_{it} = \alpha_0 + \alpha_1 RD_{it} + \beta_1 Con_{it} + \beta_2 \sum Ind_{it} + \beta_3 \sum Year_{it} \quad (3)$$

In model (3), the explained variable Cus_{it} is the customer concentration of firm *i* in year *t*, the explanatory variable RD_{it} is the technological innovation level of firm *i* in year *t*, Con_{it} is the set of control variables, Ind_{it} and $Year_{it}$ are industry and year fixed effects respectively; if α_1 is significantly positive, then *H1* is supported.

In order to test the impact of market environment uncertainty on the relationship between innovation and customer concentration, verify *H2*. and *H3*., establish model (4) and model (5):

$$Cus_{it} = \alpha_0 + \alpha_1 RD_{it} + \alpha_2 FEPCM_{it} + \alpha_3 RD \times FEPCM_{it} + \beta_1 Con_{it} + \beta_2 \sum Ind_{it} + \beta_3 \sum Year_{it} \quad (4)$$

$$Cus_{it} = \alpha_0 + \alpha_1 RD_{it} + \alpha_4 EU_{it} + \alpha_5 RD \times EU_{it} + \beta_1 Con_{it} + \beta_2 \sum Ind_{it} + \beta_3 \sum Year_{it} \quad (5)$$

In model (4), $RD \times FEPCM_{it}$ represents the crossover term between technological innovation and product market competition, If α_3 is significantly positive, then *H2* holds; In model(5), $RD \times EU_{it}$ represents the cross-product term of technological innovation and environmental uncertainty, if α_5 is significantly negative, then *H3* is supported.

In order to test the influence of the nature of property rights on the adjustment effect, verify *H4* and establish a model (6):

$$Cus_{it} = \alpha_0 + \mu_{SOE} RD_{it} + \gamma_{m_{SOE}} RD \times FEPCM_{it} + \gamma_{n_{SOE}} RD \times EU_{it} + \beta_1 Con_{it} + \beta_2 \sum Ind_{it} + \beta_3 \sum Year_{it} \quad (6)$$

In model (6), $SOE=1$ when *i* is a state-owned enterprise and $SOE=0$ when *i* is a non-state-owned enterprise. If the significance degree of μ_1 is greater than μ_0 , it means that *H4* is supported; *H4a* is supported when the significance degree of γ_{m_1} is less than γ_{n_1} ; *H4b* is supported when the significance degree of γ_{m_0} is greater than γ_{n_0} .

4. RESULTS

4.1 Descriptive statistics

The results of descriptive statistical analysis of all variables in the Model are shown in Table 1. As can be seen from Table 1, the minimum and maximum values of enterprise customer concentration are -7.13089 and 0, respectively, with a standard deviation of 0.82158, indicating that there is a large difference in customer

concentration among different enterprises, and the mean value is -1.50741, which is smaller than the median -1.43801, with a left-skewed distribution, indicating the existence of some enterprises with low customer concentration; The mean value of enterprise technological innovation is -3.76471, the median is -3.36665, the difference between them is small, the minimum value is -16.14030, the maximum value is 0.23039, and the standard deviation is 1.49925, indicating that the technological innovation of Chinese listed companies shows a good normal distribution overall, but there are large differences in the technological innovation levels of different companies; The mean value of product market competition is -0.52189, the minimum and maximum values are -0.39581 and 3.96186, respectively, and the median is -0.04747, indicating that there is a big difference in the level of product competition among different enterprises; The mean value of environmental uncertainty is 1.31802, the minimum and maximum values are 0.11282 and 23.43730, respectively, indicating that Chinese enterprises generally have environmental uncertainty, and there are large differences in the uncertainty of external business environment faced by different enterprises.

Table 1. Descriptive statistics

Variable	Mean	Std. Dev	Min	Med	Max	N
Customer	-1.50741	0.82158	-7.13089	-1.43801	0	8158
RD	-3.76471	1.49925	-16.14030	-3.36665	0.23039	8158
FEPCM	-0.52189	0.08772	-0.39581	-0.04747	3.96186	8158
EU	1.31802	1.34228	0.11282	0.97738	23.43730	8158
SOE	0.34298	0.47473	0	0	1	8158
Size	22.52339	1.25674	19.56033	22.35974	27.46772	8158
Lev	0.42715	0.19489	0.05362	0.41962	0.99075	8158
Listage	2.45627	0.50695	0	2.39790	3.33221	8158
ROA	0.03344	0.07298	-0.51721	0.03437	0.22093	8158
BM	1.11855	1.25119	0.04810	0.73636	12.53083	8158

4.2 Regression results

Before the regression model test, the variables are tested for multicollinearity in this paper, and the results are shown in Table 2. It can be seen that the mean value of the VIF value of each variable is 1.49, and the maximum value is 2.08, which is much less than 10, indicating that there is no effect of multicollinearity in the variables in the regression analysis.

Table 2. Collinear Results

Variables	VIF	1/VIF
Size	2.08	0.47971
BM	1.99	0.50374
Lev	1.75	0.57129
Listage	1.40	0.71320
SOE	1.39	0.72157
ROA	1.36	0.73716
RD	1.24	0.80425
FEPCM	1.23	0.81564
EU	1.02	0.97953
Mean VIF	1.49	

The specific model regression results are shown in Table 3, where models (3), (4), and (5) are full-sample regression results, and model (6) is a sub-sample regression result.

Table 3. Regression Results

Variables	Modle (3)	Modle (4)	Modle (5)	Modle (6)		
	Full sample				SOE=1	SOE=0
RD	0.020** (0.008)	0.022*** (0.008)	0.030*** (0.009)	0.032*** (0.009)	0.039*** (0.014)	0.022* (0.014)
FEPCM		0.086 (0.070)		0.082 (0.070)	0.164 (0.443)	0.092 (0.072)
RD×FEPCM		0.066* (0.039)		0.065* (0.039)	0.025 (0.101)	0.086* (0.047)
EU			-0.045*** (0.010)	-0.044*** (0.010)	-0.065*** (0.017)	-0.025* (0.013)
RD×EU			-0.005*** (0.002)	-0.005** (0.002)	-0.010*** (0.003)	-0.002 (0.003)
Size	-0.049*** (0.017)	-0.050*** (0.017)	-0.034* (0.018)	-0.035** (0.018)	-0.080** (0.033)	0.003 (0.021)
Lev	0.057 (0.060)	0.055 (0.060)	0.058 (0.060)	0.056 (0.060)	0.133 (0.117)	0.067 (0.071)
Listage	-0.281*** (0.072)	-0.277*** (0.072)	-0.280*** (0.072)	-0.276*** (0.072)	-0.128 (0.175)	-0.165* (0.093)
BM	-0.023*** (0.008)	-0.022** (0.008)	-0.023*** (0.008)	-0.021** (0.008)	-0.022** (0.011)	-0.035*** (0.013)
ROA	0.07 (0.073)	0.032 (0.080)	0.052 (0.073)	0.014 (0.080)	-0.045 (0.234)	0.010 (0.087)
Industry fixed effects	Controlled					
Time fixed effects	Controlled					
N	8158					

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; standard errors reported in parentheses.

From the full-sample regression results, we can see that Model (3) verifies the influence relationship between technological innovation and customer concentration, and the regression results show that enterprise technological innovation and customer concentration are significantly positive at the 5% level, indicating that technological innovation has a significant promotion effect on customer concentration, and H1 is supported that increased investment in technological innovation by companies can significantly improve customer concentration.

Model (4) verifies product market competition's influence on the relationship between technological innovation and customer concentration. The regression results show that there is a significant positive correlation between the cross-product multiplier (RD×FEPCM) of technological innovation and product market competition and customer concentration, indicating that product market competition can promote the positive effect of technological innovation on customer concentration, and H2 is supported that when product market competition becomes more and more intense, enterprises will adopt more technological innovation in order to occupy a certain market share to improve the competitiveness of their products to attract customers.

Model (5) verifies the impact on the relationship between technological innovation and customer concentration under environmental uncertainty. The regression results show a significant negative relationship

between the cross-product of technological innovation and environmental uncertainty (RD×EU) and customer concentration, indicating that environmental uncertainty will weaken the positive effect of technological innovation on customer concentration, and H3 holds that under higher environmental uncertainty, firms' attention to R&D innovation will be diverted, and customers' trust in the firm will be reduced.

Model (6) is a sub-sample regression based on the heterogeneity of property rights nature, and the total sample is classified according to whether it is state-controlled or not. The regression results show that: among state-owned enterprises, the effect of technological innovation on customer concentration is significantly positive at the 1% level, while the effect of technological innovation on customer concentration among non-state-owned enterprises is significantly positive at the 5% level, and $\mu_1 > \mu_0$, H4 holds that under the uncertainty of the market environment, the promotion effect of technological innovation on customer concentration is more obvious among state-owned enterprises, and different property rights nature has different coping abilities under the uncertainty of the market environment. Meanwhile, the moderating effects of product market competition and environmental uncertainty on the relationship between technological innovation and customer concentration also differ in intensity depending on the nature of property rights. The regression results find that the moderating effect of environmental uncertainty on the relationship between technological innovation and customer concentration is more significant in state-owned enterprises. At the same time, the moderating effect of product market competition on the relationship between technological innovation and customer concentration is more significant in non-state-owned enterprises, then H4a and H4b are supported. The results of the regression models for hypothesis testing are shown in Table 4.

Table 4. Hypotheses Test Results in Regression Model

Model	Hypothesis	Path	Test Coef.	Regression Coef.	Std. Err.	T-values	P-values	Support
(3)	H1	RD → Customer	α_1	0.02019	0.00800	2.52	0.012	yes
(4)	H2	RD×FEPCM → Customer	α_3	0.06688	0.03907	1.71	0.087	yes
(5)	H3	RD×EU → Customer	α_5	-0.00546	0.00204	-2.68	0.007	yes
(6)	H4	RD → Customer (SOE=1)	μ_1	0.03880	0.01391	2.79	0.005	yes
		RD → Customer (SOE=0)	μ_0	0.02234	0.01353	1.65	0.099	
	H4a	RD×FEPCM → Customer (SOE=1)	γ_{m_1}	0.02505	0.10058	0.25	0.803	yes
		RD×EU → Customer (SOE=1)	γ_{n_1}	-0.01019	0.00300	-3.40	0.001	
	H4b	RD×FEPCM → Customer (SOE=0)	γ_{m_0}	0.08650	0.04672	1.85	0.064	yes
		RD×EU → Customer (SOE=0)	γ_{n_0}	-0.00030	0.00290	-0.10	0.918	

5. CONCLUSION

This paper selects A-share listed companies in Shanghai and Shenzhen from 2015-2019 as the research object empirically tests the impact of corporate technological innovation on customer concentration under an uncertain market environment and further analyzes the impact of different property rights nature on the moderating effect. The main research findings are as follows: The higher the level of investment in technological innovation, the higher the customer concentration of enterprises, i.e., technological innovation and customer concentration are significantly positively correlated; The more intense the product market competition, the more pronounced the positive effect of technological innovation on customer concentration, i.e., product market competition positively

regulates the relationship between technological innovation and customer concentration; While environmental uncertainty negatively regulates the relationship between technological innovation and customer concentration, the higher the environmental uncertainty, the weaker the promotion effect of technological innovation on customer concentration of enterprises. It is further found that the contribution of technological innovation to customer concentration is more significant for SOEs than for non-SOEs under market environment uncertainty, and the moderating effects of product market competition and environmental uncertainty on the relationship between technological innovation and customer concentration are heterogeneous depending on the nature of property rights, with the moderating effect of environmental uncertainty being more pronounced for SOEs and the effect of product market competition being more pronounced for non-SOEs.

The research in this paper has certain implications: First, in the era of the digital economy, enterprises face the process of digital transformation to comply with the changes in the market environment, and technological innovation is the primary means for enterprises to carry out digital transformation. In order to occupy a place in the market and maintain a higher market position, enterprises must pay attention to technological innovation and meet the diversified and differentiated needs of customers. The core of product innovation lies in the full use of technologies. Hence, companies in technological innovation need to add digital technology to product innovation according to market customers' expectations to produce innovative products that meet the market environment's needs. Second, different from previous studies, product market competition plays a significant role in promoting technological innovation's positive effect on customer concentration. So that firms should not actively abandon long-term investment in innovation in the face of fierce product market competition but should instead increase technological innovation investment to improve their products' irreplaceability and customer service satisfaction levels. Finally, firms with different property rights react differently to the uncertainty of the market environment. Still, in general, technological innovation under such uncertainty can significantly increase customer concentration, only that the motivation and effect of technological innovation by firms with different property rights are not entirely consistent. Non-SOEs need to invest in innovation to improve their pricing power and increase the cost of switching suppliers in a competitive environment. In contrast, SOEs need to invest in technological innovation to meet their customers' diversified needs based on financing advantages under the uncertainty of the environment, especially when many force majeure factors lead to unpredictable and uncontrollable environmental uncertainties.

ACKNOWLEDGEMENT

This research was supported by the Open Research Project of The Hubei Key Laboratory of Intelligent Geo-Information Processing (KLIGIP-2019B08).

REFERENCES

- [1] Lin Zhonggao, Zhang Chunyan. Environmental Uncertainty, Customer Concentration and Enterprise Innovation Capability: Based on the Perspective of Changes in Major Customers Before and After the Listing of GEM Companies[J]. *Accounting Friends*,(2017)(16):47-53.(in Chinese)
- [2] Matt C., Hess T., Benlian A. Digital Transformation Strategies[J]. *Business & Information Systems Engineering*, (2015), 57(5): 339-343.
- [3] Sheikh Shahbaz. The Impact of Market Competition on the Relation Between CEO Power and Firm Innovation[J]. *Journal of Multinational Financial Management*, (2018)(44): 36-50.
- [4] Allen, JW, Phillips, GM Corporate Equity Ownership, Strategic Alliances, and Product Market Relationships [J]. *Journal of Finance*, (2000), 55(6): 2791-2815.

- [5] Kale, JR, Shahriar, H. Corporate Capital Structure and the Characteristics of Suppliers and Customers [J]. *Journal of Financial Economics*, (2007), 83(2): 321-365.
- [6] Inderst, R., Wey, C. Countervailing Power and Dynamic Efficiency [J]. *Journal of the European Economic Association*, (2011), 4: 702-720.
- [7] Zhang Feng. The Interactive Effects of New Product Development, Customer Relationship Management, and Supply Chain Management Capabilities [J]. *Journal of Management Engineering*, (2016), 30(03): 44-53. (in Chinese)
- [8] Rogers EM. *Diffusion of Innovations*[M]. New York: The Free Press,(1995).
- [9] Wang Jingyu, Fu Jianing, Zhang Hongliang. Product Market Competition and Enterprise Innovation: A Quasinatural Experiment [J]. *Modern Finance and Economics*, (2019), 39 (12): 52-66. (in Chinese)
- [10] AGHIONP, HOWITT. Revisiting the Relationship between Competition, Patenting, and Innovation [J]. *Advances in Economics and Econometrics*, (2013)(6):451-455.
- [11] Peters, J. Buyer Market Power and Innovative Activities[J]. *Review of Industrial Organization*, (2000), 1: 13-38.
- [12] ALLEN F, GALED. Financial Contagion[J]. *Journal of Political Economy*, (2000), 108(1):1-33.
- [13] JIANG FX, KIMKA, NOFSINGERJR, et al. Product Market Competition and Corporate Investment: Evidence from China[J]. *Journal of Corporate Finance*, (2015), 35(6):196-210.
- [14] INDUSTRIAL ENGINEERING, Liu Nan, Xiang Hui. Managerial Entrenchment, Product Market Competition, and Innovation Investment [J]. *Accounting Monthly*, (2020) (23): 91-98.
- [15] JOSEPH LC, et al. Organizational Slack and Response to Environmental Shifts: The Impact of Resource Allocation Patterns [J]. *Journal of Management*, (1997), 23 (1) :1-18.
- [16] Tung R L. Dimensions of Organizational Environments: An Exploratory Study of Their Impact on Organization Structure[J]. *Academy of Management Journal*, (1979), 22(4):672-693.
- [17] Shen Hui Hui. The Impact of Environmental Uncertainty on Earnings Management [J]. *Audit Research*, (2010) (1): 91-98. (in Chinese)
- [18] Ghosh D, Olsen L. Environmental Uncertainty and Managers' Use of Discretionary Accruals[J]. *Accounting Organizations and Society*, (2009), 34(2):188-205.
- [19] Zhang Ye, Lan Fengyun, Shen Huayu. Customer Concentration and Corporate Innovation Investment——Based on the Perspective of Customer Bargaining Power[J]. *China Circulation Economy*,(2019),33(04):76-88.(in Chinese)
- [20] Zhao Lu, Liu Heng, Zhang Jianqi. Degree of Market Perception, the Spirit of Innovation and Enterprise Property Right System: State-owned and Private Enterprises Comparative Study of Industry [J] *South. Economic*, (2014) (5): 25-41. (in Chinese)

Short Research Paper

Customer cooperative capability, perceived value and customer loyalty: a perspective of smart retailing

Jinfeng Wu, Zhiyuan Zhou, Delin Hou*

School of Management, Wuhan Textile University, Wuhan 430200, China

Abstract: Smart retailing has become an inevitable trend in the development of the retail industry, and how to enhance smart retail technology-related customer cooperative capability has become a top priority of traditional retailers who have migrated to smart retailing format. This study explores the mechanism of customer cooperative capability on customer loyalty in the context of smart retailing. Firstly, we construct a conceptual model in which customer cooperative capabilities affect perceived value and customer loyalty. Secondly, we adopt structural equation model method to verify the hypotheses. The study found that among the three dimensions of customer cooperative capability, search capability and participation capability positively affect customer loyalty through hedonic value and utility value; learning capability positively affects customer loyalty through hedonic value; the effect of hedonic value on customer loyalty is weaker than that of utilitarian value.

Key words: smart retailing, customer cooperative capability, perceived value, customer loyalty

1. INTRODUCTION

Smart retailing is “the smart use of smart retail technologies by enterprises and consumers to reshape and strengthen their role in the new service economy by improving the quality of shopping experience”^[1]. The key to the success of smart retailing lies in whether smart retail technology can enhance customer's shopping experience—perceived value. Perceived value is the overall evaluation made by customers on the utility of products or services based on the perception of what is received and given, and it has an important effect on customer satisfaction and customer loyalty^[2]. In the smart retail environment, the use of smart retail technology is an important approach for firms and customers to create value together. This view has been recognized by firms and many scholars^[3]. The service-dominant logic suggests that firms, customers and other stakeholders can co-create value in a specific experience context in an interactive way through joint investment of resources^[4]. In a smart retail environment, customer cooperative capability, as a resource invested by customers in value co-creation, should be one of the important antecedents of customer perceived value, and the relationship between the two variables is necessary to be explored. However, so far, no research has explored the impact of customer cooperative capability on perceived value from the perspective of smart retailing. How does customer cooperative capability affect perceived value in a smart retail environment? What kind of changes will this influence bring on customer loyalty? There is no answer yet.

Xie et al.^[5] suggested that customers have three cooperative capabilities related to big data technology: search capability, learning capability and participation capability. As the “intelligent” use of technology to collect real-time data (big data) of each consumer's behavior and preferences is an important feature of smart retailing^[3], the dimension of customer cooperative capability related to big data can also be applied to smart retailing, albeit with certain variation. Therefore, we divide the customer cooperative capability based on smart retailing into three dimensions: (1) search capability, which refers to the ability of customers to search for

* Corresponding author. Email: z8595313@126.com

effective information using various smart retail technologies; (2) learning capability, which refers to the ability of customers to quickly master the use of smart retail technologies; (3) participation capability, which refers to the ability of customers to develop products or services in cooperation with retailers using smart retail technologies.

2. THEORETICAL BACKGROUND AND HYPOTHESES

2.1 Effects of search capability on perceived value

Smart retail technology can provide customers with a wealth of information that is helpful for decision-making. For example, customer can receive comments from other customers on products through social media, or receive discount information from nearby stores through mobile apps, and know their total purchase expenditure through the budget monitoring function of the smart shopping cart when shopping. The stronger the search capability, the more effective information customers can obtain by using smart retail technology, making it easier to make better purchase decisions according to their own needs ^[6]. Further, the improvement of the quality of decision-making will enhance customers' perception of utility value ^[7]. Therefore, customers with strong search capabilities can perceive more utility value from retailers with "smart transformation". In addition, customers with strong search capability can use smart retail technology to access more new products or get more new experiences, which will make customers perceive more fun of exploration in shopping ^[8]. Therefore, search capabilities can enhance the hedonic value perceived by customers. Thus, we hypothesize that:

H1a: Customers' search capability based on smart retailing has a positive effect on perceived utility value

H1b: Customers' search capability based on smart retailing has a positive effect on perceived hedonic value

2.2 Effects of learning capability on perceived value

In the context of smart retailing, customers are exposed to many smart retail technologies that were unfamiliar in the past. The use of these technologies directly determines whether customers can obtain new experiences that are different from traditional retail scenarios. Through learning, customers can master the use of these smart retail technologies. The stronger the learning capability, the less time and effort customers need to invest in learning smart retail technology, correspondingly, the stronger their perception of the ease of use of smart retail technology ^[9]. Further, existing studies have shown that perceived ease of use of service technology has an important effect customer perceived value of service providers ^[7,10]. Therefore, customers with strong learning capability should perceive more customer value from retailers who migrate to smart retail format. Thus, we hypothesize that:

H2a: Customers' learning capability based on smart retailing has a positive effect on perceived utility value

H2b: Customers' learning capability based on smart retailing has a positive effect on perceived hedonic value

2.3 Effects of participation capability on perceived value

In a smart retail environment, smart retail technology provides customers and retailers with more opportunities to jointly improve products. For example, customers use mobile applications provided by retailers to feed back their opinions on new products, and use virtual reality technology to design customized products along with retailers. The stronger the customer's ability to participate in product development, the higher the degree that the final product provided by the retailer meets the customer's personalized needs ^[11], which means that customers can perceive higher utility value. In addition, customers with strong participation capability often have strong interests in participating in the process of product and service development ^[12], and therefore it is easier for them to experience the fun of designing products in this process, which means that customers can perceive more to hedonic value. Thus, we hypothesize that:

H3a: Customers' participation capability based on smart retailing has a positive effect on perceived utility value

H3b: Customers' participation capability based on smart retailing has a positive effect on perceived hedonic value

2.4 Effects of value perception on customer loyalty

The effect of customer perceived value on customer loyalty in a retail environment has been widely supported by theoretical and empirical studies. Research on online stores such as Chiu^[13] has shown that utility value and hedonic value have positive effects on customer loyalty. Therefore, the utility value and hedonic value perceived by customers in a smart retail environment should positively affect customer loyalty. Thus, we hypothesize that:

H4a: The perceived utility value of customers based on smart retailing has a positive effect on customer loyalty

H4b: The perceived hedonic value of customers based on smart retailing has a positive effect on customer loyalty

Based on the above mentioned hypotheses, we construct the conceptual model shown in Figure 1:

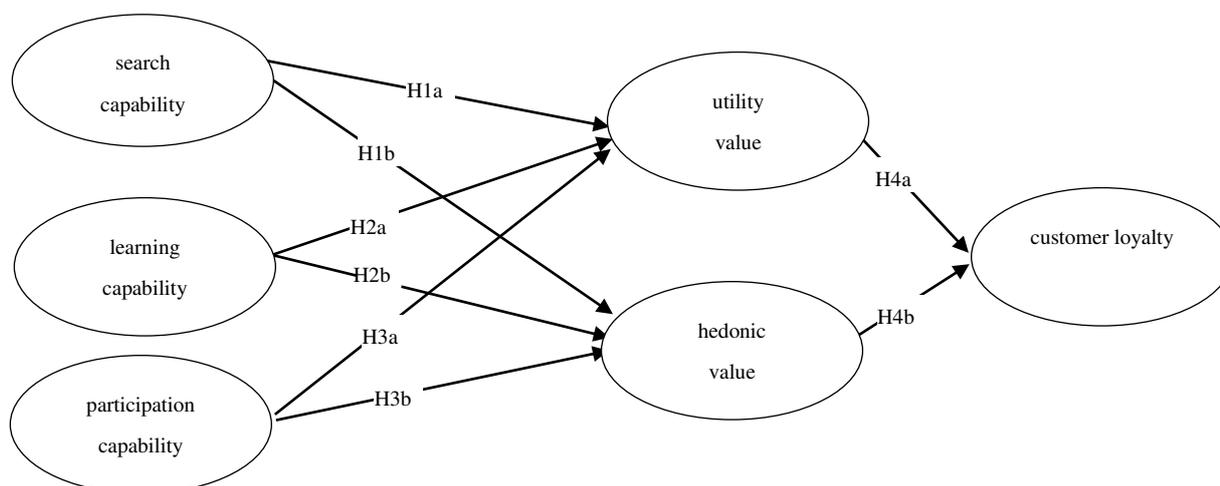


Figure 1. Conceptual model.

3. METHODOLOGY

3.1 Samples

This research conducts a formal questionnaire survey to actual retail shoppers who have used smart retail technology. Subjects from a comprehensive university who have used smart retail technology in shopping during the past 6 months were selected to participate in this study. Respondents were asked to fill in a questionnaire about their latest shopping experience in using smart retail technology. A total of 402 responses were obtained. Elimination of incomplete responses left 355 eligible responses for analysis. 41% of the respondents were men and 59% were women. 99% were between 18-25 years old. 43% of the respondents reported using smart retail technologies in more than 50% of their shopping in the last 6 months. The smart retail technologies reported by the respondents included face recognition technology, in-store interactive display screens, augmented reality, virtual reality, electronic tags, mobile APP, mobile self-checkout, and retailer's WeChat official account.

3.2 Questionnaire

Our measurement items come from two sources: scales verified by previous empirical studies and scales developed by ourselves. Before the formal survey, we conducted a pilot test with 150 college students who had prior experience with smart retail technologies to further refine the measurement items. All items were measured on 7-point Likert-type scales.

4. EMPIRICAL ANALYSIS AND RESULTS

4.1 Measurement model

Confirmatory factor analysis (CFA) enables the performance of tests regarding the convergent validity, discriminate validity, and reliability of the study constructs. The measures of overall fit mostly meet conventional standard, which suggests that our model fits the data well (CMIN=407.807, CMIN/DF=2.997, RMSEA=0.070, GFI=0.904, AGFI=0.867, NFI=0.944, CFI=0.962, IFI=0.962, RFI=0.930).

More specially, for all constructs, the composite reliability and coefficient alpha values exceed the threshold value of 0.6. Therefore, the scale for constructs appears to exhibit satisfactory internal consistency reliability. All the factor loading, which range 0.784 to 0.912, are significant ($p < 0.001$), indicating that convergent validity is achieved for all the study constructs.

The discriminate validity of construct measures was assessed on the basis of the Fornell and Larcker's criterion^[14]. All the square root of average variance extracted (AVE) are greater than interconstruct correlations, indicating that discriminant validity is supported.

Table 1 . Discriminant validity test results.

Latent variable	1	2	3	4	5	6
search capability	0.910					
learning capability	0.578	0.857				
participation capability	0.780	0.609	0.902			
perceived utility value	0.672	0.516	0.688	0.850		
perceived hedonic value	0.661	0.591	0.616	0.774	0.879	
customer loyalty	0.690	0.646	0.620	0.761	0.810	0.871

4.2 Structural model

Structural Equation modeling (SEM) was used to estimate parameters of the structural model in Figure 1, and the completely standardized solutions computed by the AMOS22.0 are reported in Table I. Goodness-of-Fit statistics, indicating the overall acceptability of the structural model analyzed, were acceptable: RMSEA is 0.082, GFI is 0.882, AGFI is 0.840, NFI is 0.928, RFI is 0.913, TLI, CFI, and IFI are all higher than 0.9, and each fitting index reaches an acceptable level.

We found that customer perceived utility value was positively related to search capability and participation capability. The relationship of search capability to perceived utility value was the strongest (0.434 t value= 5.833), next was the relationship of participation capability to perceived utility value (0.124, t value= 2.332).

We found that customer perceived hedonic value was positively related to search capability, learning capability and participation capability. The relationship of participation capability to perceived hedonic value was the strongest (0.377, t value= 4.428), next was the relationship of search capability to perceived hedonic value (0.324, t value= 5.099). The relationship of search capability to perceived hedonic value was the weakest (0.301, t value= 5.557).

The customer loyalty was positively related to perceived utility value and perceived hedonic value. The relationship of perceived utility value to customer loyalty was the stronger (0.573, t value= 11.499), The relationship of perceived hedonic value to customer loyalty was the weaker (0.379, t value= 8.122).

Table 2. Structural model estimates.

Relationship	Standardized Parameter	T value	Conclusion
Relationship of dimensions of customer cooperative capability to perceived utility value			
H1a: search capability → utility value	0.434	5.833	Significant
H2a: learning capability → utility value	0.102	1.395	Insignificant
H3a: participation capability → utility value	0.124	2.332	Significant
Relationship of dimensions of customer cooperative capability to perceived hedonic value			
H1b: search capability → hedonic value	0.324	5.099	Significant
H2b: learning capability → hedonic value	0.301	5.557	Significant
H3b: participation capability → hedonic value	0.377	4.428	Significant
Relationship of dimensions of customer perceived value to perceived customer loyalty			
H4a: utility value → customer loyalty	0.573	11.499	Significant
H4b: hedonic value → customer loyalty	0.379	8.122	Significant

5. CONCLUSION

This study explores the relationship between three dimensions of customer cooperative capability and customer loyalty through the mediating role of perceived value in the context of smart retailing. The conclusions were follows: First, customer cooperative capability based on smart retailing can be divided into three dimensions, that is, search capability, learning capability and participation capability. Secondly, search capability and participation capability positively influence customer loyalty through hedonic value and utility value. Thirdly, learning capability positively affects customer loyalty only through hedonic value. Finally, the effect of hedonic value on customer loyalty is weaker than that of utility value. The results of this research show that traditional retailers who migrate to smart retailing format can enhance customer perceived value by strengthening customer cooperative capabilities, thereby achieving customer retention and customer growth.

ACKNOWLEDGEMENT

This research was supported by the Humanities and Social Science Project of Ministry of Education of China (Grant No. 17YJA630108) and the National Social Science Fund of China (Grant No. 20BGL122).

REFERENCES

- [1] Pantano E, Timmermans H. What is smart for retailing?[J]. *Procedia Environment Sciences*, 2014, 22: 101-107.
- [2] Yang Z, Peterson R T. Customer perceived value, satisfaction, and loyalty: The role of switching costs[J]. *Psychology & marketing*, 2004, 21(10): 799-822.
- [3] Pantano E, Priporas C V, Dennis C. A new approach to retailing for successful competition in the new smart scenario[J]. *International Journal of Retail & Distribution Management*, 2018., 46(3): 264-282.
- [4] Vargo S L, Lusch R F. Service-dominant logic: continuing the evolution[J]. *Journal of the Academy of marketing Science*, 2008, 36(1): 1-10.
- [5] Xie K, Wu Y, Xiao J, et al. Value co-creation between firms and customers: The role of big data-based cooperative assets[J]. *Information & Management*, 2016, 53(8): 1034-1048.

-
- [6] Cai L A, Feng R, Breiter D. Tourist purchase decision involvement and information preferences[J]. *Journal of vacation Marketing*, 2004, 10(2): 138-148.
- [7] Keller K L, Staelin R. Effects of quality and quantity of information on decision effectiveness[J]. *Journal of consumer research*, 1987, 14(2): 200-213.
- [8] Li M L, Green R D. A mediating influence on customer loyalty: The role of perceived value[J]. *Journal of Management and Marketing research*, 2011, 7: 1.
- [9] Chong X, Zhang J, Lai K K, et al. An empirical analysis of mobile internet acceptance from a value-based view[J]. *International Journal of Mobile Communications*, 2012, 10(5): 536-557.
- [10] Sivathanu B. An Empirical Study on the Intention to Use Open Banking in India[J]. *Information Resources Management Journal (IRMJ)*, 2019, 32(3): 27-47.
- [11] Hu G, Yuan Y, Zhang S, Liao H. A Critical Review of Value Co-creation: Connotation, Evolution and Mechanism [J]. *Journal of Tianjin University of Commerce*, 2017,37(02):57-64.
- [12] Füller J, Matzler K, Hoppe M. Brand community members as a source of innovation[J]. *Journal of Product Innovation Management*, 2008, 25(6): 608-619.
- [13] Chiu C M, Wang E T G, Fang Y H, et al. Understanding customers' repeat purchase intentions in B2C e - commerce: the roles of utilitarian value, hedonic value and perceived risk[J]. *Information Systems Journal*, 2014, 24(1): 85-114.
- [14] Fornell C, Larcker D F. Evaluating structural equation models with unobservable variables and measurement error[J]. *Journal of marketing research*, 1981, 18(1): 39-50.

Full Research Paper

Digital Platform Expectation Disconfirmation and Customer Satisfaction in Sharing Economy: Moderations of Direct and Indirect Network Effects

Xiangcheng Yang¹, Yi jiang^{1*}

¹ School of Economics and Management, China University of Geosciences, China

Abstract: Digital platform in sharing economy is increasingly focusing on customer satisfaction to gain sustainable competitive advantage. However, existing understandings of how network effects shaped the creative process of customer satisfaction are not well-understood in the information system field. Based on the expectation-disconfirmation theory, we proposed a conceptual model to elaborate on the roles of direct and indirect network effects on digital platform customer satisfaction. We test our model using survey data from 250 sharing economy platform customers. The key results of structural equation modeling (SEM) provide broad support for the following: (1) the effect of customer expectation on confirmation is completely dependent on the mediation of perceived performance in digital platform of sharing economy, and (2) the direct and indirect network effects contribute as the moderator at different stages of the model. We discuss the theoretical and practical implications of how managers should effectively develop and leverage the network effects to improve the digital platform customer satisfaction in the context of sharing economy.

Keywords: Digital platform, Sharing economy, Expectation-disconfirmation theory, Customer satisfaction, Network effects

1. INTRODUCTION

Over the past few years, sharing economy provides the digital space in which service providers use their privately-owned resources (e.g. physical and nonphysical goods and services) via digital platforms to serve their customers^[1, 2]. This emergent business model also initiates re-examining traditional views on resources consumption, the flow of resources among different digital platforms has become more frequent, the theory of obtaining competitive advantages by relying on the non-fluidity or technical barrier has been greatly challenged in the context of sharing economy. Structurally, the sharing economy network is peer-like in nature, with the aim of benefiting and relying on a large number of active customers^[1]. Customers, with the explosive propagation on the platform, become the base of value creation and acquisition. Understanding the digital platform (hereon DP) customer satisfaction is critical for managers to make effective and targeted strategic decisions.

For the focus of customer satisfaction, related literature of the expectation–disconfirmation theory (hereon EDT) provide critical constructs to describe the forming process of web-customer satisfaction^[3]. However, the expectation–disconfirmation relationship of customers may change as the innovation of the business model. And previous theoretical framework limits our ability to delineate the potential new relationship and to explain the possible changes which may be occurred by interactions between digital platform and customers. Therefore, the first focus of this paper is to study the creative process of DP customer satisfaction for the sharing economy.

In the digital platform era, the term “network effects”, known as network externalities referring to an additional value of a commodity or service to others^[4], is gaining prominence. Previous research has indicated the importance of network effects on customers’ adoption attitude to the sharing economy platform^[1]. To make network effects useful for customer operations, it is crucial for platform managers to understand how to take advantage of network effects in customer attraction and retention. Even so, there is still a lack of understanding

* Corresponding author. Email: wuhanjoey@163.com

of the difference between the roles of direct and indirect network effects in sharing economy, and how customer satisfaction is affected by network effects. Thus, our second target of this study is to uncover the moderation role of network effects on the creative process of DP customer satisfaction in sharing economy.

We seek to contribute to a theory of customer satisfaction that takes into account both expectation-disconfirmation theory and network effects. Specifically, we first focus on the measurement of the constructs for developing potential links in the creative process of DP customer satisfaction by the model of EDT. Additionally, we argue that network interaction in two key dimensions——direct network effects and indirect network effects affect customer satisfaction. By identifying two types of network effects as moderators of the expectation-disconfirmation-satisfaction relationship, we propose the research framework to try to answer questions as follows: (1) Based on EDT, are there any new variable links to the DP customer satisfaction in sharing economy? (2) How do different network effects moderately affect the expectation-disconfirmation model and DP customer satisfaction respectively?

2. LITERATURE REVIEW

2.1 Customer satisfaction in IS field

Customer satisfaction is used to describe customers' emotional reaction to a product or service^[5]. Studies in the IS field currently refer to this concept as a significant determinant of IT continue use^[6]. Among the studies, the theoretical frameworks mainly include the IS success model^[7] and the EDT^[3, 8].

IS success model is proposed to indicate the basic determinants including information quality, system quality, and service quality are critical for repurchase intention^[9]. The ultimate success of IS depends on customers' continuance which is based on increasing satisfaction. McKinney et al. integrated EDT with IS success model's framework to decompose web-customer satisfaction^[3]. Recent researches on customer satisfaction in IS field usually add related constructs on the basis of IS success model framework or EDT according to specific application scenarios, such as trust, perceived usability, perceived individual benefit.

The digital platform, as an emergent phenomenon in IS field, is with reference to an online foundation or base on common components around which a company might build a series of products. Exchanges in platform-mediated networks have an interactive structure in which customers trade with each other while affiliating with platform providers. The digital platform facilitates value-creating interactions among platform participants and strong customer self-organization, which content's generation is highly distributed and decentralized^[10]. The influence of these opportunities on customer satisfaction is also different in the platform environment against the traditional IS environment^[11]. Prior studies have indicated that high-quality product information has a statistically significant impact on customer satisfaction, which is related to product information, and website design is an important determinant of customer formation^[12].

2.2 Network effects in sharing economy

With the development of digital technology and social network, more research began to discuss the role of digital platforms in the sharing economy^[13], and network effects of sharing economy is being an important theme in IS studies.

Network effects, also known as network externalities, refer to an additional value of products or services to others^[4]. Broadly speaking, the fundamental benefits provided by sharing economy are not traditional products or services, but the way to get a vast customer network. As the network grows, the level of the product quality will also improve, so that the growth can be self-feedback. Previous research indicates that network effects play the role of improving customers' usage intention in the operation of platform-based services. If customers predict that they will have a larger market share based on the criteria they currently use, they will be less willing to give up the value created by a huge customer base, and therefore less willing to switch^[14].

Network effects can generally be divided into two categories: direct network effects and indirect network effects. Direct network effects are described that: With the expansion of the customer base, the product becomes more valuable. Previous studies have indicated that the size of network can be seen as a basic characteristic of network effects, which depend on the number of other network customers interacting with the network^[14]. Besides, indirect network effects result from the availability of complementary goods and services. When a product or service network increases, it may in turn increase the value of a complementary product or network.

The value of network effects on the sharing economy platform is also well-recognized. However, most articles only study network effects as an overall conception on the platform, or even do not distinguish between them. Studies investigate direct and indirect network effects simultaneously and how two types of network effects influence sharing economy DP customer satisfaction are limited. Although sharing economy and customer satisfaction research continues increasing development, the structure and characteristics of digital platform may change the relationship of customer satisfaction creation. And previous theoretical framework limits our ability to delineate the potential new relationship and to explain the possible changes which may be occurred by interactions between platform and customers. In order to fill this gap, this study is to explore a new research model to examine the creative process of DP customer satisfaction in the context of sharing economy.

3. THEORETICAL BACKGROUND AND HYPOTHESES DEVELOPMENT

3.1 Expectation-disconfirmation theory

Expectation-disconfirmation theory (EDT) was originally designed for analyzing customers' behavior to investigate customer satisfaction and post-purchase behaviors^[8]. EDT indicated that customers first generate an original expectation of a product/service prior before purchasing. After making the purchase decision and using it for a period, customers will form the perceived performance of the product/service. When perceived performance surpasses expectations, confirmation generates a positive emotion (e.g. pleased), while negative emotion (e.g. frustrated) is the result of expectations exceeding perceived performance. EDT is applied diffusely in many fields. For example, Bhattacharjee indicated that EDT based constructs can explain the continuance intention among online banking customers^[8]. In recent IS research, theoretical and conceptual frameworks of EDT were increasingly developed. Most of which have been integrated with IS or IT variables, such as information quality, perceived usability, trust.

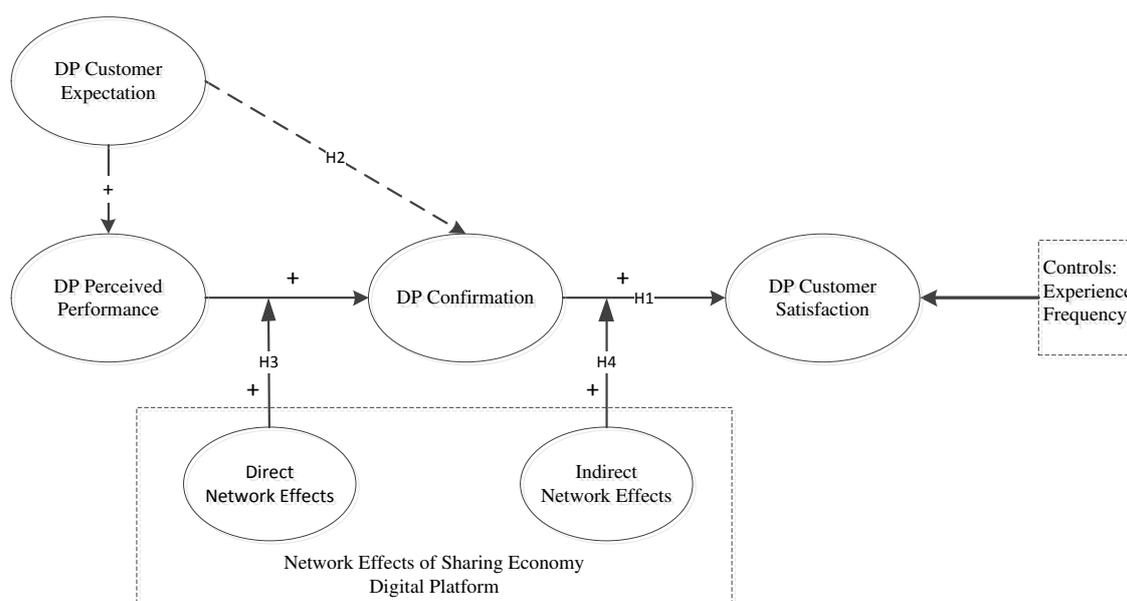


Figure 1. Research model

In this paper, we argue that IT-enabled information and resource flows among customers in digital platform trigger the expectation-disconfirmation for customer experience, and then affect customer satisfaction. We also argue that network effects can magnify the creative process of customer satisfaction. Thus, based on EDT, we integrate direct and indirect network effects into the generation process of customer satisfaction and propose a research model (see figure. 1). This model depicts that DP customer satisfaction, created through customers' "expectation-disconfirmation" variables in a digital platform, with positively moderation of direct and indirect network effects. This model also describes a full-mediation link among DP customer expectation, DP perceived performance, and DP confirmation. Experience and frequency variables are used as control variables for customer satisfaction. The definitions of the main variables are shown in table 1.

Table 1. Definitions of variables.

Variable	Definition	Reference
DP Customer Expectation	When customers consider engaging a digital platform, they form their expectations by collecting and applying previous and present information and knowledge	Venkatesh & Gopal 2010 ^[15] ; Oliver 1980 ^[5] ; McKinney et al. 2002 ^[3]
DP Perceived Performance	Customers' perception of how digital platform performance fulfills their needs, wants and desires	McKinney et al., 2002 ^[3]
DP Confirmation	Customers' subjective judgments after comparing their expectations and their perceptions of performance received	McKinney et al. 2002 ^[3] ; Bhattacharjee 2001 ^[18]
DP Customer Satisfaction	Customers' emotional state response to the entire digital platform website usage and service experience	Oliver, 1980 ^[5] McKinney et al. 2002 ^[3]
Direct Network Effects	Direct network effects arise when the benefit of network participation to user depends on the number of other network users whom they can interact	Lin & Huang 2014 ^[14]
Indirect Network Effects	An increase in usage of a product or network result in an increase in the value of a complementary product or network, which in turn can add value of the original	Zhou & Lu 2011 ^[16]

3.2 Theoretical hypotheses

Previous research on digital service has demonstrated that confirmation has a positive influence on customer satisfaction^[17]. Given that this relationship is well established in the literature, we mainly theorize about it in the context of the sharing economy digital platform. It is worth note that EDT is tested in specific contexts, however, few studies verify this relationship in digital platform scenarios. we hypothesize that:

H1: In the context of digital platform, DP confirmation has a positive impact on DP customer satisfaction.

When customers consider purchasing products, they usually collect previous information about purchasing experiences from other customers, then establish interior comparison criteria, thus forming their expectations^[5, 15]. In the context of IT usage, customers' pre-usage cognition (e.g. attitude, beliefs) is generally from second-hand information, which in turn expresses customers' initial expectations. As time goes by, when customers obtain first-hand experience using IT, they would assess their initial perceptions are consonant or inconsistent with practical experience^[18]. According to previous research, the role of perceived performance in EDT has become a comparative criterion for the degree of confirmation of expectations^[3].

However, when customers gain new information about the platform by observing other customers' experiences, they may continually adjust their expectations^[19] which may result in no direct relationship between expectations and disconfirmation^[15]. Besides, the online searching experience can also influence customer expectations, customers know what kind of service they will get in advance and are less likely to be angry or surprise at the service they received^[11]. The content in the digital platform is decentralized and customer-generated, and customers' expectations of the platform before use are mainly generated from second-hand information, which will weaken the direct impact of expectations on inconsistency. Some EDT studies take for ultimate outcome evaluations (confirmation or disconfirmation) according to the direction and magnitude of the gap between customer expectations and perceived performance, then final results are inclined to the direction of perceived performance^[20]. Thus, we deduced that DP disconfirmation or confirmation is a result of the comparison between expectations and perceived performance in the context of sharing economy. Thus, we hypothesize that:

H2: In the context of sharing economy, the effect of DP customer expectation on DP confirmation is completely dependent on the mediation of DP perceived performance.

The direct network effects caused by the increasing scale of customers have a certain influence on the reputation and the effectiveness of the platform itself. This may moderate the process by which customers compare expectations with perceived performance. Meanwhile, indirect network effects caused by increasing in complements and additional services is something that customers can't expect before they use it. This may affect the process by which customers end up generating satisfaction from confirmation or disconfirmation. Thus, we contend that two network effects act on different phases of customer satisfaction generation.

The greater the direct network effects, the greater the value of the product creates for customers. Positive network effects occur when customers spread their valuable information to others. If a product or service becomes dominant, there will be a herd effect with people rushing to purchase the product or service, especially when making choices in an information overload environment. Others' purchase behavior may highlight its strengths and encourage potential customers to imitate the purchase behavior. Customers tend to share technical knowledge by interacting with people in their social networks^[14]. Therefore, information from other customers significantly impacts the effect of DP perceived performance on DP confirmation. We hypothesize that:

H3: The higher direct network effects are, the stronger influence of DP perceived performance on DP confirmation is.

Network effects are often based on the hypothesis that: customers preferred platforms with more participants, meanwhile, participants' motivation and desire for development will be stronger for a platform with multitudinous customers. More complementary goods increase customers' demand (and vice versa), producing network effects and winner-take-all outcomes^[10]. When the diversity of products and services in one platform increases, they will be more likely to continually use this platform. Abundant complement goods produce indirect network effects and boost the network value. For example, the number of apps available in Google Play is increasing as the Android smartphone customer base grows^[14]. In the context of digital platform, payment methods, security insurance, positioning effect, promotions are significant complements provided by the ridesharing platform. Szymanski & Henard indicated that aspects related to product information are important factors of customer satisfaction. Indirect network effects bring sharing economy platform more abundant complement service, which will promote the influence of confirmation on customer satisfaction^[12]. Thus, this study proposes that a larger customer base in sharing economy platform will create and accumulate more value of complimentary service and information, which can in turn influence customer satisfaction, we hypothesize that:

H4: The higher indirect network effects are, the stronger influence of DP confirmation on DP customer satisfaction is.

4. METHODOLOGY

4.1 Construct Measurement

We confirmed the construct measurements by combing prior literature in order to make a survey and collect data on the use of sharing economy platforms. The variables were measured with multi-item, five-point Likert scales, which are illustrated in the Appendix.

DP customer satisfaction is an emotional state response to the entire digital platform website usage and service experience^[3, 5]. To create instruments for measuring the construct in this study, based on the practice of sharing economy platform, we adopt three measurement items from the factors of McKinney for web-customer satisfaction. Respondents were asked to evaluate their satisfaction with the given ridesharing platform on the following aspects: pleasure, content, delight.

Customer expectations, perceived performance, and confirmation are key constructs in EDT, which is well known in behavior research. Confirmation is measured in the EDT literature in three ways: objective, inferred and perceived. Because EDT pays attention to customers' subjective judgments of product or service, thus, this study measured perceived performance and confirmation. Based on McKinney's study, we adopt four dimensions to measure related constructs refer to access, usability, information usefulness, and service^[3].

Direct Network Effects and Indirect Network Effects. The scale of direct network effects developed by Lin & Huang, was used to assess the degree of interaction of network participation. Performance regarding the customer scale and the market occupancy was measured as factors for measuring direct network effects to construct. According to Boudreau & Jeppesen^[10], we design four items anchored as "payment method", "positioning effect", "security insurance" and "promotions" to measure indirect network effects.

Additionally, we specify usage experience and usage frequency as control variables in our research.

Usage experience. As customers' experiences in the use of the sharing economy platform increases, their prior cognition may be revised iteratively^[18]. Experienced customers may not have more potential expectations, and are therefore more likely to be satisfied with the platform.

Usage frequency. For customers in the platform, higher interaction frequency may lead to longer relationship duration. Customers with a more enduring relationship with the platform are more likely to be satisfied.

4.2 Questionnaire Development and Data collection

Given that the popularity of sharing accommodation platforms such as Airbnb is not high in China, the target population of this study is individuals that have experience using ridesharing platforms. All respondents are allowed to usage experience on at least one of the following ridesharing platforms: DiDi or Uber.

To test the research hypotheses proposed in the model, we designed a questionnaire. After pretesting the questionnaire to assess content validity and convergent validity, we sent final questionnaires through PC and smartphone to obtain data. The survey yielded 302 responses, and 250 of them were valid. We conducted a difference test on questionnaires distributed, and the results showed that there was no significant difference ($p > 0.05$) in the methods of questionnaires distributed in different ways. Most respondents own a good educational level. Students (70.8%) make up the majority of the sample. About 62 % of those respondents had a usage experience over 1 year, 61.2% of respondents use ridesharing service 1-3 times per week. According to the analysis results of descriptive statistics, most respondents are familiar with sharing economy platform. Therefore, the respondents' understanding of the questionnaire is convincing.

To exclude common method variance, this study accepted Harmon's one-factor tests to evaluate common method variance. Through the test, a single factor accounted for only 38.12% of the variance, which was less than 50%. The result suggests that a common method bias isn't considered an issue in our model.

5. RESULTS

5.1 Measurement model

We use SPSS 25 and Smart PLS 3.0 to test the measurement model by the analysis of reliability and validity. Firstly, the result indicated that KMO statistic was 0.888. Bartlett's test passed the test at the significance level of 0.001, and the cumulative explanatory variance rate of the model was 70.743%. For the reliability of items, we use Cronbach's alpha to test whether the data is reliable. Also, composite reliability (CR) is tested to ensure the reliability of the items. All results were above the acceptable value of 0.70. The validity of constructs was accessed by examining CR and Average Variance Extracted (AVE) values. All AVE values are significantly larger than the recommended value of 0.50. Table 2 showed the results of convergent and discriminant validity. Each item was loaded on its designated factors with all items' values was greater than the acceptable value of 0.40. The inter-correlation value between constructs was lower than the square root of AVE for every construct (see Table 3). Thus, the research model demonstrated satisfactory reliability and validity.

Table 2. Reliability analysis and convergent validity.

Latent Variables	Indicators	Loadings	Cronbach's α	CR
DP Customer Expectation (DPCE)	DPCE 1	0.777	0.874	0.914
	DPCE 2	0.846		
	DPCE 3	0.828		
	DPCE 4	0.817		
DP Perceived Performance (DPPP)	DPPP 1	0.731	0.835	0.890
	DPPP 2	0.667		
	DPPP 3	0.743		
	DPPP 4	0.738		
DP Confirmation (DPC)	DPC 1	0.648	0.813	0.877
	DPC 2	0.718		
	DPC 3	0.809		
	DPC 4	0.741		
DP Customer Satisfaction (DPCS)	DPCS 1	0.650	0.805	0.885
	DPCS 2	0.674		
	DPCS 3	0.601		
Direct Network Effects (DNE)	DNE1	0.762	0.799	0.908
	DNE2	0.787		
Indirect Network Effects (IDNE)	IDNE1	0.697	0.781	0.860
	IDNE2	0.717		
	IDNE3	0.680		
	IDNE4	0.518		

Table 3. Results of discriminant validity.

	DPCE	DPPP	DPC	DPCS	DNE	IDNE
DPCE	0.853					
DPPP	0.516**	0.819				
DPC	0.182**	0.487**	0.801			
DPCS	0.326**	0.516**	0.564**	0.849		
DNE	0.375**	0.446**	0.380**	0.595**	0.912	
IDNE	0.363**	0.506**	0.533**	0.650**	0.594**	0.778

Note: ** denotes $p < 0.01$. Numbers on the diagonal are the positive square roots of the average variance extracted (AVE) values; off-diagonal values are the estimates of inter-correlation between the latent constructs.

5.2 Structural model

Figure 2 shows the path coefficients of the research model. Specifically, DP perceived performance was positively linked with DP confirmation ($\beta = 0.543, p < 0.001$), DP confirmation was positively linked with DP customer satisfaction ($\beta = 0.576, p < 0.001$). DP customer expectation ($\beta = 0.536, p < 0.001$) was detected to have statistically positively correlated with DP perceived performance. However, DP customer expectation was not found to have a significant relationship toward confirmation. Thus, H1 was supported.

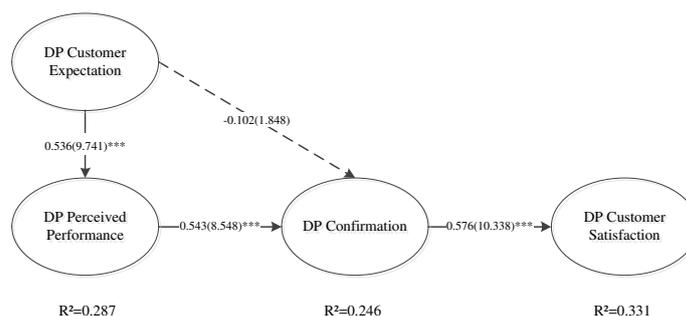


Figure 2. PLS results of research model (***p < 0.001)

DP customer expectation ($\beta = 0.536, p < 0.001$) was detected to have statistically positively correlated with DP perceived performance. However, DP customer expectation was not found to have a significant relationship toward confirmation. Thus, H1 was supported.

5.3 The mediating effect

The mediating effect of DP expectation informing DP confirmation was examined by accessing the direct and indirect effect of DP perceived performance on DP confirmation. The result is shown in table 4. Firstly, we tested a model that excluded perceived performance. The results present a significant effect of DP expectation on confirmation ($\beta=0.168, p<0.001$). Secondly, we included DP perceived performance into analysis then validated these relationships among DP expectation, DP perceived performance and DP confirmation. The result shows that DP perceived performance influences DP confirmation ($\beta=0.521, p<0.001$), however, DP expectation has an insignificant effect on DP confirmation ($\beta=-0.087, p>0.05$). The effect of expectation toward confirmation is completely dependent on the mediation of perceived performance. Thus, H2 is accepted.

Table 4. The mediating effect analysis.

path	B	SE	R ²	Sobel
DPCE->DPC(c)	0.168**	0.058	0.033	6.22***
DPCE->DPPP(a)	0.490***	0.052	0.266	
DPPP->DPC(b)	0.521***	0.063	0.244	
DPCE->DPC(c')	-0.087	0.060		

Note: Sobel= $a*b/\sqrt{(b^2*SE_a^2 + a^2*SE_b^2)}$; SE means standard error; ** denotes $p < 0.01$, *** denotes $p < 0.001$.

5.4 Moderation effect

The moderation effects of direct network effects and indirect network effects were tested. We centralized the moderating variables and dependent variables and constructed the product term with the centralized value, then it is included in the regression of dependent variables, and hierarchical regression is carried out.

Table 5 demonstrates that H3, which stated that direct network effects have a moderation effect on the relationship between DP perceived performance and DP confirmation, was supported. H4 which stated that indirect network effects moderate the relationship between DP confirmation and DP customer satisfaction was also supported. According to the results, the values of D-W statistics of each group of regression models are around 2, suggesting that there isn't a significant relationship among residuals. In addition, the VIF values of the model are far less than 10, which can prove that the model has no obvious multicollinearity. The moderation effects are shown in Figure 3 and Figure 4. As predicted, at high levels of direct network effects, confirmation rises more rapidly as perceived performance rises. Similarly, at high levels of indirect network effects, customer satisfaction rises more rapidly as confirmation rises.

Table 5. Moderation effect analysis.

	The moderation effect model of DNE			The moderation effect model of IDNE		
	Model 1 (β)	Model 2 (β)	Model 3 (β)	Model 1 (β)	Model 2 (β)	Model 3 (β)
DPPP	0.487***	0.396***	0.397***			
DNE		0.203**	0.228***			
DPPP×DNE			0.180***			
DPC				0.564***	0.304***	0.251***
IDNE					0.489***	0.494***
DPC×IDNE						0.136**
D-W	1.753	1.739	1.801	2.155	2.032	2.017
VIF	1	1.248	1.020~1.268	1	1.396	1.16~1.573
F	77.003***	45.685***	35.418***	115.668***	118.224***	83.668***
R ²	0.234	0.264	0.293	0.315	0.485	0.499

Note: **means $p < 0.01$, ***means $p < 0.001$

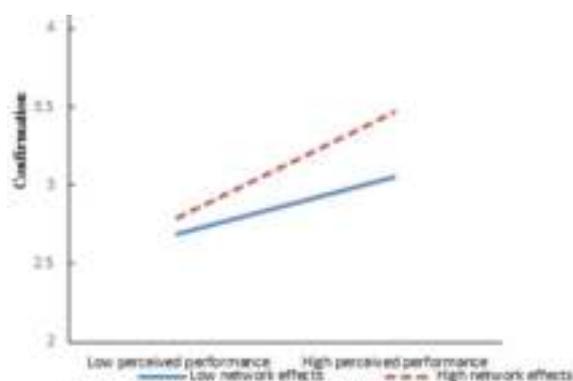


Figure 3. The moderation effect of direct network effects

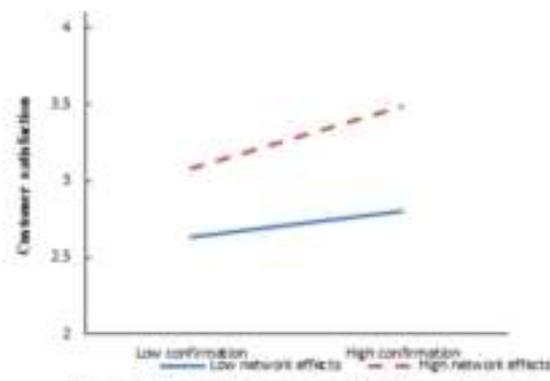


Figure 4. The moderation effect of direct network effects

In addition, the correlation of experience use on DP customer satisfaction is not significant ($\beta=0.004$; $p>0.05$), which may be the result of customers are more focused on the current stage of use, customer satisfaction will not be changed by the customer experience usage. The usage frequency's results ($\beta=0.038$; $p>0.05$) provided evidence that frequent use can't lead to higher customer satisfaction.

6. DISCUSSION

This study has sought to explain the question of how, and under what condition, the DP customer satisfaction forms in the context of sharing economy moderating by network effects. Unlike most prior perspectives which have indicated the direct relationship between DP customer expectation and confirmation on expectation-disconfirmation model [20], we argue that DP perceived performance acts as a complete mediator between DP perceived expectation and DP confirmation (Sobel=6.22, $p<0.001$) for sharing economy platform. Additionally, this study distinguished two kinds of network effects as moderators at different phases of the research model. Direct network effects were validated as a moderator between DP perceived performance and confirmation, and indirect network effects were validated as a moderator between DP confirmation and customer satisfaction. The sharing economy platform with higher direct network effects will have a more widely spread reputation and be more valuable to customers, which will magnify the results of the comparison of DP customer expectations and perceived performance. The sharing economy platform with larger indirect network effects will bring value

beyond expectation to customers who have already compared with expectation and perceived performance to generate confirmation, and the additional value will strengthen the effect of confirmation generates satisfaction.

6.1 Theoretical contributions

Our findings advanced EDT and network effect literature in several ways.

First, this study supplemented the EDT in the aspect of digital platform and validates confirmation or disconfirmation still acts as an important influence factor in the sharing economy platform context. Importantly, this study found the influence of DP customer expectation on DP confirmation is completely dependent on the mediation of DP perceived performance. In the traditional EDT^[5] and IS literature^[3, 10], perceived performance usually has a partially mediating role between customer expectation and confirmation, but which is designed for traditional business rather than sharing economy platform. Sharing economy platform is a highly standardized, the dissemination of information with high accuracy and personalized service makes the information obtained by customers before use more real than the traditional situation, which will make the customer's expectation more practical. This study suggests that DP customer satisfaction evaluations rest with the direction and magnitude of the gap between expectations and perceived performance, owning perceive performance provide critical forces to realize confirmation.

Second, our study illustrated how network effects promote the creation of DP customer satisfaction and extended their importance in the sharing economy context. We bridged the gap between EDT and network effect literature by demonstrating the moderation role of network effects plays at different phases of the model. This finding contributed to the advancement of the network effects in sharing economy by understanding the importance of focusing on different characteristics of network effects and explaining when and how they play a critical role in facilitating the creation of DP customer satisfaction. These findings extended the EDT literature and provided a new perspective for future study of the feature of sharing economy platform.

6.2 Implication for practice

In terms of practical implications for platform managers, our model would help the practitioners to clarify the forming process of sharing economy platform customer satisfaction and more accurately grasp the change mode of the customer psychological state.

The introduction of EDT for digital platform will make managers realize that the management of customer expectation alone cannot play a significant role in customer satisfaction under the sharing economy situation. In other scenarios, it is generally recommended to improve customer satisfaction by raising or lower customer expectations. The results of this study prove that this way is no longer applicable to the operation and management of sharing economy platform. Instead, managers should pay attention to the confirmation of customer expectations and improve the performance of the platform to meet or even exceed customer expectations, so as to obtain customer satisfaction and realize their continued use of the platform. This study suggests that managers stimulate the intrinsic motivation of customers such as enjoyment, fun and flow experience. These elements should be reflected in the platform website's design to enhance customers' perceived performance.

The findings of this study unveil the mechanism by taking advantage of network effects is useful to improve DP customer satisfaction. The actions of increasing direct effect, like social network marketing, award forwarding, can be more used in influencing customers. Similarly, DP customer satisfaction is affected by the moderation of the indirect network effect. It shows a positive feedback effect in the platform-mediated network. The implication for managers is that it should take effective measures to attract more service providers and expand the customer scale, encourage customers to register in the platform and use relative service. The platform can bring additional value to customers on the basis of customers' recognition of basic services by optimizing the interface, providing personalized services. This will promote customer satisfaction towards the platform, which in turn will improve the scale of platform and thus bring greater value to the platform and customers.

7. LIMITATIONS AND FUTURE RESEARCH

Several limitations should be considered when interpreting the findings. All questionnaires are filled by Chinese responders, the results may not be universal to countries with different cultures. Therefore, it's necessary for future studies to collect a more extensive customer base to test the generalizability of our results. In addition, other characteristic constructs of sharing economy platform, such as trust, may also influence the formation of customer satisfaction. Future research can consider taking these constructs into the research model.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant No. 71702176 and the Fundamental Research Funds for the Central Universities, China University of Geoscience s (Wuhan) (Grant No. 1810491C07 and No. 1910491C08).

APPENDIX. MEASUREMENT ITEMS

(5-point Like Scale, 1 = Strongly Disagree, 5 = Strongly Agree)

DP Customer Expectation

CE1: I expect that the platform provides good access.

CE2: I expect that the platform is customer-friendly.

CE3: I expect information on the platform to be useful in my purchase decision.

CE4: I expect the platform provide a dependably and prompt ridesharing service.

DP Perceived Performance

PP1: I reckon that the platform provides good access.

PP2: I reckon that the platform was customer-friendly.

PP3: I reckon that the platform performance in providing information is useful in my purchase decision.

PP4: I reckon that the platform performance in providing a dependably and prompt ridesharing service.

DP Confirmation

CONF1: Access provided by the platform is better than my expectations.

CONF2: The difficulty of using the platform is easier than my expectations.

CONF3: Information provided by the platform in my purchase decision is more useful than my expectations.

CONF4: Ridesharing service provided by the platform is better than my expectations.

DP Customer Satisfaction

CS1: After using the platform, I am very pleased.

CS2: After using the platform, I am very contented.

CS3: After using the platform, I am very delighted.

Direct Network Effects

DNE1: Compared to the traditional trip mode, I predict that the platform will have more customers in the future.

DNE2: I predict that the market share of the ridesharing network will grow constantly in the future.

Indirect Network Effects

INDE1: The platform provides a richer payment method.

INDE2: The platform provides a more convenient positioning effect, making it easier to find nearby vehicles.

INDE3: The platform provides a richer set of security insurance to ensure the security of payment and travel.

INDE4: The platform continues to launch scene-based promotions, such as shopping vouchers.

REFERENCES

- [1] Sutherland W, Jarrahi M H. (2018). The sharing economy and digital platforms: A review and research agenda. *International Journal of Information Management*, 43: 328-341.
- [2] Yang S, Song Y P, Chen S X, et al. (2017). Why are customers loyal in sharing-economy services? A relational benefits perspective. *Journal of Services Marketing*, 31(1): 48-62.
- [3] McKinney V, Yoon K, Zahedi F. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research*, 13(3): 296-315.
- [4] McIntyre D P, Srinivasan A. (2017). Networks, Platforms, and Strategy: Emerging Views and Next Steps. *Strategic Management Journal*, 38(1): 141-160.
- [5] Oliver R L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, 17(4): 460-469.
- [6] Liao C C, Palvia P, Chen J L. (2009). Information technology adoption behavior life cycle: Toward a Technology Continuance Theory (TCT). *International Journal of Information Management*, 29(4): 309-320.
- [7] DeLone W H, Mclean E R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Journal of Management Information Systems*, 3(4): 60-95.
- [8] Bhattacharjee A. (2001). Understanding information systems continuance: An expectation-confirmation model. *Mis Quarterly*, 25(3): 351-370.
- [9] DeLone W H, McLean E R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems*, 19(4): 9-30.
- [10] Brown S A, Venkatesh V, Goyal S. (2012). Expectation Confirmation in Technology Use. *Information Systems Research*, 23(2): 474-487.
- [11] Shankar V, Smith A K, Rangaswamy A. (2003). Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, 20(2): 153-175.
- [12] Szymanski D M, Henard D H. (2001). Customer satisfaction: A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science*, 29(1): 16-35.
- [13] Mantymaki M, Baiyere A, Islam A K M N. (2019). Digital platforms and the changing nature of physical work: Insights from ride-hailing. *International Journal of Information Management*, 49: 452-460.
- [14] Lin T C, Huang S L. (2014). Understanding the Determinants of Consumers' Switching Intentions in a Standards War. *International Journal of Electronic Commerce*, 19(1): 163-189.
- [15] Venkatesh V, Goyal S. (2010). Expectation Disconfirmation and Technology Adoption: Polynomial Modeling and Response Surface Analysis. *Mis Quarterly*, 34(2): 281-303.
- [16] Zhou T, Lu Y B. (2011). Examining mobile instant messaging user loyalty from the perspectives of network externalities and flow experience. *Computers in Human Behavior*, 27(2): 883-889.
- [17] Susanto A, Chang Y, Ha Y. (2016). Determinants of continuance intention to use the smartphone banking services An extension to the expectation-confirmation model. *Industrial Management & Data Systems*, 116(3): 508-525.
- [18] Bhattacharjee A, Premkumar G. (2004). Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *Mis Quarterly*, 28(2): 229-254.
- [19] Venkatesh V, Thong J Y L, Chan F K Y, et al. (2011). Extending the two-stage information systems continuance model: incorporating UTAUT predictors and the role of context. *Information Systems Journal*, 21(6): 527-555.
- [20] Brown S A, Venkatesh V, Goyal S. (2014). Expectation Confirmation in Information Systems Research: A Test of Six Competing Models. *Mis Quarterly*, 38(3): 729-U172.

Short Research Paper

Construction and Verification of Night-time Economy Development Evaluation Index System

Xiaoying Li¹, Yan Tu^{2*}

¹College of Information, Central University of Finance and Economics, Beijing, 102206, China

²College of Information, Central University of Finance and Economics, Beijing, 102206, China

Abstract: Night-time economy is an important indicator to measure the level of regional economic development. Accurately measuring and evaluating the level of night-time economy is of great significance to promote consumption upgrading and economic growth. Under the background of digitalization, e-commerce and digital technology have greatly expanded the scene and market of night consumption. In this paper, on the basis of defining the connotation of night-time economy, the key elements of its development are determined. We construct and verify the night-time economy development evaluation index system from five dimensions: Environmental Basis, Convenience of Consumption, Vitality of Consumption, Quality Evaluation of Development and Network Attention. In the selection of index data, the method of multi-source data fusion analysis is adopted to integrate statistical data, night light data and data from the Internet platform to jointly measure and evaluate the development of night-time economy.

Keywords: night-time economy, multi-index evaluation system, life service e-commerce platform

1. INTRODUCTION

"Night-time economy" was first put forward by Britain in the 1970s to improve the phenomenon of empty nest at night in the urban center. In fact, Night-time economy is an economic form based on time division, which refers to the economic and cultural activities from 6 p.m. to 6 a.m. of the next day. It is dominated by tertiary industries such as catering, tourism and entertainment. The core of night-time economy is the extension of consumption chain of urban economy in time and scene.

According to the theory of leisure time and consumer behavior, night-time economy is the product of the enrichment of consumer leisure time when the economy of a city or region develops to a certain stage. Night-time economy has become a new driving force for economic growth. As an important part of urban economy, the prosperity of night-time economy is an important way to promote the adjustment of urban economic structure and enhance the comprehensive competitive strength. At present, more than 40 cities in the world have clearly issued policies on night-time economy. As of October 2020, China has issued nearly 200 policies that are highly relevant to night-time economy. Data from China's Ministry of Commerce shows that 60 percent of consumption in China takes place at night, and the market size of the night-time economy has reached 2,643,125 billion yuan in 2019.

At present, the study of night-time economy is mainly based on qualitative analysis, which expounds the function of night-time economy and how to develop it. By constructing a reasonable measurement framework, we can understand the development status of night-time economy in different time and different regions more clearly. That is to say, it can objectively reflect the development of night-time economy from the perspective of quantitative analysis, and provide theoretical basis for the practical development of it. Given the difficulty in

* Corresponding author. Email: tuyan@cufe.edu.cn

obtaining data on the night-time economy, how to measure its level of development effectively remains a very innovative and challenging work.

The rapid development of digital technology is comprehensively restructuring the development model of various industries. The popularity of mobile Internet and e-commerce has greatly expanded the night consumption scene and promoted the digital transformation of the industry. Leading companies in the digital economy, such as Meituan, Alibaba and JD.com, are competing to enter the night-time consumer market to seek commercial value. E-commerce and social media platforms contain a large amount of data that can reflect users' online browsing, socializing and consumption behaviors at night, which provides a new data dimension to quantitatively analyze night-time economy.

In this paper, combined with the existing research basis, the measurement framework of night economic development level is constructed. We use the method of multi-source data fusion analysis to integrate statistical data, night light data and data from the Internet platform to jointly construct the evaluation index system of night-time economy development. On the basis of the index system, Analytic hierarchy process and Entropy weight method are used to comprehensively evaluate the night-time economy. Taking Beijing as an example, this paper verifies the evaluation index system and objectively evaluates its night-time economy development potential.

2. LITERATURE REVIEW

In fact, it was not until the 1990s that the night-time economy was formally incorporated into academic research ^[1]. Similar to the development course, the study of night-time economy has experienced three stages: preliminary exploration, gradual growth and rapid development. Most of the studies focus on the economic impact, social impact and development countermeasures of the night time economy ^[2]. In terms of economic impact, first of all, night-time economy can directly drive economic growth by extending the running time of economic activities ^[3]. The development of core industries in night-time economy will also drive the coordinated development of related industrial chains ^[4]. For example, the prosperity of night tourism will drive the development of catering, transportation, accommodation and retail. In addition, night-time economic agglomeration areas dominated by service industry will produce agglomeration effect and form economies of scale ^[5]. As for the social impact, the alcohol-centered night-time economy model in the past has led to a series of negative impacts threatening social security, such as urban management, crime and security at night ^[6]. With the improvement of living standard, the demand of high quality leisure and entertainment puts forward higher requirements for the development of night-time economy. Researchers began to propose a large number of policy development proposals for the high quality development of the night-time economy.

Quantitative analysis of night-time economy is still at a very slow start due to the lack of theory and the difficulty in obtaining data. Reasonable measurement of night-time economy is the primary task to carry out quantitative research on night-time economy. At present, a few scholars and research institutions have launched a preliminary exploration to the measurement of night-time economy. For example, the evaluation model for the development potential of night-time economy can be constructed from the aspects of city comprehensive strength, the ability of consumers, the business situation, the related industries, the government behavior, uncontrollable factors and so on ^[7]. However, the data are mainly from questionnaire survey, which lack the data to objectively reflect the night-time economic activity. With the development of remote sensing technology, night light data has been proved to be able to effectively reflect the economic development level of a region ^[8]. In addition, from the perspective of spatial geography, the comprehensive index of night economy can be constructed by combining

night light intensity data and POI data to analyze the level of night-time economy^[9].

In the past two years, some Internet enterprises and research institutions, such as Meituan, Alibaba and Jingdong, have released research reports on night-time economy. They use the platform's internal data to assess the vitality of the city's night-time economy from different perspectives. The "Digital Night-time Economy Development Report of Chinese Cities" constructs the digital night-time economy index from four dimensions: the development foundation of the tertiary industry, the infrastructure of night commerce, the circulation vitality of night consumption, and the initiative of the government to develop night-time economy. The "China Night-time Economy Development Quality Index 1.0" constructs the night-time economy development quality index from five dimensions: the richness of night business, the evaluation of night tour products, the activity of night consumption, the maturity of commercial facilities and the evaluation of green development. "Zhejiang Province Night-time Economy Activity Index 1.0" analyzes the difference of night-time economy in time and space from five dimensions: the richness of night business, the convenience of night consumption, the activity of night consumption, the evaluation of night consumption and the maturity of commercial facilities.

3. THEORETICAL FRAMEWORK OF NIGHT-TIME ECONOMY MEASUREMENT

Through the combing of these research reports, we can find that the environmental basis of development, the vitality of night consumption, the convenience of night consumption, the quality evaluation of development and the network attention are the key factors affecting the development level of night-time economy. Environmental basis reflects the macroscopic environment atmosphere of night-time economy development in a city or region, which is the important carrier of its development and the basic condition of night-time consumption. The convenience of consumption reflects the vitality of night-time economy from the perspective of supply side. The convenience of night consumption determines whether the supply of night-time economy can meet the needs of consumers. The core purpose of night-time economy development is to drive night consumption to expand domestic demand. Thus, the vitality of consumption is the most direct embodiment of the prosperity of night-time economy. The vitality of consumption measures the development of night-time economy from the perspective of quantity, while the quality evaluation measures its development from the perspective of quality. We evaluate the quality of night-time economy development from the two dimensions of supply and demand. The network attention of night-time economy, namely the popularity of "night-time economy" on social media, is an important factor to promote the development of night-time economy. Consumption is a kind of social behavior with strong demonstration effect. People imitate each other in their consumption. The publicity and promotion of night-time economy by social media in cyberspace can promote more people to participate in night consumption.

Drawing on the measurement experience of digital economy, this paper constructs a set of theoretical framework suitable for night-time economy measurement. First, refine and define the connotation of night-time economy. Secondly, the theoretical logic of night-time economy development is constructed. The development factors of night-time economy are subdivided into five parts: Environmental Basis, Convenience of Consumption, Vitality of Consumption, Quality Evaluation of Development and Network Attention. Thirdly, the core measurement indexes of night-time economy are screened and the evaluation index system is constructed. Next, the comprehensive evaluation method is determined. Finally, validate the evaluation development system, empirically analyze and objectively evaluate the potential of a region's night-time economy development. Figure 1 shows the theoretical framework of night-time economy measurement.

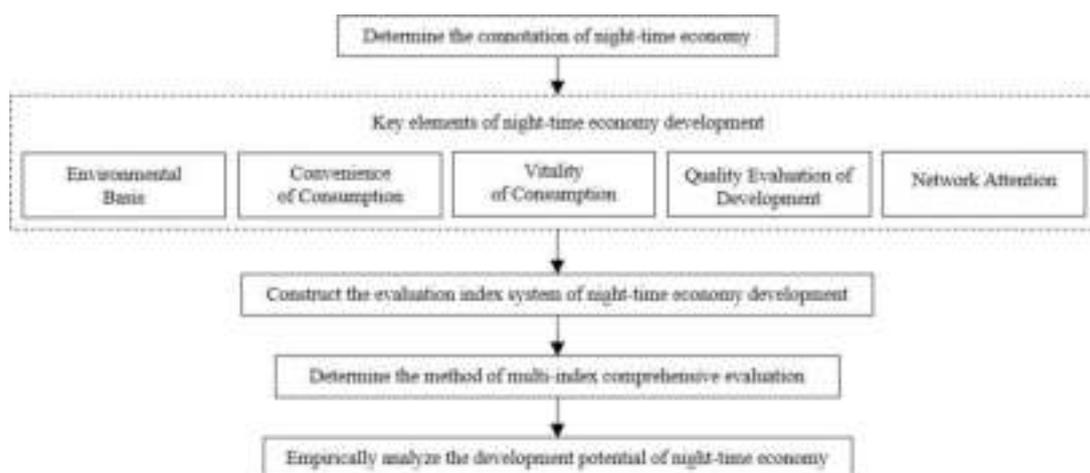


Figure 1. Theoretical framework of night-time economy measurement

4. CONSTRUCTION OF DEVELOPMENT EVALUATION INDEX SYSTEM

There are many factors that affect the development of night-time economy. In the actual measurement of night-time economy, there are usually more qualitative indicators, less quantitative indicators. The development of Internet and digital technology makes it possible to measure night-time economy reasonably.

4.1 Overall framework of the evaluation index system

On the basis of theoretical logic and research report, the evaluation index system of night-time economy development is preliminarily constructed. The index system covers the factors of night-time economy development in multiple dimensions and scenes. And it follows the principles of comprehensiveness, hierarchy and feasibility. Figure 2 shows the overall framework of the evaluation index system.

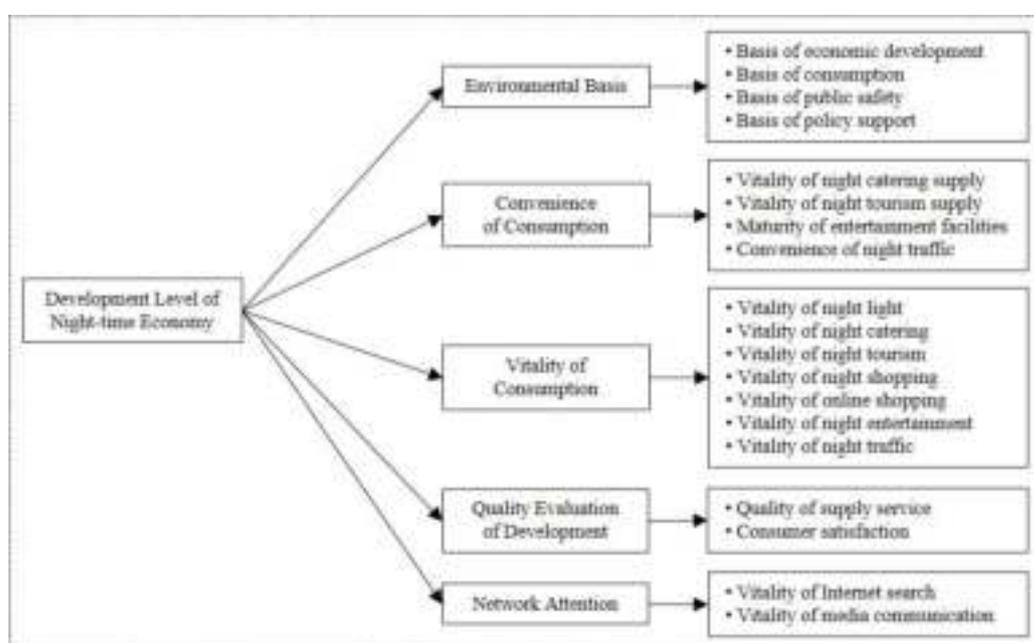


Figure 2. Overall framework of the evaluation index system

4.2 Detailed evaluation index system and index description

Table 1 shows the detailed contents of the evaluation index system.

Table 1. Evaluation index system of night-time economy development

First level indicators	Second level indicators	Description of indicators
Environmental Basis	Basis of economic development	Proportion of added value of tertiary industry
	Basis of consumption	Per capita income and consumption level
	Basis of public safety	Night public safety index
	Basis of policy support	Number of night economic policy releases
Convenience of Consumption	Vitality of night catering supply	Proportion of night catering businesses
	Vitality of night tourism supply	Proportion of night tourism businesses
	Maturity of entertainment facilities	Number of night entertainment facilities
	Convenience of night traffic	Proportion of night public transport
Vitality of Consumption	Vitality of night light	Global night light data
	Vitality of night catering	Proportion of catering consumption at night
	Vitality of night tourism	Proportion of tourism consumption at night
	Vitality of night shopping	Proportion of shopping consumption at night
	Vitality of night entertainment	Proportion of entertainment consumption at night
	Vitality of night traffic	Vitality index of public transport
Quality Evaluation of Development	Quality of supply service	Proportion of high quality night-time merchants
	Consumer satisfaction	Score for the quality of consumption
Network Attention	Vitality of Internet search	"Night-time economy" Internet search index
	Vitality of media communication	Number of information released by media platform

4.2.1 Environmental basis

Environmental Basis is the key element to support night economic development. In this paper, we divide the Environmental Basis into four aspects: basis of economic development, basis of consumption, basis of public safety and basis of policy support. The development level of tertiary industry is the necessary condition for the development of night-time economy. Consumers are the key participants of night-time economic activities. Their consumption level determines their contribution to night-time economy. Good social security and traffic environment can ensure the smooth development of night-time economy. The issue of the policy reflects the government's emphasis on and support for the development of night-time economy.

4.2.2 Convenience of consumption

"Convenience of Consumption" describes the vitality of the supply side at night. Supply of commercial infrastructure and transportation facilities is very important to the development of the night -time economy. Convenience of night traffic can provide the guarantee of traffic travel for night-time economic activities. We can describe the supply of commercial facilities in terms of the vitality of night catering, night tourism and the maturity of night entertainment facilities.

4.2.3 Vitality of consumption

This partly reflects the vitality of consumers' participation in the night-time economy. In the past, night light data could reflect to some extent the intensity of night economic activity. With the development of digitization, we can capture data reflecting users' night consumption behavior from the Internet platform. The vitality of night consumption can be measured from different consumption forms such as night catering, night traveling, night shopping and night entertainment.

4.2.4 Quality evaluation of development

This part evaluate the development of night-time economy from the perspective of quality. "Quality of supply

service" measures the development quality of night-time economy from the supply side, while "Consumer satisfaction" measures the development quality from the consumer side.

4.2.5 Network attention

With the booming development of social networks, the development dynamics of night-time economy can be quickly spread through official and personal media. "Network Attention" reflects the popularity of the "night-time economy" on social media. "Vitality of Internet search" and "Vitality of media communication" related to night-time economy can reflect this online attention.

5. VALIDATION OF EVALUATION INDEX SYSTEM

5.1 Data Collection

The data in this paper mainly includes statistical data, night light data, and data from Internet platforms such as e-commerce and social media. Statistical data are mainly from the national and provincial statistics bureaus. All the statistical data of the study area have been obtained at present. Global Night-time Light Database is derived from raw image data from the National Oceanic and Atmospheric Administration (NOAA). So far, we have taken and corrected night light data in China from 2013 to 2019. About e-commerce platform data, we use Python to crawl relevant data on the life service e-commerce platforms such as Meituan, Dianping and so on. At present, the data of some cities and regions have been obtained. In terms of social media data, "night-time economy" is used as the key word to capture data from social media platforms such as Baidu and Weibo.

5.2 Selection of research scope

As the key city of night-time economy, Beijing is suitable as the object of study in the first stage. Beijing has so far identified the first batch of four landmarks, nine business circles and nine living circles for night-time economy development. Beijing Municipal Bureau of Commerce and Meituan jointly issued the "Consumption Guide for Night City 2020".

5.3 Method of multi-index comprehensive evaluation

Multi-index comprehensive evaluation method is widely used. According to the evaluation indexes, it can be seen that some indexes are suitable for using the hierarchical structure model to abstract the problem, and some evaluation indexes meet the calculation conditions of objective weighting. Therefore, the combination weighting method is adopted to weight the evaluation indicators. The calculation formula of linear weighted combination method is as follows.

$$\theta_j = \sum_{i=1}^k b_i w_{ij} \quad (1)$$

θ_j is the combination ownership of the j-th index, b_i is the weight coefficient of the i-th method, w_{ij} is the ownership of the j-th index obtained by the i-th method.

Specifically, Analytic hierarchy process and Entropy weight method are used to comprehensively evaluate the development level of night-time economy.

5.3.1 Analytic hierarchy process

Analytic Hierarchy Process is a comprehensive evaluation method established by American operations researcher T.L.Saaty in the 1970s. It is mainly divided into the following four steps:

- Establish the hierarchical structure model and form hierarchical organization chart.
- Construct a judgment matrix and compare each element in the same hierarchy pairwise. The values of the judgment matrix reflect the relative importance of the various indicators.
- Hierarchical single sort: The weight value of each index in this level relative to an index in the upper level is calculated. The consistency test is carried out on the judgment matrix.
- Hierarchical total sort: Determine the composite weight of all indicators against the overall target for

overall sorting. The consistency test is carried out on the judgment matrix.

5.3.2 Entropy weight method

Entropy weight method is to determine the weight of each index according to the amount of information contained in it. It is mainly divided into the following steps:

- The original data of each index is normalized.
- Calculate the information entropy value of each index.
- Convert the entropy value of each index to its weight.

5.4 Rationality analysis of evaluation results

This paper constructed the evaluation index system on the basis of the existing night-time economy research report. In terms of the selection of evaluation indexes, it refers to the indexes in the existing research. We will verify the accuracy of the evaluation index system by comparing the comprehensive evaluation results with the analytical data of business reports and socio-economic statistics. If the evaluation results are highly correlated with the validation data, it would indicate that the indicator systems in this paper and other reports can achieve consistent results when assessing the level of night-time economy development in specific areas.

6. CONCLUSIONS

In this digital and intelligent era, big data, cloud computing, artificial intelligence and other digital technologies are deeply integrated with the night-time economy. The digital transformation of night-time economy related industries provides a data basis to quantitatively analyze the development of night-time economy and enriches the dimension of night-time economy quantitative research.

In this paper, we draw on the experience of economic scale measurement such as digital economy to construct the theoretical framework of night-time economy measurement under the digital background. On the basis of analyzing literatures, research reports and policies, the key elements of night-time economy development are identified. Using the method of multi-source data fusion analysis, this paper constructs and verifies the evaluation index system of night economic development. In the selection of evaluation methods, the combination of Analytic hierarchy process and Entropy weight method is used to comprehensively evaluate the development level of night-time economy.

Due to the limitation of time, we have only preliminarily constructed the framework of night-time economy measurement and the evaluation index system, which still needs to be improved with the enrichment of theories and data. At the same time, due to the heavy workload of Internet data collection, this paper does not give the results of night-time economy measurement of Beijing. In the future research, we will first complete the measurement of Beijing's night-time economy. Next, incorporate more research areas into the empirical analysis to optimize the evaluation index system and evaluation methods. And the evaluation result will be compared with the results of more research reports on night-time economy to verify the accuracy of the evaluation results. The measurement of night-time economy is the first step of its quantitative analysis. With the development of theory and practice, the driving mechanism of night-time economy to economic growth and the policy effectiveness analysis will also be included in the research scope.

REFERENCES

- [1] Montgomery, John. (1990). Cities and the art of cultural planning. *Planning Practice & Research*, 5(3), 17-24.
- [2] Mao Zhonggen, long Yanni, ye Xu. (2020). Research progress of night economic theory. *Economic trends*, No. 708 (02), 105-118(in Chinese).
- [3] Bianchini, Franco. (1995). Night cultures, night economies. *Planning Practice & Research*, 10(2), 121-126.
- [4] Olt G , Smith M K , Csizmady A , et al. (2019).Gentrification, tourism and the night-time economy in Budapest's district

- VII – the role of regulation in a post-socialist context. *Journal of Policy Research in Tourism Leisure & Events*, 1-13.
- [5] Weng C C , Wang K L . (2004). Scale and scope economies of international tourist hotels in taiwan. *Tourism Management*, 25(6), 761-769.
- [6] Brands, J. , Schwanen, T. (2014). Experiencing and governing safety in the night-time economy: nurturing the state of being carefree. *Emotion Space & Society*, 11, 67-78.
- [7] Fensha Wu.(2013).Research on the development potential assessment model of “Night-economy” Based on FCE Theory. Ms D Thesis. Dongbei University of Finance and Economics.
- [8] Wu, J. , Wang, Z. , Li, W. , Peng, J. (2013). Exploring factors affecting the relationship between light consumption and gdp based on dmsp/ols nighttime satellite imagery. *Remote Sensing of Environment*, 134(7), 111-119.
- [9] Cui Y , Shi K , Jiang L , et al.(2020) Identifying and Evaluating the Nighttime Economy in China Using Multisource Data. *Geoscience and Remote Sensing Letters*, PP(99):1-5.

Full Research Paper**What Drives Knowledge Payment in the Knowledge Payment Platform?****An Empirical Study on the Purchase Behavior in the Emerging Market***Min Zhang^{1,2*}, Jingyi Zhao^{1,2}, Weiwei Yan^{1,2}*¹Center for Studies of Information Resources, Wuhan University, Wuhan, 430072, China² Center for E-commerce Research and Development, Wuhan University, Wuhan, 430072, China

Abstract: Knowledge payment has become an emerging interactive mode. This research aims to investigate the influential factors driving knowledge demanders' knowledge payment for online knowledge in emerging markets. By integrating flow experience theory and means-end chain theory, this paper reveals the relationships between each characteristic (product characteristics, demander characteristics, and platform characteristics), and knowledge demanders' purchase intention. Hypotheses and theoretical model with product characteristics, demander characteristics, and platform characteristics of paid knowledge were proposed. 413 effective questionnaires were collected. A structural equation modeling was constructed and AMOS was used to illustrate the original estimate value of each path and correlations between items and factors. The results show that flow experience affects path dependence and perceived cost enjoyment, path dependence influences perceived cost performance. For product characteristics, perceived substitutability negatively influences perceived cost performance and purchase intention, and perceived cost performance affects purchase intention. For demander characteristics, perceived cost enjoyment positively affects purchase intention. For platform characteristics, perceived payment risk negatively influences purchase intention, while perceived payment convenience has no significant influence on purchase intention. The findings of the research provide some strategic enlightenment for content producers, knowledge payment platform managers.

Keywords: knowledge payment, purchase behavior, knowledge payment platform, emerging market

1. INTRODUCTION

From bones, tortoise shells, or bamboo slips, to papers and computers, the way people acquire knowledge is constantly changing. In emerging markets, the rapid growth of the user scale makes us pay more attention to the knowledge payment platform. Paying for knowledge is a very popular phenomenon in China. As of June 2020, in China, the scale of online payment users reached 805 million and affected by Novel coronavirus pneumonia, the number of online education users worldwide has grown rapidly^[1]. This shows the huge growth potential of knowledge payments in emerging markets such as China.

In emerging markets, knowledge, and skills sharing, as a new business form, will inevitably encounter difficulties and challenges. The emerging-market has the following characteristics: (1) In emerging markets, knowledge payment markets are highly competitive, and the free alternatives for paid knowledge content do exist widely due to the lack of copyright awareness. The demand for knowledge in emerging markets is strongly influenced by the value of free sharing^[2]; (2) In emerging markets, many of the consumers are price-sensitive, users may not be willing to pay for online content^[3]. Therefore, it is very important to study the purchase intention of users for online knowledge products in emerging markets.

Previous research on knowledge payment behavior involves four main research objects: content producers, knowledge products, knowledge payment platforms, and knowledge demanders, and these studies mainly focus on the characteristics of a single research object. There are few studies on the influencing factors of knowledge payment behavior from the characteristics of multiple research objects. Therefore, this paper constructs a

* Corresponding author. Email: zhangmin@whu.edu.cn (Min Zhang)

research model to investigate the purchase intention of users to online knowledge products in the emerging market environment. Multi-attributes, including product characteristics, knowledge demanders' characteristics, and platform characteristics have been taken into consideration to investigate essential factors that drive users' intention to pay for online knowledge. This study provides a new insight concentrating on emerging market users and may enrich existing literature in knowledge payment behavior.

2. LITERATURE REVIEW

As for knowledge producers, many scholars focus on the motivations of participants to continue to contribute to the online knowledge community^[4]. Jin et al.^[5] have studied the influencing factors of knowledge sharing from the social exchange theory perspective. We found that the research on knowledge providers is more focused on the exploration of the motivation of knowledge sharing. This kind of research is not highly relevant to our research. As for knowledge products, As an intangible information product, under the circumstances that the quality of knowledge products may not be judged directly, it is worth studying that what factors will affect consumers in the process of purchase decision-making. Zhu and Zhang^[6] examine the type of knowledge payment and the regulating effect of price. In emerging markets, previous studies paid less attention to the substitutability of knowledge products. The widespread of free content on the Internet indicates that despite a strong user base and huge market potential of knowledge payment in emerging markets, low-cost or free alternatives may be more attractive to consumers than knowledge payment. As for knowledge demanders, some researchers focused on the factors that drive consumers' online knowledge purchase intention based on the cognitive-affective-conative framework and customer value theory^[7]. Ding^[8] discussed the influence of hopes and anticipated regret on previous user behavior and user satisfaction. However, less attention has been paid to the perceived cost enjoyment of knowledge payment. As for knowledge payment platforms, payment activity in the platform plays a significant role as the last step of purchase behavior and users have to bear uncertainty and risk in this step. Considering the quality of system and knowledge, Zhou^[9] explored the influence of environmental factors on users' willingness for continuous usage. Previous studies have shown that platform characteristics do affect users' knowledge payment behavior. However, we still want to know more about the influence factors of platform characteristics on knowledge payment behavior in emerging markets.

Although there have been many researches on online knowledge payment behavior from different perspectives in the past, the research on willingness to pay for knowledge still needs a more comprehensive perspective. This research aims to explore from multiple angles which factors affect user knowledge payment behavior in emerging markets: (1) We focus on the impact of product characteristics on payment behavior in emerging markets, especially the two factors of perceived substitution and perceived cost performance; (2) We intend to study the impact of Perceived cost enjoyment on the behavior of knowledge payment from the perspective of the demander characteristics; (3) The payment process is the key point of the last step of the purchase behavior, especially in emerging markets, should consider both the payment risk and convenience factors. Therefore, we hope to find out the influence of product characteristics, knowledge demander characteristics, and platform characteristics on knowledge payment behavior, which may make the research results interesting and valuable.

3. HYPOTHESES DEVELOPMENT

3.1 Path dependence

Research in both marketing and economics proposes that consumers are more likely to choose products that they have already purchased. Kuo et al.^[10] also revealed that consumers' inertia would affect repeat purchase intention. A sunk cost is the major cause of path dependence, each transformation of behavior needs

corresponding financial and emotional cost. So, this research defines path dependence (PD) as knowledge demanders' reliance on the knowledge payment platform. In this paper, perceived cost-performance is defined to describe the cost-performance ratio in knowledge acquisition. Knowledge demanders who are used to and relied on the knowledge payment platform will financially and emotionally cost less, which means they will perceive more cost performance. Therefore, the hypothesis is as follows:

H1. Path dependence will positively affect the perceived cost performance of paid knowledge.

3.2 Flow experience

Based on consumer behavior theory, psychological factors are significant constructs to affect consumers' behavior, which has also been applied to the adoption and usage of information technology (IT). Hoffman and Novak firstly applied flow theory to the online environment and found that various online behavior such as online gaming, online shopping, and email will produce various flow experiences^[11]. According to research, flow experience will reach a peak when people seek information online which is followed by online reading and writing. This research considers flow experience (FE) as a motivational paradigm to explain users' knowledge acquisition behavior. We define flow experience as users' total involvement, concentration, and absorption in online knowledge acquisition. In conclusion, flow experience was proposed to positively affect knowledge demanders' dependence and perceived cost enjoyment towards knowledge acquisition. The hypotheses are as follows:

H2. Flow experience will positively affect the perceived cost enjoyment of knowledge payment.

H3. Flow experience will positively affect knowledge demanders' path dependence.

3.3 Product characteristics

3.3.1 Perceived substitutability

Perceived substitutability (PS) refers to individual perception of whether a specific product or service can be substituted by existing others^[12]. When a product is considered as a substitute for another one, it means they share similar physical or functional properties. More competitors in the market provide more opportunities for people to choose, and then higher substitutability appears. In the knowledge payment context, Perceived substitutability is understood as the extent to which people think they can easily find free or cheap resources. Higher perceived substitutability suggests that knowledge demanders would be more likely to find free or cheaper content from other approaches. In this research, we propose perceived substitutability will negatively affect perceived cost performance and purchase intention, so the hypotheses are as follows:

H4. Perceived substitutability of paid knowledge will negatively affect the perceived cost performance of paid knowledge.

H5. Perceived substitutability of paid knowledge will negatively affect the purchase intention of paid knowledge.

3.3.2 Perceived cost performance

Value involves higher-level goals that motivate and direct consumers' decision-making processes. According to the exchange theory in marketing researches, perceived value is considered as a precondition of purchase intention^[13]. In this study, we regard perceived value as an important dimension, and define it as knowledge demanders' overall assessment of paid knowledge. Moreover, in our research context, we integrate quality, cost and emotion into perceived cost performance (PCP). We consider that if knowledge demanders perceive higher cost performance, they will be more willing to pay for it. Thus, we propose:

H6. Perceived cost performance of paid knowledge will positively affect the purchase intention of paid knowledge.

3.4 Demander characteristics

Perceived cost enjoyment (PCE) is defined to describe knowledge demanders' emotional satisfaction with online knowledge concerning the cost that knowledge demanders' have to pay. Although cost enjoyment has not been conceptualized in a previous study, we can find this kind of description in practice. For instance, eBay applies "ratios of enjoyment per cost" to vividly express the flow experience of consumers during online shopping. Besides, some studies use perceived playfulness to present the enjoyment an individual perceives in the process of using mobile services^[14]. We hold the idea that when knowledge demanders get a sense that they get such great joy and fulfillment from online knowledge than the money they spend on it, they will have a great willingness to purchase paid knowledge in the knowledge payment platform. Hence, we propose:

H7. Perceived cost enjoyment of paid knowledge will positively affect the purchase intention of paid knowledge.

3.5 Platform characteristics

3.5.1 Perceived payment convenience

Payment activity plays a significant role as the last step of purchase behavior, and users have to bear uncertainty and risk in this step. Therefore, the dimension of perceived payment characteristics was put forward in this research. Brown Lew^[15] defined convenience from five dimensions namely time, place, acquisition, use, and execution. In detail, time dimension means providing the product when the consumer wants it; place dimension means product placement can bring more convenience to consumers; acquisition convenience means it is easy for consumers to access to products, usage convenience means the products are easy to use and execution convenience means someone provided the products. This paper adopts the usage dimension as the research variable since our research focuses on purchase intention analysis rather than payment service adoption. Thus, we propose:

H8. Perceived payment convenience of paid knowledge will positively affect the purchase intention of paid knowledge.

3.5.2 Perceived payment risk

The risk was firstly defined in the decision-making field as users' understanding of possible results and hazards, and then consumer behavior researchers proposed the concept of perceived risk. Previous studies have shown that perceived risk is one of the main factors influencing consumers' payment for finance-related mobile services^[16]. Therefore, it is appropriate to consider perceived payment risk as a construct in this study, and we only take economic risk into account for the major concern that consumer is financial loss during the process of payment. Therefore, we propose hypotheses as follows:

H9. Perceived payment risk of paid knowledge will negatively affect purchase intention toward paid knowledge.

The research model is shown in Figure 1.

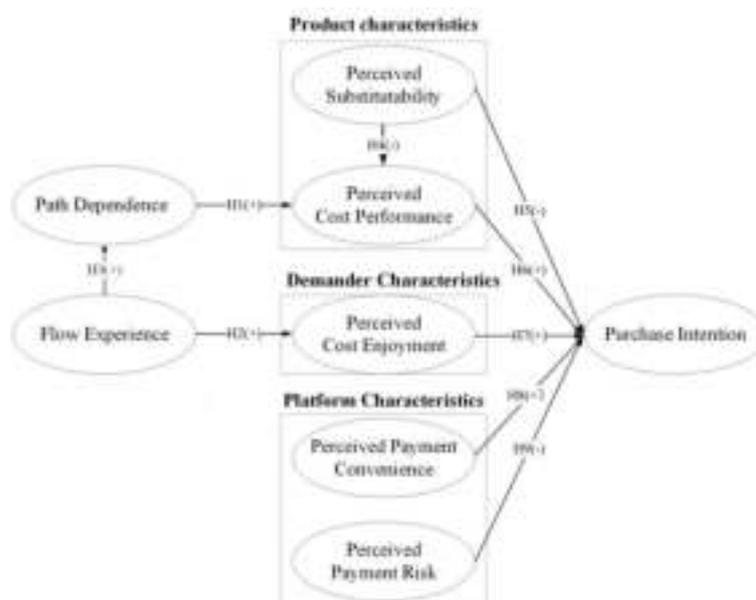


Figure 1. Research model

4. RESEARCH METHODOLOGY

4.1 Measure instrument

Our study builds on prior work, a five-point Likert scale is adopted to measure the constructs. In the respect of the measurement, items are created in English, we use a back-to-back translation to develop the items into Chinese by inviting three experts in the relating research fields. As shown in Table 1, every variable is measured by multi-item scales adapted from validated measures in past studies.

Table 1. Survey instrument

Variables	Item	Wording
Path dependence ^[10]	PD1	When I intend to acquire knowledge, my mobile devices are always my first choice.
	PD2	When I intend to acquire knowledge, switching to a new app would be bother.
	PD3	The cost of time for switching to another knowledge payment platform is high.
	PD4	The cost of effort for switching to another knowledge payment platform is high.
Flow experience ^[17]	FE1	I pay less attention to my surroundings while I am acquiring online knowledge.
	FE2	I can't feel the time passing while I am acquiring online knowledge.
	FE3	I'm very focused while I am acquiring online knowledge.
	FE4	I am totally absorbed while I am acquiring online knowledge.
Perceived substitutability ^[12]	PS1	When I need to pay for online knowledge, I would try to seek free resources from webpage.
	PS2	When I need to pay for online knowledge, I would try to seek free resources from other apps.
	PS3	When I need to pay for online knowledge, I prefer free resources from other approaches.
Perceived cost performance ^[13]	PCP1	I think the price of online knowledge is cost-effective.
	PCP2	I think the price of online knowledge is reasonable.
	PCP3	Overall, I think the cost performance of online knowledge in app is high.
Perceived cost enjoyment ^[13]	PCE1	Compared to the cost, I pay more attention to the joy online knowledge brings to me.
	PCE2	Considering the pleasure I get from Knowledge acquisition, the price of online knowledge is cheap.
	PCE3	In general, I think the cost enjoyment of online knowledge is high.
Perceived payment convenience ^[18]	PPC1	I think mobile payment is simple and easy to use.
	PPC2	I think mobile payment saved my time and effort.
Perceived payment risk ^[19]	PPR1	Transaction may disclose my personal private information.
	PPR2	Transaction may result in my finial loss.
Purchase intention ^[10]	PI1	I intend to pay for online knowledge.
	PI2	I think I would pay for online knowledge.
	PI3	I plan to pay for the online knowledge.

To make sure that the questionnaire is accurately expressed, the questionnaire was initially tested. In addition, we invited three experts whose main or research interests are user behaviors of online knowledge platforms to modify the questionnaire.

4.2 Questionnaire design and data analysis method

Knowledge demanders who have online knowledge acquisition experience on knowledge payment platforms were recruited as the subjects of this study. The questionnaire contains two sections. The first section consists of nine questions to collect demographic information. The second section consists of eight variables, including twenty-four items that measured the constructs, to assess the participants' preference and habits for

knowledge payment platforms. This study contains 8 constructs and 24 items, so the number of questionnaires gathered should be over 240, which is consistent with the theoretical effective sample size.

We chose the Chinese market as our empirical case study for two reasons. First, the China market is a typical emerging market. Second, the knowledge payment industry of China has huge market demand and broad market prospects. We first contacted 500 knowledge payment platform knowledge demanders and asked if they had the willingness to participate in our survey. Also, we sent out a questionnaire to 465 mobile knowledge demanders who were interested in our survey. After scrutinizing the responses, we removed incomplete and completely similar answer questionnaires. Finally, we got 413 valid questionnaires. Furthermore, non-response bias was examined by comparing the means of all constructs and demographics for early and late participants.

We used Confirmatory Factor Analysis (CFA) and Structural Equation Modeling (SEM) to test the relationship between factors and the project. In the process of SEM analysis, latent variables are displayed as ellipses while manifest variables are depicted as rectangles, and One-way arrows imply a hypothesis relationship between two factors. SEM includes two parts. One is a measurement model to evaluate the relations between latent variables and variables from the factor analysis. The other is the structural model to assess the relations between the latent variables.

5. RESULTS

5.1 Respondents' profiles

The demographics of our dataset are shown in Table 2. The main reason to form this kind of distribution is that the main part of mobile knowledge demanders in China is young people. They have a relatively high rate of acquiring online knowledge habits for their study and research. The data in Table 2 showed that knowledge payment platforms have widely target groups, and this kind of knowledge acquisition style is especially popular in the 18~40 age group. Furthermore, the frequency and duration of acquiring online knowledge are relatively high in this group. But we also notice that willingness to pay for knowledge content is relatively low.

Table 2. Demographic profile of respondents

Measure	Items	Frequency	Percentage
Gender	Male	196	0.47
	Female	217	0.53
Age	Under 18	3	0.01
	18 ~ 30	374	0.91
	31 ~ 40	28	0.07
	Over 40	8	0.02
Education	Junior college or less	23	0.06
	Undergraduate	255	0.62
	Master or above	135	0.33
Using frequency in the knowledge payment platform	Seldom (several times a month)	98	0.24
	Often (several times a week)	98	0.24
	Usually (once or twice a day)	48	0.17
	Always (several times a day)	150	0.36
Duration of each using	Under 15 minutes	84	0.20
	15 ~ 30 minutes	180	0.44
	30 ~ 60 minutes	93	0.23
	1 ~ 2 hours	35	0.08

Measure	Items	Frequency	Percentage
	Over 2 hours	21	0.05
Using occasion	On the car	150	0.36
	Bedtime	231	0.56
	Noontime	113	0.27
	At any time	211	0.51
Knowledge contents	Novels	214	0.52
	Magazines	124	0.30
	Academic materials	109	0.26
	News	310	0.75
	Comics	28	0.07
Purchase times of online knowledge	Never	264	0.64
	1 ~ 2	87	0.21
	3 ~ 5	36	0.09
	6 ~ 10	12	0.03
	Over 10	14	0.03

5.2 The measurement model

The measurement model should be examined before the structural model because its result can reflect the desired constructs or factors under this study. We use CFA to test the reliability, convergence validity and discriminant validity to analyze the measurement characteristics of constructs. The properties of the measurements and the descriptive statistics are presented in Table 3. The Cronbach's Alpha value of constructs ranges from 0.864 to 0.947, which shows the high (0.70-0.89) reliability of the questionnaire.

Table 3. Item reliability statistics

Variables	Item	Factor loading	AVE	CR	Cronbach's alphas
Path dependence	PD1	0.874	0.619	0.864	0.864
	PD2	0.901			
	PD3	0.679			
	PD4	0.663			
Flow experience	FE1	0.835	0.713	0.908	0.908
	FE2	0.866			
	FE3	0.818			
	FE4	0.857			
Perceived substitutability	PS1	0.867	0.812	0.928	0.928
	PS2	0.910			
	PS3	0.925			
Perceived cost performance	PCP1	0.867	0.764	0.907	0.905
	PCP2	0.884			
	PCP3	0.871			
Perceived cost enjoyment	PCE1	0.816	0.742	0.896	0.895
	PCE2	0.877			
	PCE3	0.889			

Variables	Item	Factor loading	AVE	CR	Cronbach's alphas
Perceived payment convenience	PPC1	0.908	0.812	0.896	0.896
	PPC2	0.894			
Perceived payment risk	PPR1	0.944	0.869	0.930	0.929
	PPR2	0.920			
Purchase intention	PI1	0.885	0.841	0.941	0.947
	PI2	0.932			
	PI3	0.933			

It can be assessed by the loading of two factors: Average Variance Extracted (AVE) and Composite Reliability (CR). Values of composite AVEs and CRs are considered adequate, with AVEs all above 0.6 and CRs all above 0.85. Therefore, the result satisfying convergent validity indicates that all items and variables can be considered to be acceptable and are reserved for the following research. Through comparing the square root of the AVE about a given factor and the correlations between the factors, discriminant validity can be assessed. The square root of AVE of each construct is higher than correlations between this construct with others (Table 4). This result manifests that each construct shares a higher variance with items in its factor.

Table 4. Correlation matrix of the constructs

	PD	FE	PS	PCP	PCE	PPC	PPR	PI
PD	0.787							
FE	0.384	0.844						
PS	0.135	0.352	0.901					
PCP	0.107	-0.01	-0.142	0.874				
PCE	0.09	0.233	0.082	-0	0.861			
PPC	0.081	0.21	0.408	-0.06	0.049	0.901		
PPR	-0.073	-0.19	-0.324	0.042	-0.044	-0.844	0.931	
PI	0.073	0.089	-0.058	0.356	0.381	0.139	-0.185	0.917

5.3 The structural model

A set of indices of fit goodness can be used to assess the consistency such as CFI, IFI, TLI, and RESEA, etc. The results show in Table 5, excellent fit indices with CFI 0.954 and RMSEA 0.06. In conclusion, the reliability, convergent validity, discriminant validity, and the model fit all suggest that the measured variables explain much of the variances of corresponding latent constructs.

Table 5. Model fit summary

CMIN/DF	2.498	CFI	0.954
GFI	0.898	RFI	0.914
AGFI	0.871	IFI	0.954
NFI	0.926	RMSEA	0.06

We present the model testing result by combining the path coefficient and the corresponding significance level. When the hypothesis of two constructs was effective, a larger path coefficient means stronger relations. Then, according to Figure 2, the final path coefficients and hypothesis testing were showed in Table 6. All the hypotheses except H8 were statistically significant.

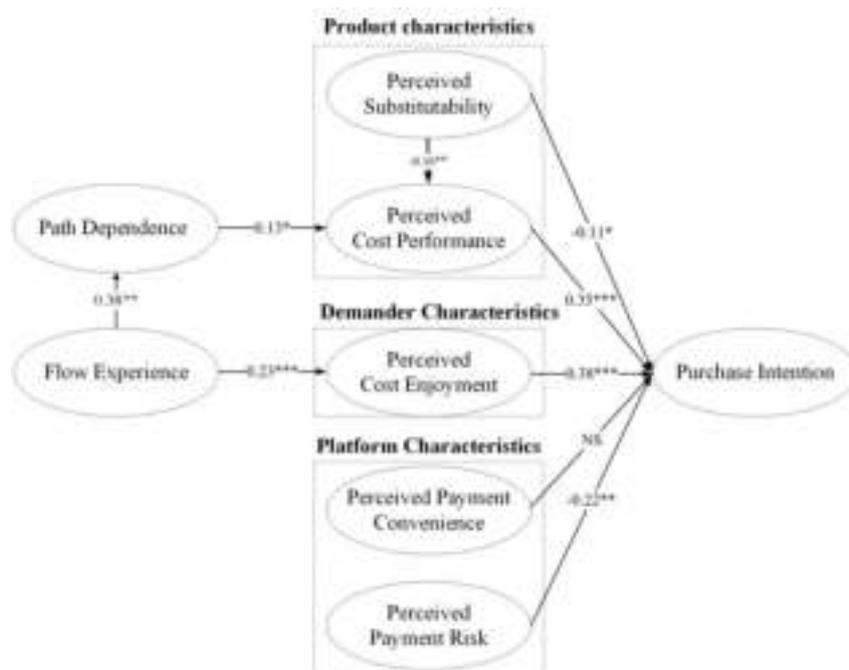


Figure 2. Results of research model

Table 6. Path coefficients and hypothesis testing

Path	Hypothesis	Path Coefficient	P-Value	Results
FE→PCE	H1	0.384	***	Supported
FE→ PD	H2	0.233	***	Supported
PD→PCP	H3	0.129	0.025	Supported
PS→ PCP	H4	-0.160	0.005	Supported
PS→PI	H5	-0.110	0.036	Supported
PCP→PI	H6	0.350	***	Supported
PCE→ PI	H7	0.381	***	Supported
PPC→ PI	H8	/	0.999	Not Supported
PPR→ PI	H9	-0/219	0.033	Supported

Notes: *p<0.05, **p<0.01, ***p<0.001.

6. DISCUSSION

6.1 Influence factors of purchase intention in the knowledge payment platform

We built a theoretical model to explain users' purchase intention of paid knowledge in the knowledge payment platform by integrating flow experience, path dependence, product characteristics, demander characteristics, and platform characteristics. To be more specific, as for product characteristics of paid knowledge, perceived substitutability has negative effects on perceived cost performance (**, p=) and purchase intention. It is consistent with the status quos of emerging markets and the real environment in China. However, the result that perceived cost performance positively affects purchase intention illustrates knowledge demanders are still economically rational and always chase maximum utility.

For demander characteristics of paid knowledge, the factor of perceived cost enjoyment has the highest direct effects on users' purchase intention. The meaning of perceived cost enjoyment in this study does not merely refer to enjoyment, but is "ratios of enjoyment per cost". The result illustrates knowledge demanders' enjoyment

feeling is the key to encourage their purchase intention, especially after comparing to the money they spend on online knowledge. It additionally supports enjoyment as a vital influencing factor to purchase intention in the online context^[14].

For platform characteristics of paid knowledge, perceived payment risk is proved to be a significant influential factor of purchase intention. The main worry is about the possibility that their private financial information might be divulged or be abused. This finding is by previous works^[16]. This research shows that perceived payment convenience has not significant effect on consumers' purchase intention. Although users may have some payment risk concerns, mobile payment has been applied and accepted in tremendous scenarios in China. Consumers, especially the young users of knowledge payment platforms, get used to mobile payment, and almost have no perceived complexity of payment. Hence, they may not be affected by transactional convenience.

Finally, the results show flow experience and path dependence do exist in knowledge acquisition. Flow experience presents a positively direct influence on path dependence and perceived cost enjoyment. Results reveal that the more joyfulness acquired from the knowledge acquisition process, the higher the hedonic value knowledge demanders perceived. While flow experience also shows the positive indirect impact on knowledge demanders' purchase intention towards online knowledge which is following the findings that flow experience can result in behavioral response. Additionally, once knowledge demanders form special inertia towards online knowledge acquisition, they will consider knowledge payment platforms more valuable and will adopt this kind of repeated behavioral pattern.

6.2 Theoretical and practical implications for knowledge payment

Consumers' purchase intention of paid knowledge in the knowledge payment platform makes great contributions to all the service providers in the online knowledge payment industry, especially knowledge demanders in emerging markets (such as China) greatly influenced by the free sharing values^[2]. We stood on the position of knowledge demanders and constructed knowledge demanders' perceived value with Chinese consumption characteristics by defining the concept of perceived cost performance and perceived cost enjoyment.

Project characteristics and demander characteristics of online knowledge reflect knowledge demanders' cognition and emotion toward knowledge in online knowledge, and platform characteristics present the last and important process of real purchase behavior. It is helpful to examine knowledge demanders' purchase intention of online knowledge, rather than the normally applied Technology Acceptance Model (TAM) framework in examining usage and continuance behavior intention^[20]. Obviously, the three kinds of characteristics all significantly contribute to knowledge demanders' knowledge payment behavior. Some managerial implications of this study might be helpful for the providers of online knowledge service to make their online knowledge service more knowledgeable and profitable.

First, knowledge producers of paid knowledge should focus on providing high-quality and original knowledge-based content and strengthen copyright protection. Meanwhile, so as to adapt to the immature and imperfect protection environment of intellectual property rights in an emerging market and reduce perceived substitutability, the knowledge producers must enforce strict operational processes to prevent the disclosure of original content.

Second, knowledge payment platform developers should enhance the quality of the platform's visual interface design to enhance users' experience. Our results show that demander characteristic (perceived cost enjoyment) is a key driving force towards purchase intention and flow experience as an antecedent of it. Providing a convenient and comfortable online environment for knowledge needs can further create an excellent process experience, thereby further enhancing the positive emotions of knowledge needs.

Finally, since the results show that platform characteristic (perceived payment risk) has significant effects on purchase intention, knowledge payment platform managers should provide trustworthy payment approach to grantee the convenience and safety of payment activity, especially some influential third-party payment platforms. Although the effect of perceived payment convenience on purchase intention was not significant in this research, the indisputable fact is that the payment environment is greatly important during the process of users making purchase decisions, especially in other emerging markets where mobile payment is not universally accepted.

7. CONCLUSIONS

In this paper, flow experience theory and means-end chain (MEC) theory are integrated into the context of online knowledge acquisition to develop hypotheses and theoretical research model. Specifically, perceived cost performance and perceived cost enjoyment are two significant positive factors while perceived substitutability and perceived payment risk are critical negative indicators. In addition, flow experience can positively influence path dependence and perceived cost enjoyment while path dependence in turn shapes perceived cost performance. The results of this study have certain practical significance, which can provide real insights for understanding the characteristics of knowledge demanders, and provide a basis for the online marketing of knowledge payment platforms and the provision of related services.

Although this paper provides meaningful findings of knowledge demanders' purchase intention of paid knowledge, there are some points that require further research. First, most of our research samples are people from high education background and young users. Future research can focus on other age groups and people with lower education levels, exploring different influencing factors about pay for online knowledge. Additionally, although our research is developed with multi-attributes, we emphasize more on intrinsic motivations. Therefore, the research elements can also be expanded from an extrinsic perspective.

ACKNOWLEDGEMENT

This study is supported by the National Natural Science Foundation of China [Grant Number: 72074174], the Humanity and Social Science Foundation of Ministry of Education of China [Grant Number: 18YJA870016] and the Fundamental Research Funds for the Central Universities.

REFERENCES

- [1] CNNIC. (2020). The 46th China Statistical Report on Internet Development.
<http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202009/P020200929546215182514.pdf>
- [2] HAN Y G J, ZHANG Z. (2016). Factors Influencing Payment for Content of Mobile Internet in Chinese Background—Empirical Study of APP Store from a TPB Perspective. *Journal of Dalian University of Technology*, 37(3):55-61.
- [3] BERGER B, MATT C, STEININGER D M et al. (2015). It Is Not Just About Competition with “Free”: Differences Between Content Formats in Consumer Preferences and Willingness to Pay. *Journal of Management Information Systems*, 32(3):105-128.
- [4] CHEN L T, BAIRD A, STRAUB D. (2019). Why do participants continue to contribute? Evaluation of usefulness voting and commenting motivational affordances within an online knowledge community. *Decision Support Systems*, 118:21-32.
- [5] JIN J H, LI Y J, ZHONG X J et al. (2015). Why users contribute knowledge to online communities: An empirical study of an online social Q&A community. *Information & Management*, 52(7):840-849.

- [6] ZHU X, ZHANG W. (2019). An Empirical Research on the Effect of Free Knowledge in the Knowledge Payment Platform. Proceedings of the 2019 2nd International Conference on Information Management and Management Sciences. Chengdu, China: Association for Computing Machinery, 80–85
- [7] SU L, LI Y, LI W. (2019). Understanding Consumers' Purchase Intention for Online Paid Knowledge: A Customer Value Perspective. *Sustainability*, 11(19):5420.
- [8] DING Y. (2018). Modelling continued use of information systems from a forward-looking perspective: Antecedents and consequences of hope and anticipated regret. *Information & Management*, 55(4):461-471.
- [9] ZHOU T. (2018). Understanding online knowledge community user continuance: A social cognitive theory perspective. *Data Technologies and Applications*, 52(3):445-458.
- [10] KUO Y F, HU T L, YANG S C. (2013). Effects of inertia and satisfaction in female online shoppers on repeat-purchase intention The moderating roles of word-of-mouth and alternative attraction. *Managing Service Quality*, 23(3):168-187.
- [11] HOFFMAN D L, NOVAK T P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing*, 60(3):50-68.
- [12] ZHANG N, GUO X H, CHEN G Q. (2011). Why adoption and use behavior of IT/IS cannot last?-two studies in China. *Information Systems Frontiers*, 13(3):381-395.
- [13] WANG Y-S, YE H C-H, LIAO Y-W. (2013). What drives purchase intention in the context of online content services? The moderating role of ethical self-efficacy for online piracy. *International Journal of Information Management*, 33(1):199-208.
- [14] HSIAO K-L, CHEN C-C. (2016). What drives in-app purchase intention for mobile games? An examination of perceived values and loyalty. *Electronic Commerce Research and Applications*, 16:18-29.
- [15] BROWN LEW G. (1990). Convenience in Services Marketing. *Journal of Services Marketing*, 4(1):53-59.
- [16] SHIN D-H. (2009). Towards an understanding of the consumer acceptance of mobile wallet. *Computers in Human Behavior*, 25(6):1343-1354.
- [17] LEE M-C, TSAI T-R. (2010). What Drives People to Continue to Play Online Games? An Extension of Technology Model and Theory of Planned Behavior. *International Journal of Human-Computer Interaction*, 26(6):601-620.
- [18] KIM C, MIRUSMONOV M, LEE I. (2010). An empirical examination of factors influencing the intention to use mobile payment. *Computers in Human Behavior*, 26(3):310-322.
- [19] LIEBANA-CABANILLAS F, SANCHEZ-FERNANDEZ J, MUNOZ-LEIVA F. (2014). The moderating effect of experience in the adoption of mobile payment tools in Virtual Social Networks: The m-Payment Acceptance Model in Virtual Social Networks (MPAM-VSN). *International Journal of Information Management*, 34(2):151-166.
- [20] HU J M, ZHANG Y. (2016). Understanding Chinese Undergraduates' Continuance Intention to Use Mobile Book-Reading Apps: An Integrated Model and Empirical Study. *Libri*, 66(2):85-99.

Short Research Paper

Research on Effects of Paid Search Advertising Based on Customer loyalty Segmentation

Jiixin Chen¹, Li Li^{1,2}, Weiwei Liao^{1*}

¹ School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

² Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094, China

Abstract: For e-commerce enterprises, it can help enterprises adjust their advertising strategies more effectively by determining the impact of paid search advertising on different customer groups. Therefore, this paper constructs a comprehensive evaluation model of immediate effect and carry-over effect of paid search advertising, and empirically discusses the impact of paid search advertising on different loyalty-based market segments. The results show that paid search advertising has the greatest impact on high-loyalty customers, while the carry-over effect on potential customers is longer. Thus, we can find that the role of paid search advertising is not only to attract new customers, but also to maintain high-loyalty customers.

Keywords: Loyalty-Based Market Segment, Paid Search Advertising, Advertising Effect Evaluation, Immediate Effect, Carry-over Effect

1. INTRODUCTION

In recent years, the penetration of Internet in marketing and e-commerce environment has affected the development of commerce. Many e-commerce enterprises also maintain and compete for market share better through online advertising. Among them, paid search advertising is the advertising displayed in response to user's query on the search engine, which is more in line with the needs of consumers and is not easy to cause aversion. This also makes paid search advertising gradually develop into an important part of enterprise marketing strategy. In order to allocate enterprise resources scientifically, Assael and Lipstein^[1] proposed that the loyalty segment must be used to test the market and determine the response of the segment to marketing variables (such as advertising). This can help enterprise managers to identify the target customer groups of advertising, so that they can adjust the advertising strategy more effectively according to the marketing objectives of the enterprise. Therefore, if an e-commerce enterprise wants to better maintain loyal customers, acquire new customers and improve marketing efficiency, it needs to more accurately evaluate the impact of paid search advertising on loyalty-based market segments.

There are many kinds of effects in online advertising. From the perspective of time characteristics, online advertising effects include immediate effect (short-term) and carry-over effect (long-term)^[2-3]. However, in the existing research, most scholars only measure the immediate effect of online advertising, and do not consider the widespread carry-over effect of online advertising^[3-6]. This may lead to the underestimation of the effect of online advertising, which leads to the unreasonable allocation of advertising budget. Therefore, in order to accurately grasp the advertising effect of paid search, it is necessary to comprehensively evaluate the immediate effect and carry-over effect of advertising. In addition, in the research involving the evaluation of advertising effect, the research tends to focus on the overall impact of advertising on the whole market, without distinguishing the differences in the impact of advertising on customer groups in different market segments, and it may lead to the deviation of enterprises' cognition of advertising effect. At present, there is no research to

* Corresponding author. Email: lily691111@njust.edu.cn (Li Li), chenjiixin@njust.edu.cn (Jiixin Chen)

discuss the difference of paid search advertising effect among different customer groups with loyalty as a market segmentation variable, which is also worthy of further exploration.

Under the current Internet environment, e-commerce enterprises can obtain more data about customer behavior, which enables us to better quantify the impact of paid search advertising. Therefore, based on the comprehensive evaluation of the immediate effect, the duration and cumulative intensity of the carry-over effect of paid search advertising, this paper constructs a comprehensive evaluation model of paid search advertising effect. Then, we empirically verify the model based on the real-time behavior data and transaction data of an online insurance agent platform to distinguish the impact of paid search advertising on different loyalty-based customer segments. The comprehensive evaluation model and method proposed in this paper can help enterprises identify their target customer groups of paid search advertising, and reasonably adjust the advertising strategy to improve the effectiveness of budget allocation.

2. RELATED LITERATURE

2.1 Evaluation of paid search advertising effect

On the issue of evaluation of paid search advertising effect, the existing research mainly focuses on the evaluation of immediate effect and carry-over effect of paid search advertising. The immediate effect of paid search advertising refers to the influence of customers on their website visit and purchase behavior in a short period of time after paying per click search advertising. Research on this aspect ^[7] is mainly based on the click through rate, conversion rate and other statistical indicators of paid search advertising to reflect the immediate effect of paid search advertising. The specific performance of the carry-over effect of paid search advertising is the impact of customers' click behavior in the current period on their later advertising click behavior and purchase behavior ^[8]. Based on this, scholars further study the carry-over effect of paid search advertising. Rutz and Buklin ^[9] use the Nerlove-Arrow goodwill model to evaluate the effect of paid search advertising, and construct a dynamic Bayesian linear model to verify the long-term impact of paid search advertising. However, this kind of research mainly focuses on the strength and long-term impact of paid search advertising carry-over effect, and cannot determine the duration of carry-over effect. In order to solve this problem, Breuer and Brettel ^[10] introduced the lag influence factor of advertising into the direct aggregation model of online advertising effect evaluation, and explored the duration and influence intensity of online advertising carry-over effect. The research compares the immediate effect and carry-over effect of banner advertising, paid search advertising, price comparison advertising and coupon advertising, and finds that the duration of carry-over effect of paid search advertising is the longest.

Through the review of the related research on the effect of paid search advertising, it is found that there are few research results on the aspect of carry-over effect, which may lead to the underestimation of the effectiveness of paid search advertising. In order to evaluate the effectiveness of paid search advertising more accurately, it is necessary to evaluate the effectiveness of paid search advertising comprehensively from the two perspectives of immediate effect and carry-over effect.

2.2 Advertising effect evaluation based on customer segmentation

In the past research on advertising effect evaluation, scholars usually focus on the overall impact of advertising on the whole market, but do not distinguish the difference of the impact of advertising on different customer groups in different market segments, which may lead to the deviation of enterprises' cognition of advertising effect. In the traditional environment, scholars have verified that loyalty is a useful segmentation variable in the study of advertising effect. Tellis ^[11] finds that the influence of repeated advertising exposure is different among customers with different loyalty, which indicates that the influence of advertising on purchase volume is moderated by loyalty.

In contrast, the research on the influence of online advertising effectiveness among different customer groups in e-commerce environment is relatively less. Blake et al. [7] explored the influencing factors of paid search advertising effect and the differences among new customers, inactive customers and active customers based on the data obtained from a series of large-scale field experiments on eBay. It is found that paid search ads with non brand keywords have a positive impact on the purchase behavior of new customers and inactive customers, but no significant impact on the purchase behavior of active customers is observed. However, due to the large proportion of active customers in enterprises, the impact of paid search advertising on customers' purchasing behavior is not significant on the whole. The results also show that the impact of non segmented paid search advertising on different customer groups will lead to cognitive bias of paid search advertising effect.

After literature review, we find that the research data used by scholars is mainly obtained through customer diary records or advertising expenditure and product sales data obtained from advertisers in the traditional environment. The problem of research based on such data is that it is unable to distinguish whether customers' purchase behavior is caused by advertising or not. Therefore, it is difficult to accurately reflect the actual impact of advertising on individual customers. In contrast, few studies have distinguished the impact of online advertising on different customer groups, and no study has discussed the difference of paid search advertising effect among different customer groups with loyalty as a market segmentation variable. Therefore, based on the evaluation of paid search advertising effect, this study further explores the difference of paid search advertising effect among different loyalty customer groups.

3. HYPOTHESIS

Based on previous research on advertising effects, Vakratsas and Ambler [12] proposed that advertising investment will bring about three kinds of intermediate advertising effects: influence, cognition and experience, and further affect customers' purchase behavior. Based on the characteristics of paid search ads and different loyalty types of customers, this paper puts forward some hypotheses about the immediate effect, duration of carry-over effect and ultimate cumulative effect of paid search advertising on different loyalty customer groups (high-loyalty customers, medium-loyalty customers, low-loyalty customers and potential customers).

Paid search advertising is an advertisement displayed when customers search for keywords related to their expected products on search engines driven by their own needs and interests. Therefore, it can be considered that paid search advertising has a positive impact on customers' purchase behavior. Generally speaking, customers with higher loyalty have more experience in purchasing and visiting, so they have a higher understanding of the enterprise and its products and services. It can also be considered that the access path from pay per click search advertising to purchase is relatively short. So we propose the following hypotheses:

H1a: Paid search advertising has a strongest positive immediate effect on high-loyalty customers.

H1b: Paid search advertising has a stronger positive immediate effect on medium-loyalty customers.

H1c: Paid search advertising has a relatively weak immediate effect on low-loyalty customers.

H1d: Paid search advertising has a weakest positive immediate effect on potential customers.

On the other hand, customers usually have specific needs before searching, however, the content presented by paid search ads is not constant, and the ranking of search engine presentation will be adjusted according to specific strategies or algorithms. Therefore, customers may not repeatedly contact the same paid search advertisement for a long time, that means the duration of the carry-over effect of advertising is relatively short. For customers, customer with relatively lower loyalty usually have a lower degree of cognition of the enterprise, but it takes longer time to understand the cognition in order to finally transform the purchase. Then we propose the following hypotheses about the duration of the carry-over effect:

H2a: Paid search advertising has a shortest duration of carry-over effect on high-loyalty customers.

relationship between the click stock of paid search advertising on day t and the click stock of every day before day t . According to the research of Bass and Leone^[14], the lag impact factor only represents the percentage of paid search advertising stock transferred from day $t-1$ to day t , while the duration of paid search advertising and the cumulative effect of carry-over effect need to be further calculated by lag impact factor λ and immediate effect parameter β . The specific calculation methods are as following formula (3) and (4):

$$T = \log(1-p) / \log \lambda \quad (3)$$

$$\gamma = \beta / (1-\lambda) \quad (4)$$

where p means the percentage of the total cumulative effect of paid search advertising unit pulses, λ is the lag impact factor of paid search advertising, T is the duration of the residual effect of paid search advertising, that is, the number of days required for the cumulative effect of paid search advertising unit pulse to reach p . According to previous research experience, if the cumulative effect of paid search advertising reaches 90%, the effect of paid search advertising is considered to be exhausted^[14], so we set p to 90% for calculate the duration of paid search advertising. In formula (4), γ represents the cumulative strength of immediate effect and carry-over effect of paid search advertisement. According to the verification of the model by previous scholars^[15], the advantage of this model is that it will not be affected by intercept deviation, and it still converges even if the lag time obtained from the lag influence factor is more than one month. So although the computational cost of the model is large, it is still used by many scholars in the study of ads' carry-over effect.

In order to solve the unknown parameters β and λ contained in the model, it is necessary to determine the value of λ . in this study, the grid search method is adopted. Grid search is an exhaustive search method, that is to cycle through all candidate parameters, try every possibility, and select the best performance parameters. In this study, we set the value range of λ between 0 and 1, with the increment of 0.05, to conduct grid search, that is, 19 values from 0 to 1 were used for multiple linear regression fitting. As for the fitting effect of the model, we need to introduce the total sum of square (TSS) to evaluate. The smaller the TSS value is, the higher the fitting degree of the model is. Thus, the parameter λ with the minimum TSS obtained by the fitting model is the optimal value of λ . After getting the value of the optimal lag influence factor, it is substituted into the model to get the estimated value of the immediate effect coefficient β .

5. EMPIRICAL RESEARCH

5.1 Data

The data in this paper are from an online insurance agency platform in Nanjing, where consumers can browse, consult and purchase all kinds of products independently. In order to improve the popularity of the platform and increase product sales, the platform guides consumers to the website by putting paid search advertisements in search engines (Baidu, Sogou, 360). Because the purpose of this study is to explore the impact of paid search advertising on different-loyalty customers, we use the user visit behavior log table in the log database to extract the click behavior characteristics and purchase behavior characteristics of paid search advertising based on single visit. The specific steps are as follows: firstly, according to the relevant keywords set in the website URL, we select the customers who enter the website through pay per click search ads, and extract the click behavior and purchase behavior of these customers in a single paid search ad; secondly, we select the relevant data of the whole year of 2018 to analyze the click times and purchase times of paid search ads of the same customer on the same day by day A total of 63741 records were obtained by data aggregation and data cleaning; finally, we associate these records with the existing customer loyalty classification table in the enterprise for subsequent analysis (customers that do not match any customer loyalty group are identified as potential customers).

5.2 Research results

According to the evaluation model of immediate effect and carry-over effect of paid search advertising, this study takes the click times of daily paid search ads as the independent variable and the daily purchase times as the dependent variable to conduct linear regression on the data grouped by loyalty. Through the grid search method, by calculating the TSS value, we find the optimal lag influence factor λ which makes the model achieve the best fitting effect. Table 1 shows the optimal lag influence factors obtained by integrity grouping.

Table 1. Optimal lag influencing factors of paid search advertising in each customer group

Parameter	Parameter interpretation	Optimal value	Standard error	Minimum standardized residual	Minimum TSS value
$\lambda_{potential}$	Potential customers	0.45	0.016***	6.53578	0.134128
λ_{low}	Low-loyalty customers	0.15	0.015***	0.43408	0.034581
λ_{medium}	Medium-loyalty customers	0.2	0.024***	0.222556	0.024864
λ_{high}	High-loyalty customers	0.3	0.028***	0.504425	0.037329

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table 1, we can see the value of lag influence factor of paid search advertisement in potential customer group is the largest, which is 0.45, followed by high-loyalty customer group, and the lowest value of paid search advertising lag influence factor of low-loyalty customer group is 0.15. Based on the optimal lag influence factor of paid search advertising in the above loyalty groups, it is substituted into the evaluation model of immediate effect and carry-over effect of paid search advertisement to obtain the unknown parameter β which represents the immediate effect of paid search advertisement. The results are shown in Table 2.

Table 2. Immediate effect of paid search advertising in each customer group

Parameter	Parameter interpretation	Estimated value	Standard error	T value	Pr(> t)
$\beta_{potential}$	Potential customers	0.23563	0.01619	14.557	<2e-16***
β_{low}	Low-loyalty customers	0.22494	0.01583	14.211	<2e-16***
β_{medium}	Medium-loyalty customers	0.24145	0.02358	10.241	<2e-16***
β_{high}	High-loyalty customers	0.41716	0.02782	14.994	<2e-16***

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

According to Table 2, we can see that paid search advertising has a significant and positive impact on customers' short-term purchase behavior. And the intensity of immediate effect of high-loyalty customer group ($\beta_{high}=0.41716$) is significantly higher than that of other groups, while there is no significant difference between potential customer group ($\beta_{potential}=0.23563$), medium-loyalty customer group ($\beta_{medium}=0.24145$) and low-loyalty customer group ($\beta_{low}=0.22494$). Thus, it support hypothesis H1a and H1c, and negate hypothesis H1c and H1d. This shows that paid search advertising can effectively promote the immediate purchase behavior of high-loyalty customers. Therefore, when high-loyalty customers enter the E-commerce enterprise site through paid search advertising, enterprises can provide certain marketing measures (such as short-term coupons) to maximize the transformation ability of paid search advertising in the short term.

After obtaining the immediate effect parameter β and the optimal lag influence factor λ of paid search advertising in each loyalty group, we can calculate the duration of carry-over and ultimate cumulative effect according to formula (3) and (4), which shown in Table 3.

Table 3. Carry-over effect of paid search advertising in each customer group

Customer group	Immediate effect	Lag factor	Duration of carry-over effect	Cumulative effect
Potential customers	0.23563	0.45	2.883610395	0.428418
Low-loyalty customers	0.22494	0.15	1.213726655	0.264635
Medium-loyalty customers	0.24145	0.2	1.430676558	0.301813
High-loyalty customers	0.41716	0.3	1.912489289	0.595943

As can be seen from Table 3, the impact of paid search advertising on potential customers is the longest, the impact on high-loyalty customers is longer, while the impact on low-loyalty customers and medium-loyalty customers is relatively low and at the same level. This showed that the hypothesis H2a, H2b and H2c are negated while H2d is verified. This means that the current click through of paid search ads is more likely to promote the purchase behavior of potential customers in the later stage. In order to improve the potential customers' recognition of new products, we can consider the possibility of new customers' access to new products when they purchase new products.

Combining the two effects, the cumulative strength of paid search advertising reaches the highest in the high-loyalty customer group, about 0.59, followed by the potential customer group. This shows that although the impact of paid search advertising on potential customers' immediate purchase is relatively weak, its carry-over effect is strong. The cumulative impact of paid search advertising on potential customers is second only to its impact on high-loyalty customers, indicating that potential customers and high-loyalty customers are more likely to search for products that meet their own needs through search engines after generating purchase demand, and more often visit enterprise websites through paid search advertising channels. Therefore, the hypothesis about cumulative effect passed the test. This means that the main customer groups of paid search advertising impact are potential customers and high-loyalty customers, and its role is to attract new customers and maintain high-loyalty customers. E-commerce enterprises can use paid search advertising as a marketing tool for potential customers and high-loyalty customers. When they put the keywords of paid search advertising, they can put the keywords according to the preferences of potential customers and high-loyalty customers, so as to save the cost of paid search advertising and improve the marketing effect.

6. CONCLUSIONS

This paper constructs a comprehensive evaluation model of the immediate effect and carry-over effect of paid search advertising. By modeling customers with different loyalty groups, we explore the differences in the impact of paid search advertising on loyalty segments. The results show that in terms of immediate effect, paid search advertising has the strongest short-term impact on high-loyalty customers, and is much higher than other customers; in terms of carry-over effect, the duration of paid search advertising's impact from long to short is potential customers, high-loyalty customers, low-loyalty customers and medium-loyalty customers; comprehensively considering the immediate effect and carry-over effect of paid search advertising, the cumulative intensity of paid search advertising to high-loyalty customers is the strongest, followed by potential customers, and the cumulative effect to low-loyalty customers is the weakest. Enterprises can refer to these characteristics, adjust the keyword delivery strategy of paid search advertising, so as to optimize the allocation of advertising resources and further improve the advertising effect.

In addition, in the previous studies on the effect of paid search advertising, most of them focused on the immediate effect of paid search advertising, while there was a lack of research on the carry-over effect of paid search advertising. Moreover, most of the research on paid search advertising effect is to evaluate its overall effect. There is no research on customer market segmentation through loyalty segmentation variables to test the impact of paid search advertising on different loyalty customer groups, and few studies based on customer behavior data in e-commerce enterprise database. Therefore, this paper further enriches the existing research on

the effect of paid search advertising, and plays a certain reference role for enterprise managers for the adjustment of paid search advertising and marketing strategy.

And there are still some limitations in this study, which need to be studied in the future:

(1) Due to the limitation of data sources, this paper only uses loyalty as a customer segmentation index to explore, and we can also use other segmentation indexes to explore the impact of paid search advertising on different customer groups in the future;

(2) This paper only studies paid search advertising, considering the need of enterprises to adjust the overall advertising strategy, we can consider more channels of advertising for comparison in the future research.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71771122 and the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant SJCX19_0045.

REFERENCES

- [1] Assael H, Lipstein B. How to reallocate marketing resources based on market response[C]. 24th Annual Meeting of the Advertising Research Foundation, New York, (1978).
- [2] Breuer R , Brettel M , Engelen A . Incorporating Long-term Effects in Determining the Effectiveness of Different Types of Online Advertising[J]. Marketing Letters, (2011), 22(4):327-340.
- [3] Rutz O J, Trusov M. Zooming in on paid search Ads--a consumer-level model calibrated on aggregated data[J]. Marketing Science, (2011), 30(5): 789-800.
- [4] Li H , Kannan P K . Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment[J]. Journal of Marketing Research, (2014), 51(1):40-56.
- [5] Zhang Y , Jansen B J , Spink A . Identification of Factors Predicting Clickthrough in Web Searching Using Neural Network Analysis[J]. Journal of the American Society for Information Science and Technology, (2009), 60(3):557-570.
- [6] Katona Z, Sarvary M. The race for sponsored links: Bidding patterns for search advertising[J]. Marketing Science, (2010), 29(2): 199-215.
- [7] Blake T, Nosko C, Tadelis S. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment[J]. Econometrica. (2015), 83(1):155-74.
- [8] Braun M, Moe WW. Online display advertising: Modeling the effects of multiple creatives and individual impression histories[J]. Marketing Science. (2013), 32(5):753-67.
- [9] Rutz O J, Bucklin R E. From generic to branded: A model of spillover in paid search advertising[J]. Journal of Marketing Research, (2011), 48(1): 87-102.
- [10] Breuer R, Brettel M, Engelen A. Incorporating long-term effects in determining the effectiveness of different types of online advertising[J]. Marketing Letters, (2011), 22(4): 327-340.
- [11] Tellis G J. Advertising exposure, loyalty, and brand purchase: A two-stage model of choice[J]. Journal of marketing research. (1988), 25(2):134-44.
- [12] Vakratsas D, Ambler T. How advertising works: What do we really know? [J]. Journal of Marketing, (1999), 63(1):26-43.
- [13] Srinivasan V, Weir H A. A direct aggregation approach to inferring microparameters of the Koyck advertising-sales relationship from macro data[J]. Journal of Marketing Research, (1988), 25(2): 145-156.
- [14] Bass F M, Leone R P. Temporal aggregation, the data interval bias, and empirical estimation of bimonthly relations from annual data[J]. Management Science, (1983), 29(1): 1-11.
- [15] Clarke D G. Econometric measurement of the duration of advertising effect on sales[J]. Journal of Marketing Research, (1976), 13(4): 345-357.

The Influence of Live Streamer Morality on Consumer Purchase Intentions from the Perspective of Identity

Cuiping Wang¹, Li Wang^{2*}

¹School of Humanities, Southeast University(Nanjing),China

²School of Marketing and Logistics Management, Nanjing University of Finance and Economics, China

Abstract: Most of the existing studies have described the characteristics of live streaming and the participation motivation of consumers using surveys. However, the relationship between consumer judgment of live streamer morality and their purchase intention has been largely ignored. Based on identity theory, this study constructs a theoretical framework for the influence of live streamers morality on consumer purchase intentions. The results show that the morality of live streamers influences consumer purchase intentions by influencing their identity. Furthermore, the entertainment value of live streaming content has a positive moderating effect on live streamers morality while the information value of live streaming content has no moderating effect. The purpose of this study is to focus on the psychological mechanism and boundary conditions of the influence of live streamer morality on consumer decision making and expand the research on live streaming, which will have important significance for consumer decision making and enterprise marketing strategy formulation.

Keywords: live streaming, live streamer morality, consumer identity, entertainment value of live streaming content, information value of live streaming content

1. INTRODUCTION

Live streaming marketing is a new online marketing form emerging due to the maturity of Internet technology and the development of e-commerce platforms. Live streamers present product information more vividly through a live streaming platform to promote consumer purchase intentions^[1]. However, recently, the "Fanche" phenomenon has frequently appeared, which means that streamers engage in unethical marketing during their live streaming. For example, Simba's team, Jiaqi Li and Yonghao Luo once sold fake or substandard products in their studio. This kind of unethical marketing behavior seriously affects consumer trust and further undermines marketing effectiveness.

In fact, the field of marketing and psychology has been paying close attention to the morality of celebrity spokesman and key opinion leaders, and points out that virtue characteristics are important factors that influence consumers' willingness to buy. And scholars have called for more attention to the morality of sales-related personnel^[2, 3]. For example, Martin(2017) indicated that the morality of celebrity spokesperson can influence people's purchasing choices^[4]. Cheng (2015) argued that ethical sales practices in e-commerce have a significant impact on consumer attitudes^[5]. At present, a number of studies on the characteristics of live streamers are focused on professionalism^[6], attractiveness^[7], popularity^[8], interactivity^[1] and so on, while there is little research focusing on the morality of live streamer. This study argues that there are two main reasons for the above limitation. First, as a new form of marketing, research on live streaming marketing is still in its infancy at present. Although many studies have pointed out current problems in the live streaming industry and given development forecasts or legal regulation recommendations, but ignored the most important core figure in live marketing - the live streamer.

* Corresponding author. Email: Lily_liwang@163.com(Li Wang) , 18851165701@163.com(Cuiping Wang)

Second, the market and enterprises pay too much attention to the immediate benefits brought by live streaming instead of the damage to the long-term benefits caused by unethical sales behavior.

Based on identity theory, this study discusses the psychological mechanism by which live streamer morality (LSM) affects consumer purchase intention (CPI) in live streaming marketing and proposes the boundary conditions of the above relationship. According to identity theory, an individual's identity is always based on some widely accepted characteristics in the social and cultural contexts. People either internalize them to form the characteristics of their own or use them to enrich and perfect their original identity when they interact with others^[9]. Stets(2011) argued that morality is one of the factors that influence identity^[10]. In addition, identity also includes the extent to which one adopts a behavior from others or groups. Wang Ning(2001) stated that in the era of consumption, consumer behavior is an important way to construct identity^[11]. According to the research of Cao (2012), consumer self-identity and social identity will have significant impacts on the brand attitude and purchase intentions of an endorsed brand^[12]. Martin (2017) proposed that the morality of celebrity spokespersons would influence consumer consumption choices by affecting their identity^[4]. According to the bidirectional construction relationship between identity and consumption^[13], it can be inferred that consumer identity(CI) may have a mediating role which live streamer morality influences consumer buying intentions.

There are two main motivations for consumption, hedonic and utilitarian^[14], that is, consumers hope to obtain a pleasant experience or useful information during consumption^[15]. Existing studies have shown that brand-related content that is informative and entertaining can affect consumer emotional responses^[16]. Similarly, if the content of live streaming is funny or useful, it will arouse the emotional reaction and attract attention, which makes it possible for consumers to deeply participate in the interaction process with live streamers and establish a "para-social relationship" with them^[17]. Therefore, we believe that the entertainment value (EV) and information value (IV) of live streaming content may moderate the influence of live streamer morality on consumer identity. In summary, a moderated mediation model was established in this study, as shown in Figure 1. First, the model reveals the mediating role of consumer identity in the relationship between live streamer morality and consumer purchase intentions. Second, the moderating effect of the information value and entertainment value of live streaming content on the first half of the mediating effect will be explored.

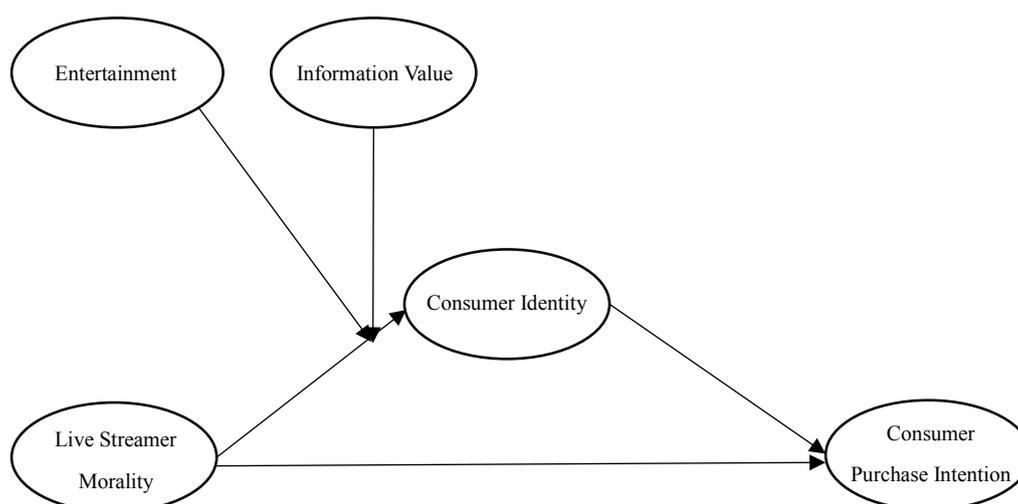


Figure 1. Conceptual model.

This study aims to open the "black box" of consumer behavior in live streaming marketing and explore the complex influence of live streamer morality on consumer purchase intentions, thus providing a new perspective for the research conclusions that are inconsistent in the existing literature. Furthermore, this study provides the corresponding theoretical basis for the selection and cultivation of live streamers in marketing practice.

2. THEORY AND HYPOTHESIS

2.1 Live streamer morality and consumer purchase intentions

Unethical behavior in marketing will damage the trust and affect attitudes and behaviors of consumers^[5]. For example, studies have shown that consumers are paying increasing attention to socially responsible brands^[18]. Dang (2020) shows that an enterprise's sense of social responsibility has a direct impact on consumer behavior^[19]. Valenzuela (2010) also believes that social responsibility can help retailers gain loyal customers and improve their customer relationships^[20]. Similarly, in the traditional marketing field, morality, as a source of meaning transmission and credibility, will increase consumer purchase intentions by increasing their identity with celebrity spokespeople^[4]. Research also shows that the credibility of opinion leaders will influence consumer purchase decisions^[21].

The morality of live streamers defined in this study refers to the reliability, honesty, sincerity and social responsibility of live streamers perceived by consumers. For example, a live streamer's moral level can be reflected by whether they objectively introduce the use effect of a product and a guarantee of after-sales service and the enthusiasm of live streamers for participating in the public welfare broadcasts. Recently, the unethical sales practices occurring in live marketing have seriously affected consumer trust and further affected marketing effectiveness. Therefore, this paper concludes that the moral level of live streamers will affect consumers' purchase intention. The more moral live streamers are, the more willing consumers are to identify with live streamers and become like them.

Based on the above analysis and the results of existing empirical studies, we propose the following hypothesis:

- H1: The morality of live streamers positively influences consumer purchase intentions.

2.2 The mediating role of identity

Baudrillard pointed out in *The Consumer Society* that consumption has gradually become the main way for people to construct their identity. Previous studies on opinion leaders also point out that the influence of opinion leaders on consumers' buying intentions is largely through identification. People can relieve the anxiety caused by the identity crisis by wanting to be the person they want to be, such as stars or key opinion leaders, because those people often represent the symbolic significance of fashion, trend, profession and so on. Therefore, buying the products recommended by them becomes a way to construct their own identity. As described earlier, the formation of identity will be influenced by some commonly recognized characteristics in social culture. Moreover, Stets(2011) argued that morality is also one of these characteristics^[10]. Erikson (1998) argued that people tend to seek a balance between their moral self and their behavior, which also implies that morality is an essential element of one's self^[22]. When consumers perceive the moral traits of celebrities, it is reasonable for them to process the information through their own moral sense and identify with celebrity spokespersons to strengthen their moral identity. In this way, celebrities shape the self-concept of consumers^[23]. Johnson (2005) found that as time passes, consumers will have a sense of identity and intimacy with celebrities after they are exposed to the news of celebrities in the media^[24]. Therefore, the celebrity's morality will increase consumer sense of identity with the celebrity. In live marketing, we assume that a live streamer can be thought of as an extension of the traditional spokesperson. The good moral character of live streamers provides a moral example for consumers, affects consumer moral judgments, and strengthens the moral identity of consumers. Therefore, live streamer morality may be the influencing factor shaping consumer self-concepts.

Based on the above analysis and the results of existing empirical studies, we propose the following:

- H2: The morality of live streamers positively influences consumer identity.

Consumer attitudes towards celebrities would shift to recognized brands^[25]. Horton and Wohl (1956) argue that a unilateral connection and the illusion of face-to-face interaction with a celebrity can be described as a para-social relationship, which reflects a deep identity that leads to changes in attitudes and behaviors^[26]. Consumers

with a higher degree of identification will correspondingly show a more positive attitude towards celebrities^[24], thus producing attachment and loyalty to the endorsed brand^[27]. In addition, the identity effect also represents that people's emotional and cognitive identification of the meaning of an event has a significant impact on their evaluation of, attitude towards and behavior towards the event. To achieve the goal of self-improvement, individuals tend to respond positively to external information consistent with their own identity^[28]. Therefore, we believe that in the context of live streaming, consumer identification of live streamer morality will affect their purchase decisions and intentions related to live streaming.

Based on the above analysis and the results of existing empirical studies, we propose the following:

- H3: Consumer identity positively influences consumer purchase intentions.

Based on H1, H2 and H3, we propose the following hypothesis:

- H4: Consumer identity mediates the influence of live streamer morality on consumer purchase intentions.

2.3 The moderating effect of the characteristics of live streaming content

Previous studies have stated that consumer motivation is essentially dualistic because consumers purchase goods and services mainly for two basic reasons: hedonism and utilitarianism^[14]. The hedonistic motivation is based on sense, that is, the desire to obtain a pleasant experience in consumption activities. Utilitarianism is based on function, that is, the evaluation of a product's functional attributes^[15]. Recent studies have also shown that consumer dual motivations are also reflected in social networking sites and online consumption activities. For example, the research results of Lin (2011) show that the most important factor influencing people to continue to use social networks is the fun of social networks, followed by the number and usefulness of peers^[29]. Kim (2016) also stated that brand-related content that is informative, practical and entertaining will affect consumer emotional responses^[16]. We believe that in live streaming marketing, consumer consumption motivation is also an important factor for consumers to choose whether to watch live broadcasts or not. The funniness and usefulness of live streaming content correspond to consumer hedonistic and utilitarian motives, respectively, because they may be important factors to attract consumer attention and arouse consumer emotional reactions. That is, the fact that interacting with live streamers can acquire happy experiences and useful information from them will encourage consumers to deeply participate in the live-streaming process^[17]. In this way, consumers are more likely to have an in-depth understanding of live streamers, identify with their moral characteristics and allow live streamers to influence their consumption choices. However, when watching live streaming fails to satisfy consumer motivation, consumers may not choose to watch it, let alone establish an emotional connection with anchors. Therefore, we believe that the entertainment value and information value of live streaming content may moderate the influence of live streamer morality on consumer identity.

Based on the above analysis and the results of existing empirical studies, we propose the following:

- H5: The entertainment value of live streaming content positively moderates the relationship between live streamer morality and consumer identity.
- H6: The information value of live streaming content positively moderates the relationship between live streamer morality and consumer identity.

3. RESEARCH METHODS

3.1 Samples and investigation process

The questionnaires in this study were distributed on the Questionnaire Star platform from July 18, 2020 to August 3, 2020. The questionnaires were distributed to the online fan groups of main live streaming, including Weibo fan groups and QQ fan groups. In this study, a total of 300 questionnaires were sent out, and 291 effective responses were received, with an effective response rate of 97%.

3.2 Research tools

The measurement items used in this study were all derived from mature domestic and international scales with good reliability and validity. To ensure a respondent's accurate understanding of the questions, the wording of the measurement items was carefully modified considering the background of live streaming marketing. All the items were measured using a 5-point Likert scale (1= strongly disagree to 5= strongly agree). Specific scale items are shown in Table 1. According to previous studies, individual statistical variables, such as age, gender, education level and monthly income, may also affect consumer purchase intentions. To exclude the influence of other relevant variables on the results, these variables were treated as control variables in this study^[30].

Table 1. Measurement items.

Variables	Items	References
Live streamer morality	I think the live streamer is honest.	Davies ^[31] , 2004
	I think the live streamer is sincere.	
	I think the live streamer is trustworthy.	
	I think the live streamer is socially responsible.	
Consumer identity	Being someone who has the characteristics of the live streamer is an important part of who I am.	Karl & Americus ^[32] , 2002
	A big part of my emotional well-being is tied in with the live streamer.	
	It would make me feel good to be a person like the live streamer.	
	Being like the live streamer is truly important to me.	
Consumer purchase intention	I consider it likely that I would purchase this service from this brand.	Reinikainen, H. ^[33] , 2020
	I am happy to wait in the studio for the live streamer to place the purchase link.	
	I consider it possible that I would purchase this service from this brand.	
	I consider it probable that I would purchase this service from this brand.	
Entertainment value of live streaming content	I think live streaming is fun.	Voss et al. ^[15] , 2003
	I think live streaming is exciting.	
	I think live streaming is delightful.	
	I think live streaming is thrilling.	
	I think live streaming is enjoyable.	
Information value of live streaming content	I think live streaming is effective.	Voss et al. ^[15] , 2003
	I think live streaming is helpful.	
	I think live streaming is functional.	
	I think live streaming is necessary.	
	I think live streaming is practical.	

4. RESEARCH RESULTS

4.1 Confirmatory factor analysis

First, in order to investigate and confirm the convergent validity and discriminant validity of each variable, we used AMOS 22.0 to conduct confirmatory factor analysis. The results show that the five-factor model of each factor in the factor loading and the t value reached a 0.05 significance level, and there was no inappropriate solution, which means a good convergent validity.

In addition, we use the model comparison method to investigate the discriminant validity of each variable. The results of the comparison between the five-factor model and four four-factor models showed that the adaptation index of the five-factor model with all variables independently separated was reasonable (the

adaptation result was $\chi^2=451.653$, TLI = 0.941, CFI = 0.949, RMSEA = 0.066, and SRMR = 0.047) and was superior to all four-factor models (see Table 2), indicating that the model has good discriminant validity. In addition, when all the items were combined into a potential factor, the single-factor model with all the indicators had a poor fitting effect ($\chi^2=2323.260$, TLI = 0.532, CFI = 0.576, RMSEA = 0.187, and SRMR = 0.126), so there was no serious common method bias in this study^[34, 35]. Finally, the corresponding mean extraction variance and comprehensive reliability coefficient were calculated. The Cronbach's α values of the variables were 0.911, 0.862, 0.922, 0.928, and 0.911, respectively. The normalized factor loadings of all dimensions were greater than 0.7 and significant at the 0.001 level. Moreover, the combined reliability coefficients of each dimension are all greater than 0.8, and the average extraction variances are all greater than 0.6, indicating that the scale used in this paper has a good reliability level.

Table 2. Comparison of measurement models.

Model	Descriptions	χ^2	df	TLI	CFI	RMSEA	SRMR
Model 1	Five factors: live streamer morality, consumer identity, purchase intentions, entertainment value of live streaming content, and information value of live streaming content.	451.653	199	0.941	0.949	0.066	0.047
Model 2	Four factors: live streamer morality and consumer identity were combined into one potential factor.	831.856	200	0.854	0.873	0.104	0.081
Model 3	Four factors: purchase intentions and consumer identity were combined into one potential factor.	810.802	200	0.859	0.878	0.103	0.074
Model 4	Four factors: purchase intentions and entertainment value of live streaming content were combined into one potential factor.	986.130	200	0.818	0.842	0.116	0.081
Model 5	Four factors: purchase intentions and information value of live streaming content were combined into one potential factor.	897.106	200	0.839	0.86	0.110	0.071
Model 6	Live streamer morality, consumer identity, purchase intentions, entertainment value of live streaming content, and information value of live streaming content were combined into one potential factor.	2323.260	209	0.532	0.576	0.187	0.126

Note. RMSEA=root mean square error of approximation, SRMR=standardized RMR, CFI=comparative fit index, and TLI=Tucker Lewis index.

4.2 Descriptive statistical results

The mean and standard deviation of each variable and the correlation coefficients among variables are shown in Table 2. There were significant pairwise correlations among the major variables. These results are consistent with our theoretical expectations and can be used for further conditional process testing.

Table 3. Means, standard deviations and correlations among variables.

Variables	M	SD	1	2	3	4	5
1. LSM	3.141	0.763	1				
2. CI	2.434	0.869	0.492***	1			
3. CPI	3.021	0.935	0.498***	0.516***	1		
4. EV	3.269	0.805	0.342***	0.375***	0.620***	1	
5. IV	3.276	0.772	0.573***	0.456***	0.630***	0.685***	1

Note. $N = 291$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

4.3 Hypothesis testing

4.3.1 Mediating effect test

According to the suggestions of Baron and Kenny, the mediating effect was tested in four steps: (1) The influence of independent variables on outcome variables. (2) The influence of independent variables on mediating variables: Based on the introduction of control variables, the influence of the live streamer morality on consumer identity was analyzed (Hypothesis 1). (3) Influence of mediating variables on outcome variables (Hypothesis 2): Based on the introduction of control variables, the influence of identity on consumer purchase intentions is analyzed. (4) Mediating effect (Hypothesis 3): The effect of live streamer morality and identity is introduced simultaneously to analyze whether the effect of live streamer morality on consumer purchase intentions disappears (complete mediation) or weakens (partial mediation) due to the introduction of identity. Model 2 showed that live streamer morality had a significant positive relationship with their identity ($\beta=0.45$, $P<0.01$), and Hypothesis 1 was verified. Model 5 shows that identity has a significant positive relationship with consumer purchase intentions ($\beta = 0.48$, $P<0.01$). Hypothesis 2 is verified. Model 6 showed that after the addition of identity, the relationship between a live streamer morality and consumer purchase intentions was still significant (β decreased from 0.45 to 0.29 from Model 3 to Model 6, respectively). Therefore, the mediating role of identity is verified. Thus, the bootstrap test showed that the mediating effect was significant, with a 95% confidence interval [0.24, 0.44], and the mediating effect was 0.15, accounting for 33.33% of the total effect (0.45). Combined with the above evidence, we believe that identity partly mediates the relationship between live streamer morality and consumer buying intentions.

Table 4. Stepwise stratified regression.

Variables	CI			CPI		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Sex	-0.02 (0.16)	-0.05 (0.14)	-0.06 (0.14)	0.12 (0.15)	0.15** (-0.15)	0.13** (0.15)
Age	0.05 (0.13)	0.06 (0.12)	0.05 (0.12)	-0.01 (0.13)	-0.04 (0.12)	-0.03 (0.12)
Education	-0.0 (0.08)	-0.05 (0.07)	-0.07 (0.07)	0.002 (-0.08)	0.01 (0.08)	0.02 (0.07)
Monthly income	0.01 (0.08)	0.01 (0.07)	-0.002 (0.07)	0.11 (0.08)	0.10 (0.08)	0.11 (0.07)
LSM		0.45*** (0.06)	0.28*** (0.07)	0.45*** (0.06)		0.29*** (0.07)
CI					0.48*** (0.06)	0.35*** (0.06)
EV			0.22*** (0.07)			
IV			0.13 (0.08)			
LSM*EV			0.22*** (0.05)			
LSM*IV			-0.11 (0.04)			
R ²	0.009	0.206	0.304	0.024	0.225	0.321

Note. $N = 291$; $\dagger p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. The numbers in brackets in the table are the standard errors for regression coefficients.

4.3.2 Moderating effect test

In this step, we conduct the moderation test of the first stage of the model. To eliminate the possible threat of collinearity, when constructing the product term of an independent variable and moderating variable, we standardized the two (live streamer's morality, entertainment value of live streaming content, and information value of live streaming content). Combined with the addition of the interaction items in Model 3, we found that the interaction items of live streamers morality and entertainment value of live streaming content had a significant impact on identity ($\beta=0.22$, $p<0.01$), and the interaction between live streamer morality and the information value of live streaming content had no significant effect on identity ($\beta=-0.11$, $p<0.05$). Therefore, it can be concluded that the entertainment value of live streaming content plays a positive moderating role in the relationship between a live streamer morality and consumer identity while the information value of live streaming content has no moderating role in the relationship between a live streamer morality and consumer identity. Furthermore, bootstrapping was used to conduct a second test on the moderating effect of the entertainment value and information value of live streaming content on identity in the first stage, and the results are shown in Table 5. The confidence interval of the determination index does not contain 0, and the test results are the same as those in Table 4. In addition, the information value of live streaming content has no significant moderating effect. The test results support hypothesis 5 but do not support hypothesis 6.

Table 5. Moderating effect.

Result variable	Moderating effect					Index			
	Effect	BootSE	BootLLCI	ULCI	INDEX	BootSE	BootLLCI	ULCI	
EV	-1 SD	0.069	0.025	0.018	0.119	0.048	0.016	0.018	0.081
	+1 SD	0.165	0.034	0.101	0.234				
IV	-1 SD	0.093	0.025	0.044	0.144	0.009	0.015	-0.019	0.041
	+1 SD	0.112	0.035	0.047	0.185				

To further confirm whether the type of moderating effect of the entertainment value of live streaming content on the relationship between live streamer morality and consumer identity meets the original expectation, we refer to the practice of Aiken and West (1990^[36]), put the entertainment value of live streaming content plus or minus one standard deviation into the regression model, and then make a graph (see Figure 2). Figure 2 shows that compared with the low entertainment value of live streaming content, a live streamer morality has a more significant impact on consumer identity under the high entertainment value of live streaming content, which is consistent with the expected hypothesis. Hypothesis 5 is supported.

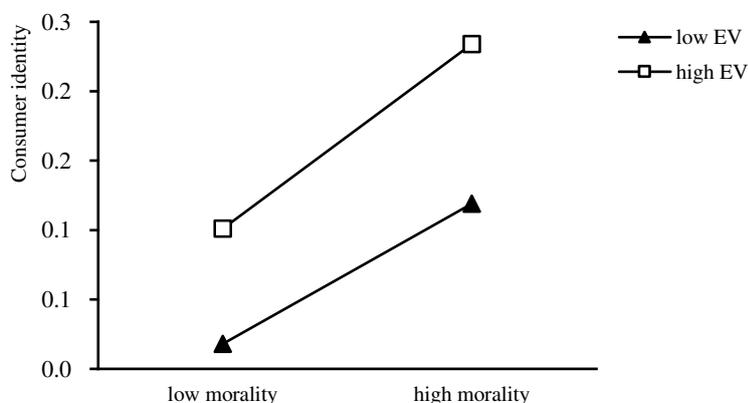


Figure 2. Interaction effect of live streamer morality and the entertainment value of live streaming content on consumer identity. High and low levels of entertainment value represent one standard deviation above and below the mean, respectively.

5. DISCUSSION

There is growing interest in live streaming marketing. For the first time, this study elucidates its internal mechanism, constructs its impact on consumers, and proposes an integrated model. This study expands previous studies on live streamer characteristics from the perspective of morality, and considers the mediating role of consumer identity. In addition, the model also proposes the possible moderating effect of the information value and entertainment value of live streaming content from the perspective of consumer consumption motivation. The results show that live streamer morality positively influences consumer identity and then influences consumer attitude and purchase intentions towards the products recommended by live streamers. The entertainment value of live streamer content positively moderates the relationship between live streamer morality and consumer identity. The results of our study are expected to have theoretical significance for researchers studying Internet marketing and live streaming marketing. The results also inform brands and consumers of effective marketing and consumption knowledge.

One of the main findings is about the role of live streamer morality in influencing consumers attitudes and the indirect effect of consumer identity. This adds to our knowledge of research on the morality of consumer opinion leaders and live streamers. This finding is consistent with the research results in the field of traditional marketing. Thus, live streamers can be seen as the extension of traditional spokespersons. Due to the particularity of Internet marketing, consumers cannot see the goods and have a direct feeling before making purchase decisions. Therefore, the final purchase decision depends on the influence of relevant factors in the Internet environment to a large extent. As the core of live streaming, the influence of live streamers on consumer attitudes can be clearly seen. However, as a new marketing form, the development of live streaming marketing is not yet mature in many aspects. For example, Unethical selling by live streamers often appears. Many scholars even point out that live streaming marketing is essentially a kind of trust camp. In many cases, consumers will buy the products recommended by live streamers out of their trust. Thus, live streamer morality is a very important influencing factor in live streaming marketing.

The second main finding is that the entertainment value of live streaming content has a positive moderating effect on the relationship between live streamer morality and consumer identity. When live streaming content is more entertaining, live streamer morality has a stronger predictive effect on consumer identity. In contrast, the information value of live streaming content has no regulating effect. For the first time, this paper proposes the possible moderating effect of live streaming content characteristics from the perspective of consumer binary motivation. When live streaming content is entertaining, this content can just meet the hedonistic motivation of consumers. When the live streaming content is informative, it can just meet the utilitarian motivation of consumers. The results show that consumers watch live broadcasts more for fun than for information provided by live streamers. We speculated that the reasons for this result are that consumers who make buying decisions on the Internet often encounter an "information explosion", which makes consumers feel resistant to all kinds of product information. Entertaining live will attract the attention of consumers and allow them to participate in the interaction with the live streamer, conditional on the morality of the live streamer to make basic judgments. In this case, consumers do not need to spend many cognitive resources to make purchase decisions based on their sense of identity with anchors.

5.1 Theoretical contribution

This study is the first to explore the internal mechanism and boundary conditions of how live streamer morality characteristics influence consumer purchase intentions, extending the framework of live streamer characteristics. This study fills the gap between changing Internet marketing practices and the lack of existing research and increases the impact of the importance of live streaming marketing. This study can also be used as a starting point for future empirical research on live streaming marketing, which has the following important

theoretical significance. First, based on the theory of identity, this paper discusses the influence of morality on consumer purchase intentions in the context of live streaming marketing. The results show that morality factors are indispensable to the effectiveness of live streaming marketing, which expands the research field of live streamer characteristics. Second, this paper seeks to explain the psychological mechanism of the influence of live streamer morality on consumer purchase intentions, that is, the mediating role of consumer identity. The results show that identity plays a partial mediating role, which enhances the importance of the construction of identity in the field of consumption. Finally, this paper focuses on the boundary conditions under which live streamer morality affects their identity, as well as the entertainment value and information value of live streaming content, which expands the research perspective of consumer consumption motivation.

5.2 Managerial implication

This study also provides some useful suggestions for those interested in live streaming marketing. For example, since consumer recognition of live streamer morality plays an important role in consumer purchase intentions, brands and enterprises should pay more attention to the moral quality and moral behavior of live streamers when cultivating or selecting live streamers to ensure that the good morality of live streamers promotes consumer recognition of live streamers and then generates consumer behavior. In addition, live streamers, as sellers, are also content creators. In addition to maintaining their professional status and attractiveness, live streamers should continue to improve their entertainment sensitivity so that they can produce more content with entertainment value. Finally, this paper also provides enlightenment for the development of live streamers. Live streamers should constantly improve their moral cultivation, strive to create interesting live content, and better win the recognition of consumers.

5.3 Limitations and prospects

The current study also has some limitations. First, the small sample size and the deviation of the sampling method may affect the reliability of the data. Future studies can consider expanding the sample to the general population to explore the universality of the conclusions. Second, we believe that the factors identified in this study are helpful for us to better understand the mechanism of live streaming marketing. In addition, we recognize that other relevant factors not included in our study will also affect this process, such as social presence during live streaming. Third, the study asked consumers of live streaming marketing about their overall acceptance of live streamer morality. However, due to the different types of live streamers themselves, for example, network live streamers can be divided into entertainment type, goods type and skills type, different types of live streamers may have different influences on consumer attitude, and future research can refine this aspect. Finally, the cross-sectional data collection method is adopted in this study, and the causal logic between variables cannot be determined. Future research can use other methods, such as longitudinal studies and experimental designs, to study the causal relationships between specific variables.

REFERENCES

- [1] LIU Y, GAO P, LI C, et al.(2019) The effect and mechanism of social presence in live marketing on online herd consumption from behavioral and neurophysiological perspectives. *Advances in Psychological Science*, 27(6): 990-1004.
- [2] ROMÁN S, RUIZ S.(2005). Relationship outcomes of perceived ethical sales behavior: the customer's perspective. *Journal of Business Research*, 58(4): 439-445.
- [3] CHEN M F, MAU L H. (2009).The impacts of ethical sales behaviour on customer loyalty in the life insurance industry. *The Service Industries Journal*, 29(1): 59-74.
- [4] MARTIN F, TAO-PENG F. (2017).Morality matters? Consumer identification with celebrity endorsers in China. *Asian Business & Management*, 16(4-5): 272-289.
- [5] YANG Q, ZHANG K, WANG X, et al.(2018).The influence of web celebrity information source characteristics on

- consumers' purchasing intention: A moderated mediation model. *Operations and Management*, 11(1): 65-68. (in Chinese)
- [6] HEO J, KIM Y, YAN J Z. (2020). Sustainability of Live Video Streamer's Strategies: Live Streaming Video Platform and Audience's Social Capital in South Korea. *Sustainability*, 12(5): 13.
- [7] JOHNSON J W M R. (2019). Live Streamers on Twitch.tv as Social Media Influencers: Chances and Challenges for Strategic Communication. *International Journal of Strategic Communication*, 13(4): 321-335.
- [8] FANG C. (2018) Research on the influence of e-commerce network anchors on consumer attitudes. Anhui University, 2018. (in Chinese)
- [9] FITZ-KOCH S. (2020). Who am I, and if so, how many? Identity Dynamics in Agricultural Entrepreneurship. *Acta Universitatis agriculturae Sueciae*, 11(4): 100-103
- [10] STETS J E, CARTER M J. (2011). The Moral Self: Applying Identity Theory. *Social Psychology Quarterly*, 74(2)
- [11] WANG N. (2001). Consumption and identity: an analytical framework for the sociology of consumption. *Sociological Research*, (1): 4-14. (in Chinese)
- [12] CAO Z, CHEN Q. (2012). Research on the influence of celebrity endorsement on purchase intention based on consumer identity. *Consumer Economics*, 28(01): 53-56. (in Chinese)
- [13] WU Q. (2007). Identity in Consumer Society. Sun Yat-sen University. (in Chinese)
- [14] BATRAR, AHTOLA O T. (1991). Measuring the hedonic and utilitarian sources of consumer attitudes. *Marketing Letters*, 2(2): 159-170.
- [15] KEVINE E. VOSS E R S A B G. (2003). Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude. *Journal of Marketing Research*, 40(3): 310-320.
- [16] KIM A J, JOHNSON K K P. (2016). Power of consumers using social media: Examining the influences of brand-related user-generated content on Facebook. *Computers in Human Behavior*, 58:98-108.
- [17] GONG W, LI X. (2017). Engaging fans on microblog: the synthetic influence of parasocial interaction and source characteristics on celebrity endorsement. *Psychology & Marketing*, 34(7): 720-32.
- [18] NYILASY G, GANGADHARBATLA H, PALADINO A. (2014). Perceived Greenwashing: The Interactive Effects of Green Advertising and Corporate Environmental Performance on Consumer Reactions. *Journal of Business Ethics*, 125(4): 693-707.
- [19] DANG V T, NGUYEN N, PERVAN S. (2020) Retailer corporate social responsibility and consumer citizenship behavior: The mediating roles of perceived consumer effectiveness and consumer trust. *Journal of Retailing and Consumer Services*, 55.
- [20] VALENZUELA L M, MULKI J P, JARAMILLO J F. (2010). Impact of Customer Orientation, Inducements and Ethics on Loyalty to the Firm: Customers' Perspective. *Journal of Business Ethics*, 93(2): 277-291.
- [21] LOU C, YUAN S. (2019). Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media. *Journal of Interactive Advertising*, 19(1): 58-73.
- [22] ERICKSON E H. (1998). Identity: Adolescence and Crisis. Zhejiang. (in Chinese)
- [23] DWIVEDIA, JOHNSON L W, MCDONALD R E. (2015). Celebrity endorsement, self-brand connection and consumer-based brand equity. *Journal of Product and Brand Management*, 24(5): 449-461.
- [24] JOHNSON A R. (2005). When a celebrity is tied to immoral behaviour: Consumer reactions to Michael Jackson and Kobe Bryant. *Advances in Consumer Research*, 32(1): 100-101.
- [25] MCCRACKEN G. (1989). Who is the Celebrity Endorser? Cultural Foundations of the Endorsement Process. *Journal of Consumer Research*, 16(3): 310-321.
- [26] HORTON D, WOHL R R . (1956). Mass communication and para-social interaction; observations on intimacy at a distance. *Psychiatry*, 19(3): 215-229.
- [27] KIM C K, RYOO J-Y, SUNG H. (2002). The Effect of Brand Personality on Word-Of-Mouth and Brand Loyalty ; Applying the Theory of Social Identification. *Journal of Consumer Studies*, 13(1): 25-43.

- [28] PETTY R E, WHEELER S C, BIZER G Y. (1999). Is There One Persuasion Process or More? Lumping Versus Splitting in Attitude Change Theories. *Psychological Inquiry*, 10(2): 156-163
- [29] LIN K-Y, LU H-P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, 27(3): 1152-1161.
- [30] ALISON B, SUSAN M. (2010) Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. *British journal of health psychology*, 15(Pt 4): 797-824.
- [31] DAVIES G, CHUN, R., DA SILVA, R. V., & ROPER, S. A . (2004). Corporate Character Scale to Assess Employee and Customer Views of Organization Reputation. *Corporate Reputation Review*, 7(2): 125-146.
- [32] KARL A, AMERICUS R. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, 83(6): 1423-1440.
- [33] REINIKAINEN H, MUNNUKKA J, MAITY D, et al. (2020) ‘You really are a great big sister’ – para-social relationships, credibility, and the moderating role of audience comments in influencer marketing. *Journal of Marketing Management*, 36(3-4): 279-298.
- [34] MALHOTRA N K, KIM S S, PATIL A. (2006) Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research. *Management Science*, 52(12)
- [35] ZHOU H, LONG L. (2004) Statistical test and control method for common method deviation. *Advances in Psychological Science*, (06): 942-950. (in Chinese)
- [36] AIKEN L S, WEST S G. (1990). Invalidity of True Experiments. *Evaluation Review*, 14(4): 374-390.

Short Research Paper

A Study on Purchasing Optimisation for Cross-border E-commerce Enterprises under Uncertainties

*Jin Lin**

College of Economics and Management, Min Jiang University (Fuzhou), 350108, China

Abstract: In recent years, cross-border e-commerce has become a new economic growth point for many countries. However, in the purchasing process, most cross-border e-commerce enterprises still use stocking strategies based on general experience. Since goods need to be reordered immediately when they are out of stock, a lack of a scientifically-based purchasing plan may result in excessive inventory and a low inventory turnover rate. Thus, the ability to determine appropriate order quantity under uncertain cross-border market demands is key to improving the inventory control of cross-border e-commerce enterprises. To ensure the quality of cross-border e-commerce products and maximise profits, it is crucial to formulate an effective and practical purchasing strategy. Based on existing research, this paper examines a cross-border e-commerce purchasing strategy under uncertainty by using the newsvendor model and setting expected profit maximisation as the goal. Specifically, this paper explores the optimal purchasing strategy for cross-border e-commerce products under discrete and continuous stochastic demands.

Keywords: cross-border e-commerce, ordering strategy optimisation, newsvendor model

1. INTRODUCTION

At present, the definition of cross-border e-commerce that is well accepted by the academic community is international business activities in which trading entities belonging to different customs borders complete transactions, payments and settlements through e-commerce platforms, and deliver goods through cross-border logistics services.

China's cross-border e-commerce trade volume has maintained an upward trend in recent years due to increasing consumption upgrading in numerous countries, the significant improvement of international logistics services and the continuous advancement of cross-border payment channels. Cross-border e-commerce has become an important driving force for the recovery of China's foreign trade. Currently, the biggest challenge faced by cross-border e-commerce businesses is inefficient purchasing processes. The purchasing pattern of cross-border e-commerce is characterised by long delivery periods, large batches and large transaction amounts, while that of traditional businesses is characterised by short delivery periods, small batches, large variety of goods and high trading frequency. Due to long transportation distances, long procurement lead times and large fluctuations in product demand, the biggest risks of cross-border e-commerce enterprises lie in problems caused by the over- or undersupply of goods.

Real-life business conditions are constantly changing. This study will explore the purchasing problems that cross-border e-commerce enterprises face under uncertainties, so as to provide such companies with valuable suggestions on purchasing optimization. By using the newsboy model, an optimal ordering plan is formulated to reduce inventory and stock-out costs, increase expected revenue, and achieve the ultimate goal of improved overall competitiveness.

1.1 Importance of purchasing strategy optimisation for cross-border e-commerce enterprises

The purchasing problems shared by cross-border e-commerce enterprises are marked by certain

characteristics. First, to reduce production costs, upstream suppliers of cross-border e-commerce enterprises usually have a minimum order quantity for each category of goods. Second, as these businesses have a large number of SKUs and high sales demand, the purchasing pattern features large order quantities, stable ordering cycles, a wide variety of products and small batches. Third, shipping times are long, and customs clearance and inspection times fluctuate. To avoid running out of stock, cross-border e-commerce enterprises usually place orders with suppliers at a quantity higher than the actual sales demand. Last, arrival times and quantities of goods are uncertain. In the case of a sudden increase in sales, a supplier may fail to deliver the required goods in one batch on time.

1.1.1 Theoretical significance

Among existing research on purchasing strategies of cross-border e-commerce businesses, most studies have focused on inventory control of the entire supply chain. As time goes on, purchasing strategies have expanded to areas such as diversified buying channels and multi-stage ordering portfolios. To view cross-border e-commerce purchasing strategies from a scientific perspective, and to improve both consumption quality of overseas consumers and business operating profits, this paper aims to find the most effective and practical purchasing strategy for cross-border e-commerce enterprises under uncertain market demands. Moreover, factors influencing profit are also studied to provide comprehensive and helpful suggestions for such enterprises. Therefore, this paper hopes to broaden current research on purchasing strategies.

1.1.2 Practical significance

This paper aims at providing decision-making references and management suggestions for cross-border e-commerce enterprises, to help them improve their customer service quality and management systems. Demand forecasting methods which are scientific, effective, standardised and practical are critical for increasing operating profits, reducing inventory loss, minimising inventory oversupply, saving social resources, and improving customer service ability and satisfaction.

2. LITERATURE REVIEW

Purchase quantity has a great effect on an enterprise's various costs and bottom line. Therefore, establishing an appropriate ordering amount is crucial from both an economic and practical standpoint. The newsboy model is often used to determine the ideal order quantity. It has been studied by many researchers, such as Cachon, who systematically examined how companies can maximize their expected returns under uncertainties by constructing a newsboy model to obtain the optimal order quantity under various circumstances.

Furthermore, many scholars have realized the importance of considering the impact of uncertain factors on both the demand and supply sides of purchasing decisions, and have focused their attention on ordering models under uncertain procurement lead-times and unpredictable supply shipment arrival times. Among related studies, most have based their work on random fuzzy theory. Many scholars, such as Chang and Yao, have explored ordering models that combine randomness and fuzziness, i.e., when the fuzzy variable is customer demand, or when the demands within the procurement lead-time are a random fuzzy variable. In another study, Yu Chunyun et al. used the random fuzzy demand expectancy theory to establish a random fuzzy programming model for both single and multiple products, with the goal of maximizing expected value.

By utilizing existing research and applying theories to real-life scenarios, this study aims to explore the purchasing processes of cross-border e-commerce enterprises. Based on related characteristics, an ordering model is created to guide cross-border e-commerce companies to form a systematic and reasonable ordering plan, reduce inventory costs and improve capital utilization.

3. INTRODUCTION OF THE NEWSVENDOR MODEL UNDER UNCERTAINTIES

The research method adopted in this paper is mainly quantitative analysis, supplemented by qualitative analysis. The characteristics, nature and influencing factors of cross-border e-commerce purchasing strategies are explained from a real-world perspective, and the newsvendor model and probability theory are applied to examine the proposed strategy under uncertain demands.

As a classic model addressing single-period inventory problems, the newsvendor model can be used to obtain the maximum expected profit under uncertainties, which fits the demand uncertainty of cross-border e-commerce products studied in this paper. In this way, the optimal order quantity of discrete and continuous cases is obtained, and the degree of influence of each variable is further compared and examined through numerical analysis. Finally, suggestions are given to cross-border e-commerce enterprises with respect to purchasing strategy.

The traditional newsvendor problem is also known as the single-period inventory problem. In the early years of business management research, the newsvendor model was applied to resolve retailers' purchase quantity problems to maximise profit under uncertain demand ^[10]. The model has since attracted great scholarly attention since 1888, when the famous economist Edgeworth ^[11] used the newsvendor model to find the ideal reserve amount banks should keep. The model has a simple framework and can deal real-life problems related to inventory, economy and production, such as the setting of safety stocks, one-time sales, and demand estimations during peak seasons .

In practical application, the newsvendor model, based on statistical analysis of past market demand, uses the order quantity Q , sales volume x , sales price a , wholesale price b , bargain price c and other parameter variables to maximise the profits of retailers or manufacturers.

The expected profit model of the newsvendor model is generally expressed as follows:

$$E(\pi) = \int_0^Q [(a-b)x - (b-c)(Q-x)]f(x)dx + \int_0^\infty (a-b)Qf(x)dx$$

By calculating the second derivative with respect to inventory Q using the formula above, we have:

$$\frac{d^2 E(\pi)}{dQ^2} = -(a-c)f(Q) < 0$$

Hence, the expected profit is a convex function with respect to Q , and the optimal purchase quantity has a unique solution, which is:

$$Q^* = F^{-1}((a-b)/(a-c))$$

As cross-border e-commerce enterprises are faced with uncertain overseas market demands and increased risks of product damage from transportation and customs clearance processes, they need to determine the optimal purchase quantity before initialising the transaction. The newsvendor model is suitable for solving the purchasing problem under uncertainties; and therefore, it is used in this paper to determine the best purchase quantity for cross-border e-commerce businesses.

4. RESEARCH ON PURCHASING STRATEGY OPTIMISATION FOR CROSS-BORDER E-COMMERCE ENTERPRISES UNDER UNCERTAINTIES

Restrained by trade policies, political environments and transportation regulations of different countries, uncertain consumer demand is a common market problem faced by many cross-border e-commerce enterprises, which has also given rise to unknown risks and difficulties for such businesses in making purchasing decisions.

4.1 The problem in detail

In the sales process of cross-border e-commerce products, since market demand is stochastic and subject to

consumer preferences, festivals, customs, etc., there may be oversupply or undersupply, which may affect business profits. Therefore, before purchasing, cross-border e-commerce enterprises usually take historical sales as a reference to predict market demands for next year. In the sales process, the enterprise first buys Q products at wholesale price b , and sets the retail price of the product as a . Then, at the end of the period, if the sales volume is lower than the purchase quantity, the remaining products will be sold at bargain price c ($c < b$). If a certain product is out of stock, there will be a shortage cost.

4.2 Model symbols and assumptions

4.2.1 Assumptions

1. The inventory is 0 when the first purchase is placed;
2. Within the sales cycle, the demand x of the cross-border e-commerce product is a random non-negative variable;
3. The seller, that is, the cross-border e-commerce enterprise, is rational, and its risk preference is neutral.

4.2.2 Symbol description

- a : the retail price per unit of the cross-border e-commerce product;
- b : the wholesale price per unit of the cross-border e-commerce product;
- c : the bargain price per unit of the cross-border e-commerce product ($a > b > c$);
- x : the daily market demand of the cross-border e-commerce product, which is a random variable, and its probability is $P(x)$;
- Q : the purchase quantity of the cross-border e-commerce product;
- $f(x)$: the density function of the demand of the cross-border e-commerce product;
- $F(x)$: the distribution function of stochastic demand of the cross-border e-commerce product, and the inverse function is expressed by $F^{-1}(x)$;
- π : the sales revenue of the cross-border e-commerce product;
- C : the sales loss, including C_1 unsalable loss and C_2 out of stock loss.

4.3 Purchasing strategy under discrete demand

4.3.1 Purchasing model under discrete demand

When the purchase quantity is Q , there are two kinds of losses:

- (1) When the supply exceeds the demand ($Q \geq x$), the excess product cannot be sold, and the expected value of loss is

$$C_1 = (b-c) \sum_{x=0}^Q (Q-x) P(x)$$

- (2) When the supply is less than the demand ($Q < x$), there is the loss of sales opportunities due to shortage of stock, and the expected value of loss is

$$C_2 = (a-b) \sum_{x=Q+1}^{\infty} (x-Q) P(x)$$

And the total loss is

$$C = C_1 + C_2 = (b-c) \sum_{x=0}^Q (Q-x) P(x) + (a-b) \sum_{x=Q+1}^{\infty} (x-Q) P(x)$$

If the optimal purchase quantity is Q^* , then to satisfy

$$C(Q^*) \leq C(Q^*+1) \quad (4-1)$$

$$C(Q^*) \leq C(Q^*-1) \quad (4-2)$$

By substituting into formula 4-1, we have

$$\begin{aligned} & (b-c) \sum_{x=0}^{Q^*} (Q^* - x) P(x) + (a-b) \sum_{x=Q^*+1}^{\infty} (x - Q^*) P(x) \\ & \leq (b-c) \sum_{x=0}^{Q^*+1} (Q^* - x + 1) P(x) + (a-b) \sum_{x=Q^*+2}^{\infty} (x - Q^* - 1) P(x) \end{aligned}$$

Similarly, from formula 4-2, we have

$$\sum_{x=0}^{Q^*-1} P(x) \leq \frac{a-b}{a-c}$$

Combining the above, we see that

$$\sum_{x=0}^{Q^*-1} P(x) \leq \frac{a-b}{a-c} \leq \sum_{x=0}^{Q^*} P(x) \quad (4-3)$$

The optimal purchase quantity under discrete demand can be determined by formula 4-3, where $(a-b)/(a-c)$ is the critical value.

4.3.2 Example analysis I

The daily demand of product A sold by a cross-border e-commerce enterprise is between 30 pieces and 40 pieces. The retail price $a = 10$ yuan / piece, the wholesale price $b = 6$ yuan / piece, and the bargain price $c = 4$ yuan / piece. The probability distribution of demand x is shown in Table 4-1.

Table 4-1 Probability distribution of the demand of product A.

Requirement x	30	31	32	33	34	35	36	37	38	39	40
P (x)	0.08	0.10	0.12	0.14	0.17	0.12	0.10	0.08	0.04	0.03	0.02

According to formula 4-3

$$\sum_{x=30}^{Q^*-1} P(x) \leq \frac{10-6}{10-4} \leq \sum_{x=30}^{Q^*} P(x)$$

We have

$$\sum_{x=30}^{Q^*-1} P(x) \leq 0.67 \leq \sum_{x=30}^{Q^*} P(x)$$

Since $P(30) + P(31) + P(32) + P(33) + P(34) = 0.61 < 0.67$; $P(30) + P(31) + P(32) + P(33) + P(34) + P(35) = 0.73 > 0.67$, the optimal purchase quantity $Q^* = 35$ (pieces).

4.4 Purchasing strategy under continuous demand

The above analysis only considers the purchase quantity problem under discrete stochastic demand; however, in many cases, continuous demand distribution is more in line with actual demand. Therefore, the purchasing strategy under continuous demand is analysed in the next part.

4.4.1 Purchasing model under continuous demand

Considering undersupply and oversupply, the daily sales profit of a cross-border e-commerce product is

$$\pi = \begin{cases} (a-b)Q & , x > Q \\ (a-b)x - (b-c)(Q-x) & , x < Q \end{cases} \quad (4-4)$$

$$E(\pi) = \int_0^Q [(a-b)x - (b-c)(Q-x)]f(x)dx + \int_Q^{\infty} (a-b)Qf(x)dx \quad (4-5)$$

To calculate the first derivative of Q with respect to the expected profit function $E(\pi)$ of the cross-border product

$$\frac{dE(\pi)}{dQ} = (a-b) \int_Q^{\infty} f(x)dx - (b-c) \int_0^Q f(x)dx \quad (4-6)$$

Let formula 4-6 = 0, then

$$F(Q) = \frac{a-b}{a-c} \quad (4-7)$$

To calculate the second derivative of Q with respect to the expected profit function $E(\pi)$

$$\frac{d^2E(\pi)}{dQ^2} = -(a-c)f(Q) \quad (4-8)$$

Since

$$\frac{d^2E(\pi)}{dQ^2} < 0,$$

$\pi(Q)$ is a convex function with respect to Q . The optimal purchase quantity Q^* of the cross-border e-commerce product exists and is unique, and can be obtained from formula 4-7

$$Q^* = F^{-1}\left(\frac{a-b}{a-c}\right).$$

4.4.2 Example analysis II

The daily demand of product A sold by a cross-border e-commerce enterprise is between 30 pieces and 40 pieces. The retail price $a = 10$ yuan / piece, the wholesale price $b = 6$ yuan / piece, and the bargain price $c = 4$ yuan / piece. The demand is a random variable and follows the normal distribution of $x \sim N(33.99, 2.48)$.

$$F(x) = \frac{a-b}{a-c} = \frac{10-6}{10-4} \approx 0.67,$$

According to the standard normal distribution table, the corresponding value of 0.67 is 0.44, then $(Q-\mu)/\sigma=0.44$, $Q=35.08$, so the optimal purchase quantity of product A is 35.08 pieces.

5. SUGGESTIONS ON PURCHASING STRATEGY OPTIMISATION UNDER UNCERTAINTIES

To reduce the loss of potential sales opportunities when demand exceeds supply, and lower the waste caused by unsellable products when supply exceeds demand, the purchase quantity of cross-border e-commerce products can be predicted by using the newsvendor model based on previous sales data, provided that the demand fluctuation is small. This section analyses the purchasing strategy using the newsvendor model by considering discrete and continuous stochastic demand variables based on the market sales of cross-border e-commerce product A, to obtain the optimal purchase quantity for profit maximisation.

First, for cross-border e-commerce products that are easily damaged in transportation, commodity inspection or customs clearance processes, a purchasing pattern featuring small batches and high frequency should be adopted. In addition, time, environment and other factors should be considered, so as to reduce inventory and management costs of cross-border e-commerce enterprises.

Second, in the process of transportation, circulation, commodity inspection and customs clearance, it is necessary to implement strict management procedures for cross-border e-commerce products according to their characteristics. Furthermore, logistics-related measures such as surface cleaning, shockproof packaging, prevention of trampling and stacking, and secondary cleaning should be adopted to ensure product quality and minimise losses caused by product mishandling and other related problems.

Third, it is necessary to synchronise inventory information in real-time to minimise losses caused by stock shortages. In addition, enterprises can reversely affect the market demand through their sales plans. For goods to

be ordered up, an enterprise may reduce the number of sales plans to leave room for the inventory control system to respond, while for unmarketable goods which requires promotion, the enterprise may increase the planned sales volume to reduce inventory pressure.

6. CONCLUSIONS

This study focused on purchasing strategy optimisation for cross-border e-commerce products under uncertainties. Limited by assumptions in this paper, the following problems need to be further studied:

(1) This paper only examined the optimal purchasing decision-making technique for cross-border e-commerce enterprises in the case of a single supplier. However, in real life, in order to minimise procurement risks, enterprises usually purchase from multiple suppliers. Therefore, in future works, the model can be further expanded to cover multiple suppliers.

(2) In the research process, other factors that may affect the cross-border e-commerce purchasing strategies were simplified, such as the promotion of cross-border e-commerce platforms, consumers' preferences and so on. Therefore, this paper can be improved upon by considering influences from multiple factors in regards to effective and practical purchasing strategies.

REFERENCES

- [1] She Liting, Lu Kebin.(2017). Research on Newsvendor Problem for the Return of Cross-Border E-Commerce Defect-Free Products Under Overseas Warehouse Mode. *Journal of Xichang University*, 31 (02): 75-79(in Chinese)
- [2] Zhou Xuejiao. (2018).Research on Joint Ordering Strategy for Cross-Border E-Commerce Bonded Warehouse. Ms D Thesis.Shandong: Shandong University.(in Chinese)
- [3] Shen Danyang, Huang Jinli, He Shiqi.(2015). Research on Cross-Border E-Commerce Logistics Mode in China . *Jiage Yuekan*, (08): 39-42(in Chinese)
- [4] Jin Hong, Lin Xiaowei. (2015).Suggestions on the Development Mode and Strategy of Cross-Border E-Commerce in China.*Research of Macroeconomics*.(09): 40-49(in Chinese)
- [5] He Jiang, Qian Huimin.(2017). Research on the Development of Cross-Border E-Commerce in China:Review and Prospect. *Science and Technology Management Research*. 37 (17): 213-220(in Chinese)
- [6] Li Xuemin, Miao Lixin, Xu Qingqing.(2008). Review of Research Progress of Newsvendor Model. *Statistics & Decision*, (17):11-14(in Chinese)
- [7] Sibong Ding, Yuan Gao.(2014). The (σ, S, S) . Policy for Uncertain Multi-product Newsboy Problem. *Expert Systems with Applications*.41(8):3769-3776.
- [8] Huang H, Ke H.(2017). Pricing Decision Problem for Substitutable Products Based on Uncertainty Theory. *Journal of Intelligent Manufacturing*. 28(3):503-514.
- [9] Erlebacher S J. (2000) . Optimal and Heuristic Solutions for Thmulti-Item Newsvendor Problem with A Single Capacityconstraint. *Production & Operations Management*. 9(3):303–318.
- [10] Gorman Michael F.(2020).Practice Summary: Le Macaron Implements Ordering Optimization Informs. *Journal on Applied Analytics*,10(22):144-150
- [11] Siegel Andrew F.;Wagner Michael R.(2020). Profit Estimation Error in the Newsvendor Model Under a Parametric Demand Distribution. *Management Science*,14(5):297-305
- [12] Wang Jiaqi,Liu Wei.(2019). Inventory Optimization of Distribution Center based on The Purchasing Model of Fuzzy Economic.In: *Proceedings of 4th International Conference on Humanities Science and Society Development VOL.328*.Xiamen: *Advances in Social Science, Education and Humanities Research*,585-589

Full Research Paper

Research on the Influencing Factors of Knowledge Sharing Behavior among Makers in IT-enabled Open Innovation Communities

Lijun Li, Jian Tian¹

School of Economics and Management, Jiangsu University of Science and Technology, Zhenjiang, 212100, China

Abstract: In recent years, Open Innovation Community has become the gathering place of enterprise creativity. From the perspective of the individual Maker, this paper explores the influencing factors of knowledge sharing behavior in the IT-driven Open Innovation Community, which is of great significance for enterprises to absorb creative solutions, improve innovation ability and performance. A total of 338 data sets were collected, and the data were analyzed and tested using structural equations and PROCESS. The results show that material interests, social interaction, self-efficacy and willingness to share knowledge have a significant positive impact on knowledge sharing behavior, while perceived personal knowledge ownership significantly negatively affects knowledge sharing behavior. Information technology support plays a moderating role in the willingness and behavior of sharing, and the more information technology support, the stronger the willingness and behavior of sharing.

Keywords: Open innovation community, it driven, maker, knowledge sharing, influencing factors.

1. INTRODUCTION

Open innovation was first proposed by Harvard Business School Chesbrough Professor in 2003, it is defined as “The integration of external resources into the internal innovation of an enterprise through the purposeful inflow and outflow of knowledge, and their application to the market and expansion”^[1]. How to acquire, store and utilize knowledge resources has become the key for enterprises to gain competitive advantages^[2]. As an important carrier of creative knowledge sharing, the Open Innovation Community (OIC) has attracted more and more attention from enterprises. Online users can directly or indirectly involved in innovation-related activities such as creativity, R&D and promotion of products within an enterprise through an open online community^[3]. Wang calls the network platform where companies use emerging information technologies to attract users to participate in internal innovation as OIC^[4]. The well-known OICs around world include P&G’s Connect & Develop, MyStarbucksidea at Starbucks, and Haier Open Partnership Ecosystem (HOPE), etc.

In recent years, the issue of knowledge sharing in the OIC has gradually become the focus of scholars at home and abroad, which can be summarized from the perspectives of enterprise innovation performance, the construction of community platform and the factors affecting the participation of the main body. Some scholars have pointed out that the key activity of Open Innovation is knowledge sharing to realize knowledge innovation^[5]. Haefliger et al. proposed that knowledge protection should be paid attention to in knowledge sharing in OIC considering the nature of enterprise boundaries and competitive advantages^[6]. Based on the internet platform, Xia enjun and others explore the impact of external capability, external innovation source, the relationship between enterprise and external innovation source and guarantee on the performance of Open Innovation^[7]. Based on the classification of Open Innovation, Stefanhrastinski summarizes the seven characteristics of the operation of Open innovation in foreign countries^[8]. Cheng and Yang found that Open innovation’s ability to make a breakthrough depends on its ability to acquire and share knowledge^[9]. Taking into

¹ Corresponding author. Email: tianjian@just.edu.cn

account the characteristics of the Internet, Chen studies the impact of user networks of OIC on innovation performance from the perspective of social capital^[10]. F.V.Briel et al, provided focus and guidelines for community managers by analyzing the problems with the failed Open Innovation case^[11]. Chen nan studies the process of knowledge sharing among participants in internet based Open Innovation platforms^[12]. Researchers point out that social benefits, learning benefits and users' identification with enterprises have an impact on knowledge sharing willingness^[13]. From the perspective of social network, Liu studies the factors that influence the community performance of OIC^[14]. Wang points out that the leading characteristics of users in online communities have a significant positive impact on the level of knowledge sharing^[15].

To sum up, knowledge sharing in the OIC has been studied from different perspectives at home and abroad, but the knowledge sharing in OIC under information technology has not been fully studied. In addition, more than 90% of enterprises believe that open innovation is of great significance in helping enterprises speed up product research and development, enhance customer intimacy and improve supply chain efficiency. However, some scholars have pointed out that most global Fortune 1000 companies have built OIC communities that do not deliver the economic benefits and rewards that companies expect^[16]. Guo believes that the members of OIC include internal employees, Makers, customers from outside the enterprise and general users in the community^[17]. It's likely that the OIC isn't working well because the company doesn't differentiate between community members in terms of operations, standardized management, or incentives, and fail to consider different individual participation motives or large differences in needs. Makers are the primary providers of innovative resources who are passionate and interested in innovation and are loyal to making use of their knowledge to make the innovation come true. The Maker is an indispensable innovation subject in the innovation activities of enterprises, the level of activity and the content of shared knowledge directly affect the quality of Enterprise's creative generation. As Xu Jiabo said, the "Maker" to hold up the "New wave"^[18]. Therefore, from the perspective of IT driving, this paper focuses on the OIC, studies the factors that influence the participation of Makers in knowledge sharing, and explores the moderating effect of IT support, it is of great significance to improve enterprise's innovation ability, Enterprise's performance, enhance enterprise's competitiveness and realize enterprise's strategic goal.

2. RESEARCH HYPOTHESIS AND MODEL CONSTRUCTION

2.1 Makers engage in extrinsic motivation and shared willingness

Material benefits, as an external stimulus, refer to the interests or encouragement after people finish a certain task. Material interests can be money, bonuses, community virtual coins, gifts, community rating, etc. According to the theory of reasonable conduct (TRA), an action will be taken only when the expected benefits and compensations of an action are greater than the perceived moral costs and risks. Koo et al. took the hotel industry as the research object, and the results showed that external material interests had a positive effect on employees' work^[19]. Based on the Social Exchange Theory, Ren Ling et al. proposed that the more benefits users in OIC gain from the exchange of knowledge sharing and material interests, the greater the willingness of individual innovative users to share^[20]. With the development of the society, the Makers are no longer pursuing the realization of self-satisfaction, and gradually become commercialized, and will also consider the investment and return. In the process of knowledge sharing, the Makers also pay some costs, such as time and energy, and they also hope to have the corresponding material rewards to compensate for their losses.

Social interaction is a process of social exchange. Through certain interaction, participants can obtain valuable resources for themselves. The motivation of individual behavior is to obtain certain benefits and have the purpose of exchanging with others^[21]. The ability of members to express their views, comments and solve problems not only enables the transfer of resources, but also increase the level of goodwill and trust among

members through frequent communication, thus facilitating more effective and long-term communication and knowledge-sharing. In the social interaction, the Maker can feel the sense of organizational identity and belonging, which is more conducive to the creation of knowledge sharing tendency. Through research, Bohari has found that interaction and sense of identity in groups are the main motivations for knowledge sharing [22]. Social interaction is generally considered to be a channel and means for the flow of resources and information, and good social interaction enhances information exchange among users[23]. The more Makers communicate in social interaction, the more social network groups will share resources, complement each other's strengths, and build consensus to create new knowledge through constant brainstorming of ideas and creativity. Therefore, material interests and social interaction driven by extrinsic motivation are taken as two measurement indexes of makers' participation in knowledge sharing behavior, and the following hypotheses are proposed:

H1: Material interests have a significant positive effect on knowledge sharing willingness;

H2: Social interaction has a significant positive effect on knowledge sharing willingness;

2.2 Makers' Intrinsic Motivation and Willingness to Share Knowledge

Self-efficacy is a kind of cognition and judgment of self-competence, which refers to the degree of confidence of an individual to take part in and successfully completing a certain job or task. Self-efficacy is a belief, believe that you can clearly express and share thoughts, experiences, and can successfully complete a certain behavior. The high-self-efficacy Makers have no pressure on knowledge sharing, and will show great interest and be willing to participate in knowledge sharing. Kim et al. indicated that individuals with high self-efficacy have higher enthusiasm in knowledge sharing willingness[24]. When problems are complex, difficult, and challenging, sharing knowledge to solve them not only increases your confidence in your ability to solve them, but also increases your sense of self efficacy[25], Makers are someone who are interested in things, loves challenges, and has relevant experience that is different from the average user and the organization's employees, they start out with a high level of creativity, curiosity, and the willingness to put their ideas into action.

The ownership in this paper refers to the individual's psychological ownership of knowledge, that is, psychological ownership, which can not be protected by law, such as the individual from the practice of experience, capacity and so on. Xu Jianzhong pointed out that when employees have higher awareness of personal ownership of knowledge, they tend to be exclusionary, unwilling to share knowledge[26]. Baer and Brown have found that individuals with high perceptual ownership of knowledge feel violated territory and loss when their knowledge is used by others[27]. The research object of this paper is Maker, as an independent individual, so only the influence of individual perceived knowledge ownership on knowledge sharing willingness is studied. From the perspective of self-interest, the Makers will take measures to protect their own "knowledge territory" whether for the purpose of realizing self-worth or obtaining corresponding rewards, and are unwilling to share with others. Therefore, self-efficacy and perceived knowledge ownership, which are driven by intrinsic motivation, are taken as the two measurement indicators of creative knowledge sharing behavior:

H3: Self-efficacy has a significant positive effect on knowledge sharing willingness;

H4: Perceived personal knowledge ownership has a significant negative impact on knowledge sharing willingness;

2.3 Knowledge sharing willingness and knowledge sharing behavior

Willingness is the tendency of an individual to take a specific behavior, and it is the prerequisite for an individual to complete a specific behavior. Scholars point out that the generation of knowledge sharing behavior is composed of motivation and execution, and knowledge sharing willingness is the only way for knowledge sharing behavior to happen. Bock and Kim studied knowledge sharing behavior from the theory of Rational Behavior, which states that the stronger a person willingness to act, the more likely they are to perform a certain

behavior^[28]. Willingness is the subjective tendency of an independent individual to take a certain action or make a certain decision. The degree of inclination directly affects the occurrence of behavior and is easy to be affected by various external factors. According to the theory of Rational Behavior, an individual's behavior needs to be determined by willingness before it occurs, which affects the execution of the behavior. Chang's research results show that knowledge sharing willingness can significantly promote the occurrence of knowledge sharing behavior^[29]. The Makers is a group of innovation passion. They are keen on innovation and have a strong sense of innovation. Therefore, the following hypotheses are proposed:

H5: Knowledge sharing willingness has a significant positive effect on knowledge sharing behavior;

2.4 The moderating effect of information technology support

In the context of Open Innovation, high-tech technologies such as information and communication in community network platforms are regarded as a necessary condition to knowledge sharing. With the emergence of high-tech technologies such as cloud computing and big data, network technology has penetrated into all aspects of people's life; And people have long been accustomed to the convenience of technology, and even rely on technology. According to Zhou Tao, information support in a virtual health community is not only about the willingness of the user to stay in touch with the provider, but also about the trust and willingness to share knowledge through communication^[30]. Open innovation refers to the global information exchange and knowledge sharing and participation in innovation creation. The universality of knowledge resources and the risk of globalization, emphasizing the degree of openness of technology strategies within the enterprise Open Innovation^[31], The breakthrough of information technology breaks the barriers of communication and improves the timeliness and convenience of information communication through multi-channel and barrier free communication. Even if the Maker is willing to share knowledge, if communication barriers result in information receiving in a timely manner, technology backwardness leads to low efficiency of sharing, difficulty in operating information and time-consuming, etc. , it will no doubt weaken the willingness to participate, and even the willingness of Makers to reject knowledge sharing. The rapid development of mobile network and other technologies has widened the channels for users to participate in knowledge sharing, the convenience and flexibility of mobile terminal greatly enhance the enthusiasm of users to participate in knowledge sharing behavior^[32]. Therefore, the support of information technology can significantly affect the transformation of knowledge sharing willingness to knowledge sharing behavior. Therefore, the following hypothesis is proposed:

H6: Information technology strengthens the relationship between knowledge sharing willingness and knowledge sharing behavior, that is, the more supportive information technology is, the stronger the willingness to participate in knowledge sharing will be, and the more it can stimulate the occurrence of sharing behavior.

Based on the above assumptions, the model construction in this paper is shown in Figure 1:

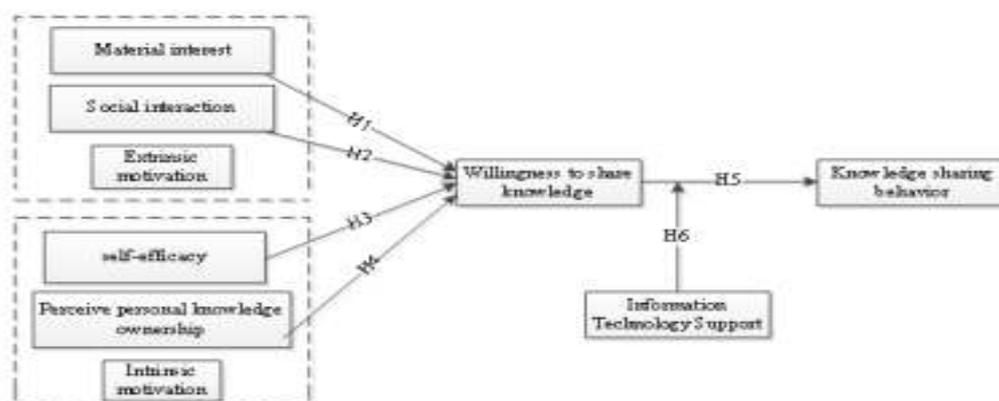


Figure 1. Conceptual model.

3. QUESTIONNAIRE DESIGN AND DATA COLLECTION

3.1 Variable measurement

In order to ensure the validity and reliability of the scale, this paper draws on the maturity measurement scales at home and abroad. However, due to the different research perspectives and theoretical models, the emphasis on the measurement indicators of the influencing factors of knowledge sharing in open innovation communities is different, and some variables cannot be directly applied by the appropriate maturity measurement scales. Therefore, most of the indicators in this paper are properly modified according to the specific situation of this paper, while a small part of the scale is designed by itself according to the connotation of independent variables. The measurement indexes of each variable are shown in Table 1. All items were measured with 5-point Likert scale, in which “1” means “Totally disagree” and “5” means “Totally agree”.

Table 1. Measurement items of Makers' Participation in Knowledge Sharing in Open Innovation Community.

Variable	Measurement index	source
Material interest(MI)	It is very important for me to get economic interests when I participate in knowledge sharing.	[33]
	Under similar conditions, I prefer to participate in open innovation communities with economic interests.	[34]
	I want my knowledge sharing behavior to earn me virtual interests (network points, forum gold, level upgrades, etc.).	[35]
	I hope that knowledge sharing can bring me some material benefits.	[36]
Social interaction(SI)	Community allows members to communicate, exchange and share knowledge and information.	[37]
	Through knowledge sharing, the communication channels between me and members have been broadened, which helps me make more friends, improve social relations, and cooperate smoothly with community members.	[38,39]
	It helps to maintain social interaction in the community through knowledge sharing.	[40]
	I am confident that I can provide valuable knowledge in the community.	
Self-efficacy(SE)	I am confident that my expertise, skills, experience, and insights will provide valuable knowledge to community members.	[32]
	I am confident that I can easily answer and supplement questions from other members of the community.	
	I believe it's easy to comment on and respond to community members' posts.	[41]
Perceiving Personal Knowledge Ownership (PPKO)	I consider the knowledge I share in the community to be my own.	[30]
	The knowledge and experience accumulated in the innovation activities of the open innovation community have nothing to do with others, but only belong to me.	[29]
	In the community, when I perceive that the ownership of knowledge is violated, I will refuse to share knowledge.	[34]
Willingness to share knowledge(WTSK)	I would like to recommend to my friends to join the open innovation community.	
	I am willing to browse the community information and actively participate in the community discussion.	[23]
	I am willing to actively share my knowledge, experience and skills in the community.	[34]
	I am happy to put forward my ideas and express my opinions in the community.	[42]

Information Technology Support (ITS)	The support of information technology in the community will improve my work efficiency and strengthen my willingness to participate, thus generating knowledge sharing behavior.	[43]
	In the community, the simpler the operation of information technology is, the more time and effort is saved, and the more willing I am to participate in knowledge sharing.	
Knowledge sharing behavior(KBS)	In the community, the better the information technology conditions, the stronger my willingness to participate in knowledge sharing.	self-designed
	In the open innovation community, I often share my creative knowledge and problem solutions with others through knowledge sharing.	[36]
	I often participate in knowledge sharing and innovation activities with community members.	
	I will often retweet and share valuable content from myself.	self-designed
	I will often share my accumulated experience, skills and my relevant views and insights with other members.	[32]

3.2 Sample survey

In this paper, we select “Xiaomi community”, Huawei’s “Pollen Club” and the creators of Haier’s HOPE platform, which are well-known enterprises in China, as the Open innovation. The questionnaire issued a total of 380, rejected the unqualified questionnaire, the final access to a valid questionnaire 338. Among the valid sample data, 50.9% and 49.1% were males, 40.7% were aged from 18 to 25, 2.1% were under 18, 22.5% were aged from 26 to 30, “31-40 years old” accounted for 7.7%, “41-50 years old” for 5.9%, “51-60 years old” for 0.9%, “61 and above” for 0.1%. The percentages of “Occupation” were 42.3%, 12.1%, 22.8%, 10.7%, 3.3% and 8.9% respectively for students, government officials, ordinary staff, professionals, manual workers and other occupations, 61.2%, 4.7%, 8.0%, 26.0%, and 74.3%, 15.7%, and 10.1%, respectively, have been involved in innovation for less than one year, 1-3 years and more than three years.

4. DATA ANALYSIS AND RESULTS DISCUSSION

4.1 The reliability analysis

This paper uses Cronbach’s consistency coefficient is used to test and analyze the reliability of the questionnaire. SPSS25.0 is used to analyze the questionnaire data, and the results are shown in Table 2. The reliability Cronbach α of all variables is greater than 0.7, indicating that the questionnaire data had good reliability and good internal consistency.

Table 2. Factor load of measurement index, Cronbach α .

Latent variables	Item quantity	Cronbach α	KMO
MI	4	0.873	0.806
SI	3	0.879	0.736
SE	4	0.923	0.837
PPKO	3	0.826	0.672
WTSK	4	0.927	0.844
ITS	3	0.882	0.744
KSB	4	0.930	0.854

MI--material interests, SI--social interaction, SE--self-efficacy, PPKO--perceived personal knowledge ownership, WTSK--willingness to share knowledge, TIS -- information technology support, KBS--behavior of sharing knowledge

4.2 Confirmatory Factor Analysis (CFA)

Table 3 shows that the factor load is greater than 0.6 and Ave is greater than 0.5, which shows that the questionnaire has good convergence validity. The diagonal of Table 4 is the AVE value of latent variable. It is observed that the average of AVE of each latent variable is larger than the coefficient of correlation between the variable and other variables.

Table 3. Convergence validity.

The path			Non-standardized factor load	S.E.	C.R.	P	Standardized factor load	AVE	CR
MI4	<--	MI	1				0.807		
MI3	<--	MI	1.078	0.065	16.572	***	0.821	0.634	0.874
MI2	<--	MI	1.022	0.067	15.351	***	0.809		
MI1	<--	MI	0.977	0.07	13.937	***	0.747		
SI3	<--	SI	1				0.854		
SI2	<--	SI	1.004	0.054	18.733	***	0.88	0.710	0.880
SI1	<--	SI	0.938	0.057	16.575	***	0.791		
SE4	<--	SE	1				0.793		
SE3	<--	SE	1.147	0.062	18.586	***	0.877	0.752	0.924
SE2	<--	SE	1.215	0.064	18.966	***	0.903		
SE1	<--	SE	1.166	0.062	18.913	***	0.892		
PPKO3	<--	PPKO	1				0.751		
PPKO2	<--	PPKO	1.295	0.09	14.323	***	0.887	0.631	0.836
PPKO1	<--	PPKO	1.074	0.085	12.603	***	0.737		
WTS4	<--	WTSK	1				0.876		
WTS3	<--	WTSK	0.986	0.045	22.078	***	0.871	0.762	0.928
WTS2	<--	WTSK	1.057	0.049	21.725	***	0.878		
WTS1	<--	WTSK	0.961	0.045	21.51	***	0.867		
ITS3	<--	ITS	1				0.869		
ITS2	<--	ITS	1.006	0.056	18.125	***	0.835	0.714	0.882
ITS1	<--	ITS	0.939	0.052	18.074	***	0.83		
KSB4	<--	KSB	1				0.916		
KSB3	<--	KSB	0.937	0.044	21.285	***	0.825	0.773	0.931
KSB2	<--	KSB	1.042	0.04	26.243	***	0.908		
KSB1	<--	KSB	0.955	0.04	23.862	***	0.864		

Note: *** stands for P<0.001

Table 4. Discriminating validity.

Variable	MI	SI	SE	PKO	WTSK	ITS	KSB
MI	0.634						
SI	0.464***	0.710					
SE	0.491***	0.445***	0.752				
PPKO	0.366***	0.477***	0.530***	0.631			
WTSK	0.487***	0.518***	0.535***	0.297***	0.762		
ITS	0.426***	0.355***	0.418***	0.180***	0.607***	0.714	
KSB	0.373***	0.376***	0.477***	0.197***	0.485***	0.484***	0.773
AVE square	0.796	0.842	0.867	0.795	0.873	0.845	0.879

Note: *** stands for $P < 0.001$, the diagonal line is AVE to evaluate the amount of variance variation extraction.

4.3 Hypothesis testing

AMOS24.0 software is used to analyze the model, and the calculated results are shown in Table 5 and Table 6. The fitting indexes and parameter estimates of the structural equation model as well as the test results are shown in Figure 2 below. The results show that the fitting degree of each index reaches the ideal state, and there are no problems such as collinearity. All research hypotheses are valid:

Table 5. Fitting index of structural equation model.

Indicators	Evaluation standard		This model values	Fitting situation	
	Can accept	ideal			
Absolute fitting exponent	CMIN/DF	[3.0,5.0)	(2.0,3.0)	2.040	ideal
	RMSEA	<0.09	<0.08	0.056	ideal
	GEI	[0.70,0.9)	>0.9	0.905	ideal
Relative fitting index	NFI	[0.70,0.9)	>0.9	0.930	ideal
	CFI	[0.70,0.9)	>0.9	0.963	ideal
	IFI	[0.70,0.9)	>0.9	0.963	ideal
Concise fitting index	PNFI		>0.5	0.785	ideal
	PGFI		>0.5	0.698	ideal

Table 6. Hypothesis testing.

The path	Nonstandardized coefficient	Normalized coefficient	S.E.	C.R.	P	hypothesis
WTSK <--- MI	0.198	0.192	0.058	3.419	***	support
WTSK <--- SI	0.272	0.297	0.052	5.195	***	support
WTSK <--- SE	0.346	0.4	0.052	6.619	***	support
WTSK <--- PPKO	-0.141	-0.142	0.057	-2.484	0.034	support
KSB <--- WTSK	0.998	0.821	0.116	8.632	***	support

Note: *** stands for $P < 0.001$

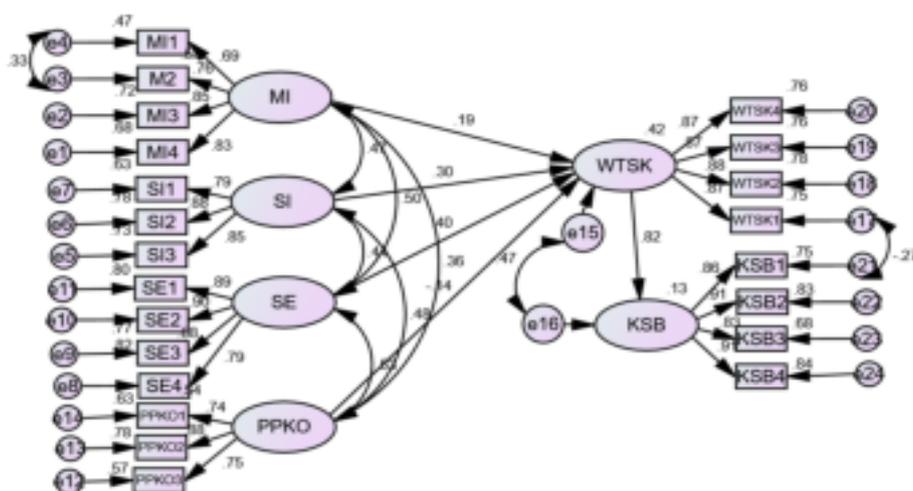


Figure. 2. Structural equation model of influencing factors of makers' participation in knowledge sharing in open innovation communities

This paper uses PROCESS to test the moderating effect of information technology on sharing willingness and sharing behavior. The results shown in Table 7 below are significant. In combination with the test chart in Figure 3, it can be seen that information technology has a significant positive moderating effect, that is to say, the support of information technology strengthens knowledge sharing willingness.

Table 7. Test results of regulatory effect.

	Model 1			Model 2		
	Beta	T	significant	Beta	T	significant
WTSK	0.367	7.478	***	0.393	8.624	***
IST	0.367	7.495	***	0.361	7.928	***
Interactive items				0.291	7.437	
R square		0.408			0.492	
F		115.249			107.725	

Note: *** stands for P<0.001

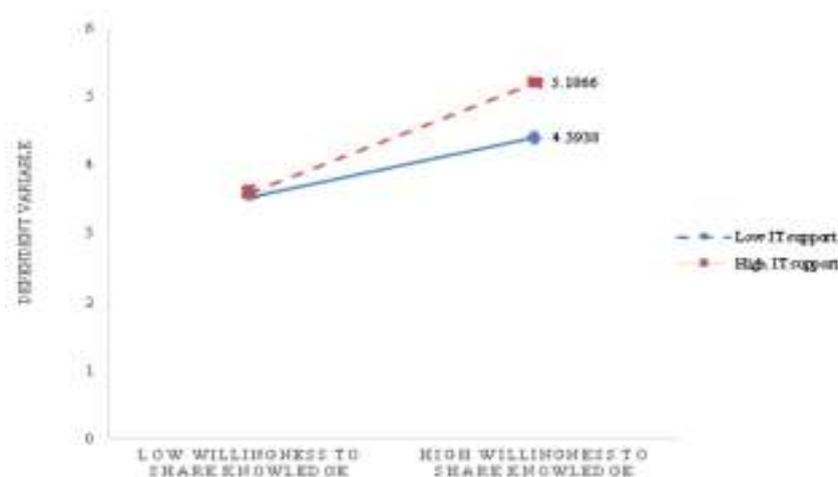


Figure3. Test diagram of regulatory effect.

4.4 Discussion of results

The structural equation fitting index shows that the research model of influencing factors on knowledge sharing behavior of IT driven OIC is reasonable and reaches the ideal state. The results are as follows:

(1) The results of the hypothesis test table show that material interests, social interaction and self-efficacy have significant positive effects on knowledge sharing willingness, while perceived personal knowledge ownership has significant negative impact on knowledge sharing willingness, and knowledge sharing willingness is a prerequisite for knowledge sharing behavior, and knowledge sharing willingness has a significant positive impact on knowledge sharing behavior.

(2) As can be seen from the structural equation path, the standardized coefficients of each variable from high to low is 0.4 of self-efficacy, 0.297 of social interaction, 0.192 of material interests and -0.142 of perceived personal knowledge ownership. That is to say, the factors that have a positive impact on the willingness to share knowledge are self-efficacy, social interaction and material interests in descending order of influence degree. The standardized coefficient of sharing willingness is 0.998, which means that the behavior must be generated through willingness. From the perspective of opportunity cost and income, Makers pay time and energy to participate in knowledge sharing behavior, but the results show, material interests has the lowest impact on Makers, while self-efficacy has the greatest impact on Makers. So they are more care about their spiritual needs, and more concerned with self-affirmation. Social interaction, as a necessary means of social interaction, not only contributes to the accumulation and sharing of knowledge, but also enhances self-efficacy in the process of sharing, which can achieve win-win or multi win, it also has a significant positive impact on the Makers; the perceived personal knowledge right has a significant negative impact on the willingness of knowledge sharing, because the Makers take the realization of creativity as a way to measure personal value, and has a stronger awareness of knowledge protection. For the knowledge sharing that may infringe personal interests, the Makers are reluctant to share knowledge because of their resistance to it, the stronger the sense of personal knowledge ownership, the greater the negative effect.

(3) It can be seen from the two slopes of high and low it supports in the moderating effect diagram that it support plays a moderating role in the willingness and behavior of knowledge sharing. Under the motivation, Makers produce the willingness to share knowledge, but the generation of willingness is not necessarily related to the occurrence of behavior, and information technology support plays a positive moderating role in it, that is, the more information technology supports, the stronger the knowledge sharing will, the more it can stimulate the knowledge sharing behavior occurrence.

5. CONCLUSION AND PROSPECT

Based on the internal and external motivation of the Makers, this paper constructs an influential factor model of IT driven knowledge sharing by the Makers in the OIC, and collects data through questionnaire survey, and uses structural equation and SPSS plug-in PROCESS to analyze the data. It is verified that material interests, social interaction, self-efficacy and knowledge sharing willingness have significant positive effects on knowledge sharing behavior, sense of personal knowledge ownership significantly negative influence knowledge sharing behavior; Furthermore, it is found that the influencing factors from deep to shallow are self-efficacy, social interaction, material benefits and perceived personal knowledge ownership; And IT support plays a positive moderating role in knowledge sharing willingness and knowledge sharing behavior. Above all, when the enterprise Open innovation community attracts the creators to participate in knowledge sharing and community management, according to the influence degree of the influence factors on the creators, the priorities are sorted out and the targeted measures are taken, in this way, enterprises can avoid taking "Detours" or "Go wrong ways" and achieve the expected goal smoothly. At the same time, we should make full use of the

moderating effect of information technology on Maker's willingness and behavior of knowledge sharing, optimize the information technology system, and increase the support of information technology, to improve the enterprise innovation ability and enterprise performance have twice the result with half the effort.

This paper only studies the factors that influence the knowledge sharing behavior of IT driven Open innovation, it does not consider the impact of the environment of OIC platform on Makers, which has certain research limitations. At the same time, the questionnaire is mainly distributed to the well-known OICs in China. In the later stage, the research perspective will be expanded, the environmental factors of community platform will be added, and the sample data will continue to be expanded, and the questionnaire will be put on foreign open innovation community platforms for in-depth- research.

ACKNOWLEDGEMENT

Fund projects: The National Social Science Fund of China(17BGL028).

REFERENCES

- [1] Chesbrough H W. (2003).Open innovation: The new imperative for creating and profiting from technology. Boston: Harvard Business Press
- [2] Grant R M. (1996).Toward A Knowledge-Based Theory of the Firm. *Strategic Management Journal*, 17(S2):109-122
- [3] Simard C, West J. (2005).Knowledge networks and the geographic locus of innovation. *Open Innovation Researching A New Paradigm*.
- [4] Wang Xinchun, Qi Guijie, Liang Yikai. (2016). Research on enhancement of knowledge creation ability of open innovation community organizations. *Journal of Intelligence*, 35(3):203-207,202(in Chinese)
- [5] Mika L A, Jukkapekka B A, Ari J B, et al. (2013).Knowledge sharing in the open innovation process — Case: Grid computing
- [6] Haefliger S, Monteiro E, Foray D, et al. (2011). Social software and strategy. *Long Range Planning*, 44(5-6):297-316.
- [7] Xia Enjun, Zhang Ming, Jia Shunan, et al. (2013).Influencing factors of network innovation performance of open innovation community: An Empirical Study Based on Internet platform. *Technical economy*,32 (10): 1-7(in Chinese)
- [8] Hrastinski S , Kviselius N Z , Ozan H , et al. (2010).A Review of Technologies for Open Innovation: Characteristics and Future Trends. In: *System Sciences (HICSS)*. Vol I. Hawaii: 43rd Hawaii International Conference on IEEE
- [9] Cheng C C J, Yang C. Sheu C. (2016).Effects of open innovation and knowledge-based dynamic capabilities on radical innovation: An empirical study. *Journal of Engineering & Technology Management*, 41:79-91
- [10] Chen Jiali. (2019). Research on the impact of open innovation platform user network on innovation from the perspective of social capital. D Thesis. Jinan: Shandong University(in Chinese)
- [11] Briel F V, Recker J. (2017).Lessons from a failed implementation of an online open innovation community in an innovative organization. *MIS Quarterly Executive*, 16(1):35-46.
- [12] Chen Nan. (2017). Research on knowledge sharing process of Internet based open innovation platform participants. Ms Thesis. Baoding: Hebei University(in Chinese)
- [13] Kosonen M, Gan C, Vanhala M, et al. (2014).User Motivation and knowledge sharing in idea crowdsourcing. *International Journal of Innovation Management*, 18(5):1450031-
- [14] Liu Jingyan, Wang Yu, Lin Li. (2020).The impact of user participation innovation on enterprise community innovation performance in OIC: A social network perspective. *Science and Technology Progress and Countermeasures*, 37(6):128-136(in Chinese)
- [15] Wang Nan, Zhang Shikai, Zhao Yuru, et al. (2019).The influence of leading user characteristics on knowledge sharing level in online communities: The mediating role of social capital. *Management Review*, 31(2):82-93(in Chinese)

- [16] Qin Min, Qiao Han, Chen Lianghuang. (2015). Research on online user contribution behavior of open innovation communities based on CAS theory: A Case Study of China's Leading Enterprise Communities. *Management Review*, 27(1):126-137(in Chinese)
- [17] Guo Dongqiang, Lin Qin. (2016). Research on the knowledge cloud dissemination mode of open innovation community. *Science and Technology Progress and Countermeasures*, 33(20):13-17(in Chinese)
- [18] Xu Jiabo. (2015). "Maker" holds up the new upsurge of "double innovation". *Contemporary economy*, 000 (015): 6-8 (in Chinese)
- [19] Koo, Yu, Chua & Han. (2020). Relationships among Emotional and Material Rewards, Job Satisfaction, Burnout, Affective Commitment, Job Performance, and Turnover Willingness in the Hotel Industry, 21(4), 371-401
- [20] Ren Ling. (2019). Research on influencing factors and development approaches of knowledge sharing in online open innovation community. *Information science*, 37 (9): 48-53(in Chinese)
- [21] Xing Caixia. (2017). Research on influencing factors of maker knowledge sharing in virtual community. Ms Thesis. Taiyuan: North University of China (in Chinese)
- [22] Bohari A M. (2016). Knowledge contribution determinants through social network sites: Social relational perspective. *International Re-view of Management and Marketing*, 6(3): 454-459
- [23] Liu Yuqing, Chang Guilin. (2018). Research on influencing factors of knowledge sharing among mobile reading users in university libraries. *Library Science Research*, (18):54-62(in Chinese)
- [24] Hannah Kim & Sung Eun Oh. (2020). Individual characteristics influencing the sharing of knowledge on social networking services: Online identity, self-efficacy, and knowledge sharing willingness. 39(4):379-390
- [25] Jennex M E. (2005). Knowledge Management Metrics: A Review and Directions for Future Research. *International Journal of Knowledge Management*, 1(2):20-32
- [26] Xu Jianzhong, Li Fengshu, Huang Jinghan. (2020). Georg Simmel connection, knowledge density, and disruptive technological innovation. *Journal of Management Science*. 41(1):79-89 (in Chinese)
- [27] Baer M., & Brown G. (2012). Blind in one eye: How psychological ownership of ideas affects the types of suggestions people adopt. *Organizational Behavior and Human Decision Processes*, 118(1), 60-71
- [28] Bock G W, Kim Y G. (2002). Breaking the myths of rewards: An exploratory study of attitudes about knowledge sharing. *Information Resources Management Journal*, 15(2):14-21
- [29] Chang C M, Hsu M H, Lee Y J. (2015). Factors influencing knowledge sharing behavior in virtual communities: A longitudinal investigation. *Information Systems Management*, 32(4): 331-340
- [30] Zhou Tao, He Lianzi, Deng Shengli. (2020). Research on the influencing factors of user knowledge sharing in open innovation community. *Modern Intelligence*, 40(3):58-64(in Chinese)
- [31] Lichtenthaler U, Ernst H. (2010). Opening up the innovation process: The role of technology aggressiveness. *R&D Management*, 39(1):38-54
- [32] Gu Meiling, Bi Xinhua. (2017). Knowledge synergy in open innovation communities in mobile environment: A perspective of knowledge ecology. *Library and Information Service*, 61(13):99-107(in Chinese)
- [33] Zhao L, Detlor B, Connelly C E. (2016). Sharing knowledge in social Q&A sites: The unintended consequences of extrinsic motivation. *Journal of Management Information Systems*, 33(1):70-100
- [34] Chen Xiangjun. (2019). Research on the influencing factors of knowledge sharing behavior in open innovation communities. Ms Thesis. Harbin: Heilongjiang University(in Chinese)
- [35] Yang Chen. (2019). Research on makers' knowledge contribution behavior and promotion strategies in open innovation community. Ms Thesis. Changchun: Jilin University(in Chinese)
- [36] Li He, Zhang Keyong, Hong Chuang. (2017). Research on the influencing factors of makers' knowledge sharing in open innovation community. *Library and Information Service*, 61(21):13-21(in Chinese)

- [37] Lin H C, Chang C M. (2018).What motivates health information exchange in social media? The roles of the social cognitive theory and perceived interactivity. *Information & Management*, 55(6):771-780
- [38] Bock G W, Zmud R W, Kim Y G, et al. (2005).Behavioral willingness formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *Mis Quarterly*, 29(1):87-111
- [39] Hsu M H, Ju T L, Yen C H, et al. (2007). Knowledge sharing behavior in virtual communities: The relationship between trust, self-efficacy, and outcome expectation. *International Journal of Human - Computer Studies*, 65(2):153-169
- [40] Geng Ruili, Shen Jing. (2019).Research on knowledge sharing behavior motivation of social network users from different cultural perspective. *Journal of Library Science in China*, 45(01):60-8(in Chinese)
- [41] Zhang Xudong. (2017). Research on knowledge sharing platform design of maker education in colleges and universities. Ms Thesis. Wuxi: Jiangnan University(in Chinese)
- [42] Chen L H. (2016). An empirical study on the influencing factors of user participation behavior in open innovation communities. Ms Thesis. Nanchang: Jiangxi Normal University(in Chinese)
- [43] Davis F D. (1989).Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319-340

Short Research Paper

The influence mechanism of product attributes on product demand in online reviews

Ying Chen¹, Li Li^{1,2}, Yulan Lin^{1*}

¹ School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China

² Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094, China

Abstract: To explore the influence mechanism of product attribute characteristics contained in online reviews on users' purchase decisions can provide decision-making basis for the improvement of products and services of e-commerce enterprises. Based on the previous empirical research and management practice, we put forward the hypothesis that each product attribute in e-commerce online reviews affects product sales, and use the method of panel data modeling to explore the influence mechanism and degree of each product attribute in online reviews on product demand. The results show that there are significant differences in the influence mechanism and degree of different attributes on product demand. In apparel e-commerce, users pay most attention to the size attribute characteristics of products. This study can guide e-commerce enterprises to improve their products and services.

Keyword: Online review, demand forecasting, panel data modeling

1. INTRODUCTION

In recent years, with the rapid development of Internet information technology, online reviews in e-commerce have been developing rapidly. Online users generate a large number of online comments recording consumer experience on the e-commerce platform, and online comments have become an important source of information that affects consumers' attitude and behavior in online shopping^[1]. E-commerce online review, as an important part of customer feedback system, is an important way for e-commerce enterprises to understand consumers' needs and wishes, as well as an important way for consumers to understand enterprise goods and services and reduce information asymmetry^[2]. Before purchasing goods, consumers often browse online reviews to aid decision-making.

On the whole, e-commerce online review has increasingly become the focus of academic circles and e-commerce enterprises. In the research of e-commerce online review, some scholars focus on the research of online review content, using text mining and algorithms to extract online review content features, then study the usefulness of online review and online review sentiment analysis; in recent years, more and more researchers began to study the impact of online review on consumer purchase decision. Senecal and Sylvain^[3] conducted an online experimental study on 487 subjects and found that consumers are more likely to buy products after viewing the comments of other users; Duanetal^[4] believed that the number of reviews has the greatest impact on consumers' purchase behavior, which is different from the content of reviews generally considered by people; Chevalier and Mayzlin^[5] have studied the impact of consumer reviews on the relative sales of Amazon and bainsannoble books, and found that customer reviews can promote the sales of books, particularly, the impact of negative reviews is greater. At the same time, many domestic scholars have also done such research. Nie Wei and Tian Juanjuan^[6] found that the quality and potency of online reviews have a positive impact on consumers' purchase intention through perceived usefulness; Li Haiqin^[7] used structural equation model to test the model, and found that the quality of online reviews and the professional ability of recipients have a significant direct

* Corresponding author. Email: lily691111@njust.edu.cn (Li Li)

positive impact on online consumers' purchase intention; Huang Hua and Mao Haifan^[8] constructed a transmission mechanism about the influence of the length, quality, quantity and timeliness of negative online reviews on consumers' purchase intention, and put forward suggestions to enterprises and consumers respectively according to the conclusions.

The existing literatures have analyzed the influence of digital reviews(including star levels, scores,etc.), quality and potency of online reviews on consumers' purchase decisions. These studies have enriched the literature content of the influence of online reviews on consumers' purchase decisions,to a certain extent. However, they have not considered the product attributes contained in online reviews. In fact, in the process of purchasing decisions, consumers will refer to various attributes of products^[9]. In the e-commerce environment, the websites will provide a large number of online comments, and these online comments often include consumers' comments on product attributes^[10], which provides convenience for research and provides data support for researchers to explore the impact mechanism of online comments on users' product needs .

Therefore, we explore the influence mechanism of product attributes on product demand in e-commerce online reviews. Under the background of big data, the content of online reviews can be recorded accurately, which provides a new idea for the study of the influence mechanism of various attribute characteristics on product demand. In this paper, we use panel data model to analyze the influence mechanism of online reviews on product demand, which has guiding significance for the improvement of products and services of e-commerce enterprises.

We make a supplement to the existing literature from three aspects: the first aspect is the data sources, most of the existing research data sources are questionnaire survey data. However,our research data comes from an e-commerce fashion women's clothing website, which is more objective; The second one is research content, although the existing literature involves the impact of Online reviews on consumers' purchase decisions, there is a lack of specific research on online reviews. The last is the research method,the existing literatures mostly use multiple regression analysis, structural equation and other data analysis methods, while we use panel data modeling method, which has a certain degree of innovation.

2. HYPOTHESIS

In the process of obtaining product information, users will often browse the comments of other users who have purchased the products or services, which will affect their purchase decision and then has effect on the sales volume of the products. The attributes of products are various, by summarizing the views of some scholars, Mark Perry proposed that product attributes mainly include the internal attributes and external attributes. The internal attributes refer to the physical composition of the products, such as material; the external attributes refer to the attributes that can be recognized when the product is not used, such as product service^[11]. For clothing product attributes, after filtering out useless variables through data preprocessing, the proposed product attribute features include size, material, breakage, loose / off button, off-line, wool dropping, threads, zipper (bad), odor, discoloration, dirtiness, and these attributes can be divided into product specification attributes, quality attributes, users' perception attributes. Then,we will explore the impact of these three types of product attributes on product sales respectively.

Product specification attributes mainly include product size, volume and other attributes. Yu Yishuang takes Taobao clothing products as the research object and empirically analyzes the impact of online reviews on clothing sales.This paper studies the influence of clothing size, color difference and other attributes on sales, and the results show that color difference and size have no significant impact on sales^[12]. However, this paper fails to specifically analyze whether there is multicollinearity between product size, color difference and other comment features. Generally speaking, when a user found that the size of a garment is wrong after purchasing a

product, he / she will write down his / her intuitive feelings on the e-commerce website. Other customers who want to buy the product will browse these comments of the product. After seeing other user's poor comments on the size, they maybe change their purchase decision. Therefore, the following hypotheses can be put forward.

H1: product size in online reviews has a significant impact on sales.

There will be differences in users' concerns about different products, such as the price and ease of use of cameras. Zhang Meng^[13] and others have found that if words such as service and transportation convenience appear in hotel reviews, they can show more real situation of the hotel. Archak^[14] and other scholars believe that the quality of product design may be an important determinant of consumers' purchase decisions. The material of clothes is often one of the product attributes that consumers most concerned. Generally speaking, the better the material of clothes, the more consumers like them. When users find that the clothes are damaged after receiving the products, it will greatly affect their views on this kind of products. The quality attributes of products also include some attributes that reflect the poor quality of products, such as loose / off buttons, off-line, hair falling and defective products. The increase in the number of these comments will often make other users who want to buy such products give up the idea of buying, thus affecting their purchase decisions. Therefore, the following hypotheses can be put forward.

H2a: the product material in online reviews has a significant impact on sales;

H2b: product breakage in online reviews has a significant impact on sales;

H2c: the product button loose / off in online reviews has a significant impact on sales;

H2d: product off-threads in online reviews has a significant impact on sales;

H2e: the product wool dropping in online reviews has a significant impact on sales;

H2f: product threads in online reviews has a significant impact on sales;

H2g: zipper in online reviews have a significant impact on sales.

Facing different types of products, consumers are eager to get different product information^[15]. For exterior products such as fashion women's wear, users want to know more about the external information related to the product, such as the color and other attributes of the product^[16]. The users' perception attributes of products include odor, discoloration and dirtiness. After receiving the product, whether the product has peculiar smell will directly affect the user's purchase experience, and thus affect the perceived value of other users and their purchase decision. Therefore, the following hypotheses can be put forward.

H3a: product odor in online reviews has a significant impact on sales;

H3b: product discoloration in online reviews has a significant impact on sales;

H3c: product dirtiness in online reviews have a significant impact on sales.

3. RESEARCH METHOD

In order to explore the impact of online review product attribute characteristics on product sales, we adopt the method of panel data modeling. Panel data refers to the sample's data that takes multiple sections in the time series and selects the sample's observation value on the section. The variables used in the panel data model include the duality of time series and cross section.

We use panel data modeling method to build a panel data model. Compared with cross-sectional data, panel data can provide individual dynamic behavior information, and the sample's size is larger. Because the time dimension of the panel data used in this study is small, it is a short panel data, and we adopt the short panel data modeling method. In the process of building the model, quarter is taken as the time variable to study the influence mechanism of product attribute characteristics in online reviews of different commodities on product demand in different quarters. Finally, the panel data model established in this paper is shown in formula (1).

$$y_{it} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_7 + \alpha_8 x_8 + \alpha_9 x_9 + \alpha_{10} x_{10} + \alpha_{11} x_{11} + \alpha_{12} x_{12} + u_{it} \quad (1)$$

Among them, y_{it} is the explained variable, which represents the product sales volume of the i -th commodity in the t period; the explanatory variable adopts the characteristics of each attribute of the product in online reviews. At the same time, in addition to the attributes of the product itself, there are other factors that may also affect the sales volume. Li Jian points out that the fluctuation of commodity price affects the purchase decision of users to a certain extent through summarizing previous literature^[17]. Therefore, we will take the product price as the control variable of the model. The explanation of each variable of the model is shown in Table 1.

Table 1. Variable description of panel data model

<i>Variable type</i>	<i>Variable name</i>	<i>Variable explanation</i>
explained variable	sales	Average daily sales volume of the i -th commodity in quarter t
	size	The proportion of size reviews of the i -th commodity in quarter t
	material	The proportion of material reviews of the i -th commodity in quarter t
	breakage	The proportion of breakage reviews of the i -th commodity in quarter t
	button loose / off	The proportion of button loose / off reviews of the i -th commodity in quarter t
explanatory variable	wool dropping	The proportion of wool dropping reviews of the i -th commodity in quarter t
	off-threads	The proportion of off-threads reviews of the i -th commodity in quarter t
	threads	The proportion of threads reviews of the i -th commodity in quarter t
	zipper	The proportion of zipper reviews of the i -th commodity in quarter t
	odor	The proportion of odor reviews of the i -th commodity in quarter t
	discoloration	The proportion of discoloration reviews of the i -th commodity in quarter t
control variable	price	The proportion of price reviews of the i -th commodity in quarter t

4. EMPIRICAL RESEARCH

4.1 The data

The data in this paper comes from an e-commerce fashion women's clothing website, where consumers can buy and comment on all kinds of products. The data mainly includes online review data and order data. Due to the complexity of corpus data, data preprocessing is needed. In the model, the main data include the control variable(price),the explained variable ,including product sales, and the explained variable,which contains the product attribute characteristics. We select 3470 kinds of products in 2017 as the sample.

We extract the product attribute features reflected in online reviews, including product specification attributes, product quality attributes and users' perception attributes, and calculating the proportion of various attribute features in the total reviews in a certain period as the explanatory variables of the product; because the product demand can not be measured, but the sales volume of the product can be obtained, and the product demand can be reflected from the side, therefore. The explanatory variable of this paper is the average daily sales volume of the product,its calculation method is the ratio of the total sales volume of the product and the corresponding date in the order data table. The control variable of the product is the price of the product. Because the product is the fashionable women's clothing, the price value is relatively large. Therefore, the price variable is obtained by taking the logarithm of the product price, and the bottom is the natural logarithm E . The logarithm processing can make the data more stable and weaken the collinearity and heteroscedasticity of the model. The individual data of panel data is the commodity code of each commodity, with quarter as the time data, January to March as the first quarter, and then every three months as the next quarter. Finally, the panel data for the study are shown in Table 2.

Table 2. Panel data

<i>Prodcode</i>	<i>quarter</i>	<i>Size</i>	<i>dirtiness</i>	<i>sales</i>
00c741a448848f680dc7a2cc2ff84c00	1	0.05	0.17	4.94
00c741a448848f680dc7a2cc2ff84c00	2	0.05	0.62	6.18
.....
00d8dcb63f3e47cb0c6728c7d4dd79de	3	0.2	0.67	4.12
00d8dcb63f3e47cb0c6728c7d4dd79de	4	0.2	0.39	2.63

We take the product attributes and average daily sales volume of different products in different quarters of online reviews as the research object. Before the panel data analysis, we make a correlation analysis between explanatory variables to avoid the problem of multiple collinearity of explanatory variables. From the correlation analysis between variables, except the correlation coefficient between threads and off-threads attributes is $0.792 > 0.7$, the correlation coefficient between other variables is small, and the correlation between most of the independent variables is significant. Therefore, in the follow-up experiments, we will eliminate the off-threads variable and reject the hypothesis of H2d.

4.2 Result

The Table 3 is the descriptive statistics of variables. Through the descriptive statistical analysis of variables, it can be found that users are most concerned about the size and material of clothing, and less about the problem of clothing wool dropping. The reason for this phenomenon may be that the wool dropping of clothing needs to be found after using the product for a period of time, and the impact on the user's experience is not as big as other factors.

Table 3. Variable descriptive statistical analysis results

<i>Variable</i>	<i>Obs</i>	<i>Means</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
Size	3470	0.606	0.299	0.000	1.000
material	3470	0.557	0.205	0.000	1.000
breakage	3470	0.093	0.151	0.000	1.000
button loose / off	3470	0.024	0.062	0.000	0.500
wool dropping	3470	0.022	0.076	0.000	1.000
threads	3470	0.379	0.083	0.000	1.000
zipper	3470	0.405	0.144	0.000	0.802
odor	3470	0.099	0.051	0.000	0.250
discoloration	3470	0.245	0.036	0.000	0.447
dirtiness	3470	0.207	0.266	0.001	0.907
price	3470	5.078	0.509	3.378	6.907
sales	3470	4.245	1.411	0.693	9.458

Because the time dimension of the panel data used in this study is small, it is a short panel data, and we adopt the short panel data modeling method. Through the panel data modeling, adding time dimension for analysis, the explanatory variable is the product attribute characteristics of different products in different quarters, the price is taken as the control variable, and the explained variable is the average daily sales volume of different products in different quarters. The data is imported into Stata for analysis and the analysis results of the panel model are shown in Table 4.

Table 4. Panel model analysis results

<i>sales</i>	<i>Coef.</i>	<i>Std. Err.</i>	<i>t</i>	<i>P> t </i>
Size	-.759	.131	-5.80	.000
material	-.844	.111	-7.59	.000
breakage	-.822	.166	-4.96	.000
button loose / off	-.129	.364	-3.10	.002
wool dropping	-.574	.295	-1.94	.052
threads	-.411	.451	-5.35	.000
zipper	-.619	.200	-4.79	.000
odor	-.359	.799	-2.95	.003
discoloration	-.711	.707	-0.31	.000
dirtyness	-.049	.085	-2.38	.000
price	.139	.045	3.10	.002
constant item	4.759	0.535	8.89	3.710

R2=0.181

Prob>F=0.000

From the test results of the model, the F test p value of all the coefficients of the model is 0, which is less than 0.05, so the fitting result of the model is good; the goodness of fit r of the model is 0.181, which shows that the fitting effect is also good. According to the weight of regression coefficient, size, material and breakage attributes are the main indicators that affect the sales volume of products, and the most important one is the product material attribute, which indicates that users are very concerned about the product material. The more comments about the product material attribute, the less users are likely to buy the product. From the results of the model, it can be seen that the test p value of the influence of wool dropping on product sales is 0.052, indicating that the influence is not significant, and the hypothesis of H2e is rejected.

It can be seen from the analysis results of panel model that, for the product specification attribute characteristics, the influence coefficient of size on the product sales is negative, and the test p value is 0.000, which indicates that the influence of product size on the sales is significant, and the hypothesis H1 is accepted; for the product quality attribute characteristics, the influence coefficient of breakage, material, button loose / off, threads, etc. on the product sales is significant, and the influence coefficient is negative, which means that the more comments about product quality problems, the less sales of products, so accept the hypothesis H2a, H2b, H2c, h2f. For the users' perception attributes, the three hypotheses of hypothesis H3 all been verified, that is, odor, discoloration, dirtiness and other product attributes have a significant impact on product sales. The control variable price has a significant impact on product sales, and the coefficient is greater than 0, which indicates that the higher the product price, the more products users buy. Because the product is positioned as a luxury fashion for women, and its users pursue higher spiritual enjoyment, high price often means high quality and fashion sense of the product.

The research results of this paper have guiding significance for the product improvement direction of e-commerce enterprises. Based on the above discussion results of the influence mechanism of online reviews on product demand, most of the product attributes have significant influence on product demand. For e-commerce enterprises, we can use the research results of this paper for reference to make some improvements to the products.

Users are very concerned about the quality attributes of products. E-commerce enterprises selling clothing can pay attention to the attributes of product materials and zippers in the user's online comments, and make

certain improvements on the product materials. The material attributes of products have the greatest impact on the product sales. Therefore, e-commerce enterprises must strictly control the raw materials of products and pay attention to the product materials when making products. E-commerce enterprises should strengthen communication with users to avoid the problem of different sizes of clothes, users' perception of products also has a significant impact on product sales, from our study, we can see that the odor, discoloration, dirtiness and other problems of clothing will have a significant negative impact on product sales. Therefore, e-commerce enterprises should pay attention to the preservation of products, so as to avoid problems such as odor, discoloration and so on.

5. CONCLUSIONS

We study the influence mechanism of e-commerce review features on product demand. Different from previous studies, the data in our paper comes from the order data and review data of an e-commerce website. Taking the specific text content of online reviews as the research object, The product attribute features in online reviews are extracted, and the influence of product attribute features on product sales is discussed through panel data modeling.

Through the analysis of the panel data model, it is found that many product attributes in online reviews do have a significant impact on product sales. The specific conclusions are as follows:

(1) Product specification attributes, such as size, are one of the most concerned product attributes when users purchase, which has a significant impact on product sales, and it is consistent with the situation in real life. When buying clothing and other products, users often try to see if they fit firstly.

(2) The quality attribute of products is also an important factor affecting the sales of products. The material of clothing products has the greatest correlation with product sales, which indicates that users are very concerned about the comfort of products; product attributes such as breakage, threads, loose / off buttons have a significant impact on product sales; the impact of wool dropping on product sales is not significant. To a certain extent, e-commerce enterprises can learn from the research results of this paper to improve the properties of clothing, such as material, button, zipper and so on.

(3) The users' perception attributes include whether the product has odor, dirtiness and discoloration. Through the analysis of the panel model, we can find that the products with odor, dirtiness and discoloration have significant effects on the product sales;

At the same time, in the process of our study, we found that there are still some limitations, some problems still need to be further explored, mainly including the following:

(1) Research object: the research object of our paper is only limited to clothing products, which has no universal significance, and can only provide certain reference for e-commerce enterprises operating clothing products; in the follow-up research, we hope to analyze more commodity types and explore the influence mechanism of online reviews on product demand;

(2) Due to the lack of some order data, the study only uses the order data and review data in 2017, the sample size of the study can be larger. The follow-up research hopes to use several years of order data and review data as samples to explore the influence mechanism of online review product attribute characteristics on product demand.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71771122 and the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant SJCX19_0045. Authors thank Made-in-China.com for the financial, business and technical supports.

REFERENCES

- [1] Li D., Yang J., Chen J.(2019). Analysis on Influencing Factors of tea e-commerce consumer satisfaction based on online reviews. *Forestry economics*, 41 (10): 70-77.
- [2] Li Y.(2017). Comparative study on online reviews of e-commerce platform users. *Modern intelligence*, 37 (07): 79-83
- [3] Sylvain S(2004). The Influence of online product recommendation on consumers' online choices. *Journal of Retailing*, 80(2):159-169.
- [4] Whinston D, A.B(2005). Analysis of herding on the internet-An empirical investigation of online software download[J]. *Journal of the Eleventh Americas Conference on Information Systems*, 7(1):15-20.
- [5] Judith A. Chevalier, Dina Mayzlin(2006). The Effect of Word of Mouth on Sales: Online Book Reviews. 43(3):345-354.
- [6] Nie W., Tian J.J.(2017). An Empirical Study on the impact of online reviews on consumers' purchase intention . *Technology and innovation management*, 38 (05): 502-506
- [7] Li H.Q.(2017). Research on the influence of online reviews on consumers' purchase intention. Information Engineering Research Institute, USA, Singapore Management and Sports Science Institute, Singapore. *Proceedings of 2017 6th International Conference on Applied Social Science (ICASS 2017) V98*. Information Engineering Research Institute, USA, Singapore Management and Sports Science Institute, Singapore: intelligent information technology application society, 2017:218-223
- [8] Huang H., Mao H.F.(2019). Research on the impact of negative online reviews on consumers' purchase intention . *Economic issues*, (11): 71-80 + 88
- [9] Hong F., Zheng H., Zhou Y.F., Ao J.Q.(2019). Research on the influence of online reviews on College Students' consumers' purchase intention . *Business economics research*, (08): 52-56
- [10] Thomas Y. Lee, Eric T. Bradlow(2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5).
- [11] Mark E. Perry(2003). *Strategic marketing management*. China financial and Economic Publishing House, translated by Li Yisong et al.
- [12] Yu Y.S.(2016). Research on the influence of negative online comments on clothing sales. *Foreign trade and economic cooperation*, (04): 124-126
- [13] Zhang M., Zhang G.Y., Ye Z.l.(2011), et al. Research on the influence of online information on hotel online reservation -- Based on the analysis of hotel online reservation data of Ctrip. *Journal of tourism*, 26 (7): 79-84
- [14] Nikolay Archak, Anindya Ghose, Panagiotis G. Ipeirotis(2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8).
- [15] Li L., Huang P.K., Wang X.H.(2009), et al. Image edge detection based on beamlet transform. *Systems engineering and electronic technology (English version)*, 20 (1): 1-5
- [16] Chen X.(2020). Research on the influence of online reviews on consumers' purchase intention. *Hefei University of technology*
- [17] Li J.(2012). Research on the influence of online commodity reviews on commodity sales. *Modern intelligence*, 32 (01): 164-167

Full Research Paper

Does Online Product Presentation Matter: An Empirical Analysis of the Effects of Listing Photos on Short-term Rental Purchases

Jiang Wu^{1,2*}, Xinru Zheng^{1,2}, Jingxuan Cai^{1,2}

¹ School of Information Management, Wuhan University, Wuhan, 430072, China

² Center for E-commerce Research and Development, Wuhan University, Wuhan, 430072, China

Abstract: Under the rapid development of the sharing economy, short-term rental has become prevalent in recent years. However, how to promote consumer purchases is an important practical issue for short-term rental platforms. This study aims to explore the promoting effects of listing photos on short-term rental purchases in a representative Chinese room-sharing platform — Xiaozhu.com. Based on information adoption theory and cue utilization theory, this paper applies the econometrics model that integrates image processing techniques to estimate the effects of listing photos. The results show that the photos taken by professional photographers from Xiaozhu.com and listing photos with high brightness and clarity can help enhance consumers' willingness to book, whereas the hue of listing photos does not significantly increase renter reservations. The findings provide useful guiding suggestions for short-term rental platforms and room providers, which can help them bring a better purchasing experience to consumers.

1. INTRODUCTION

With the rapid development of new technologies and the popularity of personalized consumption concepts, short-term rental has been chosen by more and more people. Short-term rental such as Airbnb enables “regular people” (i.e., as opposed to business entities) to rent out their spare rooms or unoccupied houses and apartments and serve tourists. Short-term rental is essentially an experiential service. Online information is the only information cue when consumers make the purchase decision. For short-term rental platforms, how to decrease information asymmetry and promote consumers to make purchase decisions have become important practical issues.

In short-term rental platforms, listing photos can directly reflect housing information, which can help consumers assess whether the listing meets its own needs and then make reasonable purchase decisions. Therefore, the effects of the presentation of the listing photo on consumers' purchase behavior should be attached great importance to. However, in the existing research on short-term rental platforms, researchers usually touch on user-generated content such as text, or numerical information provided by merchants as determinants to affect consumers' purchasing decisions. There is still a space for the study of images provided by merchants. As an important part of the merchant-generated content, the influence mechanism of listing photos on consumer behaviors is of great significance. Therefore, it is necessary to conduct in-depth research.

This study focuses on the effects of listing photos on short-term rental purchases in a representative Chinese room-sharing platform — Xiaozhu.com. Based on information adoption theory and cue utilization theory, we apply the econometrics model that integrates image processing techniques to evaluate the effects of listing photos on listing purchases, which we want to discuss from more in-depth and thorough perspectives in the study.

2. RESULTS

We conducted an empirical study to evaluate the effects of listing photos on listing purchases made on Xiaozhu.com. We found that the photos from the professional photography service on Xiaozhu.com have a

* Corresponding author. Email: jiangw@whu.edu.cn.

significant impact on the sales of listings. Compared with the photos taken by the hosts, the photos taken by professionals from Xizozhu.com are more credible, and consumers are more inclined to believe in the authenticity of the listing information, which in turn encourages consumers to make purchase decisions. The quality of the photos in our study is expressed as information entropy, hue, brightness and clarity of the image. The image information entropy does not have a significant effect on listing purchases, whereas it is significant at different price levels. The results show that listing photos containing only warm tone or cool tone do not better promote consumer purchase intentions, listing photos with cool color and warm color for different scenes may be more likely to trigger consumers' pleasure and prompt them to make purchasing decisions. The effects of the brightness and clarity of listing photos are consistent with our experience and common sense.

3. CONTRIBUTIONS

This study has the following theoretical contributions. First, this study introduces information adoption theory from online shopping to short-term rental. We conducted in-depth research on listing photos from the perspectives of image credibility and image quality, and verified the significant influence of listing photos on listing purchases, which further enrich the literature on listing purchases in the sharing economy. Second, the findings of this study also expand the research on the characteristics of listing photos in short-term rentals. There are few empirical studies on the characteristics of listing photos, and most of them only focus on the hue and brightness of the photos. We extract the text, information entropy and clarity of the photos and enrich the study. Additionally, this study extends the application of cue utilization theory in the context of the sharing economy. Based on cue utilization theory, a more comprehensive analysis of the image quality characteristics can be carried out from two aspects that affect consumer cognition and emotion.

This study also has certain practical contributions for online short-term rental platform operators and hosts. For hosts on the short-term rental platform, they should pay more attention to the importance of listing photos on the platform. Listing photos with cool color and warm color for different scenes will give consumers a better visual experience. Additionally, the host can also choose the professional photography service on Xiaozhu.com, or invite a professional photographer to complete the collection of listing photos, to enhance the trust of consumers. The research results of the control variables introduced in this study show that the host's application for real-name certification, application of as many labels as possible, and implementation of loose cancellation policy are all conducive to the growth of listing purchases. For online short-term rental platform managers, they should do a good job of management and supervise the quality of the photos provided by the host. In addition, they should fully promote the professional photography service, or invite experts to train and guide hosts.

REFERENCES

- [1] Ert E, Fleischer A. (2020). What do Airbnb hosts reveal by posting photographs online and how does it affect their perceived trustworthiness? *Psychology & Marketing*, 37(5), 630–640.
- [2] Zhang S, Lee D, Singh P V, Srinivasan K. (2016). How much is an image worth? An empirical analysis of property's image aesthetic quality on demand at Airbnb.
- [3] Nguyen L S, Ruiz-Correa S, Mast M S, Gatica-Perez D. (2017). Check out this place: Inferring ambiance from airbnb photos. *IEEE Transactions on Multimedia*, 20(6), 1499–1511.
- [4] Chi Maomao, Pan Meiyu, Wang Weijun. (2020). Influence of cue consistency on purchase decision of shared accommodation platform users. *Data analysis and knowledge discovery*.(in chinese)
- [5] Keller K L, & Staelin R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of consumer research*, 14(2), 200-213.

Short Research Paper

Research on the Influencing Factors of User's Adoption of Paid Information in Social Q&A Community

Bohong Li¹, Yanmei Zhang^{2*}, Zhibin Cui³

¹School of Information, Central University of Finance and Economics, Beijing

²School of Information, Central University of Finance and Economics, Beijing

³School of Information, Central University of Finance and Economics, Beijing

Abstract: The paid knowledge live of the social Q&A community provides users with a way to interact with content creators online. How users measure the quality of information and finally adopt the paid information has high research value. Research on users' adoption of knowledge payment is conducive to open new knowledge-paying user market. This article takes Zhihu Live as the research object, builds a theoretical framework through the information adoption model, and explores the influencing factors of users' payment information adoption behavior. We combine text analysis method to perform regression analysis on characteristic data to verify the theoretical model. The research results show that carrier richness, content live broadcast duration and content professionalism will significantly affect information quality, while content interestingness has a negative impact. Information quality and credibility of information sources will positively and significantly affect the users' final adoption decisions, while information acquisition costs will have a negative impact.

Keywords: paid knowledge, information adoption behavior, information quality, source credibility, content expressiveness

1. INTRODUCTION

Since “the first year of knowledge payment” in 2016, a new generation of knowledge-based payment platforms such as zhihu, Fenda and DXY.com have come out in succession owing to the promotion of new concepts and new technologies such as the Sharing Economy and Internet Plus. Moreover, after having a stable user base, question-and-answer websites such as Zhihu and Baidu Knows begin trying to transition from “free knowledge sharing” platforms to “paid knowledge sharing” platforms [1]. The era of “knowledge payment” in China begins.

However, the rapid development of knowledge payment leads to new problems. Paid knowledge is a kind of special information. Different from traditional information, users have to pay to acquire knowledge, and find it difficult to perceive its usefulness in a effective manner. Before adopting paid knowledge, prospective knowledge paying users tend to make a judgement on the true quality of paid knowledge, information related to Live speakers and the prices of Live. Only by analyzing the factors of users' knowledge sharing perceived usefulness can Live speakers improve the content of knowledge and attract more knowledge paying users in a targeted way [2].

In order to study the information adoption behaviour of knowledge paying users, this paper intends to find solutions to the following problems:

1. Which factors affect knowledge paying users making a judgement on the quality information of Zhihu Live?
2. Which factors affect knowledge paying users adopting a knowledge paying Live on Zhihu?

In order to explore how the expressiveness of the Live content, the characteristics of the speaker and other characteristics affect knowledge paying users' perceived usefulness, this paper takes the Information Adoption Model (IAM) and the real-time interactive knowledge paying product Zhihu Live as the research framework and

* Corresponding author. Email: zhangym@cufe.edu.cn(Yanmei Zhang) , 15618158255@163.com(Bohong Li)

the object of research respectively. This paper builds a conceptual model of the user's information adoption behaviour in knowledge paying environment from a theoretical perspective, so as to enable the traditional Technology Acceptance Model (TAM) to evolve in such environment. Moreover, this model helps the speaker further improve the quality of information, set a reasonable price and release the price signal from a practical perspective.

2. LITERATURE REVIEW

2.1 Research of knowledge payment

To date the research contents of "knowledge payment" have focused mainly on information governance^[3] and product sales^[4], but the research perspective has focused mainly on users' willingness to pay. Based on a variety of theories and in terms of system environment, knowledge acquirers and knowledge sharers, scholars explore the possible impacts exerted by the factors including the quality, value, usefulness and accessibility of knowledge service through the questionnaire. In addition, some scholars find that the attributes of both the content and the speaker affect users' paying behaviour, by exploring the data of real platforms and focusing mainly on the number of knowledge paying users^[5] and their degrees of satisfaction^[6].

2.2 Research of information adoption behaviour

According to IAM, information adoption behaviour includes the selection, evaluation, absorption and utilization of information, and both the content and the source of information affect users' perceived usefulness of information and finally affect their information adoption behaviour. Users in socialized Q&A communities tend to select and judge information and evaluate the true usefulness of information according to the principles of value matching, least effort and outcome satisfaction. Source credibility refers to the reliability degree of the source of information. In other words, does the information come from highly influential KOLS or common users? The quality of information consists of the quality of the content and service interaction. The quality of the information content includes the expressiveness of the content, while the quality of service interaction includes the quality of information service, the information exchange and interaction tools provided and information transfer ways.

IAM has been widely used in the study of online information adoption. Chen et al. explored how the recreational nature and interactivity of the answer affect information adoption on social Q&A network sites and indicated that the recreational nature of the answer exerted a highly positive impact on answer adoption^[7]. Wang Xiwei et al. built an information adoption behaviour model of online group-buying APP from the perspective of information ecology^[8]. Li Jinhua et al. found a positive correlation between the features and quality of the answer and the features of the answerer versus the voting result of answer usefulness^[9]. However, domestic and overseas research findings focus mainly on behaviour of e-commerce users or users in socialized Q&A communities, and seldom pay attention to paid information adoption behaviour.

3. RESEARCH HYPOTHESIS AND MODEL BUILDING

3.1 Quality of information

3.1.1 Expressiveness of the content

(1) Carrier richness

Users are able to get access to the information content through a variety of carrier such as text, voices, figures and videos^[10]. Liu et al. found that the content containing a variety of carrier exert a positive impact on the number of forwarding and found a positive correlation between the number of forwarding and the popularity of information^[11]. Zhihu Live shares its content through voices, but users are able to select files such as text and videos in the process of live voice broadcasting. The number of appendices contained in Zhihu Live and the duration of broadcasting are used to measure the richness of carrier. On the basis of the above analysis, we put

forward the following hypothesis:

H1a: There is a positive correlation between the richness of paid knowledge carrier and the quality of information.

(2) Content broadcast duration

Knowledge payment refers to paying for knowledge online, and offering users the quality content is why Zhihu Live exists. However, the content value of Zhihu Live provides a solid foundation for attracting users. As a real-time live broadcasting product, Zhihu Live features the duration of broadcasting. According to existing research, the Live with longer long duration may bring the users who want to study by fragmenting time much inconvenience and leave a negative impression with them^[12]. The duration of the Live is used to measure the duration of the content. On the basis of the above analysis, we put forward the following hypothesis:

H1b: There is a positive correlation between the duration of paid knowledge lives and the quality of information.

(3) Content interactivity

This paper defines the interactivity of the content as Zhihu Live's ability to respond to questions raised by users in the process of live broadcasting. In other words, Zhihu Live is able to respond to consumers' needs in a quick manner so as to suit their needs in time. Interactivity tends to affect users' perceived usefulness and further affect their information adoption behaviour. The number of interactive questions and answers in the Live is used to measure the interactivity of the content. On the basis of the above analysis, we put forward the following hypothesis:

H1c: There is a positive correlation between the interactivity of the paid knowledge content and the quality of information.

(4) Content interestingness

The interestingness of content is another key factor that affects the use of social media^[13]. The answer containing only knowledge is less liable to attract users, compared to the answer with interesting contents. Furthermore, information with recreational and social attributes are able to satisfy users' subjective needs and further stimulate the initiative of users^[14]. The humor evaluating value (stat/Ratehumor) set in "Wenxin" is used to analyze the interestingness of an answer. The higher the numerical value gets, the more interesting the answer gets. On the basis of the above analysis, we put forward the following hypothesis:

H1d: There is a positive correlation between the interestingness of the paid knowledge content and the quality of information.

(5) Content professionalism

Two ways have been used by Zhihu's official certification system to measure the professional degree of users. The first is to mark part of users as quality answerers who provide a large number of professional answers for the community. Users of this kind are specialists of a certain field and able to offer answers based on disciplinary knowledge or industry knowledge, so they tend to gain recognition. The second is to mark certain celebrities as "V" users to make visible their professionalism. Users of this kind have already gained a reputation in a certain field. In this paper, the Live shared by verified users is marked as 1, and the Live shared by the rest users is marked as 0. On the basis of the above analysis, we put forward the following hypothesis:

H1e: There is a positive correlation between the professional degree of the paid knowledge content and the quality of information.

3.1.2 Quality of follow-up service

Value-added service refers to that Zhihu Live provide both knowledge related to the theme and high level contents/experiences for users. The Knowledge payment platform is emerging, so few studies focus on the value-added service of Knowledge payment platforms. Wang Qianmin found that users perceiving the quality of value-

added service exerts a positive impact on their degree of satisfaction^[15]. In this paper, the product detail page of the Live with “Return of goods without reasons within 7 days” is marked as 1, and the product detail page of the Live without “Return of goods without reasons within 7 days” is marked as 0. On the basis of the above analysis, we put forward the following hypothesis:

H2: There is a positive correlation between the quality of paid knowledge follow-up service and the quality of information.

At the end of the Live, paying users are able to write a review and grade it, and the rating and the number of reviews mirror the the quality of information in the Live. According to IAM, the quality of information exerts an impact on users’ perceived usefulness. On the basis of the above analysis, we put forward the following hypothesis:

H3: There is a positive correlation between the quality of paid knowledge information and users’ information adoption behaviour.

3.2 Source credibility

The recognition degree of knowledge receivers determines the credibility of a answerer. The voting mechanism and the reputation system of Zhihu help answerers lay bare their credibility. Low source credibility leads to low information acceptability^[16]. If The higher recognition an answerer gain in a socialized Q&A community, the higher probability other users recognize the usefulness of knowledge shared by him/her. The computing method suggested by Song et al. (the number of likes / the number of knowledge sharing) is used to measure source credibility^[17]. On the basis of the above analysis, we put forward the following hypothesis:

H4: There is a positive correlation between paid knowledge source credibility and users’ information adoption behaviour.

3.3 Information acquisition costs

Information adoption in traditional Q&A communities is free, so the information acquisition cost is relatively low. Before the new mode knowledge payment emerges, the interrogation mode “free knowledge” has rooted deep in questioners’ mind and cognitive lock-ins form^[18]. When questioner have to pay to acquire knowledge, users’ information acquisition cost increases dramatically and they are less liable to realize the value of payment-based platforms. In this paper, the price of Zhihu Live is used to measure users’ information acquisition cost. On the basis of the above analysis, we put forward the following hypothesis:

H5: There is a negative correlation between paid knowledge information acquisition cost and users’ information adoption behaviour.

3.4 Theoretical model

Based on information acceptance theories, this paper tries to build a theoretical model of paid knowledge adoption behaviour of users in socialized Q&A communities.

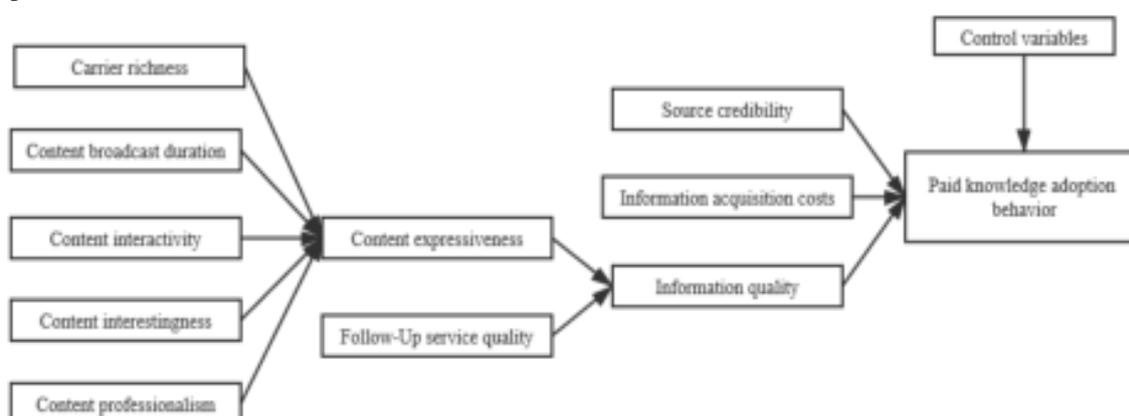


Figure 1. Theoretical model

4. RESEARCH DESIGN

4.1 Data sources

The research sample of this paper comes from the real data of Zhihu Live. The information about 733 Lives and their speakers from Jan. 2020 to Apr. 2020 at intervals of 10 days is collected by web crawlers. The data is preprocessed by removing 0 rating Lives (when the number of the participant users of a Live are relatively small, the rating will not be displayed) and the Lives that have not started yet. As a result, the rest 660 Lives have been used as research data.

4.2 Measurement and description of variable

The dependent variables set in this paper include information quality and information adoption. Paying users are able to rate the content of a Live after buying it on Zhihu Live. The number of commodities and services can reflect information quality to the full^[19]. Hence, the Live rating can be used to measure the information quality of paid knowledge. After users adopt the paid information, the number of users who are interested in the Live is used to measure the information adoption degree of paid knowledge. The control variables set include the gender of speakers and the interval of web crawlers. There are 502 male speakers that account for 76.0% of users, while there are 158 female speakers that account for 23.94% of users. A Live is still available after finishing, and users are unable to interact with the speaker once the Live is finished. Therefore, the time factor is controlled by recording the interval between the Live web crawler and the finishing time. The final description of variables is as shown in Table 1:

Table 1. Variable description and explanation

Variable name	Measured variable	Variable symbol	Min	Max	Avg	Sd	Skewness
Information quality	Live rating	Live_star	2.11	5	4.36	0.48	-1.40
Information adoption	The number of Live paying users	Live_adopt	70	55286	1873.45	3760.13	6.87
Carrier richness	The number of Live appendices	Live_attachment	0	328	20.30	26.60	5.16
Content broadcast duration	Live broadcast Duration	Live_time	0.52	154.69	50.72	23.57	1.24
Content interactivity	The number of Live interactive Q&A	Live_reply	0	299	26.28	27.70	2.78
Content interestingness	Stat/Ratehumor set in "Wenxin"	Live_entertainment	-3	2	0.223	0.486	1.85
Content professionalism	Live professional certification	Certification	0	1	0.83	0.38	-1.74
Follow-Up service quality	Return of goods without reasons within 7 days	Isrefund	0	1	0.19	0.39	1.26
Source credibility	The number of likes/knowledge sharing	Credibility	0	5.18	0.843	0.831	1.603
Information acquisition costs	Live price	Live_price	2.32	598	27.40	31.48	10.58
Control variables	Interval time	Interval_time	8	76	32.28	92.97	-0.02
	Sex	Gender	0	1	0.76	0.43	-1.22

According to Table 1, the standard deviation of the number of Live information adoption, followers and likes is relatively large. In order to meet the conditions of regression analysis, the independent variables mentioned above are given a logarithmic transformation.

4.3 Empirical models

In order to study the influence factors of information adoption, we build the following models on the basis of correlation analysis:

$$\text{Live_star} = \beta_0 + \beta_1 \text{Live_attachment} + \beta_2 \text{Live_time} + \beta_3 \text{Live_reply} + \beta_4 \text{Live_entertainment} + \beta_5 \text{Certification} + \beta_6 \text{Isrefund} \quad (1)$$

$$\ln \text{Live_adopt} = \beta_0 + \beta_1 \text{Live_star} + \beta_2 \ln \text{Credibility} + \beta_3 \text{Live_price} + \beta_4 \text{Interval_time} + \beta_5 \text{Gender} \quad (2)$$

Model 1 is the inquiry model of influence factors of information quality, the dependent variable is the rating of Zhihu Live and OLS regression is used to examine.

Model 2 is the model of information adoption, the dependent variable is the number of paying users of Zhihu Live which is the non-negative integer and does not satisfy the hypothesis that dependent variables are distributed continuously, and negative binomial regression is used to examine.

5. ANALYSIS OF EMPIRICAL RESULTS

The glm(), glm.nb() function of the MASS package in R are used to carry out regression analysis, and the final result is as shown in Table 2:

Table 2. Model regression results

Model 1			Model 2		
Variables	Coef.	VIF	Variables	Coef.	VIF
Live_attachment	0.002***	1.027	Live_star	0.085***	1.254
Live_time	0.005***	1.307	Credibility	0.166***	1.633
Live_reply	0.002	1.275	Live_price	-0.004***	1.036
Live_entertainment	-0.080**	1.020	Interval_time	0.002***	1.390
Certification	0.230***	1.054	Gender	-0.155	1.033
Isrefund	0.106	1.049			

Note: *, **, p < 0.01; ns, p > 0.05 not significant; β, regression coefficients; VIF, variance inflation factor

In order to avoid the influence of multicollinearity, the Variance Inflation Factor(VIF) is used to measure the colinearity between variables. According to Table 2, the VIF values in two models are less than 5, which shows that the colinearity between variable is relatively weak.

Model 1 explores the influence factors of Live information quality in terms of content expressiveness and follow-up service quality. The research findings show that content expressiveness has a highly positive correlation with the richness of carrier ($\beta = 0.002$), the duration of the content ($\beta = 0.005$) and the professional degree of the content ($\beta = 0.230$) respectively, a highly negative correlation with the interestingness of the content ($\beta = -0.080$), and there is no relation between content expressiveness and the interactivity of the content ($\beta = 0.002$).

H1a is possible. The richness of carrier has a highly positive correlation with Live information quality. Previous studies found that multimedia were able to stimulate users' senses so as to leave a positive impression with them^[20]. Identical with previous research findings, this paper finds that users tend to have a more positive attitude towards the content with rich carrier and the content with a variety of carrier is a key factor in improving information quality.

H1b is possible. The duration of the content has a highly positive correlation with Live information quality, which is identical with the expectation. The longer the Live broadcasting gets, the deeper and broader its content gets.

H1c is impossible. There is no relation between the interactivity of content and Live information quality. The reason may be that most of participant users buy the Live after their finishing and fail to participate in the real-time interactive session, and the interactive questions may not produce an effect on these users,

H1d is impossible. Information quality has a highly negative correlation with the interestingness of the content. The reason may be that Lives have a high threshold and high requirements for professionalism, so their content tend to more serious in style and highly recreational Lives tend to lose their core value.

H1e is possible. The professional degree of the content has a highly positive correlation with Live information quality and has the greatest influence on content expressiveness. All candidates for Live speakers are carefully vetted by providing the identity card, the certificate of employment and the schooling certificate. Therefore, verified users tend to post Live with higher quality, compared to common users, so the professional degree of the content is able to affect information quality.

H2 is impossible. There is no relation between the quality of follow-up service and Live information quality. Since the Live is a kind of non-entity experience product, its users pay much attention to the experience and feel in the process of acquiring knowledge and pay less attention to follow-up quality.

On the basis of Model 1, Model 2 transforms the dependent variable in model 1 “information quality” into an independent variable. Moreover, a model of information adoption is built in terms of Source Credibility and information acquisition costs, and Negative-binomial (NB) regression models are used to explore. According to the regression results, both the Live rating ($\beta = 0.085$) and the number of followers ($\beta = 0.166$) have a highly positive correlation with users’ information adoption behaviour, which shows that both information quality and source credibility have a highly positive correlation with users’ information adoption behaviour. Hence, both H3 and H4 is possible. Moreover, the Live price ($\beta = -0.004$) has a highly negative correlation with users’ information adoption behaviour, which shows that information acquisition costs have a highly negative correlation with users’ information adoption behaviour. Hence, H5 is impossible.

6. CONCLUSIONS AND IMPLICATIONS

This paper takes Zhihu Live and IAM as the object of research and the theoretical basis respectively and analyze the influence factors of users’ knowledge paying information adoption behaviour through OLS Regression, Negative-binomial (NB) regression and text analysis. The study finds that information quality, information acquisition costs and source credibility have an obvious influence on users’ information adoption behaviour; the richness of carrier, the duration of the content, the professional degree of the content have a highly positive influence on user’s judgment on information quality, the interactivity of the content and the quality of follow-up service has no obvious influence on user’s judgment on information quality, and the interestingness of the content has an adverse influence on user’s judgment on information quality; Live information quality and speakers’ source credibility have a highly positive influence on users’ Zhihu Live adoption behaviour, while Live acquisition costs have a negative influence on users’ Zhihu Live adoption behaviour.

For Speakers of Zhihu Live, we have three suggestions. First, the richness of carrier and interactivity of the content have an obvious influence on user’s judgment on information quality, so speakers need to take the distribution of sharing duration and interactive duration into consideration. In addition, the randomness of questions require the speakers to have a wider range of knowledge. In order to make users understand the sequence of ideas of a Live, the speaker needs to divide the content of the Live into several chapters. Moreover, the speaker can also improve the professionalism and credibility of the content through community verification and

bibliographies. Second, Live speakers should focus on content expressiveness, information quality instead of improving follow-up service quality. Third, Live speakers should improve the professionalism of the content by sharing more professional knowledge so as to attract more users and further promote influence in the process of Live broadcasting.

For suppliers of community services, we suggest that the content expressiveness of knowledge paying Lives can be improved to help users extract the value of the Lives so as to further improve the commercial value and profitability of a community. In order to highlight Live speakers' credibility and authoritativeness and display the features and advantages of the Live content to the full, suppliers should add relevant functions and services.

However, this paper has some limitations. First, this paper studies users' knowledge paying information adoption behaviour on the basis of objective data, but information adoption are also affected by psychological factors such as perception integrating degrees and satisfaction degrees under the background of knowledge payment. Therefore, these factors can be analyzed in depth by combining with questionnaires, depth interviews and quasi experiments in future studies. Second, the assessment of Live information quality made in this paper is based on the purchase page and the speaker page of Live rather than the data of the real process of the Live. Hence, the content of a specific Live can be analyzed by combining with grounded theory in future studies.

REFERENCES

- [1] Cai S, Luo Q, Fu X, et al.(2018).Paying for Knowledge: Why People Paying for Live Broadcasts in Online Knowledge Sharing Community?. pacificasia conference on information systems.
- [2] Ning Jingjing.(2014).Study on knowledge sharing situation of zhihu.Thesis.Beijing:Beijing University of Posts and Telecommunications.(in Chinese)
- [3] Zhang Anqi,Li Yuanxu.(2020). Information Overload Effect and Weakening Mechanism of Internet Knowledge Sharing Platform—A Case Study on Zhihu.Information Science,38(01):24-29+41.(in Chinese)
- [4] Cai Shun,Shi Hairong,Fu Xin,Chen Xi.(2019).Research on the Factors Influencing the Sales of Paid Knowledge Products: Taking Zhihu Live as an Example.Journal of Industrial Engineering/Engineering Management, 33(03):71-83.(in Chinese)
- [5] Zhao Yang,Yuan Xini,Li Luqi,Zhao Yu.(2018).The Impact Factors of Users' Paying Behavior for Knowledge on Social Q &A Platform Based on Social Capital Theory. Document,Informaiton & Knowledge, 2018(04):15-23.(in Chinese)
- [6] Zhang J, Zhang J, Zhang M, et al.(2019) From free to paid: Customer expertise and customer satisfaction on knowledge payment platforms. decision support systems.
- [7] Chen X,Deng S.(2014).Influencing factors of answer adoption in social Q&A communities from users' perspective: taking Zhihu as an example. Chinese Journal of Library and Information Science,7(3):81-95.
- [8] Wang Xiwei,Li Jiaying,Guo Yu,Yang Mengqin.(2015). Research on Influencing Factors of Information Adoption Behavior of Mobile Network Group Buying APP—Analysis Based on the Perspective of Information Ecology.Library and Information Service, 38(04):20-28(in Chinese)
- [9] Li Jinhua,Zhang Tingting.(2018).Research on Influencing Factors of User Perceived Usefulness of Knowledge Sharing in Social Q&A—A Case Study of Zhihu.Journal of Modern Information ,38(04):20-28(in Chinese)
- [10] McMillan S J.(2004).Effects of structural and perceptual factors on attitude toward the website[J].Journal of Advertising Research,43(4):400-421.
- [11] Liu Z, Liu L, Li H, et al.(2012) Determinants of information retweeting in microblogging[J]. Internet Research,22(4): 443-466.
- [12] Wang Qianmin.(2018).Research on the Factors of User Satisfaction in Online Knowledge Payment Products:taking Zhihu-Live as an example. Thesis.Nanjing: Nanjing University(in Chinese)
- [13] Brandtzaeg P B, Heim J.(2009) Why People Use Social Networking Sites. international conference on online communities and social computing, 143-152.

- [14] Ko H, Cho C, Roberts M S, et al.(2005).INTERNET USES AND GRATIFICATIONS: A Structural Equation Model of Interactive Advertising. *Journal of Advertising*, 34(2): 57-70.
- [15] Wang Qianmin.(2018).Research on the Factors of User Satisfaction in Online Knowledge Payment Products:taking Zhihu-Live as an example. Thesis.Nanjing: Nanjing University(in Chinese)
- [16] Grewal D, Gotlieb J, Marmorstein H, et al.(1994). The Moderating Effects of Message Framing and Source Credibility on the Price-perceived Risk Relationship. *Journal of Consumer Research*, 21(1): 145-153.
- [17] Zhao Yuxiang,Liu Zhouyin,Zhu Qinghua.(2020).From Free to Fee: Exploring Askers Switch Behavior on Online Q&A Platforms from the Perspective of Cognitive Lock-in. *Journal of the China Society for Scientific and Technical Information*, 39(05):534-546(in Chinese)
- [18] Farhadloo M, Patterson R A, Rolland E, et al.(2016). Modeling customer satisfaction from unstructured data using a Bayesian approach. *decision support systems*,1-11.
- [19] Fortin D R, Dholakia R R.(2005).Interactivity and vividness effects on social presence and involvement with a web-based advertisement. *Journal of Business Research*,58(3): 387-396.
- [20] Sabate F,Berbegalmirabent J,Canabate A,et al.(2014).Factors influencing popularity of branded content in Facebook fan pages. *European Management Journal*,32(6): 1001-1011.

Full Research Paper

Research on the Influence of Content Presentation Mode of Online Video Comments on Consumer Interaction Behavior

Hong Peng¹, Junfang Cao^{2,1}

College of Business Administration, Zhongnan University of Economics and Law, Wuhan, 400000, China

Abstract: Online video comments have become an important form of online comments on e-commerce platforms and social platforms, but there is little research on online video comments at present. In this study, 199 evaluation videos from microblogs were selected, and content analysis was used to divide the video content into three parts: daily content sharing, product testing and personal evaluation. Stata was used to analyze the influence of content presentation on user interaction (LIKE and comment). The research shows that the influence of testing product content and sharing daily content on consumers' interactive behavior is inverted "U" shape, and the increase of personal evaluation content will lead to the decrease of interactive behavior. Keywords: online video comments, product evaluation, interactive behavior

1. INTRODUCTION

In the Internet era, online shopping has gradually become a common way of shopping. Online comments are an important reference for most people before shopping, and also a hot topic for many scholars to study in recent years. Initially, online reviews were focused on various shopping platforms. However, with the rise of social e-commerce and web celebrity economy, online reviews have gradually expanded to social platforms. Many influential consumers publish personal comments on products on social platforms, which arouse more consumers' attention and interaction. Therefore, a new form of online reviews has emerged -- product review video. Min Qiu (2019)^[1] points out that product evaluation video has become a new business form of joint marketing between video publishers and brands.

According to the development history of mobile communication technology, Liting Ma (2020)^[2] points out that the forms of online comments include text, pictures and videos. At present, most scholars focus on text and pictures. For example, Mudambi and Schuff (2010)^[3], Hao Yuanyuan (2010)^[4] et al. analyzed the usefulness of online comments from the perspective of text features. Huang Simeng (2019)^[5], Wang Yue (2019)^[6] et al. analyzed the impact of online reviews on consumers' buying intention and behavior from the perspective of picture reviews. However, there are few studies on online video comments.

Pei Xu and Liang Chen (2015)^[7] and Zhang Jige (2020)^[8] compared online video comments with text and picture comments and confirmed that online video comments have a more significant impact on consumers' intention. However, few scholars have studied whether the content of online video reviews will have an impact on consumers' psychology and behavior. This paper takes product evaluation videos as the research object and divides the video content into three parts: daily sharing, testing products and personal evaluation. The different presentation ways of video content are used to study the impact on consumer psychology and then influence consumer behavior.

By dividing the specific content of the evaluation video, this paper studies the influence of the content presentation of online video reviews on consumer behavior, further enriches the research on the influence of online reviews on consumer behavior, provides new ideas and methods for analyzing the content of online video reviews, and provides certain reference for brands and video creators to attract consumers and expand their popularity and reputation by using evaluation videos.

¹ Email of Author: hpeng520@126.com

2. LITERATURE REVIEW

2.1 Online Video Comments

The concept of "online comment" was first put forward by Chatterjee, and then scholars at home and abroad defined it from different perspectives. Bickart (2001)^[9] pointed out that online reviews were mainly generated on product websites, commercial websites and personal homepages. Li Fenglin (2012)^[10] believed that the comments on goods or services that consumers share with other consumers on online media such as shopping platforms and social networking sites are all online comments, which are dominated by text and supplemented by pictures and videos.

This paper holds that online video comments refer to consumers' own comments on commodities or services published in the form of videos on online shopping platforms and social platforms. The video online comments studied in this paper are from online social platforms.

2.2 The Classification of Product Evaluation Video Content

Existing studies (Chen Yu, 2020^[11]) regard product performance testing as the functional content of evaluation videos, while the content related to the anchor is not included in the functional content category. Based on this, the content of testing product performance in the video is divided into functional content, which is recorded as testing product content, including basic information of the product such as brand, texture, color number, applicable group, and instant and interval use effect of the blogger himself. The rest is recorded as non-test product content.

Non-test product content is mainly bloggers to show users some of their daily life snippets, as well as expressed by the blogger on the product, brand personal views or feelings. These personal information related to bloggers can meet users' sensory stimulation and psychological needs (Wang Yaqian, 2020^[13]) and improve users' stickiness. Krasnova et al. (2010)^[12] defined the process of communication in which an individual transmits information about himself to a specific object as self-disclosure, and studies have pointed out that a certain degree of self-disclosure by bloggers in online videos can improve the communication object's trust in the individual and perception of the authenticity of the information (Xianya Wang, 2020^[13]). The life content and emotional expression of these bloggers can satisfy consumers' curiosity and freshness, and shorten the psychological distance between bloggers and consumers. Sullivan (2013)^[14] further subdivided self-disclosure into factual and affective. Factual self-disclosure refers to the disclosure of personal information and events, while affective self-disclosure refers to personal feelings, opinions and judgments. Therefore, this paper further split the non-product testing part, and divided the part recording the daily life of the video publisher into factual self-disclosure, which was recorded as sharing daily content. The comments and emotional expressions about products, individuals and brands made by the video publishers were classified as affective self-disclosure and recorded as personal evaluation content.

2.3 Consumer Interaction

Existing studies (Yang Yu, 2018^[15]) pointed out that in social platforms, people transmit information through chatting, LIKE, comments, forwarding and other interactive ways. Forwarding refers to users transferring the content published by others to their own content. Comment is expressing one's opinion about other people's content; LIKE is the expression of approval or approval of what someone else has posted. Zhao Changkuan (2014)^[16] analyzed the interactive behaviors of friends in blogs (sharing, commenting, leaving messages, etc.) and drew the conclusion that interactive behaviors are an important force for developing friendship and transmitting information. Yang Yu (2018)^[15] believes that forwarding, commenting and LIKE are common interactive behaviors in online social networks, and the interaction is an important reflection of the degree of trust among users. Trust means that in the social network environment, the viewer and the video publisher still believe that the other party will behave in good faith even when the viewer is unable to judge and

predict the behavior of the video publisher (Li Weimin, 2002). Lowenthal and C.Haven (1968) [17] argued that people develop intimate relationships by discovering similarities between existing and repeated interactions. Berscheid and Reis (1998)[18] defined intimacy as "a kind of closeness and emotion", which can continuously reflect one person's thoughts, emotions and behaviors towards another person. Intimacy is a measure of how close you feel. Existing studies (Tang Jiayan, 2009, Ran Jingjing, 2013 et al.) point out that intimacy is related to trust, while in social media, the establishment and maintenance of intimate relationship between users is based on well-designed credibility (Wang Yaqian, 2020[19]). When trust is generated in the evaluation blogger or video content, consumers will increase their recognition and acceptance of the video, promote the establishment and development of intimate relationship, and then trigger the interactive behavior of displaying recognition and trust. Zickmund (2007)[20] pointed out that through the Internet platform, people can tell about their emotional status, share personal information, enhance emotional intimacy and develop more intimate relationships. Therefore, in the social platform, the interaction behavior of users can reflect the intimate relationship between bloggers and users to a certain extent.

According to the researches on social networks in reality, it can be found that the importance of forwarding, commenting, and liking behaviors to the interaction situation is decreasing (Yang Yu, 2018 [15]), but the forwarding behavior is affected by time and originality, the embodiment of the consumers for product assessment video interaction behavior is not sufficient, therefore, this study selected commentary behavior and thumb up behavior as a measure of interaction situations.

2.3 Research Hypothesis

In this study, the video is the product test video, in which the content of the test product is the main body of the online video comments. The product selected in this paper is the experience-type product foundation liquid, the product information includes brand, appearance, capacity, color number, texture, actual use effect and so on. Compared with text and picture online reviews, consumers can learn more abundant and detailed information about products from online video reviews, thus reducing uncertainty and cognitive risks and improving the credibility of video content (Pei Xua, 2015 [7]).

Xianya Wang (2020) [13] used the theory of media synchronicity to point out that Vlog, which is the same as the way of evaluating video presentation, can improve viewers' trust in video content and video publishers through narrative recording and direct camera angle. In social platforms, the level of trust between users is directly proportional to their interactive behaviors (Yang Yu, 2018 [15]). At the same time, online video comments can simulate face-to-face communication, such as looking directly at the camera, tone, expression, etc., so as to enhance the intimacy between the video publisher and the viewer (Meyrowitz, 1986[22]), and thus increase the interactive behavior of users. Test product content as the main body of online video reviews, although is an indispensable part, but too high proportion may lead to video content tends to be functional, lack of interest and attraction.

Based on this, this paper proposes the following hypotheses:

H1: In online video comments, the influence of test product content on consumers' LIKE behavior is inverted "U" shaped.

H2: In online video comments, the influence of test product content on consumer review behavior is inverted "U" shaped.

Daily content sharing is a kind of factual self-disclosure that evaluates the content in the video where the blogger shows clips of his personal life to the user. User of this video contains a large number of private information content dependence is very strong (Ryoo and Koo, 2010), a blogger video scene strategy selection and expression are issued as showing the truthfulness of their private life, let users think this is the real life of the blogger, while satisfy the user's curiosity, can effectively reduce the uncertainty of user for bloggers and video

content.

Sharing daily content is the content of the blogger showing the user a fragment of personal life in the evaluation video, which is a factual self-disclosure. Users have a strong dependence on the content of this kind of video that contains a lot of private information (Ryoo and Koo, 2010^[23]). A blogger video scene strategy selection and expression are issued as showing the truthfulness of their private life, let users think this is the real life of the blogger. While satisfying users' curiosity, it can effectively reduce users' uncertainty about bloggers and video content.

These video content based on daily life scenes, the user will own life with bloggers display life connect with each other, cause the resonance of the users and the blogger (Qiu Min, 2019), effectively Narrows the perception of the distance between users and the blogger, further strengthen the identity of the user of the bloggers and their video content and trust (Wang Yaqian, 2020). The add of daily life fragments is one of the important characteristics of video online comments in social media, video of establishment of the main self disclosure is associated with higher authenticity and credibility (Daft and Lengel, 1986), helped build blogger close relationship with users (Xiao Ma, 2016), causing the user interaction with bloggers.

These video content based on daily life scenes enable users to connect their lives with the lives shown by bloggers, triggering resonance between users and bloggers (Qiu Min, 2019^[1]), effectively bringing the relationship between users and bloggers closer, further strengthening the identity of the user of the bloggers and their video content and trust (Wang Yaqian, 2020^[19]). The inclusion of daily life clips in evaluation videos is an important feature of online video comments in social media. The self-disclosure degree of bloggers in videos is related to higher authenticity and credibility (Daft and Lengel, 1986^[23]), which helps to establish the close relationship between the bloggers and users (Xiao Ma, 2016^[24]), and it also in turn triggers the interaction between them. However, since the main content of online video reviews is still to review products, too higher proportion of daily content sharing is not better, otherwise it will deviate from the topic and no longer belong to the category of online reviews.

Based on this, this paper proposes the following hypotheses:

H3: In online video comments, the influence of daily content sharing on consumers' LIKE behavior is inverted "U" shaped.

H4: In online video comments, the influence of daily content sharing on consumer review behavior is inverted "U" shaped.

The content of personal evaluation refers to the evaluation and emotion expressed by the blogger on the product and brand in the online video comment, which is a kind of emotional self-disclosure of the blogger. Personal evaluation belongs to the subjective feelings and experience of the blogger. Adding this part to the video can make the bloggers and users have an equal relationship, allowing the blogger to achieve in-depth interaction with the users through personal experience and personal subjective feelings (Qiu Min, 2019^[1]), which will have a subtle impact on target consumers, enhance consumers' awareness and trust of the brand. In social platforms, the main manifestation of the close relationship between users and bloggers is the interactive behavior of forward, comment and LIKE. However, it should be noted that if bloggers express too much feelings for products and brands, they will weaken the personal attributes of online video comments, causing users to resent and even lose users (Cheng Qia, 2020).

Based on this, this paper proposes the following hypotheses:

H5: In online video comments, the influence of personal comments on consumers' LIKE behavior is inverted "U" shaped.

H6: In online video comments, the influence of personal comments on consumer comment behavior is inverted "U" shaped.

3. RESEARCH DESIGN

3.1 Research Samples and Data Collection

The research object of this study is the online video about the "liquid foundation evaluation". This study chooses Weibo as the video source for the following reasons: First, Weibo is the leading social media in China, which supports users to upload, share, watch videos and conduct Video interactive behaviors such as LIKE, comment and forward. Second, the average number of daily active users of Weibo is 216 million, and most of them are young people. There are a large number of target audiences for evaluating bloggers and evaluating videos on liquid foundation.

Using "liquid foundation evaluation" as the key word, the researchers searched through the search function of Weibo platform. A total of 127 valid videos were found, and each selected valid video was liked as a mark to prevent repeated screening. Due to the small number of videos, the researchers found hundreds of beauty bloggers through the "Top Beauty Bloggers List", randomly entered the blogger's homepage, and searched again with the keyword "liquid foundation evaluation". In the end, 199 valid videos were found. Each video contains fields: blogger name (the Weibo nickname of the selected video's publisher), number of fans (number of fans owned by the blogger, in units of 10,000), video duration (total duration of the selected video, in seconds), release date (the date the video was uploaded to the Weibo platform), LIKE (as of the statistical date, the number of likes of the video) and the number of comments (as of the statistical date, the number of comments on the video).

The criteria for selecting videos were as follows: a video should be between five and 20 minutes long, have more than 20 LIKES, have more than 10,000 followers, and be published no earlier than January 1, 2019, and no later than November 20, 2020.

3.2 Video Content Classification Extraction

In this study, video content processing is mainly to complete the quantification of video content classification. According to the foregoing, this study divides the video content into three categories, namely, test product content, sharing daily content and personal evaluation content. Since the specific content of each category is reflected in the video, it can be quantified by the total duration of its appearance in the video. The concrete treatment method is the content analysis method.

The video content of this study is divided into three categories: test product content, sharing daily content and personal evaluation content. Combined with the description of coding content, the specific coding framework and operation definition of each category are as follows:

Test product content: including the basic information about the product displayed by the blogger (brand, appearance, texture, color number, price, etc.), the blogger's immediate and interval use effect, collectively referred to as product information. The specific operation is defined as follows: when the content about product information appears for the first time in the video, record the playing time of the video at this time, and the first cycle begins. Once the content unrelated to product information appears in the video, it means the end of the first cycle. The playing time of the video at this time is recorded again. Calculate the length of the first period by subtracting the beginning of the first period from the end of the first period. When the content about product information appears again in the video after the end of the previous cycle, the next cycle will start, and the specific operations will repeat the first cycle. Finally, the duration of each cycle of recording product information was accumulated to obtain the total duration of the content of the tested product in the video.

Sharing daily content: including the daily life content displayed by the blogger, such as eating, playing games, makeup, shopping and other life segments, recorded as the blogger's daily life. The specific operation is defined as: when the content related to the daily life of the blogger appears for the first time in the video, record the playing time of the video at this time, and the first cycle begins. Once the content that is not related to the daily life of the blogger appears in the video, it means the end of the first cycle, and the video playing time at this time is recorded

again. Calculate the length of the first period by subtracting the beginning of the first period from the end of the first period. When the content related to the daily life of the blogger appears again in the video after the end of the previous cycle, the next cycle starts, and the specific operation is repeated in the first cycle. Finally, the length of each recording blogger's daily life cycle is accumulated to obtain the total length of daily content shared in the video.

Personal evaluation content: including the feelings and evaluation of the brand or product expressed by the blogger, recorded as the blogger's personal evaluation. The specific analysis method is the same as above. Finally, the duration of each recording period of bloggers' personal comments is accumulated to obtain the total duration of personal comments in the video.

3.3 Variable Definitions

This study contains a total of 5 variables, of which the dependent variables are user LIKE behavior and user comment behavior. The data comes from the number of video LIKES and video comments. The independent variables are test product content, sharing daily content and personal evaluation content, and the data comes from the content analysis of the video. The specific variables and their definitions are shown in Table 1.

Table 1. Variables involved in this study

Variable category	Manipulate variable	Operating variable measurement standard
Dependent variable	User LIKE behavior	Number of video LIKES
	User comment behavior	Number of video comments
Independent variable	Sharing daily content	The duration of sharing daily content in the video (s)/total duration of the video (s)
	Test product content	The length of the product test content in the video (s)/total length of the video (s)
	Personal evaluation content	The length of the personal evaluation content in the video (s)/total length of the video (s)

4. DATA ANALYSIS

In this study, 199 data collected were analyzed by Statase 15 to study the relationship between different presentation modes of video online comments and user interaction behavior.

4.1 Descriptive Statistical Analysis

Using Stata descriptive statistics analysis, it is concluded that the dependent variable and independent variable characteristics of basic data in the distribution of video content, sharing daily content and personal evaluation content of the minimum value is 0, and the minimum value of test product content is 0.1983. This data is consistent with review videos, where product testing is an essential part of the review. Descriptive statistical results of specific variables are shown in Table 2.

Table 2. Descriptive statistical analysis of variables

Variables	Obs	Mean	Std. Dev.	Min	Max
User LIKE behavior	199	5307.09	6663.035	25	52000
User comment behavior	199	1994.04	2288.275	10	13000
Sharing daily content	199	0.3264704	0.1821563	0	0.7694
Test product content	199	0.5508533	0.1703525	0.1983	0.9391
Personal evaluation content	199	0.1224302	0.0776647	0	0.7652

4.2 Hypothetical Test

In this study, simple linear regression and multiple linear regression were used to analyze the impact of sharing daily content, test product content and personal evaluation content on users' LIKE behavior and review behavior. Since the dependent variables in the model are all values in ten thousand, while the independent variable

is a decimal, and there is a large numerical gap between the dependent variable and the independent variable, the logarithm of two dependent variables with large values (LIKE behavior and comment behavior) is taken in this study, so that all variables are in the same quantitative level. Statase 15 is used to draw the graph of the relationship between each independent variable and video LIKE behavior and comment behavior, as shown in Figure 1-6. The regression analysis results of this study are shown in Table 3 and Table 4.

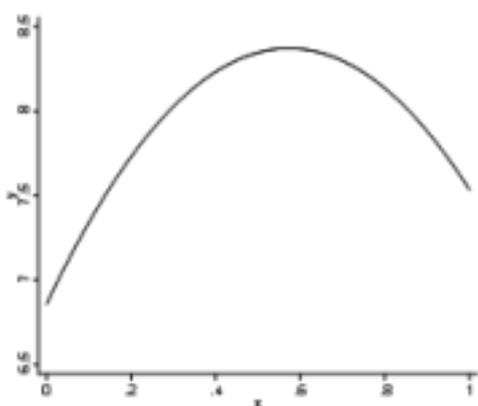


Figure 1. Sharing daily (x) and ln (like behavior) (y)



Figure 2. Test product (x) and ln (like behavior) (y)

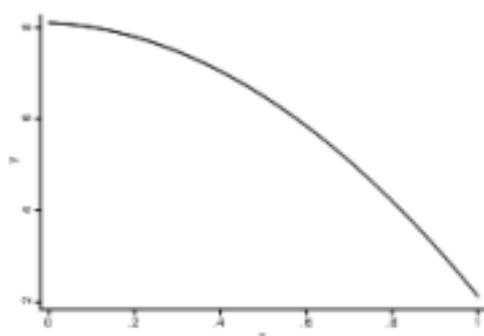


Figure 3. Personal evaluation (x) and ln (like behavior) (y)

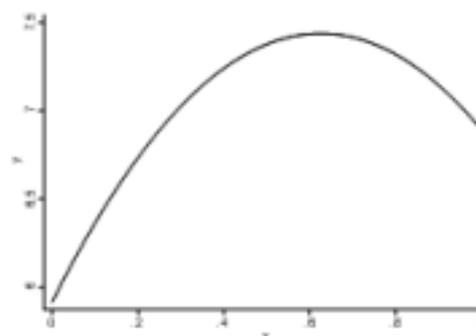


Figure 4. Sharing daily (x) and ln (comment behavior) (y)

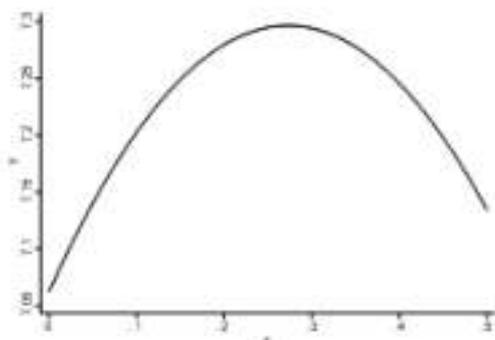


Figure 5. Test product (x) and ln (comment behavior) (y)

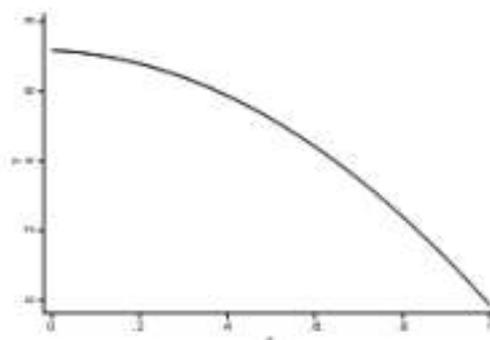


Figure 6. Personal evaluation (x) and ln (comment behavior) (y)

Table 3. Regression results of the dependent variable being the LIKE behavior

Dependent Variable: Ln (LIKE behavior)						
	<i>OLS(1)</i>	<i>OLS(2)</i>	<i>OLS(3)</i>	<i>OLS(4)</i>	<i>OLS(5)</i>	<i>OLS(6)</i>
Sharing Daily	2.2978 ***	5.283 ***				
Sharing Daily ²		-4.607 ***				
Test Product			-1.965 ***	1.299		
Test Product ²				-2.863 **		
Personal Evaluation					-3.160 ***	-0.442
Personal Evaluation ²						-5.549 **
R ²	0.1105	0.1294	0.0707	0.0758	0.0380	0.0480
_cons	7.189	6.858	9.022	8.176	8.326	8.110
n	199	199	199	199	199	199

Note: *** means P<0.05, ** means P<0.1

Table 4 The dependent variable is the regression result of comment behavior

Dependent Variable: Ln (comment behavior)						
	<i>OLS(1)</i>	<i>OLS(2)</i>	<i>OLS(3)</i>	<i>OLS(4)</i>	<i>OLS(5)</i>	<i>OLS(6)</i>
Sharing Daily	2.345 ***	4.865 ***				
Sharing Daily ²		-3.889 ***				
Test Product			-1.873 ***	1.719		
Test Product ²				-3.151 **		
Personal Evaluation					-3.847 ***	-0.517
Personal Evaluation ²						-6.795 ***
R ²	0.1128	0.1260	0.0629	0.0690	0.0552	0.0698
_cons	6.196	5.917	7.994	7.062	7.433	7.168
n	199	199	199	199	199	199

Note: *** means P<0.05, ** means P<0.1

4.2.1 Test product content and consumer interaction behavior

It can be seen from OLS (3) in Table 1 and Table 2 that the increase of the content of the test product will reduce the user interaction behavior, but this conclusion is not consistent with the actual situation. In this study, OLS (4) was obtained by further adjusting the model, at this time $\beta_1 > 0, \beta_2 < 0$, determined the best advantage of testing product content for LIKE behavior $= |\beta_1 / (2\beta_2)| = 22.68\%$; for the comment behavior, the best point of testing product content $= |\beta_1 / (2\beta_2)| = 27.27\%$. In addition, P values of quadratic term coefficients are all significant at the significance test level of 10%. Therefore, the influence of test product content on consumers' LIKE behavior and review behavior shows an inverted "U" shape, and the hypothesis H1 and H2 of this study are valid.

As mentioned above, test product content is the subject of online video reviews, with only 22.68% or 27.27% of the content appearing to be unrealistic. However, combined with the development of online shopping and social e-commerce, it is not difficult to find that consumers can easily obtain detailed basic product information on shopping platforms.

The reason why online video comments of experience-based products are popular among consumers is that online comments in the form of video can provide consumers with potential product information that cannot be obtained in other forms (text, picture), such as personal real effects and feelings of video display. This is the real core and focus of online video reviews. Therefore, the content of online video reviews should contain some non-traditional product information as much as possible, so as to help consumers to have a more comprehensive and objective understanding of products and brands while reducing uncertainties and risks.

4.2.2 Sharing daily content and consumer interaction behavior

According to OLS (1) in Table 1 and Table 2, sharing daily content in product evaluation videos can significantly improve consumers' LIKE behavior and review behavior. In order to further study the relationship between the two, the model was adjusted as OLS (2) in this study, at this time $\beta_1 > 0, \beta_2 < 0$.

For LIKE behavior and comment behavior, the turning points were 57.34% and 62.55%, respectively, and the P value of the first term coefficient was 0.001 and 0.002, and the P value of the second term coefficient was 0.041 and 0.087. Therefore, the influence of sharing daily content on consumers' interactive behavior is inverted "U" shape, and H3 and H4 are assumed to be true in this study.

Sharing daily content can significantly shorten the perception distance between bloggers and users, and the content of daily life can arouse emotional resonance between users and bloggers, making users feel that the content shown in the video is the blogger's real life state, increasing trust and recognition of video content, which induce the interactive behavior of praise and recognition. But this ratio cannot be increased indefinitely, otherwise it will deviate from the subject of online video reviews.

4.2.3 Personal evaluation content and consumer interaction behavior

Based on OLS (5) in Table 1 and Table 2, it can be seen that adding personal evaluation content in product evaluation videos will lead to a decrease in the amount of thumb up and the number of comments in the videos. In order to further verify the hypothesis of this paper, OLS (6) as shown in Table 1 and Table 2 is established. The quadratic coefficient of personal evaluation content is negative, but the coefficient of personal evaluation content is still negative, which does not meet the requirement of an inverted "U" shaped relationship between personal evaluation content and consumer interaction behavior (thumb up and comments). Therefore, the hypothesis of H5 and H6 in this study is not valid.

From the perspective of consumer psychology, if bloggers add too many positive comments on products or services in their product evaluation videos, consumers will think that bloggers have reached a deal with brand owners, which violates the principle of authenticity and objectivity of product evaluation videos and is likely to cause consumers' aversion. If the blogger adds more negative comments on the product or service in the product evaluation video, it may cause consumers to question the blogger himself, which is not conducive to the

establishment of an intimate relationship between the blogger and consumers. Therefore, too much personal subjective feeling should be avoided in the product evaluation video as much as possible, and the products and services should be displayed in a real and objective way, leaving room for consumers to make self-judgment.

5. DISCUSSION

In the past studies, there are very few literatures on online video reviews, and the research on online video reviews is even less due to the lack of research tools and other reasons. In this study, 199 online video reviews were disassembled through content analysis, which were divided into three parts: testing product content, sharing daily content and personal evaluation content. The regression model was used to analyze the relationship between the presentation of the above video content and consumer interaction behavior. The results show that: in online video comments, the effects of test product content and sharing daily content on consumer interaction behavior are inverted "U" shaped; The increase of personal evaluation content will reduce consumer interaction behavior.

For promoting users' LIKE behavior and comment behavior, the optimal proportion of test product content in the video is 22.68% and 27.27%, respectively, the content should be inclined to display non-traditional product information. Sharing daily content can improve the closeness between bloggers and users, and thus increase the interaction behavior, with the optimal proportion of 57.34% and 62.55%. However, the more content of personal comments, the less interactive behavior of consumers. In the era of increasingly intense internet marketing, consumers are also more sensitive to advertisements. Therefore, excessive personal comments in the evaluation videos will make the commercial intention of online comments of videos obvious and cause consumers' aversion.

5.1 Theoretical contribution and practical significance

At present, most of the studies on online comments focus on text and picture comments, while few scholars have studied online comments in the form of video, and all of them remain at the superficial stage of comparing three online comments, and lack of in-depth analysis on the content of online video comments. By identifying and defining the content of online video comments, this research studies the influence of the presentation of content of online video reviews on consumers' interactive behavior, promotes the research process of online video reviews, and enriches the research of online video comments.

For social platforms and e-commerce platforms, attention should be paid to platform intercommunication, which means combining social networking with e-commerce to allow consumers to obtain potential and non-traditional information about products, so as to reduce the uncertainty of information. This move can not only attract social platforms, but also bring more potential consumer groups to e-commerce platforms.

For brands, they can expand the marketing and promotion of their products on social platforms, and cooperate with bloggers to recommend products by means of online video comments and so on. However, we should pay attention not to blindly pursue product marketing while ignoring the authenticity of video content, otherwise it will have a negative impact on the reputation of the product and the image of the blogger. For review bloggers, the key to making their content popular with consumers is authenticity. Therefore, bloggers should try to show their true feelings in their videos. For example, a blogger can create a personal image by carefully designing the scene and content presentation, which can enhance the authenticity and credibility of the video content. At the same time, when bloggers cooperate with brands to promote products, they should not add too many advertisements and praises. Video content with obvious marketing intention will arouse consumers' disgust, which is detrimental to the long-term development of the relationship between bloggers and fans.

5.2 Limitation

The limitations of this study is to fix the video content on framework, which make the lack of certain innovative video content. And the online video comments this research mentioned refer to the evaluation of social platform video, which are quite different with the online video comments on shopping platform. So this study

can't provide too much reference for the traditional online video reviews.

REFERENCES

- [1] Qiu Min, Zhang Xiang. Vlog: A new form of content marketing[J]. Youth Journalist, 2019(24):81-82. (in Chinese)
- [2] Ma Liting, Liang Caiyun. The evolution of online comment form and value mining in the context of the development of mobile communication technology [J]. Science and Technology for Development, 2020, 16(05): 499-507.. (in Chinese)
- [3] Susan M. Mudambi, David Schuff. Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com[J]. MIS Quarterly, 2010, 34(1).
- [4] Hao Yuanyuan. An Empirical Study on the Impact of Online Reviews on Consumer Perception and Purchase Behavior [D]. Harbin Institute of Technology, 2010. (in Chinese)
- [5] Huang Simeng, Li Jiaqian, Zhao Xin. Research on the usefulness of image comments based on ELM model [A]. China Management Modernization Research Association, Fudan Management Award Foundation. Paper of the 14th (2019) China Management Conference [C]. China Management Modernization Research Association, Fudan Management Award Foundation: China Management Modernization Research Association, 2019:12. (in Chinese)
- [6] Wang Yue. Research on the influence of online photo comments on consumers' purchase intention [D]. Beijing University of Posts and Telecommunications, 2019. (in Chinese)
- [7] Pei Xu, Liang Chen, Radhika Santhanam. Will video be the next generation of e-commerce product reviews? Presentation format and the role of product type[J]. Decision Support Systems, 2015, 73.
- [8] Zhang Jige. Research on the Influence of Online Video Comments on Purchase Intention [D]. Hebei University, 2020. (in Chinese)
- [9] Bickart B, Schindler R M. Internet forums as influential sources of consumer information[J]. Journal of interactive marketing, 2001, 15(3): 31-40.
- [10] Li Fenglin, Liu Changping, Hu Yuan. Research on Online Reviews Searching Behavior of Online Consumers., 2012. (in Chinese)
- [11] Chen Yu. Cross-border symbiosis between self-media and toy marketing[J]. News Research Guide, 2020, 11(11): 232-233. (in Chinese)
- [12] Hanna Krasnova, Sarah Spiekermann, Ksenia Koroleva, Thomas Hildebrand. Online Social Networks: Why We Disclose[J]. Journal of Information Technology, 2010, 25(2).
- [13] Xianya Wang, Byenghee Chang. The Impact of the Audience's Continuance Intention Towards the Vlog: Focusing on Intimacy, Media Synchronicity and Authenticity[J]. International JOURNAL OF CONTENTS, 2020, 16(2).
- [14] H. S. Sullivan, The interpersonal theory of psychiatry, New York: Routledge, 2013.
- [15] Yang Yu, Hu Yazhou, Guo Yong, Deng Bo. Modeling and analysis of user trust transfer based on online social networks[J]. Computer Engineering, 2018, 44(11): 265-270. (in Chinese)
- [16] Zhao Changkuan, Li Feng, Xu Bin, et al. Research on similarity of interaction behaviors of blog friends[J]. Computer Engineering and Applications, 2014, 6. (in Chinese)
- [17] M. F. Lowenthal and C. Haven, "Interaction and Adaptation: Intimacy as a Critical Variable," American Sociological Review, vol. 33, no. 1, pp. 20-30, 1968
- [18] Berscheid E, Reis H T. Attraction and close relationships[J]. 1998
- [19] Wang Yaqian. "New Mimic Environment": A Study on Short Video Blogger's Personalized Expression and Its Impact on Users [J]. Chinese Youth Research, 2020(01): 68-75. (in Chinese)
- [20] S. Zickmund, "Deliberation, Phronesis, and Authenticity: Heidegger's Early Conception of Rhetoric," Philosophy & Rhetoric, vol. 40, no. 4, pp. 406-415, 2007.
- [21] J. Meyrowitz, "Television and interpersonal behavior: Codes of perception and response," Inter/media: Interpersonal communication in a media world, vol. 253, p. 272, 1986.
- [22] S. Y. Ryoo and C. Koo, "The Moderating Effect of Media Synchronicity in the Communication Media Use and Knowledge

Creation," *Asia Pacific Journal of Information Systems*, vol. 20, no. 2, pp. 103-124, Jun. 2010.

- [23]R. L. Daft and R. H. Lengel, "Organizational Information Requirements, Media Richness and Structural Design," *Management Science*, vol. 32, no. 5, pp. 554-571, May. 1986
- [24]Ma X, Hancock J, Naaman M. Anonymity, intimacy and self-disclosure in social media[C]//Proceedings of the 2016 CHI conference on human factors in computing systems. 2016: 3857-3869.

A New Analytic Framework for Identifying Tour Routes Service Innovation Using Knowledge Graph

Shuang Zhang¹, Minghao Li², Zhen Zhu^{1*}

¹School of Economics and Management, China University of Geosciences, Wuhan, 430074, China

²School of Geography and Information Engineering, China University of Geosciences, Wuhan, 430074, China

Abstract: In recent years, the innovation of service products in the e-commerce Red Sea area is still facing challenges due to their own content heterogeneity and indivisibility. Identifying innovation based on the characteristics of service components can help companies improve product competitiveness in increasingly intense market competition. In this paper, we present a new analytic framework to identify service innovation of tour routes that tackles the challenges of product similarity measurement. A knowledge graph-based method of service products is developed to obtain the dynamic weight of each dimension, and aiming to calculate the importance of graph network nodes through entity definition and attribute extraction after dimension division. Furthermore, we construct a small-scale knowledge graph of tour routes in the Ctrip. The results show the effectiveness of the proposed approach in terms of accurately identifying the level of service innovation, and provide a benchmark for exploring the difference in product unstructured attributes.

Keywords: Service innovation, E-commerce platform, Tour routes, Analytic framework, Knowledge graph

1. INTRODUCTION

In the context of increasingly fierce market competition, online sellers are constantly innovating and iterating their products to improve core competitiveness. Service innovation refers to the innovation taking place in the various contexts of services, including the introduction of new services or incremental improvements of existing services^[1]. Compared with tangible products, service products have four distinct characteristics: intangibility, inseparability, non-storability, and variability in quality^[2]. Meanwhile, the key components that constitute attribute characteristics of service products can't be identified like physical products, such as cell phones, refrigerators, laptops, and so on. One of the major challenges companies currently are facing is how to use a quantitative method to make reasonable and precise decisions when a large amount of text content is presented.

The calculation method of text similarity is a measure of product innovation in the field of text mining and sentiment analysis, but it is difficult to give a refined similarity judgment according to products contents available in e-commerce platform. This raises the need for a more flexible and hierarchical framework that can be used to analyze service products. The following research work collects the product data of outbound travel routes in the Ctrip platform to construct a small-scale knowledge graph and focuses on the question: *How to identify the service innovation of tour routes through the analytic framework using a knowledge graph method?*

2. METHODOLOGY

A knowledge graph is a kind of semantic network and has a good representation of structured knowledge which provides a good idea for our research. To express the multi-dimensional information of products, the innovative identification of service products needs to be established based on product structure segmentation^[3]. Accordingly, we transform the service characteristics of tour routes into relatively structured and quantifiable attributes. As shown in figure1, we develop a new analytic framework based on the knowledge graph to decompose the service features of the product into components and propose a new method for calculating the

* Corresponding author. Email: zhuzhen2008@gmail.com

importance of graph nodes to weigh the dimensions extracted by the product.

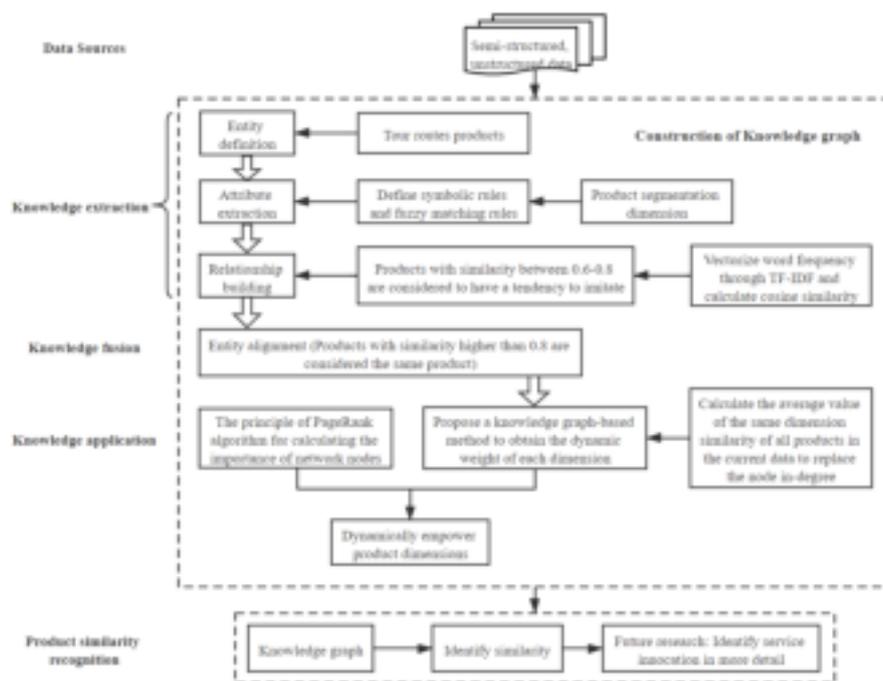


Figure 1. Analytic Framework

3. RESULTS AND IMPLICATIONS

In the actual calculation process, compared with products with obvious imitation behaviors identified by TF-IDF method, the similarity obtained by the dynamic weight similarity calculation method is more accurate. By comparing the similarity of all tour routes data in December 2017, we select the data with the similarity between 0.2-1.0, 0.4-1.0, 0.6-1.0, and 0.8-1.0 to calculate the average value. Within the same similarity range, the knowledge graph method can identify more dimensional features than the TD-IDF method, and get richer service innovation characteristics.

The identification of service product innovation in our study is closely linked to the unbundling of service products into separate components. Modularization has in turn served as a catalyst for innovation in the Ctrip platform, by providing new insights into how the individual components can be rearranged into new products. In addition, this study use the knowledge graph to obtain more accurate similarity recognition. The analytic framework of weighting the dimensions of unstructured features break through the previous method of quantifying product innovation. What's more, identifying service innovation of tour routes based on the analytic framework of the knowledge graph has the following two advantages:

- (1) Dimensional division narrows the scope of innovation identification.
- (2) Establish the connection between the service attributes of the tour routes.

REFERENCES

- [1] Durst S, Mention A L, Poutanen P. (2012). Service innovation and its impact: What do we know about?. *Investigaciones Europeas de Dirección y Economía de la Empresa*, 21(02):65-72.
- [2] Martínez T. (2013). Application of evolutionary computation techniques for the identification of innovators in open innovation communities. *Expert Systems with Applications*.
- [3] Abrahams A.S., Fan W., Wang G.A., Zhang Z., Jiao J. (2015), An Integrated Text Analytic Framework for Product Defect Discovery. *Production and Operations Management*, 24: 975-990.

Short Research Paper**Online Customer Experience at China's Double 11 Online Shopping****Festival: An Empirical Model of Antecedents and Consequences***Wen-Lung Shiau¹, Mengru Zhou^{1*}, Chang Liu¹*¹Department of Business Administration, Zhejiang University of Technology, China

Abstract: The Double 11 shopping festival has become the largest and most prevalent national shopping festival in China. However, research empirically exploring consumer experience based on the interaction between merchants and online customers in the context of Double 11 is limited. Therefore, this study attempts to empirically test a comprehensive model of the relationship between antecedents and consequents of online customer experience (OCE). Data from 451 valid samples were obtained using an online survey. The research model is assessed using partial least squares analysis. Results suggest that perceived information credibility and familiarity as antecedents of OCE positively influence two dimensions of OCE, that is, cognitive and affective experiences. As the consequence of the OCE model, customers' behavioral intention is positively affected by cognitive and affective experiences. The study makes contributions toward new knowledge and understanding of how e-retailers can provide effective online experiences for customers.

Keywords: Online customer experience, Double 11 online shopping festival, Behavioral Intention

1. INTRODUCTION

Founded on November 11, 2009, Double 11 is one of the largest online shopping carnivals established by the Tmall B2C shopping platform under the Alibaba Group. In 2020, the number of Double 11 transactions reached 498.2 billion RMB, and the total logistics order reached 2.321 billion. Evidently, the Double 11 online shopping carnival becomes increasingly influential on consumers. Therefore, we suppose that studying the determinants of consumers' purchase intentions in the context of Double 11 is of great significance.

Previous studies have demonstrated that the relationship between a retailer and its customers is affected by the results of interactive experience. Then, online shopping experience proved that effective sales strategies can create an online customer experience (OCE) and further influence customers' shopping behavior^{1,2}. Consumer experience is considered the accumulated feelings during interaction². In the Double 11 shopping carnival, diversified interactions exist between consumers and businesses. Therefore, this research explores the OCE in the context of the Double 11 shopping carnival to expand and deepen the understanding of OCE, particularly the antecedents and consequences of OCE. The main objectives of this research are threefold. The first objective is to identify the components of OCE. Second, the study aims to develop a comprehensive theoretical model of OCE that incorporates direct and indirect antecedent variables, OCE component variables and outcome variables are based on pre-existing theory of customer purchase intention. Finally, this study aims to test the OCE model empirically to find support for the proposed causal relationships.

This study starts by reviewing current definitions and dimensions of OCE. Then, the study proposes an explanatory model of OCE and provides support to the literature for the antecedent and consequent variables and the relationships between them. Moreover, the study explains the research methods adopted and then analyzes the research results. Finally, the study discusses the findings, implications, limitations of the study, and indications for future research.

*Corresponding author. Email: macshiau@zjut.edu.cn (Wen-Lung Shiau), zhoulengru1217@163.com (Mengru Zhou)

2. CONCEPTUAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

2.1 OCE

Customer experience is driven by the interactions between consumers and retailers, generates value, and shapes satisfaction and purchase intention^{3,4}. Molinillo et al. perceived that cognitive experience is the conceptualization of cognitive experience based on the concept of flow. Cognitive experience is defined as a cognitive state experienced during navigation, whereas affective experience is a set of mental processes that includes emotions, moods, and attitudes⁴. Many previous studies described OCE from two dimensions, namely, cognitive and affective, cognitive and affective experiences could affect customers' purchase intention. Double 11 is the largest shopping carnival in China, with the number of consumer transactions and participating rate reaching the peak. Therefore, exploring the relationship between cognitive and affective experiences becomes pivotal. Double 11 is favored by consumers due to the colorful preferential interactive activities and largescale discount promotions, which allow consumers to enjoy the joy and excitement of shopping spree while getting the most favorable experience. In other words, consumers have received cognitive and emotional satisfaction from participating in Double 11. Therefore, we elaborate on consumer experience from two dimensions of cognition and affection and analyze the impact on consumers.

2.2 Antecedents of OCE

2.2.1 Perceived information credibility

Perceived information credibility refers to the degree to which people perceive that a recommendation/review information is believable, true, or factual⁵. Rose et al. took information processing as antecedents of consumer experience and explained the casual relationships between them¹. Then, Fan discussed the impacts of perceived information credibility on cognition and affection⁵. Double 11 is the shopping carnival with the most discount, full reduction, and other preferential information, OCE is the cumulative effect of repeated exposure to the Double 11 platform², this study chooses the perceived information credibility as the antecedent of OCE for discussion.. When customers think that the information on the Double 11 platform is of high authenticity, then they will interact with retailers actively, thereby generating a positive customer experience.

On the cognitive dimension, Bleier believed that informativeness is the main cognitive state of OCE. Informativeness based on factual information objectively affects customer's cognitive experience and then attitude⁶. Kang considered that the perceived service quality is objective evaluation and cognitive attitude of the website, illustrating a reasonable evaluation of the customer experience in the process of interaction between consumers and retailers⁷. In the context of Double 11, when customers believe that information on the Double 11 platform is important, necessary, and meaningful to themselves, they will gain a good sense of experience. On the affective dimension, according to the literature of the technology acceptance model, customers value entertainment when using information systems. Several studies confirmed the importance of information to arouse customers to feel happy, contented, relaxed, and others. Kang indicated that the information load created by the cognitive and emotional stimulation in the web service environment evokes a powerful user experience⁸. In the process of interaction between customers and retailers, if customers perceive that the information provided by the Double 11 platform is true, then they will gain a sense of satisfaction and pleasure. Then, this real sense may facilitate the formation of a more positive affective experience. Based on these arguments, the following hypothesis is proposed:

H1: The perceived information credibility will positively contribute to the cognitive experience.

H2: The perceived information credibility will positively contribute to the affective experience.

2.2.2 Familiarity

Familiarity refers to the consumer's degree of acquaintance with the selling entity, which includes

knowledge of the vendor and understanding its relevant procedures, such as searching for products and information and ordering through the Website's purchasing interface^{9,10}. Familiarity has brought good experiences with vendors in the past, which will bring more positive emotions and behaviors to consumers in the future⁹. Familiarity is considered a subjective experience generated by the repeated interaction between consumers and businesses¹¹. Over time, interactions will increase familiarity and can lead to a deeper familiarity, which will facilitate creating much more customer experience in Double 11.

On the cognitive level, several studies pointed out that familiarity has a positive effect on cognitive experience^{5,7}. On the affective level, in Fan's view, familiarity is a necessary foundation for establishing and maintaining an emotion, positively influenced by interaction. In the process of interacting, deep feelings are generated, and customer experience is increased⁵. Kang stressed that interactivity has an emotional or hedonic effect, which contains diverse emotions, such as happiness, content, and excitement⁸. The Double 11 shopping carnival has been continuously held for more than 10 years. With the increase in interaction between consumers and the Double 11 platform, consumers have become more familiar with Double 11. Simultaneously, the consumer experience has been formed in this interactive process. Thus, the following two hypotheses are proposed:

H3: A consumer's familiarity with a selling party positively affects the consumer's cognitive experience during the Double 11 shopping festival.

H4: A consumer's familiarity with a selling party positively affects the consumer's affective experience during the Double 11 shopping festival.

2.3 Consequence of OCE

Cognitive experience refers to the concept of flow that captures the effect of cognitive aspects, such as cognitive curiosity and concentration, utilitarian features, informativeness, effectiveness, perceived ease of use, and perceived usefulness⁴. Lemon attempted to combine customer experience and customer behavior in terms of customer satisfaction and service quality and found a positive impact of customer experience on behavioral intention¹². Affective experience refers to the subjective feelings experienced, including affects, emotions, and attitudes⁴. Zhao suggested that when customers' experience matches their expectations, they will take a positive attitude toward the vendor's activities and encourage them to participate¹³. In the Double 11 shopping carnival, when consumers participate in the carnival at a preferential price, they will obtain a cognitive experience, which stimulates their intention to buy. Meanwhile, in the period of Double 11, factors such as the interactive games provided by the platform and the carnival atmosphere attract and encourage consumers to obtain a sense of satisfaction and pleasure. Then, such factors enhance consumers' purchasing intention.

Although the relationship between the affective and cognitive dimensions of experience is not consistent across the literature, in the present study, following Bagozzi¹⁴, we consider that a person's emotional experience may influence cognition experience. Kang showed that affective experience influences the consumer's cognitive experience^{4,8}. In addition, Laumer et al. demonstrated that affective experience has positive impacts on the cognitive dimension of the experience¹⁵. Based on these arguments, the following hypotheses are therefore proposed:

H5: Cognitive experience can significantly affect purchase intention. As the degree of cognitive experience increases, the purchase intention of consumers becomes stronger.

H6: Affective experience can significantly affect purchase intention. As the degree of affective experience increases, the purchase intention of consumers becomes stronger.

H7: The affective experience of customers has a positive effect on their cognitive experience.

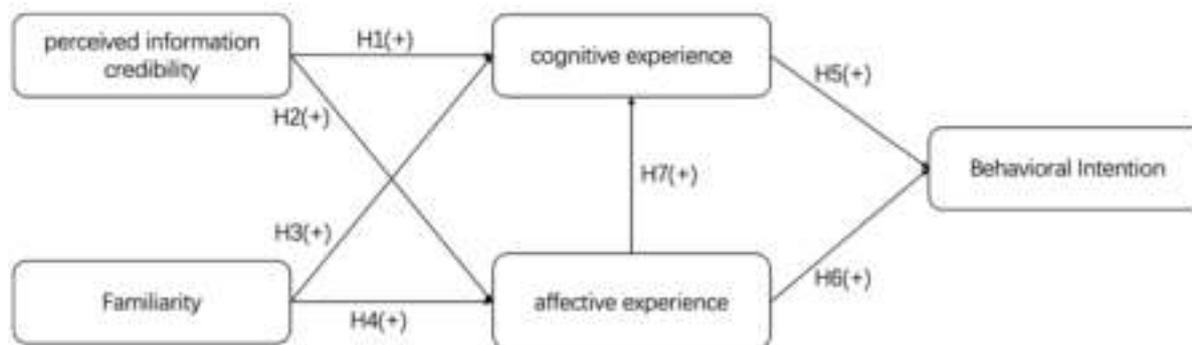


Figure 1. Proposed model

3. METHODOLOGY

3.1 Measuring Instrument

To ensure the reliability and validity of the research model, all measurement items are adopted from extant literature and adapted according to the Double 11 shopping festival. To achieve the research objectives and evaluate the proposed model, an empirical study was conducted through an online survey. Table 2 shows the questionnaire items used in the data collection for the evaluation of the research model. Prior to the data collection, a pilot test was conducted to ensure the validity of the questionnaire. The constructs were measured on a seven-point Likert-type scale.

3.2 Sample

The data were collected from November 4, 2020, to December 18, 2020, through an online questionnaire survey. A total of 583 samples were collected, and of them, 451 were valid questionnaires. Of these 451 participants, 152 were men (33.7%), and 299 were women (66.3%). The majority of respondents were aged between 21 and 35 (78.7%), which is consistent with the age distribution of Double 11 users. Table 1 shows the respondent demographics.

Table 1. Sample characteristics

Demographic variable	Classification	No.	Percentage
Gender	Male	152	33.70%
	Female	299	66.30%
Age	20 and below	29	6.43%
	21-25	237	52.55%
	26-30	62	13.75%
	31-35	56	12.42%
	36-40	33	7.32%
	41-45	21	4.66%
	46-50	6	1.33%
	51 and above	7	1.55%
Double 11 shopping expense	Below 200	50	11.09%
	200-499	94	20.84%
	500-999	148	32.82%
	1000-2999	124	27.49%
	Above 3000	35	7.76%

4. RESULTS

4.1 Measurement model analysis

This study employs the structural equation modeling tool Smart partial least squares (PLS) for confirmatory factor analysis to estimate the measurement model using PLS analysis to test construct reliability and validity. The individual reliability of each item was evaluated by factor loadings. Values above 0.7 indicate that the shared variance between the item and its construct is greater than the error variance¹⁶. Table 2 shows that all item loadings exceeded the recommended minimum value. The internal consistency of each construct was evaluated by the factor's composite reliability. The constructs presented values above the recommended minimum of 0.7. Convergent validity was also guaranteed as all the latent variables had an average variance extracted (AVE) higher than the recommended minimum value of 0.5¹⁷.

To assess discriminant validity, the Fornell-Larcker criterion¹⁷ and the heterotrait-monotrait (HTMT) ratio¹⁸ were used. First, the square root of the AVE of each construct must be greater than the interconstruct correlations of the model. Then, the HTMT ratio of correlations between two constructs should be below 0.9. Both criteria were met, so the discriminant validity of the measurement model is confirmed (Table 3).

Table 2. Descriptive statistics and measurement scales

	No.	Measurement item	Factor loading	Mean	S.D
PIC (AVE:0.698;CR:0.920)	PIC1	I think the information is credible.	0.782	5.266	1.454
	PIC2	I think the information is believable.	0.886	5.259	1.426
	PIC3	I think the information is trustworthy.	0.829	5.002	1.232
	PIC4	I think the message is truthful.	0.811	4.951	1.256
	PIC5	I think the message is reliable.	0.866	4.993	1.254
FAM (AVE:0.705;CR:0.905)	FAM1	I am familiar with "Double 11 shopping festival"	0.785	5.242	1.354
	FAM2	I am familiar with searching for products on the website during "Double 11 shopping festival"	0.855	5.386	1.336
	FAM3	I am familiar with buying products on the website during "Double 11 shopping festival"	0.863	5.388	1.316
	FAM4	I am familiar with the processes of purchasing products from the website during "Double 11 shopping festival"	0.854	5.501	1.214
CE (AVE:0.765;CR:0.929)	CE1	I have ever felt strongly immersed in an application.	0.866	4.741	1.571
	CE2	I felt strongly immersed when using an application.	0.889	4.849	1.520
	CE3	Most times I use an application I feel strongly immersed.	0.868	4.654	1.491
	CE4	Every time I use an application I feel strongly immersed.	0.877	4.670	1.549
AE (AVE:0.846;CR:0.965)		Double 11 shopping festival is:			
	AE1	Very Happy	0.921	4.951	1.586
	AE2	Very Content	0.902	4.825	1.522
	AE3	Very Pleased	0.926	4.911	1.529
	AE4	Very Excited	0.924	4.800	1.577
	AE5	Very Stimulated	0.925	4.825	1.632

BI (AVE:0.609;CR:0.861)	BI1	I am likely to purchase the products(s) on “Double 11 shopping festival”.	0.814	5.761	1.204
	BI2	I intended to purchase the products(s) on “Double 11 shopping festival” this year.	0.740	5.492	1.369
	BI3	I am likely to recommend “Double 11 shopping festival” to my friends.	0.844	5.188	1.372
	BI4	I am likely to make another purchase from this site if I need the products that I will buy.	0.715	5.373	1.286

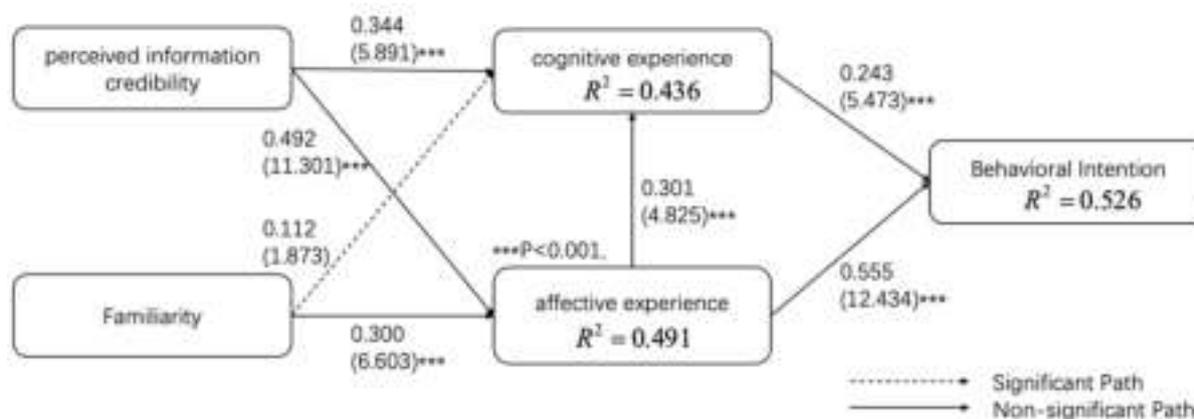
Table 3. Fornell-Larcker Criterion and HTMT Ratio

	AE	PIC	FAM	CE	BI
AE	0.920	0.707	0.623	0.635	0.804
PIC	0.653	0.835	0.610	0.667	0.894
FAM	0.564	0.537	0.840	0.529	0.744
CE	0.589	0.601	0.466	0.875	0.670
BI	0.698	0.750	0.612	0.570	0.780

4.2 Structural model

The bootstrap resampling method (5,000 resamples) was applied to determine the significance of the structural model paths. The path coefficient and significance of each hypothesis were examined. The explained variance (R^2) of each dependent construct was calculated.

The model explains 52.6% of variance for behavioral intention, 43.6% of variance for cognitive experience, and 49.1% of variance for affective experience. The results support H1 ($\beta = 0.344$, $p < 0.001$) and H2 ($\beta = 0.492$, $p < 0.001$), thereby confirming the positive association between perceived information credibility and cognitive and affective experiences. Then, H4 which indicates that familiarity positively associated with affective experience is also confirmed ($\beta = 0.300$, $p < 0.001$). However, H3 is not supported, which suggests that familiarity not positively associated with cognitive experience ($\beta = 0.112$, $p > 0.05$). The findings also provide support for H5 ($\beta = 0.243$, $p < 0.001$) and H6 ($\beta = 0.555$, $p < 0.001$), cognitive and affective experience are positively associated with behavioral intention. Finally, the findings confirm H7 ($\beta = 0.301$, $p < 0.001$). Figure 2 shows the results.



Note: *** $P < 0.001$, significant at the 0.001 level

Figure 2. Results of the structural model

5. DISCUSSION AND CONCLUSION

By reviewing several studies, we find the antecedents and consequences of OCE in the context of Double 11 and then build a comprehensive OCE model. In addition, we acquire relevant data from consumers participating in Double 11 to conduct empirical analysis for testing the mechanism of the OCE model in the context of Double 11. The study discusses the findings, implications, limitations and indications for future research.

In antecedents of OCE, perceived information credibility having significant positive effects on cognitive and affective experiences are consistent with those of Lee¹⁹. It is possible that online shopping lacks exposure to physical products, and consumers understand that only the product information can be presented through the Double 11 platform rather than real products. Thus, the authenticity of the information is crucial. When customers perceive that the information on the Double 11 platform is true, they will have the intention to participate, and affective and cognitive experiences will be generated. From the initial price reduction of products to the rule of spending 300 and getting 40 reductions, in the addition to interactive games, activities related to the Double 11 have been continuously enriched. When consumers are familiar with the novel rules, they interact with friends in the process of participating in the game, feeling happy and pleased. Thus, familiarity have a positive impact on the affective experience. But consumer's familiarity has no significant effect on the cognitive experience, it is possible that when consumers are not fully familiar with continuous novel shopping rules, they may also believe that participating in Double 11 will bring them appreciable benefits. Therefore, perceived information credibility has significant positive effect both on cognitive and affective experience. Familiarity has positive effect on consumers' affective experience, while not on cognitive experience.

In consequence of OCE, we find that cognitive and affective experiences positively affect behavioral intention. In other words, when consumers regard Double 11 as a meaningful shopping festival cognitively and emotionally, they will participate in Double 11 and purchase products. These findings are entirely consistent with the results of Rose et al.^{1,2}. This study discusses the impact of perceived information credibility and familiarity on cognition and affective experience and analyzes the effects of these factors on consumers' behavioral intentions. Exploring the antecedents and consequences of OCE contributes to helping merchants understand consumers' buying psychology and launch more effective strategies to increase sales and market share. Time and resource constraints do not allow iterations of data collection to observe customer intent over time and determine any long-term impact of the factors discussed on user intent and behavior. Future research should consider a longitudinal approach to validate and extend the current research model of OCE behavioral intention.

REFERENCES

- [1] Rose, S., Hair, N., & Clark, M. (2011). Online Customer Experience: A Review of the Business-to-Consumer Online Purchase Context. *International Journal of Management Reviews*, 13(1), 24–39.
- [2] Rose, S., Clark, M., Samouel, P., & Hair, N. (2012). Online Customer Experience in e-Retailing: An empirical model of Antecedents and Outcomes. *Journal of Retailing*, 88(2), 308–322.
- [3] Alnawas, I., & Aburub, F. (2016). The effect of benefits generated from interacting with branded mobile apps on consumer satisfaction and purchase intentions. *Journal of Retailing and Consumer Services*, 31, 313–322.
- [4] Molinillo, S., Navarro-García, A., Anaya-Sánchez, R., & Japutra, A. (2020). The impact of affective and cognitive app experiences on loyalty towards retailers. *Journal of Retailing and Consumer Services*, 54(April).
- [5] Fan, H., & Lederman, R. (2018). Online health communities: how do community members build the trust required to adopt information and form close relationships? *European Journal of Information Systems*, 27(1), 62–89.

- [6] Bleier, A., Harmeling, C. M., & Palmatier, R. W. (2019). Creating effective online customer experiences. *Journal of Marketing*, 83(2), 98–119.
- [7] Kang, Y. J., & Lee, W. J. (2018). Effects of sense of control and social presence on customer experience and e-service quality. *Information Development*, 34(3), 242–260.
- [8] Kang, J. Y. M., Mun, J. M., & Johnson, K. K. P. (2015). In-store mobile usage: Downloading and usage intention toward mobile location-based retail apps. *Computers in Human Behavior*, 46, 210–217.
- [9] Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, 44(2), 544–564.
- [10] Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, 4(October).
- [11] Ba, S., & Pavlou, P. A. (2013). Evidence of Evidence of the Effect of Trust Building Technology in Technology in Electronic Markets: Electronic Markets: Price Premiums and Premiums and Buyer Behavior'. *MIS Quarterly*, 26(3), 243–268.
- [12] Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96.
- [13] Zhao, Y., Chen, Y., Zhou, R., & Ci, Y. (2019). Factors influencing customers' willingness to participate in virtual brand community's value co-creation: The moderating effect of customer involvement. *Online Information Review*, 43(3), 440–461.
- [14] Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science*, 27(2), 184–206.
- [15] Laumer, S., Maier, C., Eckhardt, A., & Weitzel, T. (2016). User personality and resistance to mandatory information systems in organizations: A theoretical model and empirical test of dispositional resistance to change. *Journal of Information Technology*, 31(1), 67–82.
- [16] Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), 106–121.
- [17] Fornell, C., & Larcker, D. F. (1981). Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and.pdf. XVIII(February), 39–50.
- [18] Henseler, J., Ringle, C. M., & Sarstedt, M. (2014). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- [19] Lee, J., Chambers, E., Chambers, D. H., Chun, S. S., Oupadissakoon, C., & Johnson, D. E. (2010). Consumer acceptance for green tea by consumers in the United States, Korea and Thailand. *Journal of Sensory Studies*, 25(SUPPL. 1), 109–132.

Short Research Paper**Understanding Barriers to Service Robot Usage: A****Qualitative Study in Hotels***Xiaojiang Zheng¹, Shixuan Fu^{1*}, Ge Liu²*¹School of Tourism Science, Beijing International Studies University, China²School of Information Security, Chongqing Police College, China

Abstract: Robot applications are gaining momentum in many fields, especially in service-oriented industries. However, the robot usage rate in service industry is relatively low. Thus, we conducted 23 semi-structured interviews in upscale and luxury hotels. Accordingly, we identified barriers to service robot usage from three stages, including the pre-usage stage, the encounter stage, and the post-usage stage. Moreover, based on Expectation-disconfirmation Theory (EDT), we investigated the influential mechanisms of identified barriers. In the pre-usage stage, negative organizational valence and hoteliers' unreadiness would trigger hoteliers' low initial expectations to service robot usage. In the encounter and post-usage stage, robot unfriendly design, robot low working efficiency, and hoteliers' technostress would jointly influence the expectation-performance disconfirmation. The disconfirmation would cause the hoteliers' dissatisfaction to service robot usage. Consequently, hoteliers are reluctant to use service robots. The findings extend the literature of service robot studies and Expectation-disconfirmation Theory in the context of hospitality management as well as help guide the development of service robots.

Keywords: Service Robot, Barriers, Robot Usage Stages, Expectation-disconfirmation Theory

1. INTRODUCTION

Service robots are defined as “actuated mechanisms programmable with a degree of autonomy and humanness, moving within its environment, to finish specific works” [1]. Nowadays, service robot usage is developing rapidly, diffusion into a broad of service-oriented industries including public service, healthcare, tourism, etc. For example, in airports, service robots could provide travelling information to travelers. In many restaurants, service robots are used for food preparation. The robot working progress would be visible to customers as they could see how service robots providing the foods in the kitchen [1]. In the post-COVID-19 era, service robots could provide contactless service to meet the need for epidemic prevention and control [2]. Noticing the robot development momentum, the International Federation of Robotics (IFR) avowed that the market for professional service robots grew rapidly in 2020 by 32% from USD 8.5 billion to USD 11.2 billion. Meanwhile, the global service robotics market is expected to reach USD 63.8 billion by 2025 [3].

However, it's noteworthy that there are still tremendous service companies not using service robots [4]. Besides, although some companies have successfully adopted robots, the robot workforces were still reduced. Therefore, it is imperative to investigate why individuals are reluctant to use service robots. Previous studies have preliminarily revealed the challenges faced by service robot manufacture, for example, the scarcity of government regulations would impede service robot manufacture [1][5]. Meanwhile, some studies also explored customers' perceptions and indicated that customers' distrust of service robots was a major impediment to service robot usage [6][7]. However, considering employees are major users and decision-makers [8], it still lacks a holistic consideration for barriers to service robot usage from the employee perspective in an organization [1]. What salient barriers eventually influence service robot usage and in what way this influential mechanism in an organization is largely unknown.

* Corresponding Author. Email: fsx8888@163.com (Shixuan Fu)

Noticing the above research gaps, firstly, this study identified barriers to service robot usage holistically based on three robot usage stages, including pre-usage stage, encounter stage, and post-usage stage. These stages are covering the whole processes of service robot usage in an organization and could provide a set of useful theoretical explanations for understanding the barriers in different service robot usage levels. Secondly, we investigated the influential mechanisms of the barriers by integrating the Expectation-disconfirmation Theory (EDT), which is used for explaining individuals' technology using intentions^[9]. Thus, EDT is resonating with our study purpose. Moreover, this study proposed a theoretical framework of barriers to service robot usage. Accordingly, we conducted semi-structured interviews of 23 hoteliers in upscale and luxury hotels in China. Upscale and luxury hotels are typical service enterprises applying service robots and more agile to new technology compared with other organizations^[10]. The research questions are as follows:

RQ1: What are salient barriers to service robot usage from different stages in the hotel context?

RQ2: How could these barriers jointly impede service robot usage in the hotel context?

The following sectors of this paper are structured as follows. Firstly, we present a literature review of the related studies. Secondly, we post the methodology and data collection process. Thirdly, we show the preliminary findings and the theoretical framework. Finally, we draw a conclusion of this paper, including theoretical contributions, practical implications, and future work directions.

2. LITERATURE REVIEW

2.1 ICTs and robot in the service industry

Information and Communication Technologies (ICTs) have been conceptualized as all devices, networking components, applications, and systems that allow humans to interact with the world^[11]. ICTs have been dramatically influencing the hospitality industry. Nowadays, ICTs have become pervasive as a routine part of daily operations of collecting customers' information, conducting business, and making managerial decisions.

As a form of ICTs, service robots are different from other traditional ICTs. The novelties of service robots include the following perspectives. Firstly, the capability of sensing, learning, and adapting without human intervention is the critical point that distinct service robots from traditional automation technologies^[12]. Only those automations that perform a certain level of intelligence could be regarded as robotics, which are enabled by advanced mechanical engineering and AI technologies, e.g., image recognition, automatic speech recognition, and natural language processing. Thus, tremendous ICTs are considered as automation but not robotics. Secondly, robots have anthropomorphic characters and afford a higher level of human-technology interactivity^[13]. Many service robots have human-like appearances and can mimic human motion like tilting their head and making gestures.

Drawing on previous studies related to service robot usage, studies paid attention to the following perspectives: Firstly, studies illustrated the drivers to service robot usage. Service robots could work continuously without needing a rest or any salary. Meanwhile, service robots could handle simple tasks fast and accurately^[14], thus, service robots would release employees from repetitive or dangerous works. Thus, service robot usage is beneficial to increase operational efficiency. Furthermore, individuals' curiosity also drives them to use service robots. Individuals would think service robots are more advanced, stylish, as individuals have yet to interact with robots^[15].

Secondly, current scholars preliminarily explored the individuals' robot usage experience after interacting with service robots. Drawing on individuals' acceptance of service robots, studies revealed that perceived ease of use (PEOU) and perceived usefulness (PU) are two major indicators of individuals' acceptance^{[16][17]}. Besides, studies indicated that individuals would evaluate robot embodiments, human emotions, human-oriented perceptions, the feeling of security, and co-experience during what human-robot interaction process^[18].

2.2 Robot usage stages

Previous studies revealed the robot usage process in the service industry context^{[19][20]}. Generally, the service robot usage could be divided into three stages, including pre-usage stage, encounter stage, and post-usage stage as follows^{[19][20]}.

The pre-usage stage is the starting condition. In this stage, individuals would realize the existence of service robots and find the relevant information of those advanced facilities.

The encounter stage involves the interaction between individuals and service robots. In this stage, individuals and service robots co-create the service delivery process, and individuals would preliminarily evaluate the service robot design (e.g., technology interface, functions).

The post-usage stage is the process that individuals would appraise the technology performance and effectiveness (e.g., service quality, working efficiency) based on using experiences in the robot-encounter stage. Then, users would compare the evaluation results with the initial expectations toward service robots.

2.3 Expectation-disconfirmation theory (EDT)

Expectation-disconfirmation theory (EDT) was developed by Bhattacharjee^[21] to systematically explain individuals' expectations and purchasing intention in marketing across many products and services. In technology usage contexts, EDT explains why individuals are reluctant to use the new technology^[22], such as outsourcing platforms^[23], mobile applications^[24], and augmented reality (AR)^[25]. From previous studies, EDT has four key elements, including expectation, expectation-performance disconfirmation, satisfaction, and users' behavioral intentions^[21]. Expectations are individuals' beliefs and images before using the technology. Expectation-performance disconfirmation is a subjective post-usage comparison that thinking the technical performance was worse than individuals' expectations, while individuals' performance in the workplace could not be improved by technology usage. Especially, satisfaction is an essential dependent variable that represents a hotelier's emotional state following the usage experience. The behavioral intention is an individuals' determination to behave in a certain way^[22].

3. METHODOLOGY AND DATA COLLECTION

3.1 Methodology

In this research, we chose a qualitative methodology as this research is an early attempt to investigate the impediments of service robot usage in hotels. Corbin and Strauss^[26] revealed that qualitative methods are preferable to study the topics in the nascent stage of theoretical development. Accordingly, this study adopted the grounded theory and "grounding" the theory by analyzing research data. we conducted semi-structured interviews to gather first-hand data and developed systematic analysis procedures based on grounded theory. From Corbin and Strauss^[26], the systematic procedures focus on the individuals' perceptions and experience adhere to a strict and rigorous protocol. we strictly followed the coding processes including open coding, axial coding, and selective coding to better understand the impediments to service robot usage from hoteliers' perspectives. Furthermore, we applied the constant comparative during the coding processes to achieve maximum rigor. It is an iterative process and requires the researchers to go back and forth when analyzing each interview to the next and back again to continually compare the data and identify the similarity and differences from each statement.

3.2 Data collection

Data were collected from February 2020 to April 2020. The sample consists of 23 hoteliers who were working in upscale and luxury hotels. Meanwhile, studies verified that hoteliers could provide indicative insights about hotel technology usage^[8]. We selected hoteliers with a balanced ratio of men and women. Considering the different robot usage stages, we chose hoteliers who yet to work with service robots or stop using service robots. Their job position including general managers, operation directors, middle managers, and grass-root staffs.

At the beginning of each semi-structured interview, we invited interviewees to give a short self-introduction. Meanwhile, we introduced the service robot development and application. In the interview process, we left enough time for the interviewees to think deeply about each question. Each interview lasted over 30 minutes. With interviewees' permissions, the interviews were well recorded and transformed. As interviewees explained their perceptions, they were asked the current robot usage situation and why hotels yet to use robots or stop employing the robots, followed by more probing questions about their experience related to service robot usage. Every respondent was labeled after the interview.

4. PRELIMINARY FINDINGS

4.1 Barriers to service robot usage in hotels

The following part is our preliminary findings regarding barriers to service robot usage in hotels. The barriers were identified from interview transcripts based on the grounded theory approach. The salient barriers in the pre-usage stage include negative organizational valence and hoteliers' unreadiness. Unfriendly design is the most salient barrier in the encounter stage. In the post-usage stage, low working efficiency of robots and hoteliers' technostress are the salient barriers.

● The pre-usage stage

Barrier 1: Negative organizational valence. Negative organizational valence means that hoteliers believe the organization (hotel) will not be benefited from the service robot usage [27]. Hoteliers would search relevant information about service robots before adoption, and then preliminarily evaluate the service robot according to the collected information. In this study, respondents mentioned that hearing from current news and hotel experts, some adopted hotels would face the potential legal risks introduced by service robot usage, as the data in the robot system would be leaked. Besides, hoteliers emphasized that the hotel would face financial distress. According to previous cost estimates, hoteliers would evaluate the potential financial cost to service robot usage. If the service robot is too expensive for the hotels, the hotel would suffer from a cash-flow shortage based on cost estimates, then the hotel would not obtain benefits from service robot usage. Consequently, hoteliers would refrain from adopting service robots.

Barrier 2: Hoteliers' unreadiness. Hoteliers' unreadiness is the condition or state in which a hotelier is unprepared for working and collaborating with the service robot before interacting with service robots. Respondents proposed that service robot usage requires hoteliers to achieve skills to qualify for robot-related tasks. However, respondents believed that they are incapable to operate and maintain the service robot. Thus, hoteliers' unreadiness is another resistant factor for service robot usage.

● The encounter stage

Barrier 3: Unfriendly design. The design of service robots could be generally divided into visible components and invisible components. The visible components are robot morphological features, such as color, shape. Although previous studies revealed that service robots have anthropological design cues, i.e., human-like appearance [12], respondents reported that service robots still lack human-like characters. They even used words like "scary", "terrible" to describe the robots' appearance. The invisible components are robotic functions. Respondents indicated that they would evaluate whether service robots could provide suitable functions in different scenarios. However, most service robots are designed for a fixed purpose in a targeted scenario rather than setting multifunctions for different scenarios, like concierge robots are incapable to clean the guests' rooms. Thus, service robots are hard to provide high-quality service during the interaction process due to the limited functions.

● The post-usage stage

Barrier 4: Robot low working efficiency. Respondents would evaluate the quantity and quality of tasks

finished by service robots, i.e., robot work efficiency. However, robots' low working efficiency was complained about by many respondents. Service robots are incapable to provide the service accurately and fastly. Respondents would be prone to negative attitudes due to the service robot frequently crash and restart, low technical quality, unreliable at critical times.

Barrier 5: Hoteliers' technostress. Technostress is the stress caused by service robot usage [28]. Respondents stated that they were forced to spend more time and effort on robot-related tasks. Thus, the service robot usage increased the amount of workload and the level of time pressure. Furthermore, in the post-usage stage, hoteliers faced the fast pace of service robot upgrades, which triggered hoteliers' anxiety. Every new upgrade means that hoteliers have to adapt to new situations. Meanwhile, hoteliers had to consume more time and energy to learn the new functions.

4.2 Influential mechanisms

We also roughly proposed the influential mechanisms of barriers to service robot usage as follows:

● The pre-usage stage

The low initial expectation to service robots is jointly influenced by negative organizational valence and hoteliers' unreadiness in the pre-usage stage. Respondents believed that the hotel would not benefit from service robot usage. Meanwhile, respondents are still not fully prepared for the service robot usage. Thus, they held low initial expectations of service robot usage in hotels. Consequently, the low initial expectation negatively influences their usage intention, as hoteliers would be reluctant to use service robots in the pre-usage stage.

● The Encounter/post-usage stage

Expectation-performance disconfirmation is affected by the robots' poor design, low working efficiency and hoteliers' technostress. The actual design and performance of robots could not meet hoteliers' expectations, while perceived technostress would trigger work-exhaustion in the workplace. Meanwhile, respondents stated that perceived technostress was detrimental to hoteliers' working performance. Thus, the disconfirmation between expectation and (robot/hotelier) performance would trigger hoteliers' dissatisfaction with service robots, consequently, hoteliers would be reluctant to use service robots.

4.3 Theoretical framework

Figure 1 displays the theoretical framework of this study. Firstly, we identified the contingent barrier-related perspectives based on the interview data, including organizational valence, hoteliers' readiness, unfriendly design, robot performance, and technostress. Secondly, we adopted the EDT to present the influential mechanisms of the robot-related barriers on hoteliers' initial expectation and expectation-performance disconfirmation. Thirdly, we preliminarily propose the robot usage barrier framework by integrating barriers to service robot usage and EDT.

5. EXPECTED CONTRIBUTIONS

This study investigated barriers to service robot usage and the influential mechanisms of the barriers. The expected contributions as follows: Firstly, different from existing studies mainly focused on the positive value of service robots [18], we focused on the dark sides of service robot usage in the service industry. Moreover, current studies explored service robot challenges macroscopically [1][5] and customers' perceptions of service robot usage [18]. Differently, we focused on the employees' perceptions and identified the perspectives of barriers to service robot usage based on three usage stages. Accordingly, this study will further refine barriers to service robot usage. Secondly, few studies have investigated the influential mechanisms of barriers to service robot usage [2][6]. We adopted EDT and discussed the low initial expectation and the gap between users' expectations and actual technical performance. We will further discuss the influential mechanisms with research continues. Thirdly, we extended the EDT to the service robot usage context. Meanwhile, we integrated different perspectives, e.g., organizational valence, hoteliers' readiness, robot design into EDT, as the integrative approach would improve

EDT illustrative capacity [22].

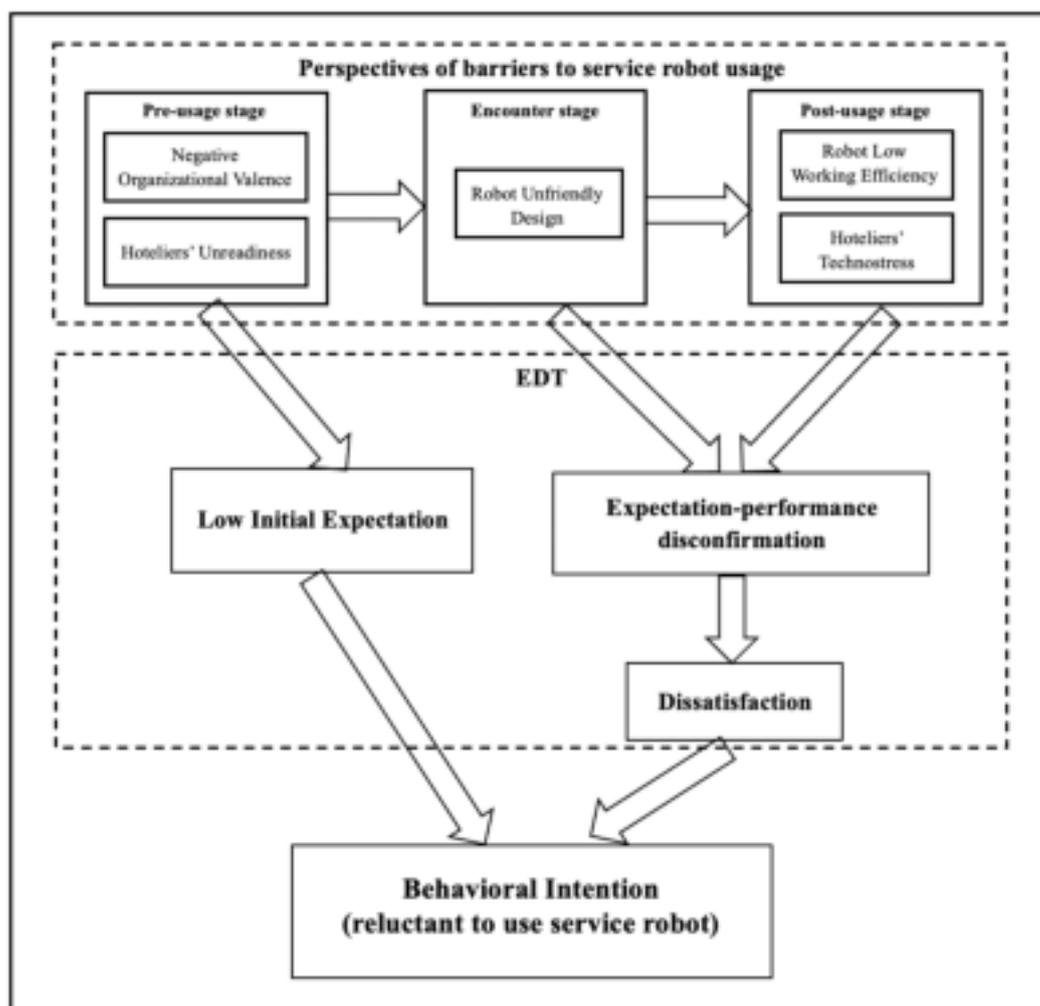


Figure 1. Barriers to service robot usage framework.

Practically, the contributions from this study will help hoteliers and robot manufacturers to find the influential factors of robot usage and better promote this advanced technology strategically. It would be beneficial to maximize the robot value and enlarge the target audiences of service robots.

6. LIMITATIONS AND FUTURE WORK

However, this study only focused on robot usage in upscale and luxury hotels from hotelier perspectives. In the future, we will increase our sample sizes and consider the characters of multi-type hotels and other service industries. Secondly, to increase the illustrative capacity, we will examine the proposed framework in the context of other technologies used in hotels or other service scenarios. Thirdly, we will adopt quantitative research to verify the proposed model.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 72001005.

REFERENCES

- [1] Zhao Y Y, Bacao F. (2020). What factors determining customer continuingly using food delivery apps during 2019 novel coronavirus pandemic period? *International Journal of Hospitality Management*, 91: 1-12.
- [2] Ivanov S. (2019). Progress on robotics in hospitality and tourism: a review of the literature. *Journal of Hospitality and Tourism Technology*, 10(4): 489-521.
- [3] IFR. (2021). Executive Summary World Robotics 2020 Service Robots. <https://ifr.org/free-downloads/>
- [4] Xiao L, Kumar V. (2019). Robotics for Customer Service: A Useful Complement or an Ultimate Substitute? *Journal of Service Research*, 24(2): 1-21.
- [5] van Pinxteren M E. (2019). Trust in humanoid robots: implications for services marketing. *Journal of Services Marketing*, 33(4): 507-518.
- [6] Ho T H, Tojib D, Tsarenko Y, (2020). Human staff vs. service robot vs. fellow customer: Does it matter who helps your customer following a service failure incident? *International Journal of Hospitality Management*, 87: 1-10.
- [7] Larivière B. (2017). "Service Encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research*, 79: 238-246.
- [8] Brown S A, Venkatesh V, Goyal S. (2012). Expectation Confirmation in Technology Use. *Information Systems Research*, 23(2): 474-487.
- [9] Lam C, Law R. (2019). Readiness of upscale and luxury-branded hotels for digital transformation. *International Journal of Hospitality Management*, 79: 60-69.
- [10] Law R, Leung D, Chan C, (2019). Progression and development of information and communication technology research in hospitality and tourism A state-of-the-art review. *International Journal of Contemporary Hospitality Management*, 32(2): 511-534.
- [11] Murphy J, Gretzel U, Pesonen J. (2019). Marketing robot services in hospitality and tourism: the role of anthropomorphism. *Journal of Travel & Tourism Marketing*, 36(1): 1-12.
- [12] Oztemel E, Gursev S. (2020). Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, 31(1): 127-182.
- [13] Choi Y. (2020). Exploring the influence of culture on tourist experiences with robots in service delivery environment. *Current Issues in Tourism*, 23: 1-17.
- [14] Chan A P H, Tung V W S. (2019). Examining the effects of robotic service on brand experience: the moderating role of hotel segment. *Journal of Travel & Tourism Marketing*, 36(4): 458-468.
- [15] Lin H, Chi O H, Gursoy D. (2019). Antecedents of customers' acceptance of artificially intelligent robotic device use in hospitality services. *Journal of Hospitality Marketing & Management*, 28: 1-20.
- [16] Gursoy D. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Management*, 49: 157-169.
- [17] Tung V W S, Au N M. (2018). Exploring customer experiences with robotics in hospitality. *International Journal of Contemporary Hospitality Management*, 30(7): 2680-2697.
- [18] Tsiotsou R H, Wirtz J. (2015). *The Three-Stage Model of Service Consumption*. Cheltenham: Edward Elgar, 105-128.
- [19] Meuter M L. (2005). Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies. *Journal of Marketing*, 69(2): 61-83.
- [20] Bhattacharjee A. (2001). Understanding information systems continuance: An expectation-confirmation model. *Mis Quarterly*, 25(3): 351-370.
- [21] Lankton N K, McKnight H D. (2012). Examining Two Expectation Disconfirmation Theory Models: Assimilation and Asymmetry Effects. *Journal of the Association for Information Systems*, 13(2): 88-115.
- [22] Wolverton C C. (2020). Outsourcing success in the eye of the beholder: Examining the impact of expectation confirmation theory on IT outsourcing. *Information & Management*, 57(6): 1-53.

- [23] Tam C, Santos D, Oliveira T, (2020). Exploring the influential factors of continuance intention to use mobile Apps: Extending the expectation confirmation model. *Information Systems Frontiers*, 22(1): 243-257.
- [24] Li H. (2020). Who will use augmented reality? An integrated approach based on text analytics and field survey. *European Journal of Operational Research*, 281(3): 502-516.
- [25] Creswell J. (2006). *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, Thousands Oaks: SAGE Publications.
- [26] Corbin J M, Strauss A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*. 13(1): 3-21.
- [27] Zimmermann A, Lioliou E, Oliveira J S. (2020). How do offshoring-related changes in job characteristics affect onshore managers' affective organizational commitment? The moderating role of perceived organizational valence. *Journal of Information Technology*, 35(4): 316-336.
- [28] Ragu-Nathan T S, Tarafdar M, Ragu-Nathan B S. (2008). The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. *Information Systems Research*, 19(4): 417-433.

Full Research Paper

A Study on the Impact of Short-Video Product Placement Advertising on Viewers Ad Adoption Intention: A Perspective of ELM and Social Learning Theory

Yuzhen Liu^{1*}

¹School of Information Management, Wuhan University, Wuhan, 430072, China

Abstract: [Purpose/Meaning] Short-video platform is currently a new mainstream social media entertainment platform. The way of placing product advertisements in short videos is increasingly accepted by more people. Exploring the influencing factors of viewers' intention to adopt short video product placement ads is of great significance for further mining the commercial value of short-video advertising and promoting the maturity of the e-commerce module in short-video platforms. [Method/Process] Based on ELM and social learning theory and combined with the characteristics of short-video product placement ads, this study constructs a model about the impact of product placement ads on viewers' ads adopt intention. The empirical research obtained 304 samples through way of questionnaire, and verified the model with SmartPLS 3. [Results/Conclusions] First, product-celebrity matching degree and plot-realistic matching degree positively affect the usefulness of perceived advertising information, and perceived usefulness positively affects viewers' advertising adoption intentions. Second, viewers may ignore discount their own beliefs and imitate others through observational learning, and both discount own beliefs and imitating others positively affect the intention of advertising adoption. Based on the research conclusions, this research provides corresponding marketing suggestions for short video makers and short-video platforms.

Keywords: Short-video product placement ads, ELM, Social learning theory, Ad adopt intention

1. INTRODUCTION

With the popularity of short-video worldwide, such as Douyin and Tik Tok, a new social media advertisement in the form of short-video is becoming more and more popular. Especially in China, short-video advertising is becoming the mainstream form of marketing in China's mobile Internet era. More and more short video bloggers are selling products by embedding product advertisements in entertaining short-videos^[1]. Short-video platforms also need a new e-commerce form to achieve the closed loop of the business model and attract more high-quality internet celebrities to settle in.

Whether in China's Douyin, Kwai, or overseas Tik Tok and other short-video platforms, there are a large number of Internet celebrity bloggers who provide funny short-video content with plots to capture viewers' attention. The product placement ads in this study means: when the internet celebrities reach a large number of followers, they will try to implant product ads in their entertainment short-video plots to guide everyone to pay attention to the selling point of the product or directly guide them to purchase, which is a good way for them to obtain profit. This type of product placement ads were usually used in movies, TV series or variety shows in the past, but now it is creatively applied in short-video with plot. Compared with short-video advertisements that directly promote products, this type of ad hidden in plots of short-video is more entertaining and causes less disturbance to viewers. In the meantime, compared with the product/brand placement ads of traditional social media such as Instagram, WeChat, and Weibo, the new short-video placement ads have the following advantages: first, traditional social media's native embedded advertisements have a strong and obvious marketing purpose,

* Corresponding author. Email: 741091201@qq.com(Yuzhen Liu)

and the ways of placement is rigid, so that it is difficult to be favored by the majority of users. However, product placement ads in short-video display the target product in a more entertaining and vivid way, which can effectively increase the willingness of viewers to accept similar ads; Second, in the era of "short-video fever", attractive enough short video advertisements will be considered as high-quality content to be pushed by platform to viewers of very large quantity. This is an opportunity not to be underestimated for these Internet celebrities.

However, this kind of marketing in short-video platform is not yet fully mature. How to promote more viewers to pay attention to product placement ads is still a difficult problem for us. Only by encouraging viewers to adopt the persuasive ads provided, will they continue to learn more about the products, search for the products actively and even pay for them. Therefore, in order to further understand the internal process of viewer' adoption of short-video placement ads and explore which factors will significantly affect viewers' intention to adopt, we conducted an empirical study on it.

Until now, online video advertising marketing methods mainly include the ads inserted in the interval between plots in videos, native video ads, product or brand placement ads and so on. Many scholars have conducted various researches on online video marketing. In recent years, scholars have focused their research on the following aspects: users' attitudes towards advertising or intention to accept online video advertising^[2], discussing the inconvenience brought by advertising^[3], the impact of video marketing on consumers' online purchase intentions^[4], the quality of experience in online video advertising^[5], and so on. Shon M(2020) analyzed the unfavorable factors that caused consumers to block the video ads, and results show that the use of personal information has the greatest impact on it^[3]. With the background of beauty and health products, Alharthey B K(2021) considered customer satisfaction as an intermediary to study the influencing factors of online video marketing on consumer purchase intentions, and the results showed that the ads have a positive impact on customer satisfaction, which in turn has a positive impact on customers' online purchase intentions^[4]. With the goal of optimizing the quality of online video advertising experience, Bulkan U(2020)proposed a novel model for inserting advertisements into online videos^[5]. The proposed model performed well in the evaluation of a controlled test environment.

There is less research on short-video marketing in emerging short-video platforms such as Tik Tok and Douyin. Liu G(2019)explored the impact of short-video marketing on consumer brand attitudes in their research^[6]. Chen X(2020) studied how to choose the right KOLs(Key Opinion Leader) and arrange their advertising campaigns for short-video marketing of products in order to maximize the marketing effect^[7]. Mou J B(2020) compared the video marketing strategies of two video platforms, Tik Tok and YouTube, to propose an effective social media video marketing strategy for the company^[8]. But there is almost no research on product placement in short-videos. In the past, most research perspective of product placement focused on placement in movies, TV series, games and variety shows^[9]. Therefore, this research focused on studying the viewers' intention to adopt advertisements information for product placement in short-video plots, so that the results can expand the research related to short video marketing, and we also can provide effective marketing suggestions for enterprises.

Regarding advertising effects, past scholars usually used TAM (Technology Acceptance Model) or UTAUT (Unified Theory of Acceptance and Use of Technology) model to explain the internal reason and process of users adopting or accepting advertisements^{[10][11]}. Meanwhile, some scholars believed that ELM(Elaboration Likelihood Model) in the field of information processing can also be used to explain the influence of persuasive advertising information on consumers. People will process advertising information from both the central path and the edge path, and promote the change of the final advertising attitude^[4]. The essence of short video placement advertising is content marketing that conveys persuasive information and advertisers expect to show the selling point of the target product through vivid short video content. Therefore, this article used ELM as the

basis to explain the advertisement information processing when viewers facing short video placement ads. In addition to the characteristics of traditional product placement ads, short-video placement ad also has its own characteristics from short-video platforms: 1) You can clearly learn the identity information of the celebrity who post the ad video and browse his past short-video works; 2) You can see the thumbs-ups, comments and reposts of many other users on this short-video placement ad. Both of these two characteristics are not available in traditional scenarios. Based on the above analysis, this article studied the influencing factors and internal mechanism of short-video product placement ads on viewers' intention to adopt ads from three aspects: the advertising content itself, the characteristics of the internet celebrity and the interactive atmosphere.

2. THEORETICAL BACKGROUND

2.1 Information adoption model(IAM) based on ELM.

The information adoption model is developed from the Elaboration Likelihood Model (ELM) proposed by psychologists Petty and Cacioppo. The theory of ELM believes that individuals will view the information to be processed from two perspectives, the central path and the edge path, thereby their attitudes towards information will be changed^[12]. The central path refers to an individual's judgment on the central content and quality of the information leading to a change in attitude, while the edge path refers to individual identifies the credibility of the information relying on the assistance of other factors closely related to central information. Sussman and Siegal combined the theories of ELM and TAM to propose a new information adoption model, as shown in Figure 1^[13]. This theory believes that when facing new information, individuals will judge the usefulness of information based on the quality of the information (central path) and source credibility (edge path), and finally decide whether to adopt the information.

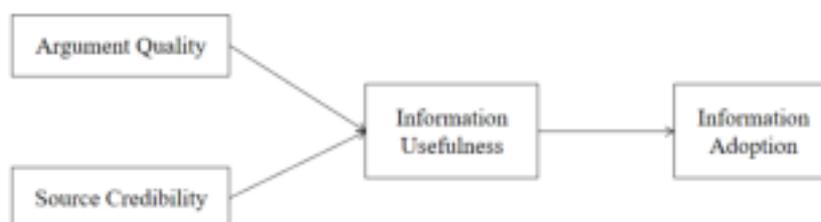


Figure 1. Model of information adoption.

2.2 Social learning theory(SLT).

SLT emphasizes the restrictive effect of social variables on human behavior, and believes that individuals will learn from important social references to determine individual behavior^[14]. The social learning process includes two types: observational learning and reinforcement learning. Observational learning means that the individual learns by observing the behaviors of people around the environment and their behavioral results before acting, so as to avoid unnecessary errors. Therefore, observational learning often leads to herd effect^[15]. There are two main conditions for herding behavior: uncertainty and observation^[16]. In short-video product ads, the content of tens of seconds is very limited for the presentation of product characteristics. Therefore, viewers often have uncertainty about the target product, moreover the interactive mechanism of short-video platform presents users a good way of observing and learning. In such condition, viewers are likely to have herd behavior.

On the other hand, reinforcement learning refers to individuals learning from the consequences of their own actions, emphasizing the experience of the product search process^[17]. As for short-video platform as the main entertainment and social platform, the main motivation of users is to absorb short-video content rather than shopping. And the process of users encountering product placement ads is random rather than active search.

Therefore, this research did not consider the reinforcement learning process and focused on the impact of the observation learning process on users.

3. RESEARCH MODEL AND HYPOTHESES

Based on the above theories and the characteristics of short-video product placement ads, a model of the impact of short-video product placement ads on viewers' intention to adopt is constructed, as shown in Figure 2. This research mainly studied the influencing factors of viewers intention to adopt short-video product placement ads from three aspects: matching, celebrity's reputation and observational learning.

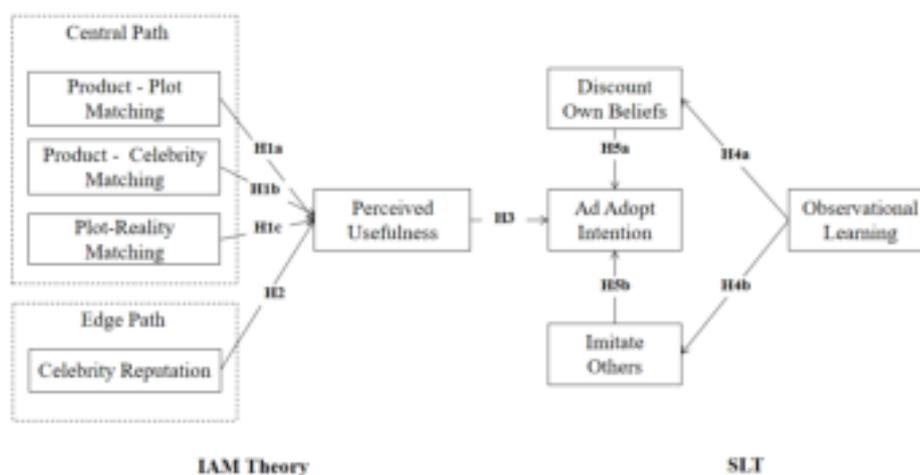


Figure 2. Research model.

3.1 The impact of matching on perceived usefulness.

According to IAM theory, information quality as a central path will affect viewers' perceived usefulness of information. The matching hypothesis was first used to discuss the relevance of the spokesperson's image and the endorsed product or brand^[18]. In the context of short-video product placement ads, Internet celebrities who place product ads in short-videos are similar to previous spokespersons. The difference is that the former involves the matching characteristic between products, plots, Internet celebrities, and reality in short-video content. Therefore, this article used matching as an indicator for viewers to evaluate the central quality of short-video product placement ads to affect perceived usefulness, and we subdivided the matching into three aspects for specific research.

(1)The match between product and plot. In short-video product placement ads, the scenarios where products are used are always often embedded in the original short-video plots to achieve product placement. When the characteristics of the product itself fit well with the plot itself, the viewers will consider it appropriate, thereby increasing the acceptance of the information conveyed by the advertisement. Otherwise, it will easily have a negative impact^[19].

(2)The match between product and celebrity. The Internet celebrity image that appears in the plot serves as the "Transmission Ambassador" of the product image in the embedded advertisement, so that the matching degree between them will significantly affect the viewers' attitude towards advertising and products^[20]. When the celebrity's image is in contrast with the product's image, it will make viewers feel abrupt and affect their perception of the information conveyed by the advertisement. Otherwise, it will make it easier for viewers to accept it.

(3)The match between plot and reality. The plot in the short-video is used to show the product usage scenario. When the viewer thinks that the plot is quite different from reality, it is difficult to form a sense of

substitution for using the product, thereby affecting the perception of advertising information^[19].

Based on the above analysis, the hypotheses of matching are proposed below.

H1a: The degree of match between the product and the plot positively affects the perceived usefulness of short-video product placement advertising information.

H1b: The degree of match between the product and the credibility positively affects the perceived usefulness of short-video product placement advertising information.

H1c: The degree of match between the plots and the reality positively affects the perceived usefulness of short-video product placement advertising information.

3.2 The impact of celebrity reputation on perceived usefulness.

According to IAM theory, the credibility of the source as an edge path will affect the usefulness of the user's information perception. Viewers are always not familiar with most bloggers on short video platforms and cannot directly judge the credibility of them. But they can observe the reputation of celebrities to determine the credibility through number of followers on the homepage, the past short-video works of celebrities and comments on the celebrity by other viewers in the comment area. In the research about the influence of advertising information, it is found that the advertisements of well-known celebrities have a stronger persuasive effect^[21]. Therefore, this study inferred that in the context of short-video product placement in advertising, the reputation of the celebrity will be considered as edge path to affect the viewers' perceived usefulness of advertising information, so the following hypothesis is proposed:

H2: The degree of reputation of the celebrity positively affects the perceived usefulness of short-video product placement advertising information.

3.3 The impact of perceived usefulness on intention to adopt the ads.

According to IAM theory, users' perceived usefulness of advertising information will positively affect advertising adoption behavior. In the context of short-video product placement ads, perceived usefulness refers to the viewer's perception of whether the delivered product advertising information is helpful for them. According to Davis (1989)'s Technology Acceptance Model (TAM), behavioral intention directly positively affects actual behavior^[22]. This study used the intention to adopt the ads instead of the behavior, that is, when viewers think that the advertisement information delivered is useful for them, they will have a intention to adopt the advertisement. Therefore, the following hypothesis is proposed:

H3: Perceived usefulness positively affects the intention to adopt the short-video product placement ads.

3.4 The impact of observational learning.

According to the observational learning process in SLT, in order to avoid unnecessary mistakes and expensive costs, individuals will observe and study the existing behaviors of people around them before acting, and will produce herd behavior when they feel sufficient positive feedback in the learning process. In the short-video platform, observational learning refers to observing the interaction between other viewers and this short-video ad, including the number of likes/comments/reposts and the specific content of comments. Herd behavior mainly includes two aspects, imitating others and discount one's own beliefs^[16].

In short-video product placement ads, imitation refers to observing that when other viewers have a high number of likes/comments/reposts on the short video, and the comment content is biased towards positive adoption of the ad, then the viewer will also choose to adopt the ad. This imitation behavior can save one's unnecessary time and effort, that is, the viewers think that the ads has been tested many others so that they no longer need to be suspicious of the advertisement and they can choose to adopt it directly.

discount one's own beliefs refers to ignoring the information you have obtained about the product and ignoring your degree of preference for the product when you choose to adopt the advertisement. It means

that in the decision-making process, viewers rely more on the observation of other viewers' behavior rather than their own ideas. They only make simple processing of information before making the adoption decision. Based on the above discussion, the following hypotheses are proposed:

H4a: Observational learning positively influences discount one's own information .

H4b: Observational learning positively influences imitating others.

H5a: Discount one's own information positively affects the intention to adopt the ads.

H5b: Imitating others positively affects the intention to adopt the ads.

4. METHODOLOGY

4.1 Measure

This paper obtained data with a questionnaire for empirical research. The questionnaire is divided into two parts. The first part sets up the research context for the participants and presents the measurement items of each variable. The context guides participants to recall the last short-video product placement ads experience and then answer the following items. The second part aims to collect individual demographic information. The measurement items of the study are adjusted with reference to the maturity scale and combined with the characteristics of short-video advertisements, and the items have undergone multiple theoretical verification according to the research purpose to ensure rigor. The final items are shown in Table 1.

Table 1. Measures of constructs.

Factors	Items	Reference
Product-Plot Matching (PPM)	PPM-1: I think the product in the short video is highly relevant to the plot. PPM-2: I think the product appearing in the short video is highly related to the scene presented. PPM-3: I think the product appearing in the short video closely match the characters in the plot.	[23]
Product-Celebrity Matching (PCM)	PCM-1: I think the celebrity is suitable for selling the products recommended in the short video plot. PCM-2: I think the celebrity matches the product sold in the ad.	[20]
Plot-Reality Matching (PRM)	PRM-1: The scene of the product in the short video plot is consistent with the scene it appears in reality. PRM-2: The use of the product in the short video plot is consistent with the way you use it in real life.	[19]
Celebrity Reputation (CR)	CR-1: The celebrity has a strong charm to me. CR-2: The celebrity has a high reputation on the short-video platform. CR-3: The celebrity has a high influence in a certain field.	[24][25]
Perceived Usefulness (PU)	PU-1: I think this short-video ad is helpful for me to some extent. PU-2: This short-video ad helps me make better shopping decisions. PU-3: This short-video ad is useful for me.	[13][26]
Ad Adopt Intention (AAI)	AAI-1: I would like to watch this short-video with ad. AAI-2: I would like to produce like/comment/repost behavior for this short-video with ad. AAI-3: I would like to adopt the information described in this short-video ad as a reference for shopping decisions.	[27]
Observational Learning (OL)	OL-1: I see a lot of people produced like/comment/repost behavior for this short-video ad. OL-2: I can feel that this short-video ad and product are very popular from the comment area of this short video ad.	[28]

Factors	Items	Reference
Discount Own Beliefs (DOB)	DOB-1: My acceptance of this ad content in this short-video does not fully reflect my true preferences DOB-2: I may not accept this advertisement based all on my own beliefs.	[16]
Imitate Others (IO)	IO-1: The most products recommended in the short-video platform are accepted by the majority of users, so I will also try to accept. IO-2: I will follow the views of other viewers to try to accept the product in this ad. IO-3: Many viewers accept this short-video ad and even purchase the product, so I will try to accept it.	[16]

4.2 Data Collection

This study used the Likert 7-point quantitative measurement method, and 1-7 indicate the range of "strongly disagree" to "strongly agree" in turn. The questionnaire was published on online and offline channels in various age groups. Individuals who participate in the questionnaire can receive a small payment to show our gratitude. The questionnaire collection time lasted from 2021.1.09 to 2021.1.14. A total of 356 questionnaires were collected in 6 days. After excluding invalid questionnaires and questionnaires of participants who had never met short-video product placement ads, 304 valid questionnaires were finally obtained.

4.3 Sample Description

The basic statistics of the samples are shown in Table 2. The results show that the ratio of men and women is not much different. The ages are mainly concentrated in the 18-30 years old. The education background is mainly undergraduate. The samples cover all parts of the frequency of use of short-video APP. Most participants have used short-video APP for 6 months to 2 years.

Table 2. Demographic profile of the respondents.

Variables	Item	Frequency	Percent%
Gender	Male	159	52.3%
	Female	145	47.7%
Age	Under 18	53	17.43%
	18-30	149	49.02%
	31-40	63	20.72%
	Over 40	39	12.83%
Highest education	High school or below	106	34.87%
	Undergraduate	139	45.72%
	Master degree or above	59	19.41%
The time of using short-video apps every day	Within 30 min	90	29.61%
	30min-1hour	86	28.29%
	1-2 hours	74	24.34%
	More than 2hours	54	17.76%
How many years have been using short video apps	Within 6 months	63	20.72%
	6months-1year	98	32.24%
	1-2years	81	26.64%
	More than 2 years	62	20.39%

5. EMPIRICAL ANALYSIS

5.1 Reliability and validity analysis.

This study used the PLS Algorithm of SmartPLS 3 to verify the reliability and validity of the measurement model. The Cronbach's Alphas (α) value and Composite Reliability (CR) value can reflect the reliability of the model, and the Average Variance Extracted (AVE) value can reflect the aggregate validity of the model. The test method for discriminative validity is to compare the square root of AVE and the absolute value of the correlation coefficient between this variable and other variables. Table 3 and Table 4 respectively show the reliability, aggregate validity and discriminative validity of the model.

It can be seen from Table 3 that outer loadings of all items are greater than 0.8, CR values are greater than 0.9, and α values are between 0.822 and 0.921, indicating that the model has good reliability. The AVE values of all variables are between 0.785 and 0.860, indicating that the model has good aggregation validity. It can be seen from Table 4 that the correlation coefficient between the variables (the values in the columns below the diagonal line) is less than the square root of the AVE value of each variable (the value of the diagonal line), indicating that the model has good discrimination validity.

Table 3. Related indicators of reliability and aggregate validity.

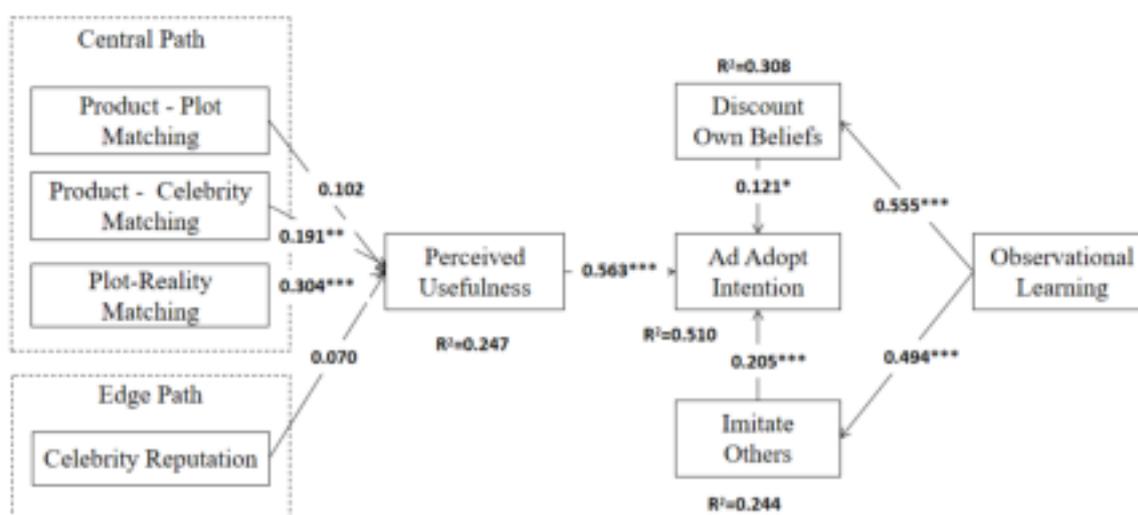
Variable	Scale items	Outer loading	AVE	CR	α Vale
PPM	PPM-1	0.910	0.788	0.918	0.866
	PPM-2	0.879			
	PPM-3	0.874			
PCM	PCM-1	0.931	0.852	0.920	0.827
	PCM-2	0.915			
PRM	PRM-1	0.941	0.858	0.924	0.836
	PRM-2	0.911			
CR	CR-1	0.932	0.855	0.946	0.921
	CR-2	0.945			
	CR-3	0.896			
PU	PU-1	0.897	0.785	0.916	0.863
	PU-2	0.900			
	PU-3	0.860			
OL	OL-1	0.940	0.860	0.925	0.839
	OL-2	0.915			
DOB	DOB-1	0.919	0.849	0.918	0.822
	DOB-2	0.923			
IO	IO-1	0.925	0.826	0.935	0.895
	IO-2	0.900			
	IO-3	0.902			
AAI	AAI-1	0.895	0.804	0.925	0.878
	AAI-2	0.914			
	AAI-3	0.881			

Table 4. Related indicators of discriminative validity.

	PPM	PCM	PRM	CR	PU	OL	DOB	IO	AAI
PPM	0.888								
PCM	0.401	0.923							
PRM	0.323	0.324	0.926						
CR	0.192	0.333	0.407	0.925					
PU	0.291	0.354	0.428	0.277	0.886				
OL	0.173	0.189	0.251	0.257	0.392	0.928			
DOB	0.167	0.239	0.268	0.208	0.348	0.555	0.921		
IO	0.125	0.180	0.357	0.216	0.286	0.494	0.475	0.909	
AAI	0.349	0.334	0.487	0.253	0.664	0.433	0.415	0.423	0.897

5.2 Model verification

This study used the Bootstrapping algorithm of SmartPLS 3 to test the significance level of each path for 1000 sampling times, and verifies the hypotheses of each path in the model based on the path coefficient. The model verification result is shown in Figure 3.



Note: ***, ** and * respectively represent $p < 0.001$, $p < 0.01$, $p < 0.05$.

Figure 3. Model verification.

From the results in Figure 3, we can see that the hypotheses H1b, H1c, H3, H4a, H4b, H5a, and H5b are valid, and the H1a and H2 are not valid. R^2 reflects the level of interpretation of the dependent variable by the independent variable. In the field of user behavior research, when the R^2 value is greater than 0.2, it is considered that the variable has a high degree of interpretation. In this study, the R^2 values for perceived usefulness, discount one's own beliefs, imitating others, and ad adopt intention are 0.247, 0.308, 0.244, and 0.510, respectively, which are all greater than 0.2, indicating that the model constructed in this study has a good fit.

5.3 Result analysis

In the hypotheses about matching, H1a is not significant but both H1b and H1c pass the significance test and are both positively affected, indicating that viewers will pay more attention to product - celebrity matching and plot-reality matching when evaluating the usefulness of advertising information. In the meantime,

product-plot matching is not necessarily the main reference standard for viewers. It may mean that the viewers have a high tolerance for the level of product-plot match or even they are used to watching the short-video ads with mismatched product-plot. When the viewers think that the Internet celebrity is suitable for promoting such product and the plot is not too absurd, they will still be willing to accept the information in the advertisement.

The celebrity reputation as the edge path is not significant in this study, which means that viewers will not generally judge the value of the advertisement based on the reputation of the blogger when evaluating a short-video product placement ad. This result may be due to the emergence of a lot of small Internet celebrities in the short-video era in recent years. They will spend most of their attention on improving the quality of their short videos in order to occupy the attention of broad viewers from other celebrities. Therefore, in such an environment, viewers are also concerned more about the quality of the video content than who create this short-video.

The path coefficient of H3 is positive and significant, indicating that when viewers believe that the advertising information is useful to them, they are likely to have a willingness to adopt the advertisement.

In the herd effect test based on SLT, H4a, H4b, H5a and H5b are all significant, which shows that the convenient interaction mechanism set up by the short video platform does have a very significant impact on the users, and they generally consider viewers' interactive response as the criterion for deciding whether to adopt the advertisement. Even if the target product in the advertisement may not be completely in line with their own beliefs, they will think that the product that has been tested by a large number of other people deserves some attention.

6. RESEARCH INSPIRATION

Based on the above research results, the marketing suggestions for short-video platforms and short video advertisement makers are as follows.

The first suggestion is for Internet celebrities who have a large number of fans and have a product marketing plan through short-videos. On the one hand, you need to choose products that suit your own image and temperament for marketing, rather than consider the popularity of product as the only main reference standard. If it does not meet the expectations of consumers for your image to sell this product, you may not get good results even if the product is hot now. On the other hand, although the creativeness and novelty in advertising helps to form a deep impression in people's mind, do not pursue the novelty too much and lead to the video ad is absurd and divorced from reality, which still can not win the trust of consumers.

Secondly, the Internet celebrities should not relax their control over the quality of video content because of the large number of followers they already have. Under the current distribution mechanism in short-video platforms, good content will be distributed to a larger viewers, so that your video ad will have a chance to get more likes, comments and reposts. Driven by the viewers herd mentality, the Matthew effect will be generated, making the marketing effect of your short video advertisements beyond your imagination.

Finally, for short video platforms. If they want to achieve a closed-loop business model through e-commerce promotion, 1) they must use technical methods to help short-video makers to create higher quality and more innovative short-video ads. 2) To guide users to more actively interact with short-videos through optimization of APP functions. 3) To push the short-video ads to more accurate viewers to improve marketing efficiency through algorithm optimization.

REFERENCES

- [1] Li J, She Y, Lin L, et al. An Approach of Short Advertising Video Generation Using Mobile Phone Assisted by Robotic Arm[C]//Computer Graphics International Conference. Springer, Cham, 2020: 157-168.

- [2] Bi X , Nam I . An Empirical Test of Model on the Characteristics of Online Personalized Video Advertisement Affecting on Chinese Consumers' Advertising Acceptance Intention[J]. THE JOURNAL OF GLOBAL CULTURAL CONTENTS, 2019, 41:193-214.
- [3] Shon M, Shin J, Hwang J, et al. Free contents vs. inconvenience costs: Two faces of online video advertising[J]. Telematics and Informatics, 2020, 56: 101476.
- [4] Alharthey B K. How Online Video Marketing Can Lead to Consumer Online Purchase Intention of Beauty and Healthcare Products in KSA[J]. International Journal of Online Marketing (IJOM), 2021, 11(1): 14-38.
- [5] Bulkan U, Dagiuklas T, Iqbal M. Modelling quality of experience for online video advertisement insertion[J]. IEEE Transactions on Broadcasting, 2020, 66(4): 835-846.
- [6] Liu G, Gao P, Li Y, et al. Research on the Influence of Social Media Short Video Marketing on Consumer Brand Attitude[C]//2019 5th International Conference on Social Science and Higher Education (ICSSHE 2019). Atlantis Press, 2019: 433-438.
- [7] Chen X, Ji L, Jiang L, et al. More Bang for Your Buck: Effective Kol Marketing Campaign in Emerging Short-Video Markets[J]. Available at SSRN, 2020.
- [8] Mou J B. Study on social media marketing campaign strategy--TikTok and Instagram[D]. Massachusetts Institute of Technology, 2020.
- [9] Guo F, Ye G, Hudders L, et al. Product placement in mass media: a review and bibliometric analysis[J]. Journal of Advertising, 2019, 48(2): 215-231.
- [10] Zhao J, Wang J. Health advertising on short-video social media: a study on user attitudes based on the extended technology acceptance model[J]. International journal of environmental research and public health, 2020, 17(5): 1501.
- [11] Muk A, Chung C. Applying the technology acceptance model in a two-country study of SMS advertising[J]. Journal of Business Research, 2015, 68(1): 1-6.
- [12] Petty R E, Cacioppo J T. The elaboration likelihood model of persuasion[M]//Communication and persuasion. Springer, New York, NY, 1986: 1-24.
- [13] Sussman S W, Siegal W S. Informational influence in organizations: An integrated approach to knowledge adoption[J]. Information systems research, 2003, 14(1): 47-65.
- [14] Bandura A. Social learning theory of aggression[J]. Journal of communication, 1978, 28(3): 12-29.
- [15] Zhang J, Liu P. Rational herding in microloan markets[J]. Management science, 2012, 58(5): 892-912.
- [16] Sun H. A longitudinal study of herd behavior in the adoption and continued use of technology[J]. Mis Quarterly, 2013: 1013-1041.
- [17] Cheung C M K, Liu I L B, Lee M K O. How online social interactions influence customer information contribution behavior in online social shopping communities: a social learning theory perspective[J]. Journal of the Association for Information Science and Technology, 2015, 66(12): 2511-2521.
- [18] Carlson B D, Donovan D T, Deitz G D, et al. A customer-focused approach to improve celebrity endorser effectiveness[J]. Journal of Business Research, 2020, 109: 221-235.
- [19] Lee M, Faber R J. Effects of product placement in on-line games on brand memory: A perspective of the limited-capacity model of attention[J]. Journal of advertising, 2007, 36(4): 75-90.
- [20] Till B D, Busler M. The match-up hypothesis: Physical attractiveness, expertise, and the role of fit on brand attitude, purchase intent and brand beliefs[J]. Journal of advertising, 2000, 29(3): 1-13.
- [21] Hsu C, McDonald D. An examination on multiple celebrity endorsers in advertising[J]. Journal of Product & Brand Management, 2002.
- [22] Davis F D, Bagozzi R P, Warshaw P R. User acceptance of computer technology: A comparison of two theoretical models[J]. Management science, 1989, 35(8): 982-1003.

- [23] Russell C A. Investigating the effectiveness of product placements in television shows: The role of modality and plot connection congruence on brand memory and attitude[J]. *Journal of consumer research*, 2002, 29(3): 306-318.
- [24] Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness[J]. *Journal of advertising*, 1990, 19(3): 39-52.
- [25] Pradhan D, Duraipandian I, Sethi D. Celebrity endorsement: How celebrity–brand–user personality congruence affects brand attitude and purchase intention[J]. *Journal of Marketing Communications*, 2016, 22(5): 456-473.
- [26] Venkatesh V, Morris M G, Davis G B, et al. User acceptance of information technology: Toward a unified view[J]. *MIS quarterly*, 2003: 425-478.
- [27] Zhu T, Lu Y, Gupta S. How do network externalities affect customers' adoption intention in mobile app store: from a perception of consumer perceived value[J]. *International Journal of Mobile Communications*, 2018, 16(5): 493-512.
- [28] Zhang K Z K, Hu B, Zhao S J. How online social interactions affect consumers' impulse purchase on group shopping websites?[C]//PACIS. 2014: 81.

Full Research Paper**An Empirical Study of the Impacts of Digital Marketing Contents on User****Engagement in Social E-commerce Platform***Ronglin Hou¹, Fei Wang^{1*}, Zhen Zhu¹, Bingchen Guo¹, Jing Zhao¹*¹ School of Economics and Management, China University of Geosciences, Wuhan 430074, China

Abstract: This study empirically investigates the impacts of digital marketing contents on user engagement behaviors. Using digital storytelling as theoretical foundation, this paper identifies three categories of choices for designing digital contents, i.e., fundamental elements, structural elements, and presentation elements. Then, the study explores how these design choices enhance three different user engagement, including likes, comments, and subsequent purchases. With 12,548 posts in JD WeChat shopping circle as the empirical dataset, this study tests the hypothesized by combining machine learning method with regression analysis. The empirical results show the design choices discriminatively impact three kinds of user engagement behaviors. The theoretical and practical implications are discussed.

Keywords: digital marketing contents, digital storytelling, user engagement, social e-commerce platform

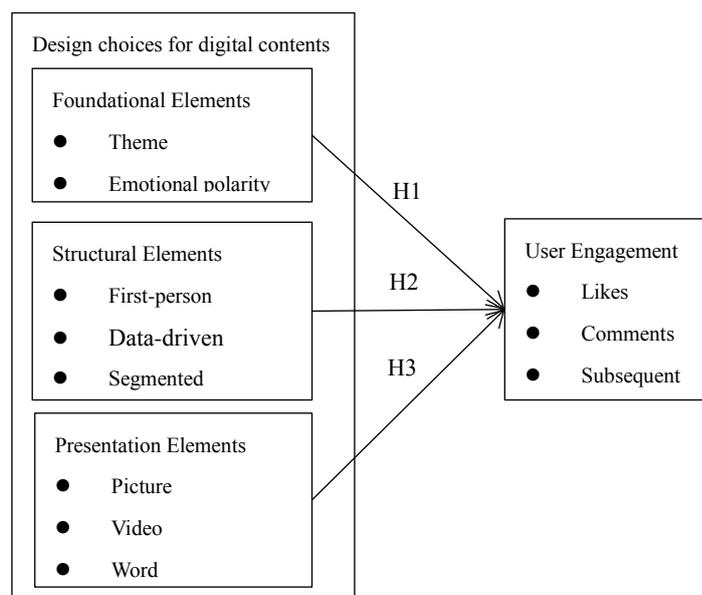
1. RESEARCH QUESTION

In order to attract customers and promote product sales, more and more firms are using social media platforms to design and release digital marketing contents such as essays, graphics, short videos, and live broadcasts. These digital contents provide customers with an immersive shopping experience, and thus enhance user stickiness and stimulate shopping needs^[1]. But how to design effective digital contents to achieve marketing goals becomes a new challenge for firms using social medial platforms.

Against above background, the objective of this study is to examine the impacts of digital contents on user engagement on social business platform. Accordingly, the research questions of this paper include: (1) what are the design choices of digital marketing contents on social e-commerce platform? (2) how these different design choices impact different user engagement behaviors?

2. THEORETICAL BACKGROUND

Although existing research explored the economic impacts of user-generated content (UGC) and firm-generated content (FGC)^[2], there are two limitations in explaining how to design digital marketing contents. First, current UGC and FGC research mostly focus on a series of contents, rather than a piece of digital content. For example, prior studies explore the overall quantity, valence, and richness of UGC and FGC. Such coarse-grained analysis cannot help digital content generators to design effective contents. Second, previous UGC and FGC research have not differentiated different user engagement behaviors, and lacked empirical research on the impacts of single text content on users actual

**Figure 1. Research model**

* Corresponding author. Email: wangfei086071@163.com

purchase behavior.

In order to fill up above knowledge gap, we employed digital storytelling as our theoretical foundation. Pavlik (2017) identified the three important components of digital storytelling in the digital media environment: (1) foundational elements (theme characteristics and emotional characteristics), (2) structural elements (whether first-person, whether segmented narrative, and whether to quote numbers), and (3) presentation forms (video, picture and word)^[3]. As shown in Figure 1, our research model presents how three categories of design choices impact three user engagement behaviors.

3. DATA COLLECTION AND ANALYSIS

We choose JD WeChat shopping circle as the research context of our empirical study because it provides data on digital content, product purchase links, user engagement behaviors such as comments, likes, and actual purchases of products after reading the content. We selected the makeup honey circle to collect data because it is one of most active shopping circles. The data includes 12548 posts from September 2016 to November 2019. We use ordinary least squares (OLS) regression analysis with robust standard error to test hypotheses. Table 1 gives the results. Model 1 only involved control variables, and Model 2 further included all independent variables.

4. CONCLUSIONS

Our empirical study found that the foundational elements, structural elements and presentation forms of digital contents can discriminatively promote users engagement behaviors. Theme features of foundational elements have the most significant impact. The foundational elements and presentation forms of content characteristics can affect the three engagement behaviors of consumers like, comment and purchase, while structural elements only affect consumers' like behavior. Our study has both theoretical and practical implications.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71702176

REFERENCES

- [1] Yang M, Ren Y, Adomavicius G. Understanding user-generated content and customer engagement on Facebook business pages[J]. *Information Systems Research*, 2019, 30(3): 839-855.
- [2] David Godes. Commentary—Invited Comment on “Opinion Leadership and Social Contagion in New Product Diffusion”[J]. *Marketing Science*. 2010, 30(2). 224-229.
- [3] Pavlik J V, Pavlik J O . Understanding Quality in Digital Storytelling: A Theoretically Based Analysis of the Interactive Documentary[M]. in *Digital Transformation in Journalism and News Media*, M. Friedrichsen and Y. Kamalipour (eds.), pp. 381–396.

Table 1. Regression analysis results

Variable	Model 1			Model 2			
	Like	Com. month	Pur. chase	Like	Com. month	Pur. chase	
Dependent Variable							
Control Variable	Follower	.362*** (1.29)	.323** (0.98)	.291*** (1.11)	.289*** (2.26)	.232** (1.13)	.192*** (-1.29)
	Level	.029** (2.13)	.273 (2.09)	.132* (1.94)	.094** (5.08)	.221 (2.81)	.102 (0.10)
	Price	-.361 (0.85)	-.633 (1.86)	-.372 (2.05)	-.278 (4.20)	-.798 (3.16)	-.232 (1.44)
	Length			.098* (1.32)	.009 (2.21)	.185 (4.21)	.185 (2.01)
	Valence			.317 (3.76)	.311 (4.28)	.185 (-2.01)	.185 (-2.01)
	Picture			.102** (0.185)	.392 (1.94)	.155* (3.19)	.155* (3.19)
	Video			.093* (1.21)	.382 (4.23)	.382** (0.28)	.382** (0.28)
	Date-driven			.044* (2.68)	.243 (4.32)	.04 (0.98)	.04 (0.98)
	First-person			.152* (-3.53)	.852 (-4.13)	.102 (3.29)	.102 (3.29)
	Layer			-.231 (3.42)	.113 (2.13)	.012 (0.14)	.012 (0.14)
Independent Variable	Topic 1			.009** (1.13)	.193* (3.24)	0.127** (4.23)	0.127** (4.23)
	Topic 2			.291 (0.271)	.852 (0.271)	0.118 (-2.09)	0.118 (-2.09)
	Topic 3			.001 (0.992)	.242 (1.27)	.209 (3.63)	.209 (3.63)
	Topic 4			.004** (2.13)	.187* (2.03)	.019** (1.93)	.019** (1.93)
	Topic 5			.013** (4.27)	.051 (9.20)	.003 (-2.09)	.003 (-2.09)
	Topic 6			.397 (0.181)	.118 (0.583)	.208** (1.02)	.208** (1.02)
	Topic 7			.102** (6.21)	.102** (1.291)	.203*** (-8.87)	.203*** (-8.87)
	Topic 8			.113** (3.58)	.229** (0.173)	.382** (2.07)	.382** (2.07)
Intercept Term	.273*** (2.13)	1.93*** (2.01)	-2.01*** (6.67)	.283** (1.98)	.111** (0.299)	-.812*** (1.21)	
N	12548	12548	12548	12548	12548	12548	
Adjusted R ²	0.118	0.128	0.108	0.182	0.199	0.248	
Change R ²				0.064	0.071	0.135	

Note: * p<0.1, ** p<0.05, *** p<0.01.

Short Research Paper

Why Do Employees Choose to Share Information Publicly? The Evidence from Enterprise Social Networks

Yu Wang¹, Daqing Zheng^{1,2†}

¹School of Information Management and Engineering, Shanghai University of Finance and Economics,
Shanghai, 200433, China

²Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics,
Shanghai, 200433, China

Abstract: More and more interpersonal communication is carried online. In this study, we distinguished two information sharing states: public sharing and private sharing, and identified the factors that influence users' states. Using communication privacy management theory (CPM), we developed a framework to explain the effects of context, motivation, risk-benefit ratio, and some additional characteristics on users' choice of the information sharing states in the context of enterprise social networks (ESNs). The results show that most metrics of context, motivation and risk-benefit ratio are significantly influence the information sharing states. As the first empirical study to classify and compare the information sharing states among users on ESNs, our study enriches the existing literature and opens new avenues for researchers and social networking platforms.

Keywords: enterprise social networks, information sharing, communication privacy management theory, multilevel mixed-effects logistic regression

1. INTRODUCTION

After the outbreak of COVID-19, popular enterprise social networks (ESNs) including Yammer (Microsoft Teams) and Ding Talk have received widespread attention. ESN is a type of SNs used in a business environment to build communication channels between organizations and employees^[1]. This study focuses on enterprise internal private social networks (EIPSNs), which are mainly used for internal communication and collaboration among employees. In the process of applying EIPSNs, enterprises often need to interface with other types of SNs, which exposes enterprises to the risk of confidential information leakage, so employees must have a higher sense of privacy protection when using ESNs for information sharing.

“Information sharing on ESNs” is an information behavior, in which two actors with specific relational connections collaborate in information exchange through ESN platforms to achieve individual or common interests^[2]. Existing research focuses on the consequences and antecedents of information sharing^[3, 4], they pay little attention to the user decisions at the time of information sharing and there is little research on how users choose whom to share with. However, the reality is that users not only need to make decisions about whether to share information, but also need to make careful decisions about with whom they share information^[5]. This study further explores the selectivity of the users' information sharing states: public sharing or private sharing. Based on Petronio's communication privacy management theory^[6], we capture the factors that influence the users' information sharing states, including context, motivation, risk-benefit ratio, and some additional factors, and use them to develop a selective model to conduct our study.

[†] Corresponding author. Email: zhengdaqing@sina.com

2. LITERATURE REVIEW

2.1 Information sharing on ESNs

ESNs allow employees within an enterprise to work and self-manage in a similar way to general SNs in order to achieve efficient, transparent and convenient communication and collaboration among employees. ESNs have empowered employees to participate in editing and creating network content, enabling employees to gradually evolve from recipients and viewers of network information to creators, publishers, sharers and disseminators^[7]. Compared with SNs, users consider information privacy and security issues more when sharing information on ESNs. Previous studies have tended to repeat the consequences and antecedents of information sharing behavior, they studied the impact of information sharing on enterprise value co-creation^[8], or discussed the endogenous and exogenous factors affecting information sharing^[9]. However, they ignored the fact that users also need to decide carefully with whom to share when judging whether to share information or not. The interviews results of Wang and Fussell^[5] acknowledged that users consider audience issues when sharing, they suggested that the key difference between public or private sharing is the user's perception of the information to be shared, but limited by the interview format of their study, they have no way of knowing whether this view is scientific.

Fortunately, the factors considered by existing information sharing studies are still important for our study. These key factors include the reputation of the user^[10], the language style of the user^[11], and the presentation of the information content^[4]. Xu and Yang^[12] summarized four categories of factors affecting information sharing, which are also applicable to the study of information sharing states.

2.2 Communication privacy management theory

Communication privacy management (CPM) theory^[13] is a very valuable privacy theory in social network communication. Petronio defined privacy boundaries as the process of sharing information from a completely public to a completely private process. Privacy boundaries are governed by privacy rules, which in turn are determined by many factors such as cultural values, gender orientation, motivation, context, and risk-benefit ratio. Among them, cultural values indicate the specific cultural context (e.g., China vs USA^[14]); gender orientation implies that men and women have different privacy boundaries; motivation refers to users are driven by clear interests (e.g., get more feedback^[15]); context points out the specific situations that users are currently in^[16]; and individuals make "mental calculations" by weighing benefits and risks when deciding whether to share information publicly. CPM theory has been used in research related to SNs widely applied and is considered applicable to the study of user information sharing management on ESNs^[17].

3. RESEARCH MODEL AND HYPOTHESES

CPM theory suggests that the factors users use to determine what information is publicly shareable are: cultural values, gender orientation, motivation, context, and risk-benefit ratio. Because the data in this study are derived from a single cultural context, we constructed a model based on other categories of factors, as shown in Figure 1.

3.1 Motivation

CPM theory suggests that the users may publicly share information if they are driven by clear interests^[13]. Users' motivation to share information publicly on ESNs is mainly due to the pressure to interact with information, if users have more friends, then the user will feel obligated to share information publicly because of the pressure of interaction, making the number of information publicly shared by the user higher^[18].

Hypothesis 1a. The user's social network size on ESNs is positively related to the public information sharing.

Wang and Fussell^[5] pointed out that users are expecting feedback when sharing information, more feedback from others strengthens users' resolve to share information further publicly, because they believe that, more feedback means that others find the information they share interesting or useful^[19].

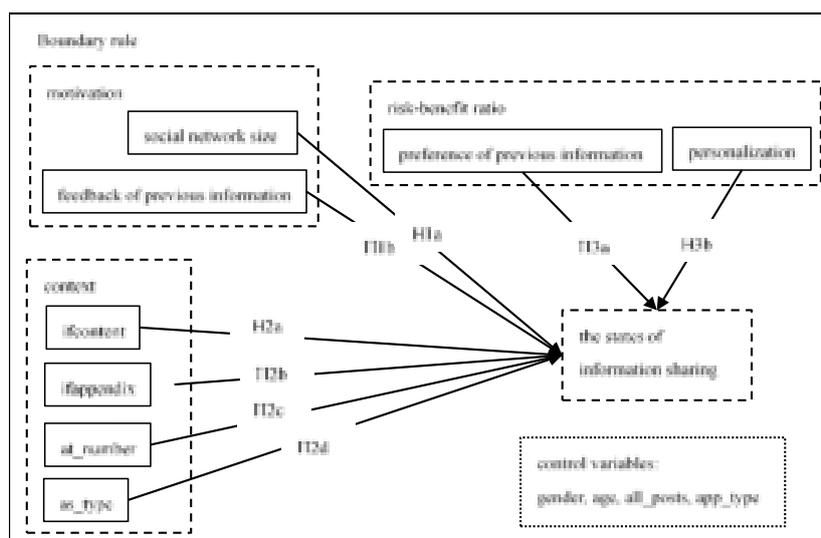


Figure 1. Research model

Hypothesis 1b. the feedback of user's previous information are positively related to public information sharing on ESNs.

3.1 Motivation

CPM theory suggests that the users may publicly share information if they are driven by clear interests^[13]. Users' motivation to share information publicly on ESNs is mainly due to the pressure to interact with information, if users have more friends, then the user will feel obligated to share information publicly because of the pressure of interaction, making the number of information publicly shared by the user higher^[18].

Hypothesis 1a. The user's social network size on ESNs is positively related to the public information sharing.

Wang and Fussell^[5] pointed out that users are expecting feedback when sharing information, more feedback from others strengthens users' resolve to share information further publicly, because they believe that, more feedback means that others find the information they share interesting or useful^[19].

Hypothesis 1b. the feedback of user's previous information are positively related to public information sharing on ESNs.

3.2 Context

Through literature combing, we propose the contextual element to refer to the specific situation in which the user is at the current moment, which is reflected by the information content^[16]. Four main points related to specific information are given and corresponding hypotheses are proposed. First, whether the information contains textual content is an important observation. If a user shares information that contains textual content, then the privacy of that information response is likely to increase and the user is more likely to choose to share privately. Second, whether the information shared by the user carries appendixes is also widely discussed, if the information contains images, videos or links, it may significantly affect the users' information sharing states. Third, the higher the number of users are mentioned in the shared information, indicating that the user has a very clear target of information sharing, the more likely the user will choose to share privately. Finally, in Twitter-related studies, whether the tweet is a retweet is highlighted^[12], original information and forwarding information significantly differ in affecting users' information sharing states. Hence, we reasonably speculate the following four hypotheses.

Hypothesis 2a. Whether a user's information shared on the ESNs contains textual content is negatively related to information sharing publicly.

Hypothesis 2b. Whether a user's information shared on the ESNs carries appendixes is negatively related to information sharing publicly.

Hypothesis 2c. The number of other explicitly mentioned users in the information shared by users on the ESNs is negatively related to information sharing publicly.

Hypothesis 2d. The post type of the information shared by users on the ESNs is significantly related to information sharing publicly.

3.3 Risk-benefit ratio

CPM theory tells us that users must perform a benefit-risk assessment when they choose to share information publicly^[13]. On ESNs, users' attitudes toward benefit-risk are related to previous information sharing experiences^[19]. If users choose to share publicly on a regular basis upfront, it indicates that users' perceived benefits are higher than risks, so they will prefer share publicly in the subsequent information sharing process.

Hypothesis 3a. Users' previous information sharing preferences on ESNs are positively related to the information sharing publicly.

Personalization is generally defined as “the ability to provide tailored content and services to individuals based on knowledge of their preferences and behaviors”^[20]. However, the more personal information users provide at ESNs, the higher the risk of exposing privacy, which means personalization is considered an important expression of users' perceived benefit-risk.

Hypothesis 3b. Users' personalization preferences on ESNs are positively related to the information sharing publicly.

4. DATA AND EMPIRICAL RESULTS

4.1 Data description

Our data comes from the EIPSN of an IT company, spanning from April 2013 to April 2018. After data cleaning, we select 93,281 information shared by 125 users as research data in this study. Combining the previous hypotheses and models, the descriptive statistics of the variables used in this study are shown in the table 1.

Table 1. Descriptive statistics of variables

<i>Variables</i>	<i>Abbreviations</i>	<i>Mean</i>	<i>Std.Dev</i>	<i>Min</i>	<i>Max</i>
<i>information sharing states</i>	<i>states</i>	0.20	0.40	0	1
<i>uid</i>	<i>uid</i>	71.82	40.93	1	125
<i>t</i>	<i>t</i>	2693.62	3631.80	1	16781
<i>social network size</i>	<i>size</i>	14.61	13.18	1	123
<i>feedback of pre-share information</i>	<i>fopi</i>	3727.71	5775.04	0	27103
<i>ifcontent</i>	<i>ifc</i>	0.71	0.45	0	1
<i>ifappendix</i>	<i>ifa</i>	0.26	0.44	0	1
<i>at_num</i>	<i>atn</i>	1.11	3.06	0	43
<i>as_type</i>	<i>ast</i>	0.51	0.50	0	1
<i>preference of pre-share information</i>	<i>popi</i>	0.45	0.50	0	2
<i>personalization</i>	<i>per</i>	6.24	2.77	0	9

4.2 Research method

Our research method is multilevel mixed-effects logistic regression (meqrlogit), which are useful for modeling within-cluster correlations in both longitudinal and panel data. This study considers a two-level model, as shown in equation (1).

$$\Pr(y_{ij} = 1|u_j) = H(x_{ij}\beta + z_{ij}u_j) \quad (1)$$

For $j = 1, \dots, M$ users, user j consists of $i = 1, \dots, n_j$ observations. The response value is the

dichotomous variable y_{ij} . The $1 \times p$ row vector x_{ij} is the fixed-effect covariate, the $1 \times q$ vector z_{ij} is the covariate corresponding to the random effect.

4.3 Empirical results

This study used stata15 for data analysis. We examined the correlation between the variables, and the variance inflation factors (VIF) results for all variables are strictly less than 10, show that there was no multicollinearity between our independent variables.

Table 3. Empirical results and hypotheses testing results

<i>Hypotheses</i>	<i>Odds Ratio</i>	<i>P> z </i>	<i>Supported</i>
<i>H1a: size</i> → states	1.097	0.000	Yes
<i>H1b: fopi</i> → states	1.000	0.000	No
<i>H2a: ifc</i> → states	0.359	0.000	Yes
<i>H2b: ifa</i> → states	1.024	0.010	No
<i>H2c: atm</i> → states	0.775	0.000	Yes
<i>H2d: ast</i> → states	0.531	0.000	Yes
<i>H3a: popi</i> → states	16.967	0.000	Yes
<i>H3b: per</i> → states	1.041	0.706	No

Table 3 shows the results after using multilevel mixed-effects logistic regression. We would like to highlight several interesting findings. First, when making information sharing states decisions, the more direct contacts a user has on the ESNs, the more likely the user is to choose public sharing. Second, users are more likely to choose private sharing if the information they share on ESNs contains exact textual content, and non-original information is more likely to be shared publicly compared to original information, these two results reflect users' concern for personal privacy information. Third, as the number of users directly mentioned when a user shares information on ESNs increases, the users are more likely share the information privately. This is relevant to the context of our study, where users within ESNs want to both ensure that as many users as possible can receive it accurately and avoid the drawbacks of public sharing. Finally, *popi* plays a key role in the decision-making process of users' information sharing states, but as a trichotomous variable, we need to give further analysis and interpretation.

Moreover, the underlying reason why *H1b* is not supported is that the odds ratio is not well explained, which requires in-depth analysis. *H2b* and *H3b* likewise need further explanation and discussion, which will be added in the next study.

5. CONCLUSIONS

This study applies CMP theory in the context of ESNs to elucidate the influence of motivation, context, and risk-benefit perceptions on user information sharing states, providing some evidence for key antecedents of user information sharing states. Using multilevel mixed-effects logistic regression, this study partially validates the importance of the relationship between influencing factors and users' information sharing states.

This study shifts the focus of information sharing behavior research to users' information sharing states decisions, and explores the influence of some key factors on users' information sharing states in the relatively emerging ESN context. It is expected to provide new ideas and theoretical support for information sharing and ESN-related research, and open up a new way for academia and industry to understand users' information sharing behavior, which is conducive to maximizing the value of enterprise knowledge management.

This study also has some limitations and directions for further research. For example, the measurement of some variables is still crude, the user's personalization (*per*) is obtained by summing multiple binary variables,

but the information lack of secondary data makes the measurement unsatisfactory. Therefore, future studies can try to find more appropriate proxy variables to drive more accurate empirical results.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China (Grant No. 20BGL287).

REFERENCES

- [1] Wang Y., Zheng D., Huang L.(2018). The Analyzing on the Connotation, Properties and Research Trend of Enterprise Social Media. in The 17th Wuhan International Conference on E-Business(WHICEB2018). Wuhan, China. 432-439
- [2] Bao X., Bouthillier F.(2013). Information Sharing: As a Type of Information Behavior. in Proceedings of the Annual Conference of CAIS / Actes Du congrès Annuel De l'ACSI
- [3] Martini A., Massa S., Testa S.(2013). The firm, the platform and the customer: A “double mangle” interpretation of social media for innovation. *Information and Organization*, 23(2): 198-213
- [4] Townsend C., Kahn B. E.(2014). The “Visual Preference Heuristic”: The Influence of Visual versus Verbal Depiction on Assortment Processing, Perceived Variety, and Choice Overload. *Journal of Consumer Research*, 40(5): 993-1015
- [5] Wang L., Fussell S. R.(2020). More Than a Click: Exploring College Students' Decision-Making Processes in Online News Sharing. in Proceedings of the ACM on Human-Computer Interaction. New York, USA. 1-20
- [6] Petronio S., Durham W. T.(2015). *Communication privacy management theory*: John Wiley & Sons, Inc.
- [7] Riemer K., Stieglitz S., Meske C.(2015). From top to bottom: Investigating the changing role of hierarchy in enterprise social networks. *Business & Information Systems Engineering*, 57(3): 197-212
- [8] Liu Q.(2019). The Influencing Mechanism of Information Sharing Level and Inter-Firm Value Co-Creation: An Empirical Study Based on Synergy Effect. *American Journal of Industrial and Business Management*, 9(4): 821-833
- [9] Rode H.(2016). To share or not to share: The effects of extrinsic and intrinsic motivations on knowledge-sharing in enterprise social media platforms. *Journal of Information Technology*, 31(2): 152-165
- [10] Kügler M., et al.(2015). Connect me! Antecedents and impact of social connectedness in enterprise social software. *Business & Information Systems Engineering*, 57(3): 181-196
- [11] Ludwig S., et al.(2014). Take their word for it: The symbolic role of linguistic style matches in user communities. *MIS Quarterly*, 38(4): 1201-1217
- [12] Xu Z., Yang Q.(2012). Analyzing user retweet behavior on twitter. in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Istanbul, Turkey. 46-50
- [13] Petronio S.(1991). Communication Boundary Management: A Theoretical Model of Managing Disclosure of Private Information Between Marital Couples. *Communication Theory*, 1(4): 311-335
- [14] Litt E.(2013). Understanding social network site users' privacy tool use. *Computers in Human Behavior*, 29(4): 1649-1656
- [15] Moser K. S.(2017). The influence of feedback and expert status in knowledge sharing dilemmas. *Applied Psychology*, 66(4): 674-709
- [16] Li H., Sarathy R., Xu H.(2010). Understanding situational online information disclosure as a privacy calculus. *Data Processor for Better Business Education*, 51(1): 62-71
- [17] Chakraborty R., Vishik C., Rao H. R.(2013). Privacy preserving actions of older adults on social media: Exploring the behavior of opting out of information sharing. *Decision Support Systems*, 55(4): 948-956

- [18] Young A., Quan-Haase A.(2009). Information revelation and internet privacy concerns on social network sites: A case study of facebook. in Proceedings of the Fourth International Conference on Communities and Technologies. PA, USA. 265-274
- [19] Fiesler C., et al.(2017). What (or who) is public? Privacy settings and social media content sharing. in Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. Portland, USA. 567-580
- [20] Adomavicius D., Tuzhilin A.(2005). Personalization Technologies: A Process-Oriented Perspective. Communications of the ACM, 48(10): 83-90

Full Research Paper

The Benefit Coordination Mechanism of Cloud Platforms for Rural Cultural Tourism Based on a Case Analysis of TikTok, AAuto, and Bilibili

Yueying Li¹, Guoyin Jiang^{1,2,*}, Wanqiang Yang¹

1.School of Public Administration, University of Electronic Science and Technology of China, China

2.School of Management and Economics, University of Electronic Science and Technology of China, China

Abstract: The benefit coordination mechanism of cloud platforms directly affects platform enterprises' competitiveness and plays a vital role in enterprises' sustainable development. In this study, we design a benefit coordination mechanism of cloud platforms for rural cultural tourism by summing-up the two dimensions of the proportion of online and offline user time and the degree of commercialization of TikTok, AAuto, and Bilibili. Research shows that cloud platforms' four benefits are platform traffic, advertising penetration, community ecology, and commercial monetization capabilities. Their interaction is the key to benefitting from cloud platforms. The provided benefit mechanism of the cloud platform for rural cultural tourism can provide theoretical reference and decision-making support for platform enterprises.

Keywords: cloud platform, benefit coordination mechanism, scale effect life, cycle theory

1. INTRODUCTION

In China, the rural cultural tourism industry has become a powerful driver to improve agricultural economic interests, increase farmers' income, and inherit the spirit of rural culture. In the context of digitalization, the cultural tourism industry has continued to promote the quality and upgrade of cultural tourism consumption, the short video cloud platform that integrates content production and brand effect has rapidly emerged, and the substantial use of cloud platform has made the operation of cultural tourism industry more intelligent. From the perspective of information production, the popularity of short videos and live broadcasts has lowered the threshold for content production, especially for people in the sinking market. Thus, users have more opportunities to express themselves, and more users use cloud platforms to access the content. From the perspective of information consumption, a short video is inexpensive and readily accepted and created by everyone^[1]. Online traffic monetization promotes the further strengthening of the commercial monetization capabilities of online and offline communities. The use of cloud platforms seems to increase the benefits of major stakeholders. How their benefits are coordinated and how cloud platforms' sustainable development are promoted are rarely answered in the literature. This study uses dimensions to fill these gaps: the proportion of user time and growth rate and the degree of online and offline commercialization of cloud platforms, the design of the benefit mechanism, and the improved profitability in the rural cultural tourism sector. For early-scale TikTok, AAuto, and Bilibili comparison and analysis, this study uses industries as case studies to investigate the platforms' benefit coordination between the commercialization of public and private domain traffic monetization, advertising penetration, and community ecological cycle domains. This study aims to reduce costs, improve profits and liquidity, and promote the arrival of the era of prosperity of the overall commercialization of the cloud platform of rural cultural tourism.

* Corresponding author. Email: jiangguoyin@uestc.edu.cn (Guoyin Jiang)

2. LITERATURE REVIEW

The use of advanced digital technologies has become increasingly valuable in the business sector. The rural cultural tourism industry is no exception. In particular, the use of cloud platforms has accelerated the process of smart tourism in the country. Digital marketing tools, such as cloud live streaming and cloud viewing, allow consumers to understand deeply, explore destinations, and gain access to multiple unique experiences, such as travel stories^[2]. The rural cultural tourism industry is creating business value with the “new infrastructure,” constantly reforming the management of the rural culture and tourism cloud platform^[3], optimizing the service process, innovating the product supply, and promoting the optimization and upgrading of the industrial structure. Currently, China has become the world’s largest tourism market. The rural cultural tourism industry has great potential for economic development. They have become critical in promoting the upgrading of residents’ consumption, implementing the rural revitalization strategy, and supporting high-quality development in the new era^[4]. The cloud platform is one of the services provided by cloud providers to cloud users, giving consumers the ability to deploy cloud infrastructure applications created or acquired using programming languages, libraries, services, and tools supported by the providers^[5]. They have significant advantages, such as reducing the burden of the relevant beneficiaries’ complex life cycle^[6], reducing deployment and processing times, and improving communication and collaboration among decision makers^[7]. It promotes the accessibility of the cloud platform and reduces costs^[8]. From a subjective perspective, people have changed their values and are more in pursuit of spiritual satisfaction^[9]; relevant APPs emerge at the right moment, such as the cloud platforms TikTok, AAUTO, and Bilibili. Users can conduct real-time online interaction on cloud platforms and can “travel everywhere” without leaving home. From an objective point of view, given the COVID-19 epidemic, the development of the rural cultural tourism industry puts forward new requirements for constructing cloud platforms^[10], which require accurate algorithms and exclusive cloud service modules^[11].

“Rural cultural tourism” originated in France in 1885. The significant difference between it and ordinary tourism lies in the regional characteristics of the “countryside”. The prosperity of cities after China’s reform and opening up and the acceleration of urbanization have made “rural cultural tourism” in the ascendant. At present, the development of rural cultural tourism has begun to take shape and shows a trend of diversified development. Since the 1990s, China’s rural cultural tourism has started to develop rapidly. In the 21st century, rural cultural tourism has entered a period of comprehensive development. With the increase in tourist attractions, the expansion of scale, the expansion of functions, and the expansion of distribution, a new trend of vigorous development emerges^[12]. Relevant scholars have paid attention to China’s future rural cultural tourism, and the research direction is the development and application of cultural tourism platforms. Although much attention has been paid to the rural cultural tourism industry and cloud platforms in theory, in-depth discussion on the combination of the two lacks in some research fields. Most scholars still based on the supply chain system and explored the impact on the cost of benefit^[13]. In practice, although relevant cloud platforms continue to emerge, problems such as conflicts between stakeholders because of unfair returns^[14], imperfect benefit systems^[15], and short platform life cycles have emerged^[16]. Therefore, based on the literature review, this study selects the rural cultural tourism industry’s cloud platform, which improves its benefits and has greater influence to find its core benefit factors and lay the foundation for the future development of relevant cloud platforms.

3. RESEARCH METHODS

The present analysis draws from a sample of three cloud platforms, which have been scaled in rural cultural tourism in 2018–2020 based primarily on the following criteria: a) Data sources: the official websites of the three

cloud platforms, the China Tourism Research Institute, and tourism industry reports. Qualitative data were collected, and a total of more than 80 attributes, including benefits, rural cultural tourism, and cloud-service-related information, were collected and collated by extracting keywords. b) Research method: Two researchers examined each platform independently and used content analysis to summarize the data. A large amount of information about the platforms was consulted with the emergence of proper nouns to analyze them more effectively. c) Reasons for sample selection: First, the products and services of TikTok, AAuto, and Bilibili, which are the three typical cloud platforms with economies of scale, have a flywheel effect. They all experience the following evolutionary trends: the explosion of the industry, the influx of players, the industry's intensification, the exit of long-tail competitors or annexation by other cloud platforms, the increase in market share of cloud platforms, and the scale effect. Second, they have high fixed costs but low variable costs^[17], lowering the threshold for relevant stakeholders to join the platform and creating conditions for increased traffic and advertisers to join. Finally, they have active and quality communities. Key user-side data are one of the key points to increase benefits. Their preliminary development is shown in the following table.

Table 1. The development history of the rural cultural tourism sector.

Year	TikTok	AAuto	Bilibili
2018		A poverty alleviation office was set up to explore a new model of "short video, live broadcast + poverty alleviation."	
2019	National tourism content recording platform "700 million +" annual total number of videos related to the tourism industry "38.4 billion +" annual thumbs up on videos related to the culture and tourism industry "2.6 billion +" annual comments on videos related to the tourism industry "800 million +" annual sharing times related to the cultural tourism industry	The "Ten Thousand Villages Anchor Training Plan" is launched. The total benefit of rural cultural tourism users is more than 20 billion yuan. The number of rural cultural and tourism board users is more than 5 million.	The tourism plate has a scale at the beginning. The number of UP owners in the tourism sector continues to rise. More than 50 million videos are on rural tourism.
2020	POI videos as a new way to remember travel "700 million +" card tourism destination "1 trillion +" clocking in video playback volume Rural cultural tourist destinations have become major clocking destinations.	The revitalization of rural cultural tourism is promoted. More than 50 million e-commerce transaction orders for farmers, rural areas and farmers; 650 million + daily playback of short videos; 2.2 million + daily live viewing hours	"Cloud" rural tourism has become a hot spot. UPs help the development of rural cultural tourism and promote rural live broadcasting with goods. The rural cultural tourism sector still needs to be further developed.

Source: Questmobile¹, 199IT², official websites and news reports of three cloud platforms³

¹ <https://www.questmobile.com.cn/research/report-new/128>

² <http://www.199it.com/archives/993865.html>

³ <https://e.kuaishou.com/#/e/report/949>

This study provides a dynamic study of the sustainability and operation of cloud platforms' benefits, using economies of scale and life-cycle theory to assist in the case analysis. Economies of scale can drive the cloud platform stakeholders to obtain the "long tail" benefits from the distribution curve, reducing the risk of diminishing returns to the traditional supply chain model. The cloud platform's benefits interact and influence each other in the introduction stage, growth stage, maturity stage, and virtuous cycle stage. The accumulation of traffic, the penetration of advertising, the ecological cycle of community, and commercial capability realization are a gradual process. The purpose of detailing each stage's benefits is to make the concepts more concrete and turn them into feasible practices^[18].

4. RESEARCH RESULTS

4.1 Traffic realization of cloud platform

In recent years, cloud platforms have been equipped with infrastructure based on public and private domain traffic monetization, such as e-commerce, advertising, brand entry, and promotion. The cloud platform's benefit source has transformed from the initial PGC-based benefit model to the UGC-based model of content platform traffic benefit. The combination of easy-to-produce videos and the recommendation logic of information algorithms have greatly increased the UGC duration and the number of users. As of June 2020, the scale of deduplication users of the three cloud platforms is shown in the following figure.

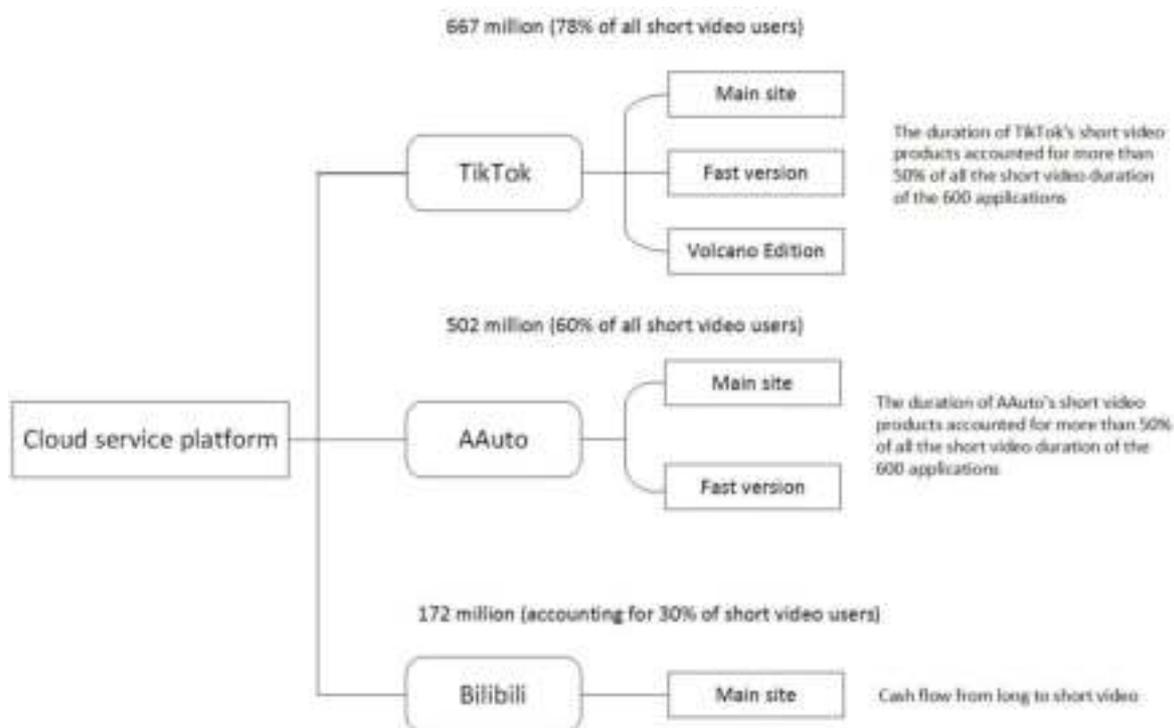


Figure 1. User size of the three cloud platforms.

As shown in the figure, we can conclude that, with the upgrading of information technology, TikTok, AAuto, and Bilibili have appeared in rich and diversified forms to meet multi-faceted information consumers' needs. The development of Internet content creation media shows a changing trend from long to short video traffic.

Furthermore, our research found that the three cloud platforms' benefit level is directly proportional to the traffic. The larger the user scales and the longer the use time, the more profitable the cloud platforms can be. In terms of content consumption, the user growth rate of the three UGC content cloud platforms, TikTok, AAuto, and Bilibili, is faster than the growth rate of the entire network.

Table 2. MAU and DAU of cloud platform.

Index \ Type	TikTok	AAuto	Bilibili
MAU (million)	527	447	119
yoy	14.0%	36.7%	23.9%
MAU (million)	296	211	35
yoy	26.1%	18.9%	41.5%
DAU/MAU%	56.2%	47.2%	28.9%
Per capita usage time per day (min)	93	87	77
yoy	40.4%	17.7%	15.9%
Per capita daily usage times (time)	14.2	19.8	11.2
yoy	36.6%	5.6%	11.7%

Source: Official websites and news reports of three cloud platforms⁴

Based on the above research of QuestMobile, China's mobile Internet has a net increase of 7.96 million monthly active users in the first nine months of 2020, surpassing the 7.04 million annual net increase in 2019. The total duration of TikTok, AAuto, and Bilibili increased by 98%, 73%, and 100%, respectively, which are all higher than the average market.

The ecological cycle of flow realization (Figure 2):

- The three cloud platforms paid more attention to private domain traffic value in the initial stage and were more monetized based on talents. The purpose was to reduce platform costs and increase user scale.
- With the improvement on the infrastructure of cloud platforms and the enhancement of information technology^[19], TikTok, AAuto and Bilibili opened up more public domain traffic.
- The deep integration of a large number of anchors and advertising benefits has led to the formation of an ecologically healthy development.

4.2 Monetization of advertising penetration

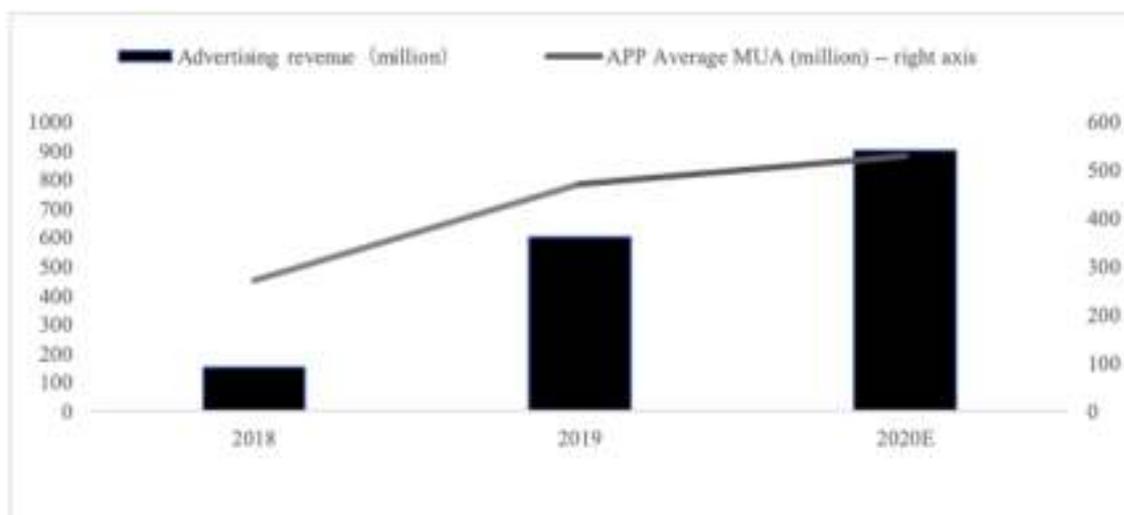
A typical representative of cloud platforms' monetization method is advertising, whose major forms are the selling of resources by cloud platforms to advertisers. TikTok, AAuto, and Bilibili use the fan base and the video content production capabilities of talents and anchors to help advertising companies and brand owners in content marketing and promotion. Thus, the rentals of the beneficiaries associated with being stationed in the cloud platform are increased, and the talents and anchors can attract traffic in exchange for commissions, further making the ads and brands more popular to achieve a win-win situation of benefits.



Figure 2. Ecological cycle of flow realization

⁴ <https://www.pmtemple.com/academy/2312/>

4.2.1 TikTok has a high advertising penetration rate and maintains rapid benefit growth.



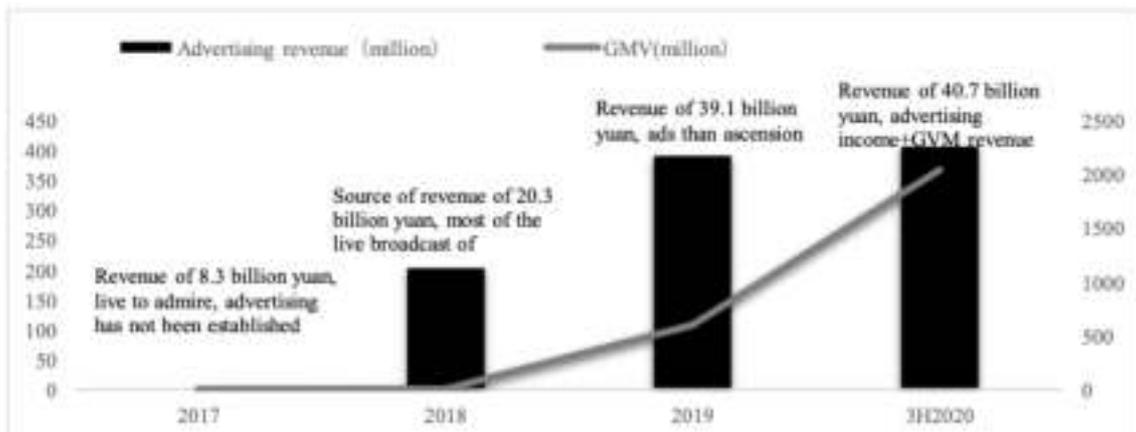
Source: The annual ecological report of TikTok's official website⁵

Figure 3. Advertising benefit of TikTok.

TikTok's advertising benefit continues to rise in 2018, 2019, and 2020 at 15 billion, 60 billion (YOY +300%), and 90 billion yuan (YOY +50%), respectively. TikTok began to lay out its advertising system in 2017. From the fourth quarter of 2020, the layout of TikTok's brand advertising, information flow effect advertising, and search advertising has been completed. TikTok fully rolled out on Blue V2.0, carrying out super challenges, setting up screens, recruiting fixed-space ads, putting up brand areas, and designing shopping carts. Its advertising benefit has maintained rapid growth. In 2020, Bytedance increased its investment, the online shopping cart function became an essential strategy for attracting advertising and traffic benefit, TikTok cut off external third-party links, and product purchases can only be made from TikTok stores. According to official data on the website, in 2020, the overall GMV of TikTok e-commerce companies increased by 12 times, the GMV of TikTok stores increased by 44.9 times, and the number of businesses that opened stores increased by 17.3 times. The increase in merchants and the deepening of advertising penetration have continuously reduced monetization unit cost, and the cloud platform has formed a considerable profit dividend.

⁵ <https://open.douyin.com/platform>

4.2.2 AAuto opens public domain traffic, and advertising benefit continues to increase.



Source: AAuto's marketing platform and annual report⁶

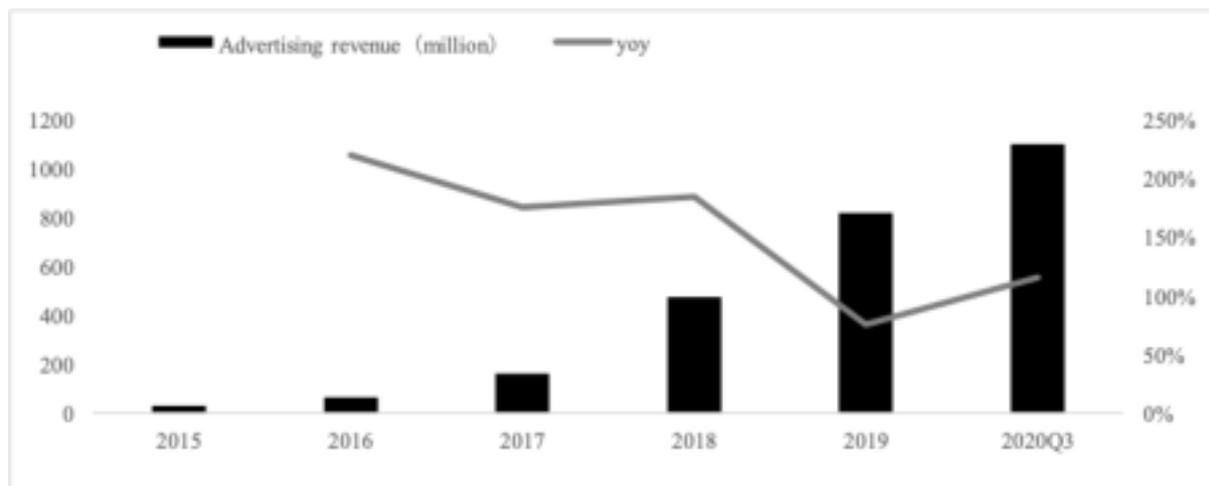
Figure 4. Advertising benefit of AAuto.

Before 2018, AAuto had most of its benefits from live streaming rewards. It paid much attention to private domain traffic while public domain traffic was weak. The cloud platform upholds the concept of neutrality and non-disturbing user values. Most of these content distributions are given to big data algorithms and have less controllable and operable traffic.

From 2019 to 2020, AAuto launched information flow advertising based on commercial technology and continuously upgraded its advertising system, opening up more public domain traffic for commercialization. Its advertising benefits have grown rapidly. Benefits were from 1.7 billion yuan in 2018, 7.4 billion yuan in 2019, and 40.7 billion yuan in the first three quarters of 2020.

AAuto improved the ecology of anchors and invested more in its cloud platform's infrastructure, such as standardizing the management of anchors and unblocking advertising promotion channels.

4.2.3 Bilibili transformation and upgrading, advertising benefit play a new pattern.



Source: QuestMobile and Bilibili's annual report⁷

Figure 5. Advertising benefit of Bilibili.

⁶ <https://e.kuaishou.com/#/e/report/952>

⁷ <https://www.bilibili.com/read/cv9100644/>

Bilibili's main product is a long video that emphasizes community attributes and user experience, but its weakness is that advertising is inefficient. Thus, in 2016, Bilibili held an advertising solicitation meeting, proposing three advertising forms within Bilibili, namely, regular hard ads, infomercials, and UP owner-created video ads. In 2017–2019, the AD TALK advertising and marketing promotion campaign fully integrated resources. Its advertising benefits grew from 61 million in 2016 to 1.12 billion yuan in the first three quarters of 2020 (yoy+112%), and the turning point for its rapid benefit growth was the recommendation algorithm for video length realizations (Efficiency of realizing long videos = consumption of production costs per unit of time / value gained by users). Consequently, Bilibili increased the promotion of short videos in 2020, creating an efficient traffic attraction effect. This practice has achieved the UP owners and increased the overall benefits of Bilibili.

4.3 Community ecological circulation

With the upgrading of information technology, the spiritual world of users is constantly enriched. Given the stratification of people's interests, various community cloud platforms need more personalized expression. The essence of community is to co-create, distribute, and socialize themes based on personal interests and promote interaction between users and cloud platforms. TikTok, AAuto, and Bilibili have launched creator support programs, incentivizing creator content production through cash, exposure of traffic, and protection of originality and thus increasing the benefits of cloud platforms.

Table 3. Community growth of cloud platforms.

Year \ Tape	TikTok	AAuto	Bilibili
2018			Create incentive plan
2019	DOU Zhi plan; VLOG billion traffic support plan, and TikTok game star group plan; Creator growth plan, live broadcast of "Dark Horse Plan," TikTok live broadcast of the creator conference, Online "Creator Academy"	Photosynthetic plan, Million game creator support plan, Education ecology partner plan, Gorgon plan, AAuto Yinyuetai music anchor support policy, Kuaizhi plan, 2020 Shuochengren training plan, ACG photosynthetic creator conference	Creation incentive rookie award, Vlog star project; The fourth rising star project, Small universe rising star project, The fifth rising star project, Music star project, Curiosity project, UP master activation project
2020	DOU growth incubation camp; TikTok Musicians 100 Million yuan subsidy plan, Educational live IP activity "Teacher Please Come on Stage," and 2020 TikTok creators conference	Original preservation program; Original music billion yuan incentive plan, Starfish plan, Cultural tourism photosynthetic plan, Virtual idol, "AVI alliance plan"	Dance treasure hunt plan, Dance talent recruitment plan, 360 line recruitment plan

Source: Official websites⁸ and news reports of three cloud platforms⁹

The above table shows different stages of community growth. In the first stage, the community needs to determine the theme, algorithm, direction, and other basic work to ensure the community's basic benefits and survival. In the second stage, the cloud platform needs to explore community content, cultivate community culture, expand the audience's scope, and increase the cloud platform's benefit. In the third stage, the cloud platform needs

⁸ <https://union.bytedance.com/open/>

⁹ <https://ir.bilibili.com>

to coordinate flow allocation and benefit coordination to realize the community's self-evolution and development based on the large enough creators, consumer groups, and capital inventory of the community.

The richer form of content on the cloud platform attracts creators to flock to the community, forming an economy of scale effect. Users are willing to spend to exchange different services and products and continuously meet users' demand for content consumption, while the cloud platform also gains benefits^[20].

4.4 The highlighted value of commercialization

Commercialization is an integral part of maintaining sustainable benefit growth of cloud platforms, and its core principle is to provide the corresponding commercial realization path according to the highest realization efficiency^[19]. The entry into commercialization stages requires the support of public and private domain traffic, advertising penetration rate, and community growth degree. The platform's commercial benefits will only grow exponentially after its traffic reaches a certain level. For example, when the traffic of a cloud platform is small, advertisers are only willing to put good ads; when the traffic of the cloud platform reaches a specific scale, advertisers will invest more money and increase the budget according to the considerable influence and communication ability generated by the cloud platform. As the community grows, the brand will tap the benefit potential for the cloud platform, promoting commercialization benefits and forming the closed loop of benefits of the cloud platform^[17].

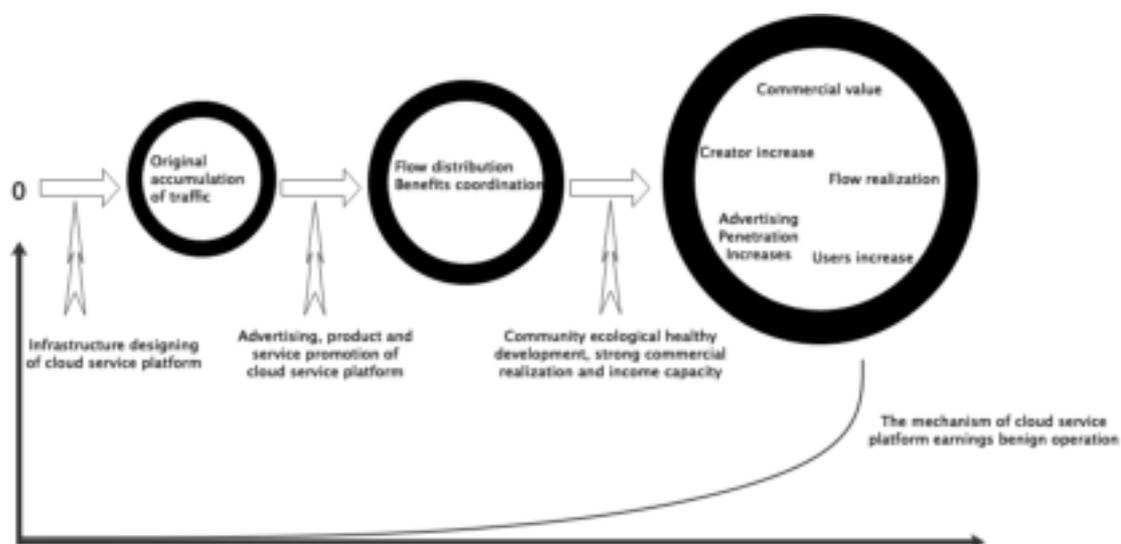


Figure 6. Benefit closed loop of the cloud platform.

Users of TikTok, AAuto, and Bilibili have stabilized at the level of hundreds of millions of MAU. The advertising penetration rate is high, and the ecological situation of the community is improving, entering the accelerated stage of commercialization.

TikTok has started its commercial business layout in the first quarter of 2018, launching shopping carts and advertising functions of millions of community goods alliances. In 2019, it launched the Community Goods Alliance and reached agreements with platforms such as JD and Vipshop. From August 2020, the cloud platform has cut off external links to products. All video and live broadcast products can only come from TikTok stores. The platform charges a 20% service fee, and TikTok stores charge a 5% service fee. According to the official data released by TikTok from January to November 2020, the overall GMV of TikTok e-commerce has increased by 11 times, the GMV of TikTok stores has increased by 44.9 times, and the number of stores opening stores has increased by 17.3 times.

In 2017, AAUTO began to build commercial functions, and live broadcasters spontaneously formed a phenomenon of bringing goods and rewards to become the main source of AAUTO's benefit. In 2018, the shopping cart function was started, and AAUTO stores were launched to achieve GMV 100 million in benefit. In 2019–2020, AAUTO developed rapidly, forming a commercial ecosystem, training and supporting mid-waist anchors. The effect of traffic benefits generation is obvious from GMV 60 billion in 2019 to over GMV 100 billion in the third quarter of 2020.

Bilibili opened the UP main commercialization functions test of 2018 and launched the “Reward Program” public test, allowing UP owners to earn benefits by placing advertisements or product links under the video. Bilibili launched the “Takeoff Plan” in May 2020. UP owners purchased FeiCoin and recommended content manuscripts and live broadcast rooms to the appropriate community, achieving accurate work recommendation and drainage effects. In July 2020, it launched the online “Huahuo Platform” to connect with owners and advertisers. Advertisers can send invitation orders to the UP hosts of the cloud platform and pay them commissions. The cloud platform only charges 5% service fees, realizing a win-win situation.

5. CONCLUSIONS

According to the case analysis of the cloud platforms TikTok, AAUTO, and Bilibili, we can summarize the design of the benefit mechanism of the cloud platform of this study as “1+2+3+4+⊕.” Through the data and content analysis of the three cloud platforms, this study designs a reasonable benefit coordination mechanism for rural cultural tourism cloud platforms by summarizing two dimensions of online and offline user time ratio and commercialization degree with the help of economy of scale and life cycle theory. The findings show that the cloud platform's four main advantages are platform traffic, advertising penetration, community ecology, and commercial monetization ability. The interaction between them is the key to influence the benefits of the cloud platform. The ultimate goal is to provide theoretical and decision-making references for the sustainable development of rural cultural tourism cloud platforms.

This study reveals two aspects of significance. In terms of theory, our study enriches the connotation of life cycles theory and economic scale theory and focuses on designing a suitable benefits mechanism for the research object of the context of the times. It also gives new momentum to China's rural revitalization and provides services for future policy formulation. On the practical side, it addresses China's main contradictions, forms a gain mechanism of cloud platform with Chinese characteristics, and provides measures and guidance for the guarantee of gain and the specific practice of gain for relevant interest subjects.

6. FUTURE PROSPECTS

The design of the benefit coordination mechanism of the rural cultural tourism cloud platforms should be adjusted and upgraded according to economic development trends and technological changes. It must have the concept of coexistence of risks and benefits and formulate a benefit mechanism from the perspective of long-term development. Great attention should be paid to the formulation of anti-monopoly and anti-unfair competition policies in the superstructure to form a strong regulatory environment, build a good atmosphere in the economic market, and avoid vicious competition among cloud platforms for rural cultural tourism.

ACKNOWLEDGEMENT

This work was partially supported by a grant from the National Key Research and Development Program of China (No. 2019YFB1405600), and the National Natural Science Foundation of China (Nos.

72071031,71671060).

REFERENCE

- [1] Chatterjee, D., Grewal, R., & Sambamurthy, V. (2002), Shaping up for E-Commerce: Institutional Enablers of the Organizational Assimilation of Web Technologies. *MIS quarterly*, 26(2): 65-89
- [2] Zysman, J., & Kenney, M. (2018), The next phase in the digital revolution: intelligent tools, platforms, growth, employment. *Commun. ACM*, 61(2): 54–63
- [3] Li, H., Shen, Q., & Bart, Y. (2018), Local Market Characteristics and Online-to-Offline Commerce: An Empirical Analysis of Groupon. *Management Science*, 64(4): 1860-1878
- [4] Chen, W., Gu, B., Ye, Q., & Zhu, K. X. (2019), Measuring and Managing the Externality of Managerial Responses to Online Customer Reviews. *Information systems research*, 30(1): 81-96
- [5] Keskin, T., & Taskin, N. (2015), Strategic Pricing of Horizontally Differentiated Services with Switching Costs: A Pricing Model for Cloud Computing. *International journal of electronic commerce*, 19(3): 34-53
- [6] Liu, X., Zhou, Y.-W., Shen, Y., Ge, C., & Jiang, J. (2021), Zooming in the impacts of merchants' participation in transformation from online flash sale to mixed sale e-commerce platform. *Information & Management*, 58(2): 103409
- [7] Wang, S., Cavusoglu, H., & Deng, Z. (2016), Early mover advantage in e-commerce platforms with low entry barriers: The role of customer relationship management capabilities. *Information & Management*, 53(2): 197-206
- [8] Guo, Z., Li, J., & Ramesh, R. (2019), Optimal Management of Virtual Infrastructures Under Flexible Cloud Service Agreements. *Information systems research*, 30(4): 1424-1446
- [9] Bhattacharyya, S., & Bose, I. (2020), S-commerce: Influence of Facebook likes on purchases and recommendations on a linked e-commerce site. *Decision support systems*, 138: 113383
- [10] Ma, D., & Seidmann, A. (2015), Analyzing Software as a Service with Per-Transaction Charges. *Information systems research*, 26(2): 360-378
- [11] Zhang, Z., Nan, G., & Tan, Y. (2020), Cloud Services vs. On-Premises Software: Competition Under Security Risk and Product Customization. *Information systems research*, 31(3): 848-864
- [12] He, J., Liu, H., & Xiong, H. (2016), SocoTraveler: Travel-package recommendations leveraging social influence of different relationship types. *Information & Management*, 53(8): 934-950
- [13] Ghoshal, A., Hao, J., Menon, S., & Sarkar, S. (2020), Hiding Sensitive Information when Sharing Distributed Transactional Data. *Information systems research*, 31(2): 473-490
- [14] Tian, F., Susic, G., & Debo, L. (2019), Manufacturers' Competition and Cooperation in Sustainability: Stable Recycling Alliances. *Management Science*, 65(10): 4733-4753
- [15] Ray, J., Menon, S., & Mookerjee, V. (2020), Bargaining over Data: When Does Making the Buyer More Informed Help? *Information systems research*, 31(1): 1-15
- [16] Setia, P., Bayus, B. L., & Rajagopalan, B. (2020), THE TAKEOFF OF OPEN SOURCE SOFTWARE: A SIGNALING PERSPECTIVE BASED ON COMMUNITY ACTIVITIES. *MIS quarterly*, 44(3): 1439-1458
- [17] Dhanorkara, S. (2019), Environmental Benefits of Internet-Enabled C2C Closed-Loop Supply Chains: A Quasi-Experimental Study of Craigslist. *Management science*, 65(2): 660-680
- [18] Van Looy, A. (2021), A quantitative and qualitative study of the link between business process management and digital innovation. *Information & Management*, 58(2): 103413
- [19] Retana, G. F., Forman, C., Narasimhan, S., Niculescu, M. F., Wu, D. J., Cornell, U., . . . School, I. B. (2018), Technology support and post-adoption IT service use: Evidence from the cloud. *MIS quarterly*, 42(3): 961-978
- [20] Morath, F., & Münster, J. (2018), Online Shopping and Platform Design with Ex Ante Registration Requirements. *Management Science*, 64(1): 360-380

Full Research Paper (Extended Abstract)

How does Knowledge Association of Online Travel Reviews Affect Product Sales? Based on the Knowledge-based Theory and Social Network Theory

Chang Wang¹, Shengyou Liu¹, Ping Wang², Zhiping Hou^{2}, Wei Luo²*

¹Business School, Central South University, Changsha, 410083, China

²Business School, Guilin University of Technology, Guilin, 541004, China

Abstract: To better understand the impact mechanism of online travel reviews, this study aims to investigate how the knowledge association of online travel reviews influences the sales volume of online tourism products from the perspective of knowledge-based theory and social network theory. We crawled online reviews of 314 tourism products from the leading OTA of China and constructed knowledge networks of online travel reviews by using the semantic association of thematic words. Our results show that both knowledge association and knowledge heterogeneity have a U-shaped curvilinear relationship with sales volume. Moreover, knowledge heterogeneity partially mediates the relationship between knowledge association and sales volume. In addition, knowledge cohesion has a negative moderating effect on the relationship between knowledge association and knowledge heterogeneity. The findings of our study have important theoretical value and managerial implications for tourism e-commerce research.

Keywords: online travel reviews, semantic association, knowledge network, social network analysis

1. INTRODUCTION

Although online travel reviews have brought convenience to tourists and tourism enterprises, the redundant and complicated information contained in review texts makes buyers and sellers fall into the contradiction of “information loss” and “information overload”. Thus, scholars have focused on how to effectively mine the knowledge value of online travel reviews [1]. Based on knowledge-based theory, online travel reviews are the embodiment of consumer knowledge, and the identification and combination of knowledge are also of great value. The smallest knowledge element of online travel reviews is the thematic word. Each thematic word is a knowledge node in the massive number of review texts, and a large number of thematic words is set to form knowledge. Knowledge extraction of online travel reviews depends on not only the literal meaning of thematic words but also the semantic association logic among thematic words. In the process of knowledge recognition and combination, knowledge network is constructed by the co-occurrence of semantic association between the thematic words. In the semantic association knowledge network of online travel reviews, consumers can obtain valuable information through the visibility of knowledge in the network, and further influence the purchase decision of tourism products.

The visibility of knowledge in the network will affect product sales, and this effect may often be nonlinear [2]. Therefore, to reveal the effect of knowledge networks of online travel reviews on sales volume, this study integrates natural language processing technology, social network theory and knowledge-based theory to explore the following three questions: (1) From the perspective of knowledge nodes, what is the mechanism of the impact of knowledge association on the sales volume of online tourism products? (2) From the perspective of knowledge content, what is the relationship between knowledge heterogeneity and knowledge association, and how does it affect sales volume? (3) From the perspective of knowledge units, what role does knowledge cohesion play in the relationship that knowledge association affects the sales volume of online tourism products through knowledge heterogeneity.

* Corresponding author. Email: hzp@glut.edu.cn (Zhiping Hou), changw1000@163.com (Chang Wang)

2. METHODS

As our data source, we choose Tuniu (<https://www.tuniu.com>), which is a leading OTA platform in China [3]. In the choice of cities, we select 163 cities as tourism destinations; they were all included in the “Charm city with Chinese characteristics 200” list in the 2016-2018 period. And a pre-test was conducted to determine the appropriate review threshold, which was at last set to no less than 416. From July 5 to July 30, 2019, we set up a crawler to obtain the online travel reviews of 370 products at all time points. After data cleaning, we obtained 314 samples of online tourism products.

In this paper, we construct the knowledge network of online travel reviews by using the method of Chinese word segmentation processing and semantic association analysis proposed by Hou et al. (2019). For each listing, first, we extract the thematic words by using the Jieba toolkit and TextRank algorithm. Second, Python and the Natural Language Toolkit (NLTK) are used to calculate the frequency of thematic words and to extract bigram phrases. Then, the knowledge network is constructed through the co-occurrence of thematic words. Finally, we calculate the eigenvector centralization, modularity index and E-I index of the knowledge networks by using the social network analysis software Gephi and Ucinet to test and verify the research model.

3. CONCLUSIONS

Based on social network theory, knowledge-based theory, natural language processing and other multidisciplinary theories and methods, this paper extracts core thematic words from online travel review texts and constructs knowledge networks of online travel reviews through the semantic association analysis of thematic words. The main conclusions are as follows: First, the core thematic words in online travel reviews are the embodiment of consumer knowledge, and the network constructed by knowledge association is the carrier of the acquisition and integration of knowledge resources. Second, this study proves the existence of a U-shaped relationship between knowledge association and sales volume; that is, compared with low-level and high-level knowledge association, sales volume can be low at a medium level of knowledge association. Third, the empirical test results support the U-shaped relationship between knowledge heterogeneity and sales volume and the mediating effect of knowledge heterogeneity between knowledge association and sales volume. Finally, the results support the negative moderating effect of knowledge cohesion between knowledge association and knowledge heterogeneity.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (NSFC No. 71962005, 72064008, 71562009).

REFERENCES

- [1] Hou Z, Cui F, Meng Y, Lian T, Yu C. (2019).Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management*, 74: 276-289.
- [2] Wang C, Rodan S, Fruin M, Xu X. (2014).Knowledge networks, collaboration networks, and exploratory innovation. *Academy of Management Journal*, 57(2): 484-514.
- [3] Zhang C, Zhou Q. (2018).Online investigation of users' attitudes using automatic question answering. *Online Information Review*, 42(3): 419-435.

Full Research Paper**Online Marketing of Cultural Tourism: A Case Study of a 5A Ancient Town***Jiali Li^{1,a,*}, Bangchen Pan^{2,b}*¹School of Business Administration, Zhongnan University of Economics and Law, Wuhan, 430073, China;²Qiannan Normal University for Nationalities, Duyun, 558000, China

Abstract:For Destination Marketing Organizations (DMOs), the use of social media marketing has become their most important marketing tool. This paper discusses the online marketing process of cultural tourism through the thematic analysis method, and reveals the online marketing content of cultural tourism from three stages: pre-tour marketing, in-tour marketing and post-tour marketing. Research shows that the marketing of cultural tourism on social media is mainly through the self-established IP of tourism-related stakeholders and tourists to automatically generate or forward high-quality cultural tourism content, generate certain influence and attract a group of fans, from which trust can be built by interacting with uploaders. The results imply that the high quality cultural tourism content is the key to the success of online marketing, therefore, both the tourism destination organizations and the tourism enterprises should do a good job in the content generation.

Keywords: Cultural tourism experience; Online marketing; Thematic analysis; Tik Tok

1. INTRODUCTION

Rapid developments in the Internet services as well as information and communication technologies (ICT) can dramatically change the manner in which the tourism industry and tourism organisations conduct business [6]. Network and social media analytics are promising avenues for new methods of online marketing and sales [8]. The main advantage of social media data is that it is shared voluntarily and is often freely accessible online, free of legal restrictions on reuse. It is usually available in real time and can be processed at a relatively low cost [12].

Most marketing for cultural attractions have started to use social media in some manner, although social media still tends to be used more for marketing more than for analysis and for input into impact assessments. Still, many cultural tourism destinations have expressed interest in making greater use of social media data for analysis and planning in the future [1]. Moreover, cultural attractions are aggressively using different social media platforms in their future plans [11].

Tourism was the sector hit particularly hard by the Covid-19 crisis, and feedback from cultural tourism experts and the regions surveyed indicated that the development of cultural tourism destinations required more attention [8]. Therefore, the government and tourism managers need to understand, in a timely manner, the relevant information of cultural tourism departments, tourism enterprises, and tourists in order to develop policy tools. All of these make the application of open and big data in cultural tourism management more attractive. User-generated big data from social media, web searches, and website visits constitute another promising source of data for monitoring and planning cultural tourism. The data can be used both to analyse tourist flows and to understand how visitors perceive and use tourist attractions. As big data is generally public, available in real time, and freely available online, the barriers to use of this data source are low in terms of access and legal restrictions. Numerous studies have shown that the quality and reliability of the data approach official tourism statistics, is creating a method to replace expensive tourist surveys through analyzing people's online behaviour at a lower cost with less time [12][6][1]. Therefore, the use of Internet and social media data to monitor and assess

¹ *Corresponding Author Name: Jiali Li; Corresponding Author mail: leeqi16@163.com

travel has become mainstream.

Based on the importance of online data marketing to cultural tourism, this paper takes the successful online marketing case of Qingyan Ancient Town' Da Ming Zhi events on the social media platform -Tik Tok in 2020 as an example to explore the online marketing process of cultural tourism destinations through theme analysis.

2. LITERATURE REVIEW

2.1 Cultural tourism

The World Tourism Organization (UNTWO) offers a broad definition of “cultural tourism”, arguing that “cultural tourism includes all aspects of tourism, from which tourists can learn the history and heritage of others, as well as their contemporary life and thoughts...”^[14]. At the same time, it also provides a narrower definition, that is, “the movement of people for cultural motivation, such as research travel, performing arts, cultural travel, visiting historical sites, nature research, folk and art, religious pilgrimage travel, travel to festivals and other cultural events”^[14]. To realize special cultural feelings, that is, tourists travel for cultural motivation. In this process, tourists observe, feel, and experience foreign or heterogeneous cultures, or have in-depth experience of the cultural connotation of tourism resources, to obtain all-round spiritual and cultural enjoyment^[15].

From a cultural tourism perspective, user-generated data and content are a promising source for collecting macro flow statistics on travel destinations, as well as for understanding (and potentially predicting) flow patterns associated with specific cultural attractions and events. In addition, Kalvet et al^[8] comment that text data posted by visitors as comments on Twitter messages or photos shared on platforms, such as Flickr, can help enterprises analyse people's emotions and engagement with specific points of interest, reveal visitors' interests, and help understand the activities that visitors engage in at different sites. Such information may assist in the development of management and marketing strategies around existing cultural tourist attractions, but can also reveal new interests for new attractions and activities, all of which can be derived much quicker than any official statistic report can grasp the same trends.

2.2 Online experience marketing

At the tourism planning and decision-making stage, the relationship between tourists and representatives of destinations is communicated online and bidirectional through multiple official and unofficial platforms. The interactions that take place on these online platforms (such as blogs, social media, and websites) form the so-called “pre-experience” associated with the destination, which is believed to be the basis for generating tourists' motivation to travel^[3]. Among the official platforms used for online destination promotion, the destination's website and its social media accounts stand out. To be successful and to distinguish themselves from any possible competitors, destination brands must deliver the promise of a memorable destination experience^[7]. Following this line of thought, authors have identified that destinations have a positive impact on visitor satisfaction, willingness to revisit, loyalty, and future word-of-mouth recommendations^{[3][9]}. Recently, however, the development of information and communication technology has enabled tourists to have direct contact with the culture, heritage, and tourism services as well as the scenery of their destinations through multimedia. As such, destination platforms are a key element in providing a destination brand online experience that encourages potential visitors to visit or interact with a destination. Through the means of seeing, listening, using, and participating, experiential marketing can fully stimulate and mobilize consumers' senses, emotions, thinking, actions, associations, and other emotional and rational factors^[5]. Brakus discussed the dimensions of experiential marketing from the perspective of brand cognition and proposed four dimensions of experiential marketing. In his opinion, brand experience can be divided into Sensory Experience, Affective Experience, Intellectual Experience, and Behavioral Experience.

2.3 Social media technology

Social media technologies (SMTs) include a wide range of Internet-based applications, including software designed primarily to support collaborative communication, interaction, and synchronous or asynchronous networks, such as social networking sites, content repositories, video games, and virtual reality. SMTs are considered to be an important channel for tourism branding and marketing ^[9]. Moreover, SMTs constitutes a low-cost technology that can be cheaply maintained through existing free platforms ^{[10][5]}.

Social media platforms are considered to be the most important destination marketing communication tool in the information age. Short video platforms are used for creation and dissemination, which has achieved good social effects and created new social values. Large travel companies aggressively market and promote their travel activities through social media reviews, reading reviews, opinions, photos, and videos (including corporate websites), often capturing a significant share of the local and international tourist market compared to smaller companies that rarely use such platforms ^[4]. As a result, social media technology is indispensable to the online visibility and accessibility of small travel businesses. The continued awareness and visibility of newly launched travel brands depends on the continued aggressive marketing and promotion of these products and services through interactive platforms, such as SMTs. Kalvet ^[8] praised the communication and transactional power of social media, especially their ability to reach new audiences and strengthen the connection between enterprises and their existing customers.

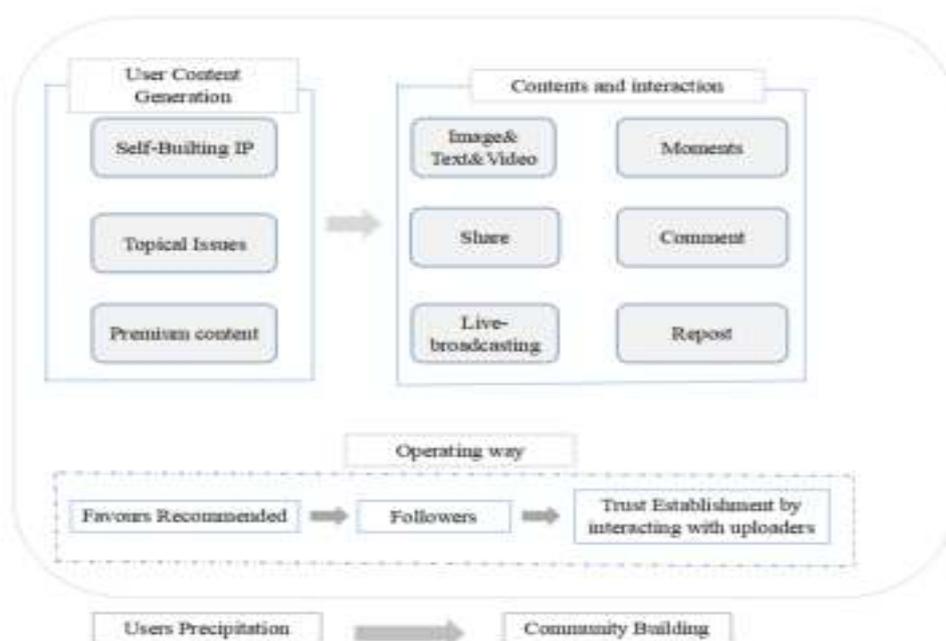


Figure 1. The deep construction process of social media platform content and experience

The social media technology in this study is Tik Tok, which is a social software for short videos of music creativity incubated by Toutiao. The software was launched on September 20, 2016 and is a community platform for all ages. Users can choose images, text and videos, and form their own contents through this software. What's more, users can repost, share and comment on the premium content they like. Under the personalized recommendation mechanism, Tik Tok users can see their own personalized videos in their personalized recommendation pages. This is in line with the user's personal and unique short video watching needs. Users do not need to filter the contents they do not like artificially. The platform and algorithm have replaced users by packaging their social networking, location, work, habits, interests and other data, removing the videos that do not conform to users' taste, and only recommending the videos that users may like to watch.

In order to build Tik Tok into a popular media platform and show its responsible attitude, Tik Tok has been

working hard in government affairs and the media industry since 2018. By inviting and guiding government organs and media organizations to building official IP on Tik Tok, Tik Tok's media platform positioning and content diversification have been realized ^[17].

Now, Tik Tok has become an important method of online marketing for many travel destinations. By August 2020, Tik Tok's daily active users exceeded 600 million, and by December 2020, the average daily video search times of Tik Tok exceeded 400 million. In 2020, Tik Tok will help the online economy, and it helps Hubei Province to restore Jiangsu Province. 42,779 merchants in Hubei Province will have 4.1 billion yuan worth of Tik Tok, Kuaishou, B and other new traffic platforms to build content and experience in the circle by means of depth, realizing the growth of both user scale and usage time, and continuously improving the commercial value. The cultural tourism poverty alleviation plan of Tik Tok, for example, promotes the popularity of characteristic tourist attractions in poor areas through the form of short videos, and makes a large number of previously "unattractive" scenic spots popular, such as Daocheng Yading in Garze Prefecture and Hefeng County in Enshi City, Hubei Province, which promotes the rapid development of local economy ^[17].

Through the study, it is found that Tik Tok has such a large data influence mainly due to its following characteristics:

- **Content experience marketing**

Authoritative user content production, increasing trust mechanism; Content diversity enhances the user experience.

Tik Tok has been polishing its content products since its launch. After a year, Tik Tok greatly increased its resources. Once it was put into the market, it was quickly upgraded to Toutiao strategic product, which directed all kinds of stars' promotion resources to Tik Tok. At the same time, Toutiao's core algorithm advantage is used to add an algorithm recommendation model to Tik Tok products to ensure the efficiency of content distribution. In addition, seven museum, the Beijing municipal public security bureau of anti-terrorism, special police corps and the communist youth league central institutions, as well as the central enterprises group to join, add a lot to trill platform is ability, authoritative content, in entertainment, funny, beautiful singing and dancing show as the main content of short video field to bring a clean, under the guidance of the government, many spread the content of socialist core values began to pop up on the short video platform.

- **LINK marketing**

Link marketing is a marketing model that matches the original content of Tik Tok with the marketing demands of customers. There are two modes. The first mode is conventional link, which reaches target groups based on a reasonable match between Tik Tok's native video traffic and brand marketing appeals. Tik Tok also makes use of the advantages of big data algorithms. It sends everyone what each individual user is interested in and wants to see, presenting the effect of "echo wall" and strengthening users' attention to the platform through displaying their own interest points to them. The other is function link, based on brand marketing preferences, to provide its own video content (talent video, star video, etc.) to add components, driving the transformation ^[17].

- **Broadcast content is extensive, short, practical and interesting**

Tik Tok's use of live streaming is more conducive to improving user engagement and exploring commercial boundaries (advertising, delivery, e-commerce and games). Tik Tok's live streaming has already broken through the single dimension, including more diversified and rich content. The current live streaming content on Tik Tok includes live teaching of makeup and musical instrument performance. More local folk customs are able to be shown live, online classes power closed is not suspended, while videos can be used to help sell goods which can support farmers, for example, as well as public welfare, the arts, humanities, education, and nature. In the face of live their own iteration, can grasp the product dynamic keen, influential platform can truly leverage amplification products value. The expansion of different content dimensions levered

the social promotion effect of Tik Tok and live broadcasting. The Greenpeace South Pole Live is another attempt by Tik Tok to break the content boundary on the basis of “Live Plus”. The power of Tik Tok lies in its “short, flat” entertainment that is satisfying and fills up a lot of fragmented user time ^[17].

3. METHODS

The study data analyzed by qualitative content analysis included actual interviews with 26 offline tourists and 586 comments from Tik Tok platform. The questions in the offline interview include:

- Where did you get the information of “Da Ming Zhi” Cultural Activities?
- How do you feel if you get information from social platforms?
- Did you share it on social media when you participated in the “Da ming zhi ”? Why share?
- What's your feeling after attending the “Da Ming Zhi”?

According to Schreier ^[13], thematic analysis is data or text driven and aims to identify hidden patterns. So an inductive thematic analysis method used. In the analysis phase, researchers search for differences and similarities in data collected through online and offline surveys, and then describe them at different levels of abstraction in the topic. Therefore, a study can achieve theoretical understanding from data. In other words, researchers can convert concrete textual information into abstract general concepts or knowledge. Thematic analysis has been used frequently in tourism research ^{[11][16]}. Therefore, we hope to understand the process of online marketing of cultural tourism destinations through the method of theme analysis.

We generated initial codes and subject classification for the data. These included “content experience”, “culture”, and “stakeholders”. Finally, according to the different stages of the data, the presentation states of different themes are divided into three stages: pre-tour marketing, in-tour marketing, and post-tour marketing.

4. RESULTS

Thematic analysis reveals the online marketing process of the cultural tourism experience of Qingyan Ancient Town on social media, specifically Tik Tok.

4.1 Brief introduction of the case

Qingyan Ancient Town is located in the southern suburb of Guiyang City, Guizhou Province, China. It is the first national 5A tourist attraction in Guiyang City, National AAAAA (5A) scenic spots are scenic spots classified by the quality of tourist attractions in China. There are five levels, and 5A is the highest level of China’s tourist attractions, representing the country’s world-class boutique tourist attractions ^[376]. Qingyan has a history of more than 600 years, with profound cultural and historical background and charming regional characteristics. It was first built in the 11th year of Hongwu in the Ming Dynasty (1378 AD). It is a mountain military city evolved from the military city defence and is known as the “Southern Gate” of Guiyang. The “Da Ming Zhi” event is a series of immersive activities of ancient Chinese culture held by Qingyan Ancient Town to attract tourists, including a series of performances of Generals’ orders, Etiquette, Mid-Autumn Festival reunion and so on. During the cultural activities of “Da Ming Zhi”, Spread through the Tik Tok platform, only from 18:00 to 21:00 in August, 2020, the number of visitors reached 155,238, an increase of 52.1% compared with July, which brought a large amount of direct income for the cultural and tourism industry of the scenic spot. Especially during the National Day period of “Da Ming Zhi· Etiquette Forever”, Qingyan Ancient Town Scenic Spot received 235,500 tourists and achieved a comprehensive income of 28.2297 million yuan ^[20].

4.2 Online marketing process

The online marketing process of Qingyan Ancient Town’s cultural tourism experience activities on Tik Tok platform is mainly divided into three stages.

4.2.1 Pre-tour marketing

Culture setting: in the early stage of the marketing, based on the analysis of the scenic spot in the research, through culture mining, made for the construction of the ancient city of the Ming dynasty, Qingyan ancient town, witnessed war years, your tender feelings, such as unique trait, lets visitors back to the Ming dynasty set a unique life stories, to a security guard, Da Ming town, army general, hanfu cruise, han culture immersion performance and so on into the tourism experience, to show performance form. At the same time, there are also many web celebrity small punch points in the scenic area, such as the Ming barracks, tea shops, wine shops, Dingguang ferry, and the rain pavilion. Through the theme of the Ming Dynasty and the multi-depth interaction with tourists and businesses, the cultural theme atmosphere of the whole scenic area is constructed in a static and three-dimensional way.

Various stakeholders joint participation: innovation “Da Ming Zhi” series theme night, online forms to participate in, Official departments, such as the government and scenic spot management companies, established IP through Tik Tok platform and actively produced and forwarded promotional videos related to “Da Ming Zhi” in the early stage of “Da Ming Zhi”. Qingyan party committee government take the lead in wearing ancient costume, playing ancient instruments, making momentum,,Organizing and holding merchant meetings, publicizing and guiding merchants in ancient towns to participate in the atmosphere building spontaneously in the scenic spot. Industry authorities timely coordinate relevant sites and give maximum support to the event. Merchants actively cooperate with the night atmosphere, extend the business hours of shops, learn the etiquette culture of the Ming Dynasty, and take multiple measures to promote the diversification of industrial forms in ancient towns.

After the theme of the cultural activity was determined, Tik Tok’s short video presented a strong cultural experience to tourists and created great influence in a very short time. In one day, 5,565 thumb ups, 880 comments and 578 reposts were created^[19].



Figure 2. Tik Tok short video for pre-tour publicity

4.2.2. In-tour marketing

Content Experience Marketing: Through the “Da Ming Zhi· Etiquette Forever”, visitors will be able to revisit the excellent etiquette of ancient China. Some characters and various interactive performances are used to bring visitors directly into the theme atmosphere with the main line of “Da Ming Zhi · General Order”, through

the Ming Dynasty military tents, soldiers, golden clothes and bodyguards and other elements. The Double Seventh Festival falls on August 25. Therefore, with the Ming Dynasty as the background, 77 pairs of single men and women, spend a fantasy Tanabata Festival, which hand in hand successfully two teams, to promote a beautiful love story.

In addition, to carry forward the Chinese traditional culture, ancient civilization, the ancient town of all crowded punch points, arranged for a variety of excellent performances of traditional art, vivid, puppet show, play the shaoxing opera performances, guzheng lusheng ensemble, skills of acrobatics, allowing visitors to feel the quintessence and the charm of traditional art.

Through the matrix rapid spread of Tik Tok platform, the “Da Ming Zhi”series of activities in Qingyan became a hot spot that quickly gathered popularity and continued to attract attention. During the one-month activity, there were tens of millions of person-times of the whole network traffic, among which, CCTV news reported once and live broadcast once. The communication benefits of web celebrity cannot be underestimated. Through the launch of “Da Ming Zhi”series of activities in Qingyan ancient town, the scenic spot has injected new vitality and brought new traffic, and also explored the innovative business model brought by “network content marketing and grand tourism” to the tourism industry, realizing the double harvest of economic benefits and web celebrity communication benefits. During the activity, CCTV financial channel (China central television (CCTV - 2), radio and television stations, guiyang in guizhou radio and television stations in major media outlets are reporting on “Da Ming Zhi” lies in the emphasis, the network media, WeChat, short video platform, since the media platforms are a large number of foreign communication,the number of topics to discuss activities aggregate exposure of more than 2380, 12 games during the live broadcast of the number up to more than 20, accumulative total of more than 1700 watch, greatly improving the “reason” of reputation and visibility ^[19]. Tourists present their own travel experience (user-generated content) through Tik Tok live broadcast during travel, forming their own experience circle. Therefore, the activity features an immersive and deep experience tour with a strong sense of immersion, so that tourists who come to Qingyan Ancient Town and are attracted by the novel attractions they see online. After feeling the charm of the new tour of the ancient town, users will actively generate content to help the scenic spot to promote during the tour.

Tourists 1: “Da Ming Zhi”activities are so interesting, I got the knowledge from shortvideo of Tik Tok, there is a soldier tour, theatrical performances and the etiquette activities, etc., are attracted to the scenic spot to attend in person, a lot of landmarks are show ancient culture.So I came here and participated in these activities.That's very impressive...

Tourists 2: I saw this activity on Tik Tok, and it was so close to me. After coming here, I found that everywhere in the scenic spot, I could feel the traces of the Ming Dynasty culture, as if I had travelled back to the Ming Dynasty. As a loyal fan of Tik Tok, I must let everyone feel the charm of our Chinese culture with me...

Tourists 3: It says on Tik Tok that tourists can enter the scenic area for free if they wear Hanfu, so I came here wearing Hanfu. There are many tourists in the scenic area wearing Hanfu as well. It feels like ancient times...

Tourists 4:By Ming dynasty chancellorsville, soldiers, the scene element such as security, I feel back to the Ming dynasty, at the same time, through the scene saw all kinds of interactive performance, I know the Qingyan ancient Ming dynasty built station troops and historical background, it's great for the experience of this form, I also took a lot of video on the trill, recommend to my friend's trill, hope everybody can be in such a way to learn about Chinese ancient culture...

Tourists 5: It's worth it. You can go to Qingyan to feel the atmosphere of Ming Dynasty. It's very meaningful to hold this activity in Qingyan ancient town. I took my children to participate in the activities of

reciting classics. Three Character Scripts and Disciple Gui are all enlightenment books for children's etiquette. Children can participate in learning through this way, and our children like it very much...

In tourism marketing, the scenic spot innovatively uses Tik Tok platform to support and promote effective interaction between S2T (scenic spot to tourist) and T2T (tourist to tourist) to jointly create its online marketing



cultural content. Tourists share their experiences of cultural events through Douyin and use it as a social online link with other tourists to co-create their cultural experiences.

Figure 3. The experience shared online by tourists during the tour

4.2.3 Post-tour marketing

The 100-day immersive theme experience of “Da Ming Zhi” in Qingyan Ancient Town 2020 is finished. According to the comments on Tik Tok, tourists are still enjoying the experience and are looking forward to a richer cultural experience.

The attempt of “Da Ming Zhi” cultural activity is not only of profound significance in boosting the night-time economic development of the ancient town and promoting the recovery and upgrading of consumption, but also in adjusting the depressed psychological state of people during the epidemic through such immersive cultural experiences.

Therefore, Tik Tok platform used its data influence to help the online marketing of Qingyan Ancient Town in a very short period of time, and promoted the smooth progress of the “Da Ming Zhi” cultural experience activities.

5. CONCLUSION AND DISCUSSION

The marketing strategies and methods of this case study can be used for reference by marketers, destination management agencies, tourism policy makers, and tourism service providers in the management, promotion, planning and development of cultural tourism destinations.

2020 of Qingyan ancient town, “Da Ming Zhi” activities, through offline culture setting, interactive experience culture, the promotion of online social media, content marketing, tourist experience content independently generated (live), and other methods to build the shock condition, development of night tour Qingyan ancient town, the town economy, promoting the regional tourism recovery at night.

First, Through content marketing, “Da Ming Zhi” attracts many followers. Relying on social media technology, it builds an online interactive platform and integrates online and offline with data-driven integration,

thus arousing emotional resonance and relationship identification of tourists and realizing group interaction and meaning sharing.

Secondly, in the overall marketing process, the local government organizations, industry authorities, and merchants of the scenic spot participates in the online marketing process by actively building official IP, reposting and sharing campaign videos. Through these active assistance and participation, as well as the high-quality content shared online in real time by tourists during the experience, the “Da Ming Zhi” cultural activity quickly attracted many followers on the Internet and generated huge influence.

Finally, As mentioned in the interview materials, the deep immersive experience of cultural activities in Qingyan Ancient Town is deeply loved and recognized by tourists. The results imply that the high quality cultural tourism content is the key to the success of online marketing, therefore, both the tourism destination organizations and the tourism enterprises should do a good job in the content generation. The online marketing strategies of the “Da Ming Zhi” cultural activities can help manage tourism enterprises and assist marketing efforts of target destinations.

Research, of course, also have some disadvantages, for example, in the field of culture and experience marketing, just choose a social media platform to do the research, however, a lot of tourist destination will choose different social media platforms while they doing marketing, but different social media platforms have different effects on how they interact with users. Therefore, future research should choose a variety of social media platforms to study the differences in the effective communication of their brand value.

REFERENCES

- [1] Ananda, A., Hernández-García, A., & Lamberti, L. (2015). Social media marketing in Italian luxury fashion. Retrieved from [http:// oa.upm.es/41536/1/INVEMEM_2015_200853.pdf](http://oa.upm.es/41536/1/INVEMEM_2015_200853.pdf)
- [2] Beckman, E., Kumar, A., & Kim, Y. K. (2013). The impact of brand experience on downtown success. *Journal of Travel Research*, 52(5), 646 - 658.
- [3] Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet-The state of eTourism research. *Tourism Management*, 29(4), 609 - 623
- [4] Erdemir, A. S. (2015). Digital pr in Turkey:How Turkeybrands integrate content marketing,social media and seo in pr campaigns . *Journal of Business and Behavioral Sciences*,27(2), 38-54.
- [5] Carson, M., Maxwell, L. Lyons, L., & Roy, A. (2016). Underskilled and out-of-time: How SMEs are marketing through social media. *Journal of Business Administration*, 1-7.
- [6] Hede, A., & Kellett, P. (2011). Marketing communications for special events. *European Journal of Marketing*, 45(6), 987-1004. doi:<http://dx.doi.org/10.1108/03090561111119930>
- [7] Huertas, A., & Marine-Roig, E. (2016). User reactions to destination brand contents in social media. *Information Technology & Tourism*, 15(4), 291-315.
- [8] Kalvet, T., Olesk, M., Tiits, M., & Raun, J. (2020). Innovative Tools for Tourism and Cultural Tourism Impact Assessment. *Sustainability*, 12(18), 7470.
- [9] Kumar, V., & Kaushik, A. K. (2018). Destination brand experience and visitor behavior: The mediating role of destination brand identification. *Journal of Travel & Tourism Marketing*, 35(5), 649-663.
- [10] Levinson, J. C., & Gibson, S. (2010). *Guerrilla Social Media Marketing: 100+ Weapons to Grow your Online Influence, Attract Customers and Drive Profits*. Entrepreneur Media Inc., Canada.
- [11] Medhekar, A., Wong, H. Y., & Hall, J. E. (2020). Health-care providers perspective on value in medical travel to India. *Tourism Review*, ahead-of-print(ahead-of-print).
- [12] Rambe, P. . (2017). Social media marketing and business competitiveness: evidence from south african tourism smmes. *Problems & Perspectives in Management*, 15(2-2), 411-423.

- [13] Schreier, M. (2012). *Qualitative Content Analysis in Practice*.
- [14] W Munsters. *Cultural Tourism in Belgium* [A]. G Richard. *Culture Tourism in Europe* [C]. CAB International, 1996. 136
- [15] Xu Jufeng. (2005). *Tourist Culture and Cultural Tourism: Some Issues in Theory and Practice*. *Tourism Tribune*(04), 67-72. (in Chinese)
- [16] Youssef, K. B., Leicht, T., & Marongiu, L. (2019). *Storytelling in the context of destination marketing: an analysis of conceptualisations and impact measurement*. *Journal of Strategic Marketing*, 27(8), 696-713.
- [17] Zhang Da . *Research on Positive and Negative Effects of Tik Tok's Contents* [D]. Hebei University, 2019. (in Chinese)
- [18] China National Tourism Administration. (2003). *Standard of rating for quality of tourist attractions*. Retrieved from <http://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=511242B712ED75EAFB5FFD34C8154E1B>.
- [19] Tengxun .(2020). Retrieved from <https://xw.qq.com/amhtml/20200914A0ASBX00>.
- [20] People.cn.(2020). *Qingyan ancient town punch hot*. Retrieved from <http://gz.people.com.cn/n2/2020/1217/c371756-34479327.html>

Full Research Paper

An Empirical Study of the Impacts of Digital Marketing Contents on User Engagement in Social E-commerce Platform

Ronglin Hou¹, Fei Wang^{1*}, Zhen Zhu¹, Bingchen Guo¹, Jing Zhao¹

¹ School of Economics and Management, China University of Geosciences, Wuhan 430074, China

Abstract: This study empirically investigates the impacts of digital marketing contents on user engagement behaviors. Using digital storytelling as theoretical foundation, this paper identifies three categories of choices for designing digital contents, i.e., fundamental elements, structural elements, and presentation elements. Then, the study explores how these design choices enhance three different user engagement, including likes, comments, and subsequent purchases. With 12,548 posts in JD WeChat shopping circle as the empirical dataset, this study tests the hypothesized by combining machine learning method with regression analysis. The empirical results show the design choices discriminatively impact three kinds of user engagement behaviors. The theoretical and practical implications are discussed.

Keywords: digital marketing contents, digital storytelling, user engagement, social e-commerce platform

1. INTRODUCTION

In order to attract customers and promote product sales, more and more firms are using social media platforms to design and release digital marketing contents such as essays, graphics, short videos, and live broadcasts. These digital contents provide customers with an immersive shopping experience, and thus enhance user stickiness and stimulate shopping needs. For example, Chinese e-commerce platform JD.com collaborated with social media platform WeChat to build shopping circles for publishing digital contents. JD WeChat shopping circles released more than 10,000 pieces of digital contents during 618 promotion period in 2019, which drove product sales up 180%.¹

However, designing digital marketing contents that better engage users is not easy. On the one hand, being different with consumer online reviews that have structured formats such as ratings and textual descriptions, digital contents on social media platforms are open-ended expressions [1]. In other words, there are many different design choices for digital content generators to engineer digital contents. Content generators can offer different information in digital contents, including product scene, product materials and detailed functions etc. They can also use different narrative techniques to attract consumers' attention, such as data-driven persuasion and different personal perspectives. Digital contents also can be displayed in different formats, including short video, live broadcast, graphic and other forms. On the other hand, facing with digital contents that adopting different design choices, users can produce different engagement behaviors, such as thumb up, comments and purchase. Therefore, how to design effective digital contents to achieve marketing goals becomes a new challenge for firms using social medial platforms.

Against above background, the objective of this study is to examine the impacts of digital contents on user engagement on social business platform. Accordingly, the research questions of this paper include: (1) what are the design choices of digital marketing contents on social e-commerce platform? (2) how these different design choices impact different user engagement behaviors? To answer above two questions, we employed digital storytelling perspective to theoretically identify three categories of design choices for digital marketing contents.

* Corresponding author. Email: wangfei086071@163.com

¹ <http://dy.163.com/v2/article/detail/EIJJ5MP80511CTRL.html>

Then, we empirically investigated their impacts on three different user engagement behaviors with 12,548 posts in JD WeChat shopping circle by combining machine learning method with regression analysis.

2. LITERATURE REVIEW AND THEORETICAL FOUNDATION

2.1 Literature review

Current research on social media marketing can be divided into two main stream from the perspective of content producers: user-generated content (UGC) and firm-generated content (FGC) . Most of UGC research focuses on the economic benefits produced by UGC. Previous study conducted a lot of research on the impact of one-time purchases of products such as movies ^[2], books ^[3] and games ^[4], on the UGC of online comment communities on different online platforms. Prior studies show that quantitative aspects of UGC, such as the number of reviews or ratings, have significant and positive impacts on product sales. In order to explore the economic impact of UGC text features, researchers often use text mining methods to extract the qualitative information. Goh (2013) used machine learning methods such as topic modeling and sentiment analysis to extract the information richness and valence of the UGC information dimension and persuasion dimension of the Facebook community to explore the effect on consumers' repeated purchases ^[5].

Recent studies attempt to clarify the effects of FGC. Ordenes (2018) took the brand community on Twitter as an example to explore the impacts of the language style and rhetoric of the tweets on user sharing behavior. Kumar (2016) proposed three characteristics of FGC, including emotional valence, ease of acceptance and customer perception. They also explored the impacts of these characteristics on firm performance, such as marketing return on investment, customer cross-buying, and enterprise profitability ^[6].

Although existing research explored the economic impacts of UGC and FGC, there are two limitations in explaining how to design digital marketing contents. First, current UGC and FGC research mostly focus on a series of contents, rather than a piece of digital content. For example, prior studies explore the overall quantity, valence, and richness of UGC and FGC. Such coarse-grained analysis cannot help digital content generators to design effective contents. Second, previous UGC and FGC research have not differentiated different user engagement behaviors, and lacked empirical research on the impacts of single text content on users actual purchase behavior.

2.2 Theoretical foundation

In order to fill up above knowledge gap, we employed digital storytelling as our theoretical foundation. Storytelling originated in pedagogy and educators usually use storytelling as a way to help students understand and master new knowledge ^[7]. Subsequent marketing research argue that storytelling plays a key role in driving consumer decision-making ^[8]. As a new and novel form of product recommendation, story narration can arouse consumers' emotional reactions, thereby prompting them to make decisions ^[9]. Therefore, marketers usually package product information through digital narrative, use story narration as a scenario for product recommendation, and convey suggestions to individuals in the form of themes. Pavlik (2017) identified the three important components of digital storytelling in the digital media environment: (1) foundational elements (theme characteristics and emotional characteristics), (2) structural elements (whether first-person, whether segmented narrative, and whether to quote numbers), and (3) presentation forms (video, picture and word) ^[10].

3. RESEARCH MODEL AND HYPOTHESIS

3.1 Research model

From a perspective of digital storytelling, we regard a piece of digital marketing content as a process of digital storytelling. Digital content generators can build a personalized digital story and offer consumers all-round product information to prompt consumers engagement. According to digital storytelling perspective,

there are three kinds of design choices for digital contents: foundational elements (i.e., thematic features, emotional features), structural elements (i.e., whether segmented, whether there are digital examples, whether the first person), and the form of presentation (i.e., text, picture, video). We also distinguish three different user engagement behaviors: number of likes, number of comments, and number of subsequent purchases after reading the content. As shown in Figure 1, our research model presents how three categories of design choices impact three user engagement behaviors.

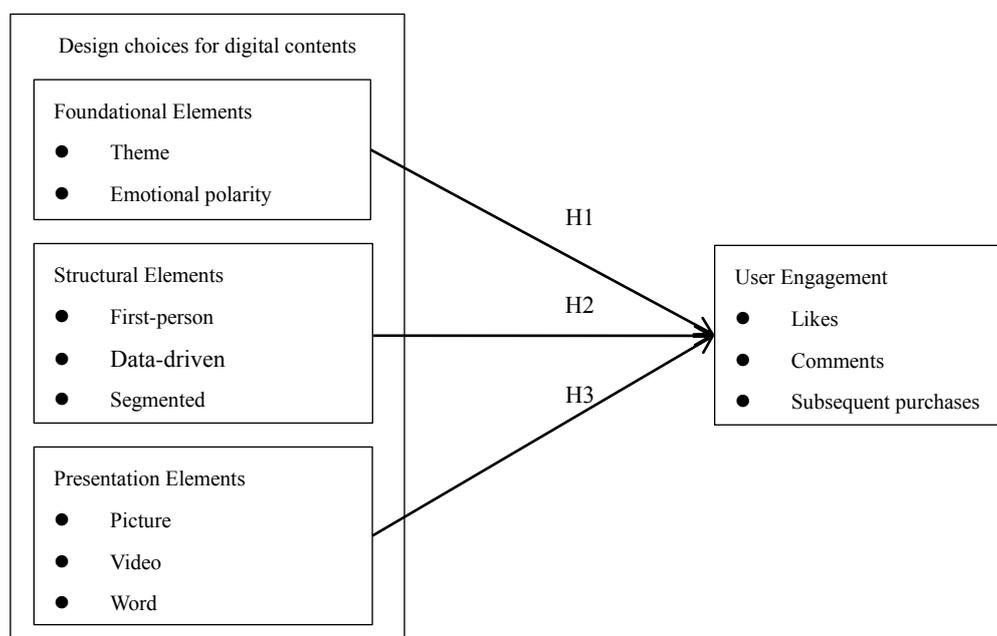


Figure 1. Research model

Table 1. The definitions of research variables

Variable	Definition	Reference
Theme	A set of short descriptions using text content	Xiang zheng ^[11]
Emotional polarity	The emotional orientation of the generators contained in the text	Xiang zheng ^[11]
Data-driven	The content user the numbers to describe products	Pavlik ^[10]
First-person	Narrative in the first person	Pavlik ^[10]
segmented	Use a segmented narrative to enhance persuasiveness	Pavlik ^[10]
Picture	Content generators show products through pictures	Liu Wenjing ^[12]
Video	When content generators design content, they can also display products in the form of videos	Pavlik ^[10]
Word	Display product information to consumers in digital content	Pavlik ^[10]
Likes	Express users' positive and affirmative emotions about existing content	Swani ^[13]
Comments	Need users to spend time and cognitive ability to write, express their opinions, emotions, etc.	Swani ^[13]
Purchases	The economic activity of goods or services obtained by consumers with money or work	ZhangWei ^[14]

3.2 Hypothesis development

3.2.1 Foundational elements

The foundational elements of digital contents include theme features and emotional features. The foundational elements are the basic components of digital contents, that is, “what to convey to consumers”. Theme can help consumers understand the product information more comprehensively, reduce their information

search costs and time costs, and encourage consumers to produce positive feedback behaviors. Emotional polarity is the emotional tendency of the content recommender towards the product, which can affect consumer engagement by enhancing trust in the emotional dimension [5]. Xiang (2017) found that the positive emotions of content recommenders can significantly and positively affect consumer feedback [11]. We therefore hypothesize that:

H1: The foundational elements of digital contents significantly affect user engagement behaviors.

H1a: Theme of digital contents significantly affects user engagement behaviors.

H1b: Emotional orientation of digital contents significantly affects user engagement behaviors.

3.2.2 Structural elements

The structural element of digital contents is the narrative structure of the content, that is, “how to arrange the content” [10]. The structural elements mainly include first-person narration, data-driven, and segmented narration. First-person narration can narrow the distance between content recommenders and consumers from the emotional dimension [8], enhance consumers' trust in content recommenders, and trigger consumer feedback behavior. Digital drive means to use actual numbers to corroborate one's own views in the content description process to enhance the objectivity and authenticity of the content [4], thereby triggering consumer feedback behavior. Unlike the previous two, segmented narration can improve the legibility of the text, make it easier for consumers to understand the key points and themes of each part of the description, accelerate the understanding of the content, and promote consumer feedback. We therefore hypothesize that:

H2: The structural elements of digital contents significantly affect user engagement behaviors.

H2a: The first-person narration significantly affects user engagement behaviors.

H2b: The segmented narration significantly affects user engagement behaviors.

H2c: The digital-driven narration significantly affects user engagement behaviors.

3.2.3 Presentation elements

The presentation form of content features focuses on “how the content is conveyed to consumers” [10]. The presentation forms of content mainly include pictures, videos, and words. Pictures make users feel that “seeing is believing”, and that pictures are not restricted by the type of text and the user's cultural level, and have the advantage of communication [12]. This is especially true for video, which can fully display product information, and is accompanied by voice commentary, creating an immersive marketing experience, and enhancing the possibility of feedback. We therefore hypothesize that:

H3: The presentation form of digital contents significantly affects user engagement behaviors.

H3a: Picture display significantly affects user engagement behaviors.

H3b: Video display significantly affects user engagement behaviors.

4. METHODOLOGY

4.1 Data collection

As shown in Figure 2, *JD WeChat shopping circle* was chosen as the research context of our empirical study because it provides data on digital content, product purchase links, user engagement behaviors such as comments, likes, and actual purchases of products after reading the content.

We selected the *makeup honey circle* to collect data because it is one of the most active shopping circles. The data includes 12548 posts from September 2016 to November 2019. We design a crawl software to obtain the nickname of the generators, the time of posting, the content of the post, the number of likes and comments, whether there are pictures, whether there are videos and follow-up purchases, etc. Meanwhile, the generators' personal information may also affect consumers' engagement on the content [15]. Therefore, we also obtained the publisher's gender, rank, number of fans, growth points and other fields by clicking the link at the publisher's



Figure 2. JD WeChat shopping circle

nickname.

4.2 Measurement

4.2.1 Dependent variables

Likes and Comments. The like behavior is a “lightweight, one-click” feedback behavior [16], which is usually used to express the user’s positive and affirmative emotions. Comments are a deliberate form of “steady communication” that requires users to spend time and cognitive ability to write [16]. Therefore, we distinguish them and try to explain the impact of digital content on users’ like and comment behaviors.

Subsequent purchases. Previous studies mostly used the overall sales of e-commerce products on the platform for a period of time. But the sales may be affected by keyword bidding, advertising, etc. And it is difficult to clarify the role of content. Fortunately, JD WeChat shopping circle offers the purchase volume of goods generated by each post. The purchase volume only comes from clicking on the post link and generates the actual purchase behavior.

4.2.2 Independent variable

Topic distribution probability. We use topic modeling method to obtain the topic probability distribution of the post. We use the LDA algorithm of Gensim toolkit for topic modeling. The first step is to determine the number of topics. We apply the perplexity value proposed by Blei to determine the appropriate number of topics [17]. The number of topics in the makeup circle is before 8, the perplexity of the topic drops dramatically. And after using the number of topics mentioned above, the value of perplexity tends to stable. Therefore, we select 8 as the appropriate number of modeling topics. After that, we use the LDA model and set the corresponding number of topics. We iterated 500 times to get the theme distribution of the makeup honey circle, as shown in Table 2.

Table 2. The topic of the makeup honey circle

Topic 1. 礼物		Topic 2. 卸妆		Topic 3. 头发		Topic 4. 潮流		Topic 5. 彩妆		Topic 6. 功效		Topic 7. 品牌		Topic 8. 成分	
词语	得分														
情人节	0.020	卸妆	0.054	超市	0.025	色彩	0.013	腮红	0.040	滋润	0.024	阿玛尼	0.027	有机	0.013
七夕	0.013	卸妆水	0.029	用品	0.019	女明星	0.010	唇彩	0.014	美肌	0.010	植村秀	0.025	无害	0.008
约会	0.009	眼线	0.016	头发	0.015	流行	0.009	男色	0.013	防晒	0.007	圣罗兰	0.013	果酸	0.008
生日	0.008	眼线笔	0.013	洗发水	0.010	女神	0.008	隔离霜	0.011	补水	0.007	欧美	0.013	角质	0.007
我爱你	0.007	眼妆	0.013	头皮	0.008	气质	0.007	口红	0.008	隔离	0.006	兰芝	0.009	石榴	0.007
送到	0.006	眼唇	0.011	护发素	0.007	很美	0.006	润唇膏	0.007	紫外线	0.005	碧欧泉	0.009	香气	0.006
魅惑	0.005	眼型	0.009	年度	0.006	驾驭	0.005	唇膏	0.006	职场	0.004	专柜	0.008	最火	0.006
海边	0.005	彩妆	0.009	秀发	0.006	最火	0.005	面膜	0.005	很美	0.004	凡士林	0.005	艾草	0.005
遇见	0.005	刺痛感	0.008	瘙痒	0.005	不失	0.004	去角质	0.005	露华浓	0.003	自营	0.005	常用	0.005
很美	0.005	干净	0.007	价值	0.004	种草	0.004	蜜桃	0.004	遇见	0.003	流行	0.005	购买	0.004

Emotional polarity. We calculate and obtain the emotional polarity of contents through SnowNLP. Sentiment analysis is a process used to discover emotional content in text in the process of text analysis and mining [18].

All variables' measurements are shown in Table 3.

Table 3. Variable measurement

Variable	Measurement	Reference
Dependent Variable		
Purchases	The number of product sales produced by each post	
Likes	The number of likes each post	Swani ^[13]
comments	The number of comments each post	Maslowska ^[16]
Control Variable		
Fansum	Followers number of the post publisher	Adamopoulos ^[19]
Level	Rank of the post publisher	Adamopoulos ^[19]
Price	Price of the item recommended by the post	Forman ^[15]
Independent Variable		
Length	The text length of the post	Xiang zheng ^[11]
Valance	Emotional polarity of the post	Tan ^[20]
Picture	Number of pictures contained in the post	Liu wenjing ^[12]
Video	Whether the post contains video, 0, 1 variable	Pavlik ^[10]
Data-driven	Whether the post contains a video, 0, 1 variable	Pavlik ^[10]
First-person	Whether the post is narrated in the first person, 0, 1 variable	Pavlik ^[10]
Layer	Whether the post is described in segments, 0, 1 variable	Pavlik ^[10]
Topic	Get topic probability distribution through topic modeling	Xiang zheng ^[11]

5. DATA ANALYSIS AND RESULTS

5.1 Descriptive statistics and correlation analysis

Table 4 shows the descriptive statistics and correlation analysis of all variables. In order to prevent the multicollinearity problem between the variables, we performed the VIFs test. The results show that the VIF

among all variables does not exceed 1.5, which are much smaller than the standard value of 10. Thus, the variable is considered no multicollinearity.

Table 4. Descriptive statistics and correlation

Variables	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13
Purchases	27	143.77	1.00												
Likes	415.05	1257.34	0.17**	1.00											
Comments	229.01	1429.25	-0.02	-0.01	1.00										
Fansum	56395	40846	-0.12**	008	-0.01	1.00									
Level	2.38	2.803	0.21*	0.03*	0.18*	0.06	1.00								
Price	553.74	1199.65	0.21*	0.032	-0.03	0.09	0.16	1.00							
Length	167.7	177.15	-0.09	0.221*	0.34	0.08	0.25*	0.34	1.00						
Valance	0.93	0.21	0.11	0.14**	0.186	-0.06	0.18*	-0.05	-0.01	1.00					
Picture	2.53	37.71	0.15**	0.18*	0.19**	-0.01	-0.01	-0.10	0.06	0.05	1.00				
Video	0.38	2.73	0.12**	0.07*	0.31**	-0.03	0.28*	0.02	-0.01	0.14	0.12	1.00			
Data-driven	0.59	0.27	0.30	0.25*	0.28	0.01	0.37*	-0.15	0.16	0.26**	-0.04	0.17	1.00		
First-person	0.28	0.65	0.22	0.05	-0.1	0.01	0.04	0.35	0.21*	0.02	0.12	0.03	-0.20*	1.00	
Layer	0.82	12.76	-0.27	0.16	0.21	-0.08	-0.15	-0.60*	-0.3*	-0.01*	-0.10	-0.03	0.06	-0.30*	1.00

Note: * $p < 0.05$, ** $p < 0.01$.

5.2 Hypotheses testing

We used ordinary least squares (OLS) regression analysis with robust standard error to test hypotheses for the following reasons: (1) the research data is cross-sectional data, rather than time-series data, (2) after logarithmic processing, the distribution of dependent variables obeys a normal distribution, (3) the results of White Test and BP Test show that there is heteroscedasticity in our data. Thus, as shown in Table 5, we adopted a step-by-step approach to perform OLS with robust standard error. Model 1 only involved control variables, and Model 2 further included all independent variables.

First, in model 1, the coefficients of the content generators' attributes on user likes are positive and significant. After including variables about design choices for digital contents, the R2 of Model 2 has been significantly improved. The content length ($\beta=0.038$, $p < 0.1$), number of pictures ($\beta=0.102$, $p < 0.05$) and whether there is video ($\beta=0.093$, $p < 0.1$) have a significant positive effect. Among the structural elements, the coefficients of data-driven ($\beta=0.014$, $p < 0.1$) and firstperson ($\beta=0.152$, $p < 0.1$) are positive and significant, while segmented narration have no significant effect. The results suggest that using actual numbers in the content to explain product parameters and attributes, and first-person narration can affect consumers' like behavior. As for theme feature, the content including gift, trend, make-up, brands and ingredient can better affect consumers' like behavior. Among them, brand ($\beta=0.102$, $p < 0.05$) and ingredient ($\beta=0.113$, $p < 0.05$) have the most significant impacts. The results show that when consumers select cosmetic products, they pay more attention to the brand, composition and whether they are in line with the trend.

Second, Model 1 in Table 5 shows that the coefficient of fansum ($\beta=0.323$, $p < 0.05$) on the number of comments is significant and positive. But user level and product prices have no significant impact on consumers' comment behavior. We can find that the level of content publishers has different effects on comments and likes. In Model 2, the topics about content theme have positive effects on comments, including

gift, trend, brand and ingredient. The most significant of these topics are brands ($\beta=0.102$, $p<0.05$) and ingredient ($\beta=0.229$, $p<0.05$). The results suggest that the consumers' feelings and the popularity of the product can prompt consumers to comment and express their opinions. In addition, the R^2 has been significantly improved from 0.128 to 0.199.

Third, the results show compared with likes and comments, digital contents have different impacts on subsequent purchases. Model 1 in Table 5 shows that both the coefficients of fansum ($\beta=0.192$, $p<0.01$) and user level ($\beta=0.102$, $p<0.1$) can positively and significantly affect consumers' purchasing decisions. In Model 2, the R^2 has been significantly improved from 0.108 to 0.243. The length of the content has no obvious effect on the purchase behavior. But the coefficients of picture ($\beta=0.155$, $p<0.1$) and video ($\beta=0.182$, $p<0.05$) are significant and positive. A plausible explanation is that a comprehensive display of products through pictures and videos can enhance consumers' overall awareness of the product, and thus obtain better product sales. When users buy the makeup products, they pay more attention to whether the product can be used as a good gift ($\beta=0.127$, $p<0.05$), whether it is in line with the trend ($\beta=0.019$, $p<0.05$), product efficacy ($\beta=0.208$, $p<0.05$), brand ($\beta=0.203$, $p<0.01$) and ingredient ($\beta=0.182$, $p<0.01$), while others are not significant.

Table 5. Regression analysis results

Variable	Model 1			Model 2			Negative binomial regression		
	Likes	Comments	Purchases	Likes	Comments	Purchases	Likes	Comments	Purchases
Dependent Variable									
Fansum	.362*** (1.29)	.323** (0.98)	.291*** (1.11)	.289*** (2.26)	.232** (1.13)	.192*** (-1.29)	1.28** (2.35)	2.17** (2.13)	1.97*** (2.05)
Control Variable									
Level	.029** (2.13)	.273 (2.09)	.132* (1.94)	.014** (5.08)	.221 (2.81)	.102* (0.10)	2.09** (1.02)	1.36 (0.94)	2.01* (1.15)
Price	-.561 (0.85)	-.653 (1.86)	-.372 (2.05)	-2.78 (4.20)	-.786 (3.16)	-.232 (1.44)	.13 (4.09)	.16 (1.43)	0.83 (1.94)
Length				.038* (1.32)	.009 (2.21)	.165 (4.21)	2.25* (3.08)	.196 (4.09)	0.97 (3.51)
Valance				.117 (3.76)	.111 (4.28)	.185 (-2.01)	1.31 (4.52)	3.32 (1.62)	1.82 (2.48)
Picture				.102** (0.193)	.192 (1.94)	.155* (3.19)	.38** (1.38)	3.14 (1.56)	4.09* (0.99)
Video				.093* (1.21)	.182 (4.23)	.182** (0.28)	2.26* (3.39)	2.93 (3.04)	0.77** (1.57)
Data-driven				.014* (2.68)	.243 (4.32)	.04 (0.98)	2.46* (3.13)	3.01 (4.44)	0.17 (3.14)
First-person				.152* (-3.51)	.852 (-4.13)	.102 (3.29)	.21* (3.29)	1.39 (1.92)	1.59 (2.63)
Layer				-.231 (3.42)	.113 (2.13)	.032 (0.14)	2.17 (1.69)	3.26 (2.12)	1.41 (5.92)
Independent Variable									
Topic 1				.019** (1.13)	.193* (3.24)	0.127** (4.23)	5.18** (0.75)	1.32* (1.10)	1.81*** (1.67)
Topic 2				.231 (0.271)	.932 (0.271)	0.118 (-2.09)	2.17 (3.08)	.971 (3.54)	3.03 (1.92)
Topic 3				.001 (0.992)	.242 (1.27)	.209 (3.63)	3.16 (4.09)	.187 (1.94)	0.73 (4.24)
Topic 4				.004** (2.13)	.187* (2.03)	.019** (1.93)	.98** (4.35)	1.17* (3.04)	1.09** (1.07)
Topic 5				.015** (4.27)	.051 (9.20)	.003 (-2.09)	1.47** (1.07)	.97 (1.42)	0.25 (2.14)
Topic 6				.997 (0.181)	.118 (0.381)	.208** (1.02)	5.37 (.69)	.197 (0.96)	2.08** (9.04)
Topic 7				.102** (6.21)	.102** (1.291)	.203*** (.887)	1.01** (3.05)	1.09** (3.25)	1.96*** (0.63)

Topic 8				.113**(3.58)	.229** (0.173)	.182**(2.07)	.88** (1.21)	1.07** (4.34)	2.05** (3.14)
Intercept Term	.273*** (2.13)	1.93*** (2.01)	-2.01*** (6.67)	.283** (1.98)	.111** (0.293)	-.812*** (1.21)	.43*** (1.24)	.38** (3.18)	2.16*** (3.01)
N	12548	12548	12548	12548	12548	12548	12548	12548	12548
Pseudo R ²	0.118	0.128	0.108	0.182	0.199	0.243	0.131	0.136	0.145
Change R ²				0.064	0.071	0.135	0.013	0.08	0.037

Note: * p<0.1, ** p<0.05, *** p<0.01; t value reported in parentheses.

Table 6 summarized the results of our research hypothesis.

Table 6. Summary of research hypothesis verification

Hypothesis	Hypothetical content	Results
H1	The foundational elements of digital contents significantly affect user engagement behaviors.	Support
H1a	Theme of digital contents significantly affects user engagement behaviors.	Support
H1b	Emotional orientation of digital contents significantly affects user engagement behaviors.	Not support
H2	The structural elements of digital contents significantly affect user engagement behaviors.	Support
H2a	The first-person narration significantly affects user engagement behaviors.	Support
H2b	The segmented narration significantly affects user engagement behaviors.	Not support
H2c	The digital-driven narration significantly affects user engagement behaviors.	Support
H3	The presentation form of digital contents significantly affects user engagement behaviors.	Support
H3a	Picture display significantly affects user engagement behaviors.	Support
H3b	Video display significantly affects user engagement behaviors.	Support

5.3 Robustness checks

We conducted several robustness checks to validate our findings. First, in order to avoid the problem of model stability caused by the number of samples, we expand the samples of makeup honey circle for regression analysis. The model analysis results are basically consistent with Table 5. Second, because our dependent variables were count data which has the nature of discrete and non-negative integers, we used negative binomial regression to further test the hypothesis model. The results are shown in Table 5. Compared the results with the OLS regression results, the correlation and significance of the variables have not changed significantly, which indicated our findings were robust.

6. DISCUSSION AND IMPLICATIONS

Our empirical study found that the foundational elements, structural elements and presentation forms of digital contents can discriminatively promote users engagement behaviors. Theme features of foundational elements have the most significant impact. One possible explanation is that the theme features include the full information of the content recommender from purchase to use of the product (e.g., product appearance, material, brand, usage scenario, usage method, usage experience, etc.). These “immersive” content can provide consumers with a comprehensive and dynamic understanding of the product and stimulate consumers' purchasing needs.

The foundational elements and presentation forms of content characteristics can affect the three engagement behaviors of consumers like, comment and purchase, while structural elements only affect consumers' like behavior. The Foundational elements and presentation forms of the content focus on the information related to the product itself. They can provide consumers with the experience of the product from the purchase to the after-sale and help them make judgments to generate feedback behavior. The structural

elements can narrow the distance between users and content generators at the emotional dimension, and thus enhance consumer' trust and affirmation of content, and trigger consumers' likes.

This study provides several contributions for theory. First, on social e-commerce platforms, the open-ended formats of digital marketing contents will cause consumers to have different feedback behaviors such as likes, comments, and purchases, which make marketing effects uncertain. Based on the theoretical perspective of digital story narration, we divide content characteristics into three categories: foundational elements, structural elements and presentation forms. We use machine learning methods such as topic model to quantitatively extract the personalized characteristics of contents. Second, previous research on social media marketing used content community as a research unit to explore the impact of the overall amount, score, and text richness of text content on brand communication and consumer buying behavior, but it was impossible to explain the design of different content from a fine-grained dimension. This paper takes each piece of text content as the research unit, and discovers the differentiating effects of the content's theme features, structural features, and presentation forms on the three feedback behaviors of consumers, such as likes, comments, and purchases.

Our study has two main practical implications. First, for content generators, our research suggests that they need to master their own professional knowledge and product identification to attract users with higher quality content, promote consumer likes and comments, and increase the conversion rate of content to purchase. Second, for cooperative firms, our findings suggest that firms should try their best to choose content creators who have the "right to speak" in this field for cooperation. This is conducive to fully exposing their products and enhancing brand communication and influence.

Our study has some limitations and important implications for future research. First, we explore the impacts of the content characteristics on consumer feedback behavior. Although the JD WeChat shopping circle provides a good observation site for this study, the data source is relatively single. Future research should establish a cross-platform data collection strategy. The conclusions of this article will be cross-validated through future research on other social e-commerce platforms such as Xiaohongshu and Taobao content recommendation community. Second, we don't explore the differential impacts of content characteristics on consumer feedback in the time dimension. Future research can expand the sample time window and adopt appropriate measurement methods to explore the impact of content characteristics on consumer feedback over time.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71702176 and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan).

REFERENCES

- [1] Yang M, Ren Y, Adomavicius G. Understanding user-generated content and customer engagement on Facebook business pages[J]. *Information Systems Research*, (2019), 30(3): 839-855.
- [2] Chintagunta P K , Gopinath S , Venkataraman S . The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets[J]. *Marketing Science*, (2010), 29(5):944-957.
- [3] JA Chevalier, D Mayzlin. The effect of word of mouth on sales: Online book reviews[J].*Journal of marketing research*. (2006), 43(3). 345-354.
- [4] Zhu F , Zhang X . Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics[J]. *Journal of Marketing*, (2010), 74(2):133-148.

- [5] Goh K Y , Heng C S , Lin Z . Social Media Brand Community and Consumer Behavior:Quantifying the Relative Impact of User- and Marketer-Generated Content[J]. *Information Systems Research*, (2013), 24(1):88-107.
- [6] Kumar A, Bezawada R, Rishika R, et al. From social to sale: The effects of firm-generated content in social media on customer behavior[J]. *Journal of Marketing*, (2016), 80(1): 7-25.
- [7] Herskovitz S, Crystal M . The essential brand persona: storytelling and branding[J]. *Journal of Business Strategy*,(2010), 31(3):21-28.
- [8] Lund N F, Cohen S A, Scarles C. The power of social media storytelling in destination branding[J].*Journal of Destination Marketing & Management*, (2018),8(1):271-280.
- [9] Escalas J E. Self-Referencing and Persuasion: Narrative Transportation versus Analytical Elaboration[J]. *Journal of Consumer Research*, (2007), 33(4):421-429.
- [10] Pavlik J V, Pavlik J O . Understanding Quality in Digital Storytelling: A Theoretically Based Analysis of the Interactive Documentary[M]. in *Digital Transformation in Journalism and News Media*, M. Friedrichsen and Y. Kamalipour (eds.), pp. 381 - 396.
- [11] Xiang Zheng, Du Qianzhou, Ma Yufeng.A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism[J].*Tourism Management*,(2017),58:51-65.
- [12] Liu Wenjing, Liang Hui'e. The effect of image dissemination of fashion criticism in the era of reading pictures[J]. *Art and Design: Theory Edition*,(2007)(10):165-167.(in Chinese)
- [13] Swani K, Milne G, Brown B P. Spreading the word through likes on Facebook: Evaluating the message strategy effectiveness of Fortune 500 companies[J]. *Journal of Research in Interactive Marketing*, (2013), 7(4):269 - 294.
- [14] Zhang Wei, Li Xiaodan, Guo Lihong. Research on the Influence of Different Weibo Marketing Channels on Product Sales: A Comparison of the Path of Brand Owned Media VS Third Party Media[J].*Nankai Management Review*,(2018),21(02):43-51(in Chinese)
- [15] Forman C, Ghose A, Wiesenfeld B. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets[J]. *Information Systems Research*, (2008), 19(3):291- 313.
- [16] Maslowska E, Malthouse EC, Collinger T. The customer engagement ecosystem[J].*Marketing Management* , (2016),32(5-6):469-501.
- [17] Blei D M, Ng A Y, Jordan M I, et al. Latent Dirichlet Allocation[J]. *Journal of Machine Learning Research*, (2003), 3:993-1022.
- [18] Pang, Bo, Lee, Lillian. Opinion Mining and Sentiment Analysis[J]. *Foundations & Trends in Information Retrieval*, 2(1 - 2):1-135.
- [19] Adamopoulos P, Ghose A, Todri V. The impact of user personality traits on word of mouth:text-mining social media platforms[J]. *Information Systems Research*, (2018), 29(3): 612-640.
- [20] Tan S, Zhang J. An empirical study of sentiment analysis for chinese documents[J]. *Expert Systems with Applications*.(2008), 34(4):2622-2629.

Full Research Paper

Augmented Reality and Marketing: A Systematic Review

Jun Liu¹, Zhao Du^{1,2*}

¹ School of Sports Business, Beijing Sport University (Beijing), China

² The Key Laboratory of Rich-Media Knowledge Organization and Service of Digital Publishing Content, Institute of Scientific and Technical Information of China, Beijing, China

Abstract: Augmented reality (AR) has been seen as the next big thing to change the way people perceive and interact with the world. Many researchers have attempted to explore how AR can be combined with different scenarios to create an immersive experience for the customers. However, previous studies on AR were fragmented, which may lead to lacking a holistic understanding of it. Our study intends to figure out the main themes and future trends of AR in marketing by conducting a systematic review of 118 papers. We divide the application of AR in marketing into four categories such as advertising, gaming, retailing, and experiences of place (tourism, museum, and sport events). In each category, we summarize current research progress and propose future research agendas for researchers. They can address the current research gap and open up new research perspectives based on our work. This study also provides practical implications for system developers and managers to design and apply this technology in their field.

Keywords: augmented reality, marketing, advertising, game, retailing, tourism, museum, sport events

1. INTRODUCTION

Augmented reality (AR) projects virtual information into consumers' physical surroundings providing embodied experience for them to interact with superimposed content. Managers in different fields such as games and retailing have leveraged this technology to increase sales and implement effective communication. Meanwhile, scholars have also provided evidence for the advantage of AR in different aspects including AR's characteristics, customers' experience, and purchase decision. Accordingly, this study realizes the necessary of conducting a systematic review to reveal the latest development and seek breakthrough of AR in marketing.

Taking "augmented reality brand" OR "augmented reality customer" OR "augmented reality marketing" OR "augmented reality experience" as search terms in Web of Science, Scopus, Ebscohost, and Elsevier ScienceDirect, we identified 118 papers that published before 2021 in the final analysis. We first conduct a descriptive analysis of these literature. Then, we explore the relationship between AR and marketing in different fields. Finally, theoretical and managerial implications are discussed.

2. RESULTS

2.1 Descriptive analysis

Based on published year of the papers, we can conclude that AR's application in marketing is still a frontier research direction since there is no closely relevant paper until 2014 and 89% (n=105) of the studies were published between the year of 2017 and 2020. Among the 59 source journals, Journal of Retailing and Consumer Service is the largest source of publications, accounting for 12.71% (n=15). In accordance with journal type, 53.39% (n=63) of the studies focus on retailing, and the number of the studies in other fields in order are tourism, games, advertising, museum, and sport events. With regard to research method, quantitative research occupied 76.6% (n=88) of the selected papers. Besides, 7% (n=8) of the studies employed a multimethod approach, of which the mix of interview and questionnaire was most used.

* Corresponding author. Email: duzhao@gmail.com (Zhao Du) , junliu_0929@163.com (Jun Liu)

2.2 Studies on AR's marketing in different fields

In the field of advertising industry, audience's perception of AR advertising (ARA) is affected by the familiarity of ARA. The perceived informativeness and the attitude towards the ARA will be weakened by the increase of familiarity [1]. Blending business strategies with digital economy era, Snapchat's brand filter is an innovative advertising pattern allowing users to express themselves with AR technology. Studies on AR games showed that users' usage intention is driven by hedonic, emotional, and social needs, while their continuous usage is less affected by social factors [2]. It emphasizes more on user experience and the calculus of rational benefits and risks. When it comes to tourism industry, tourists who are high-sensation seekers are more likely to be satisfied with AR [3], and culture background also has impact on the preference of AR. Besides, for both tourism and sports industry, visual appeal is a predictor of behavior intention [4]. Tourists' adoption of AR is also affected by the quality of AR, tourists' personal factors such as personal innovativeness, and tourists' perception of AR. Studies on museum presented that wearable AR is a driving factor to influence visitors' willingness to a museum. Applying AR in online retailing help customers better make purchase decision and enhance their shopping experience [5]. But studies on the benefits of AR on brand didn't form a consensus. Many psychological characteristics such as body consciousness and cognitive style have also been confirmed that they will lead to different perception of AR. By contrast, studies on offline stores were fewer. Current studies mainly focused on smart experience, which is driven by hedonic, functional, social, and epistemic value.

3. CONTRIBUTIONS

By reviewing AR's application in different fields, it shows that each field has the similar research content, 1) the advantages of AR compared with prior approach; 2) the influence of AR on consumers' perception and how does it vary in different consumers; 3) how does AR solve the traditional problems in each field. It is also worth noting that though the characteristic of AR is definite, it can be interpreted differently based on the application context. Additionally, the combination of virtual and reality enables users to gain not only fun but also knowledge and health from AR-related products or experiences.

This study reveals the current research status of AR's application in marketing, it also points out the main function of AR in different fields and presents future research agendas. From a practical standpoint, it provides managerial recommendations about how to leverage AR to engage consumers with their service.

ACKNOWLEDGEMENT

This paper is supported by the National Natural Science Foundation of China (71901030) and the fund of the key laboratory of rich-media knowledge organization and service of digital publishing content (ZD2020/09-07).

REFERENCES

- [1] Yang S, Carson J R, Chen S. (2020). How augmented reality affects advertising effectiveness: The mediating effects of curiosity and attention toward the ad, *Journal of Retailing and Consumer Services*, 54(C)
- [2] Hamari J, Malik A, Koski J, & Johri A. (2018). Uses and Gratifications of Pokémon Go: Why do People Play Mobile Location-Based Augmented Reality Games?, *International Journal of Human-Computer Interaction*, 35(9): 804-819
- [3] Park S, Stangl B. (2019). Augmented reality experiences and sensation seeking, *Tourism Management*, 77:104023
- [4] Rogers R, Strudler K, Decker A, Grazulis A. (2019). Does augmented reality augment the experience? A qualitative analysis of enjoyment for sports spectators, *Media Watch*, 10(3): 664-674
- [5] Heller J, & Chylinski M, de Ruyter K, Mahr D, & Keeling D I. (2019). Let Me Imagine That for You: Transforming the Retail Frontline Through Augmenting Customer Mental Imagery Ability, *Journal of Retailing*, Elsevier, 95(2): 94-114

Full Research Paper

Formation Mechanism of Regional Stigma —— Impact of Major Public Health Emergencies

Jiayi Xu^{1*}, Keyi Zhang², Cuicui Liu¹, Sijia Guo¹

¹Evergrande School of Management, Wuhan University of Science and Technology, Wuhan, 430065, China

² School of Management, Huazhong University of Science and Technology, Wuhan, 430074, China

Abstract: The COVID-19 was taken as an example to study the mechanism of regional stigmatization —— the impact of major public health emergencies. By collecting the data of Sina Weibo, a social media platform, and using the rooted theory method, the negative comments on the major public health emergency related to the epidemic situation of new crown virus pneumonia were coded. The model of the impact of major public health emergencies on regional stigma is constructed, and the theoretical saturation test is carried out. It is found that the topic events of major public health emergencies will produce regional stigma through the attribution of responsibility and negative emotions of Internet users. In order to reduce or eliminate the negative effects of regional stigma, this study suggests that the strategy and suggestion of de-stigmatization should be adopted for different subjects and different levels.

Keywords: Major public health emergencies; Regional stigma; Grounded theory; Attribution of responsibility

1. INTRODUCTION

A series of major public health emergencies have occurred frequently in recent years, such as the H7N9 avian influenza virus epidemic in 2013, the Ebola epidemic in West Africa in 2014, and the Zika virus epidemic in 2016. The COVID-19 at the end of 2019 was the second major public health emergency in Chinese mainland since the SARS of this century. Major public health emergencies have the characteristics of poor foresight, fierce coming, wide spread and uncertain future development, which has caused serious health threats and physical and mental injury (An Lu et al, 2018)^[1]. The information about the occurrence and development of the epidemic was read and spread quickly by Internet users. On January 23, 2020, Wuhan, with a population of 10 million, was closed, which caused a huge network public opinion. In the process of spreading public opinion, a large number of netizens comment and forward with personal emotions, which intensifies the breeding of various negative emotions. Among them, some netizens spread false remarks on social platforms, calling the new crown virus "Wuhan virus", saying Wuhan is the birthplace of the virus. Including not only the US, led by Trump, and other Western dignitaries, but also the domestic and foreign people, who are full of discrimination against Wuhan and people from Wuhan, full of panic about the epidemic, and increasingly stigmatized Wuhan. Stigmatization not only brings psychological shadow to people living in Wuhan area, damages the city image of Wuhan, causes difficulties to resume work and production, but also reduces the investment of enterprises in Wuhan area, the number of tourists coming to Wuhan, and the deterioration of Wuhan business environment, which has a huge negative impact on Wuhan's economic and social development.

Most of the existing articles on the impact of public health emergencies focus on the construction of medical mechanism, early warning mechanism and so on. This paper uses the rooted theory research method to capture, refine and summarize the primary data of Sina Weibo, constructs the formation mechanism model of the influence of major public health emergencies on regional stigma, and complements the research on public health emergencies, network public opinion and stigma. The study discusses the impact of major emergencies, analyzes the causes of regional stigma, and puts forward strategies to eliminate stigma, which has certain practical

significance for guiding citizens to carry out correct self-cognition and improve group identity. It has certain reference significance for the media to construct a good public opinion information communication environment, the government and relevant departments to formulate and implement practical policies to weaken regional discrimination and conflict and promote the harmonious development of urban economy and society.

2. LITERATURE REVIEW

2.1 Major public health emergencies

According to the definition of emergency response law of the people's Republic of China, public health emergencies refer to major infectious diseases, mass unexplained diseases, major food and occupational poisoning and other events that seriously affect public health that suddenly occur and cause or may cause serious damage to public health. According to the nature of public health emergencies, the degree of harm and the scope involved, public health emergencies can be classified into four levels: especially significant (grade I), major (grade II), larger (grade III) and general (grade IV). Therefore, major public health emergencies refer to public health emergencies of class II and above. Because of its negative impact on public health, economy and society, it is often the focus of attention, especially the new coronavirus pneumonia epidemic.

In recent years, the research on public health emergencies is increasing with the continuous occurrence of events. The research in medicine mainly focuses on exploring the source of emergencies and developing vaccines. The research in the field of public management mainly focuses on coping strategies of emergencies and establishing joint prevention and control mechanism. The research on communication and marketing mainly focuses on the communication of emergencies, the development of public opinion and the impact. Zheng Yu (2017) studied the influencing factors of public opinion information dissemination tendency in public health emergencies from the perspective of negative emotions of the public. The research shows that negative emotions play a regulatory role in the impact of emergencies on dissemination tendency^[2]. Wang Lin et al. (2019) studied the communication and evolution of emergencies in social media, and believed that information with strong timeliness and novel content would be more widely spread. If the official media's articles showed certain subjectivity, the amount of forwarding and attention would be greater, and put forward the corresponding public opinion monitoring and public opinion guidance scheme^[3]. Li Yuelin and Wang Shanshan (2020) believe that the information release of emergencies should be timely and standardized, form a multi-pronged dynamic release mechanism, establish two-way communication channels with the public, strengthen the investigation and attention of public information needs, ensure the accuracy of epidemic information release, and effectively meet the information needs of the public^[4].

2.2 Geographical stigma

2.2.1 Definition of geographic stigmatization

The word "stigma" originated in ancient Greece. It refers to marking slaves, criminals or traitors by portraying special symbols in the human body. With the enrichment and development of the research theory of stigma, it is generally believed that stigma refers to the derogatory and insulting label that the public places on individuals or groups with some kind of social undesired and dishonorable^[5]. Sociologists Goffman (1967) believe that stigmatization is the process of undermining the social status of individuals or groups because they have certain characteristics that society does not expect^[6]. Kurzban (2001) think that stigma is the inevitable result of evolution for the survival race^[7]. Corrigan P.W. (2004) further study holds that public stigma and self-stigma are a unified whole, and public stigma is a derogatory stereotype of stigmatized groups by other social groups. Self-stigma is self-deprecation and self-denial of stigmatized people after public stigma. Former definitions are limited by specific situations or understanding perspectives, too much attention to individuals and neglect of the corresponding social and cultural environment, which is not scientific and strict^[8]. Linket

al.(2001)redefined the concept of stigma as referring to the simultaneous occurrence of labelling, stereotype, loss of status and discrimination in society ^[9].Yang et al.(2007)believe that stigma is essentially based on linguistic relationships, discourse power, and the process of imposing derogatory labels on vulnerable groups or individuals ^[10].

Influenced by the activities of western countries and international organizations to eliminate stigma, stigma research has first attracted the attention of medical psychiatry in China. In addition, some scholars pay attention to the domestic identity and group-based stigma.From the perspective of social representation, Guan Jian (2006) conducted a qualitative research on migrant workers in Tianjin n area to understand the construction and formation mechanism of social stigma ^[11].The research of stigma is mainly focused on sociological psychology in China, and the research from the angle of marketing management is relatively few. At present, the academic circles have not put forward the representative authoritative concept of regional stigma.Synthesizing the above research on stigma, this paper holds that regional stigma refers to the social public putting a derogatory or insulting label on the region. Such negative labels will make the relevant people in the region suffer unfair treatment, even exclusion and discrimination in the network and even in social life.

2.2.2 Causes and Characteristics of Regional Stigma

With regard to geographical stigma, there are still few domestic studies. But from the existing literature can explore the causes and characteristics of regional stigma. Chen Xuewei (2017) pointed out that the reasons for regional stigmatization are: the label has spatial displacement, the communicator's main body cognition is different, and the threshold of public domain under the participation of audience is low.In the context of the new media, it is easy for the public to form a certain prejudice and impression on a certain region, and the communicators continue to publish negative news about a certain region. The region is easy to be "pan-regional label ", today's network environment has a certain right to freedom of expression, some netizens tend to blindly follow the topic of regional label, will be irrational emotional catharsis ^[12].Li Qiang et al.(2008) believed that in addition to the biological needs of stigma (raising self-esteem, gaining personal superiority, relieving anxiety), it is also influenced by culture (sociopolitical, historical, cultural background, etc.), that is, stigma is the product of the interaction biological innate tendencies and cultural forces ^[13]. Zhang Le and Tong Xing (2010) believe that in emergencies, people will form a sense of crisis and perceive risk, and the consequences of risk perception are often accompanied by "stigmatization ".Stigma has the following characteristics: The first is destructive, distorting people's normal senses, so that it brings a kind of belittling color, it will cause social, political order and other chaos to a certain extent; the other is rapid pollution, which is often widely spread before the source of risk is identified; and the third is not easy to eliminate, once stigma is formed, it develops into stereotypes and cultural prejudices, and even if much is done, it is difficult to eliminate ^[14].

2.2.3 Impact of geographical stigma

Yang Xinde et al .(2009) believed that there is a close relationship between stigma and segregation, prejudice and discrimination. "Regional stigma" is not conducive to the employment and life of people in the stigmatized areas, affects regional coordination, and even affects the process of building a harmonious society. In general, geographical stigma can lead to prejudice against the region, social exclusion and discrimination, which may cause dual psychological and physical harm to the people of the region. Secondly, regional stigma can lead to group members feel inferiority complex for their identity and produce self-stigma for group culture, which is similar to the "Rosenthal" effect of psychology, which will cause psychological pressure on people in the contaminated area. There are phenomena such as low self-esteem, low mood and self-abandonment ^[15]. At the same time, regional stigma will also damage the image of the region and have a negative impact on tourism and economic development. Wang Rong and Xie Changqing (2019) found that the negative emotions such as prejudice and anger caused by long-term regional will cause contradictions and conflicts between stigmatized

groups and contaminated groups, and will affect the of social stability^[16]. In terms of moral ethics, regional stigma distorts the outside world's cognition of the region and violates the social ethics of fairness and justice. Finally, regional stigma can also damage the international image, today's world is a global integrated, pluralistic world, the stigma of a region must damage the overall image.

2.3 Review of literature

From the content of the study, the study of major public health emergencies around its definition, characteristics, diffusion cycle and its impact on public opinion and public sentiment has achieved more fruitful results. These studies provide a theoretical basis for the study of public opinion on major public health emergencies in the social media environment. However, the existing research on major public health emergencies caused relatively little regional stigma. With regard to stigma research, at present, more attention is paid to the stigmatization of people, such as "migrant workers", which belong to the typical population stigma. Although some research have begun to pay attention to the issue of regional stigma, there is little research on the impact of major public health emergencies on regional stigma. From the research of view of research methods, most of the existing regional stigma is studied by literature review method, scenario simulation method and so on, which lacks real data. The Weibo text data has the characteristics of large quantity, timely, accurate and easy to obtain. Therefore, based on the rooted theory, this paper analyzes and studies how major public health emergencies cause regional stigmatization by influencing public psychology through the data of Weibo platform. In order to put forward some de-stigmatization strategies for different subjects such as the public, media and government, repair the image of the city, maintain social harmony and stability and economic development.

3. RESEARCH DESIGN

This paper mainly studies the influence of major public health emergencies on regional stigma. Based on the rooted theory, this paper grabs data from Sina Weibo platform and collects relevant data. Sina Weibo is one of the most influential social media platforms in China, with a monthly activity of nearly 500 million. Since the outbreak of the COVID-19, related incidents have been concerned, causing heated discussion among the majority of Internet users. Sina Weibo from January to April 2020, "COVID-19" topic reading reached 2.12 billion, was discussed 445000 times. A large number of comments on the topic of epidemic related events have fully expressed the attention of netizens to the epidemic situation of COVID-19, the real emotional state of netizens, and provided a large number of raw materials for studying the mechanism and influence of major public health emergencies on regional stigma. Therefore, with "COVID-19" as the topic, using the octopus collector to grab CCTV news, People's Daily, China News Network, Southern weekly and other mainstream news media Weibo account related events user comment data as a sample. This study extracted the following nine user interaction rate (number of likes and comments) high, overheated search caused hot topic information for data analysis, see Table 1. Because of the large number of comments, the comments are coded according to the following criteria: first, delete comments unrelated to this topic event, delete comments that are not materially helpful or too simple; third, delete similar content published in the same ID.

Table 1. Data capture.

Media	Date	Topic
The Paper	25 January 2020	85 hotels in Wuhan to support medical staff
China News Network	25 January 2020	Statement issued by Hubei Red Cross Society in Wuhan
People's Daily	7 February 2020	16 provinces, one province, one city to support Wuhan and other cities in Hubei
Southern Weekly	23 February 2020	Five foreign athletes of Wuhan Military Games are suffering from malaria
China News Network	26 February 2020	Details of patients with the earliest COVID-19 in Wuhan
CCTV news	9 March 2020	The Foreign Ministry responded to the U.S. Secretary of State's statement

Media	Date	Topic
Time video	24 March 2020	The woman was isolated from the landing of a Hubei nationality
China Newsweek	26 March 2020	Studies show that the new crown virus has spread in Italy since January 1
China Newsweek	30 April 2020	Wuhan has become the city where Internet users want to travel most

4. STUDY RESULTS

4.1 Coding and modelling

First of all, open coding, through the analysis of nine topics in Table 1, mining the original comments related to the research topic from the data, according to the sample event topic of the comments of the sentence serial number. The data are divided into coded data and reserved test data, which are distinguished by the second number of codes, such as the first comment of the first topic event that encodes 1-1-1, and the first comment that encodes 1-2-1 represents the first topic event of reserved data. Use short words or sentences to summarize the contents of the comments, develop them into concepts, and compare the concepts to classify the categories to sum up the categories. Finally ,34 concepts are abstracted from the original comment data and summarized into 11 categories (annotated with B1...Bn).

Then the main axis coding is carried out, and to connect the category formed in the open coding is by cluster analysis. Seek clues and analyze whether there are potential connections in each category at the conceptual level. Through the spindle analysis, it is found that there is a certain logical and causal relationship between the comments and topic events of Internet users. According to Bandura's three-dimensional interaction theory, the internal factors, external environmental factors and human behavior factors of human beings are independent and mutually determined after major emergencies ^[17]. Human internal and behavioral factors can be summarized as individual level (risk perception, individual behavior); the main external factors are divided into social level (media, social organization) and government level (government response, international dialogue). Finally, five main categories can be summed up from 11 categories (annotated with C1...Cn), as shown in Table 2.

Table 2. Spindle encoding of comment data.

Fundamental Category	Subcategory
C1 Topic classification	B1 Event topic subject
C2 Attribution of responsibility	B2 The virus came from Wuhan
	B3 Failure of the Government and bodies
	B4 Improper media communication
C3 Request for assistance	B5 Other cities in Hubei request support
C4 Emotional response	B6 Negative emotional reactions
C5 Geographical stigma	B7 Self-stigmatization
	B8 Psychic trauma
	B9 Discrimination against others
	B10 Regional blackening
	B11 Urban image impaired

Then the selection coding is carried out to explore the core category from the main category, to connect it systematically with other categories, and to describe the whole phenomenon and event in the form of story line. On the basis of cognitive attribution ^[18] and emotional attribution ^[19] theory, it is found that can be classified into two core categories: One is the formation of event —— attribution —— behavior. When a major emergency

occurs, internet users will discuss the topic of the incident and find out the responsibility for the origin of the matter. After the outbreak of the epidemic, some netizens attributed the responsibility to Wuhan people, and some netizens attributed the responsibility to the government and relevant departments. These responsibilities were attributed to the regional stigmatization, which seriously damaged the image of Wuhan. Second, situational stimulation — the formation of emotional — behavioral patterns. The topic of major emergencies as a situational stimulus can trigger a variety of negative emotions, including anger, resentment, indignation, disappointment, sadness and so on. Negative emotions can not only directly lead to negative behaviors such as stigma related areas, but also the attribution of negative emotions is prone to self-stigma, psychological trauma, personnel discrimination, regional blackness and so on, which makes the image of the city damaged.

Finally, the two core categories were merged into a core category of "stimulating — organism — response" to establish a model of the impact of major public health emergencies on regional stigma with topic classification (C1), responsibility attribution (C2), emotional response (C4) and regional stigma (C5) as core variables, as shown in figure 1.

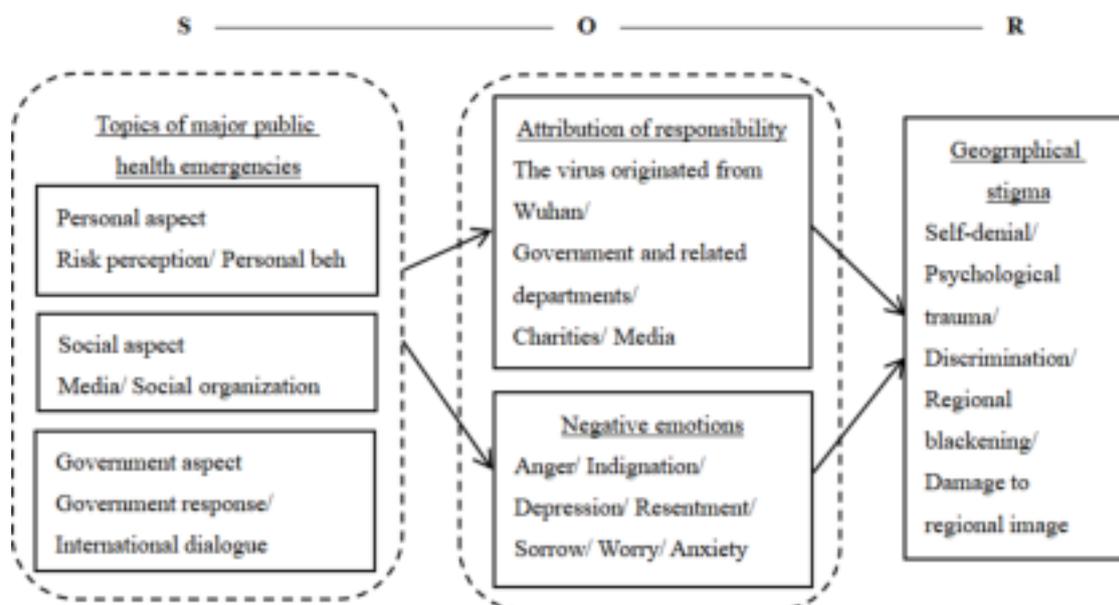


Figure 1. Model of Impact of Major Public Health Emergencies on Regional Stigma.

4.2 Theoretical saturation test

A third of the collected samples were reserved for saturation test, and 3 topics were reserved, as shown in Table 3. The collection method and processing of sample data are the same as the data used for modeling. After analyzing the reserved data, all the comments are within the scope of coding, and there are no new categories, which indicates that the theory is saturated.

Table 3. Data reservation.

Media	Date	Topic
China News Network	13 March 2020	The epidemic situation of other cities and counties in Hubei except Wuhan is low risk
CCTV News	26 March 2020	Geng Shuang batch of Pengliao reversion Wuhan virus
News Head	1 May 2020	Government report on the gathering epidemic situation in Heilongjiang

5. RESEARCH CONCLUSIONS

Taking the epidemic situation of new coronavirus pneumonia as an example, using the method of taking root theory, this paper grabs the comments on new coronavirus pneumonia from Sina Weibo from January to April. The theoretical model of the influence of major public health emergencies on regional stigma is

established by open coding, spindle coding and selective coding, and the saturation test of the model is carried out. The study found that major public health events produce a variety of topics under social media conditions, which can be divided into three types: first, personal level, mainly including personal risk perception and personal behavior; second, social level, Mainly media and social organizations; third, government level, mainly government response and international dialogue. These topics lead to regional stigmatization through two paths, and the first makes the public attribute responsibility to the topic, thus creating regional stigmatization. Another path is that the topic creates negative emotions for the public, leading to regional stigmatization. The harm caused by regional stigma can be summarized as follows: self-stigma, psychological trauma of people in contaminated areas, discrimination and geographical blackness of personnel, damage to the image of contaminated areas, affecting the employment and life of people in contaminated areas, and hindering the economic recovery of contaminated areas. This includes many negative economic impacts, such as reduced investment, lower tourism revenues and lower employment rates.

6. COPING STRATEGY

In general, regional stigma can have many negative effects and is extremely unfavorable to regional development. In order to reduce the negative impact of stigma, all walks of life should give full play to their strengths and make concerted efforts to work together for the cause of stigma removal in Wuhan. Therefore, we should be put forward of targeted de-stigmatization strategy for different subjects: At the individual level, we should keep sober, resist rumors, spread positive energy, insist on the national and collective interests as the most important, and appeal for individual interests through reasonable channels, and fully believe in the decisions made by the state and the government. At the social level, we should do our best to overcome the difficulties together with the city. Among them, the media should have conscience, resist rumors, unblock public opinion, propagate positive energy, and reshape the image of the city. The media can also invite stars, network celebrities and other webcast, to help the city with goods, to help the city restore and develop the economy. Social organizations should do their own work, not like the Hubei Red Cross to use power for personal gain. Enterprises should also actively assume social responsibility, actively donate donations, prohibit the development of difficult wealth, do not smear the city. At the national level, the government should actively monitor and guide the network public opinion, in the event of causing a greater stir, the government should disclose information to the public in time to clarify the facts; improve the ability of comprehensive management, open online government platform, actively listen to public opinion, understand the people, better serve the people, make public management more convenient and efficient; establish a reward and punishment mechanism to reward those who contribute during the epidemic period, and severely punish those who violate the law and discipline; actively organize the resumption of work, provide more preferential policies, and promote urban economic development; To oppose the politicization of the virus, for the insult from the west, we should take effective counter measures to enhance national self-confidence and social cohesion.

ACKNOWLEDGEMENT

This research was supported by 2020 Provincial College Students Innovation and Entrepreneurship Training Program under Grant S202010488069.

REFERENCES

- [1] Anlu, Liang Yanping. A Study on Weibo Topic and User Behavior Choice in Public Health Emergencies Data analysis and knowledge discovery ,2019,3(4):33-41.(in Chinese)
- [2] Zheng Yu. The influencing factors of public opinion information dissemination tendency in public

- emergencies—Based on the research perspective of people's negative emotion[J].Intelligence Theory and Practice ,2017,40(7):80-87.(in Chinese)
- [3] Wang Lin, Wang Ke, Wu Jiang. Public opinion Communication and Evolution of Public Health Emergencies in Social Media—A case study of Vaccine events in 2018[J].Data analysis and knowledge discovery ,2019,3(4):42-52.(in Chinese)
- [4] Li Yuelin, Wang Shanshan. Analysis of the Characteristics of Information Dissemination for Public Health Emergencies [J]. and Books and Intelligence ,2020,(1):27-33.(in Chinese)
- [5] Yuan Huixia. An Analysis of Stigma Communication in the Comment Area of News Client[J].Media,2016,(15):55-56.(in Chinese)
- [6] Goffman. Stigma: Notes on the management of spoiled identity[J]. American Journal of Sociology, 1969,45(527): 642-643.
- [7] Kurzban R, Leary M R. Evolutionary origins of stigmatization: the functions of social exclusion.Psychological Bulletin,2001,127(2):196-202.
- [8] Corrigan P W. How stigma interferes with mental health care[J].Am Psychol,2004,59(7):614-625.
- [9] Link, Phelan. Conceptualizing stigma[J].Annual Review of Sociology,2001,27(2): 363-385.
- [10] Yang, Kleninman, Culture and stigma: adding moral experience to stigma theory[J].Social Science& Medicine,2007,64(7): 1524-1535.
- [11] Guan Jian. The Construction and Social Characterization of Identity Stigma—Taking Migrant Workers in Migrant Workers in Tianjin N Area [J].China Youth Studies ,2006(3):21-27.(in Chinese)
- [12] Chen Xuwei.Regional Labeling in New Media Context[J].Young Journalists ,2017,(3):39-40.(in Chinese)
- [13] Li Qiang. A Review of the Theory of Stigma Formation in Mental Illness[J].Advances in Psychological Science ,2008,16(4):582-589.(in Chinese)
- [14] Zhang Le, Xing Tong. Stigma: an in-depth analysis of the consequences of emergencies[J].Social Sciences Research ,2010,(6):101-105.(in Chinese)
- [15] Yang Xinde, Peng Lihui, Huang Ying.Stigma and its effective coping strategies[J]. Journal of Ningbo University (Humanities Edition)(1):135-139.(in Chinese)
- [16] Wang Rong, Xie Changqing. The Social Influence and Harm of "Stigma" Communication in the Media Age—Taking Chongqing bus falling River as an example[J].Current Media (academic Edition)27(6):60-63.(in Chinese)
- [17] Zhu Daiqiong, Wang Guohua. The influencing factors and mechanisms of Internet users' social emotions in emergencies—A qualitative comparative analysis of multiple cases based on ternary interaction determinism (QCA)[J]. Journal of Intelligence ,2020,39(3):95-104.(in Chinese)
- [18] Lacznia R N,Decarlo T E,Ramaswami S N. Consumer's responses to negative word-of-mouth communication: An attribution theory perspective[J].Journal of Consumer Psychology,2001,11(1): 57-73.
- [19] Li Zijian. An Analysis of Weiner's Theory of Emotional Attribution[J].Dietary Care,2016,3(8):255-256.(in Chinese)

Research on the Location of Nucleic Acid Detection Institutions Based on Coverage Model

Jiangping Wan ^{1*}, Xuejian Li ¹, Chunxiao Yang ², Hexin Zhang ³

¹South China university of technology, Business administration college, Wushan Road No. 381, Tianhe Guangzhou Guangdong, 510640, China

² Guangdong twenty commercial chain co., LTD., Linhe East Road No. 128, Tianhe Guangzhou Guangdong, 510610, China

³ The board of directors office of Guangdong province company of China national tobacco corporation, Linhe East Road No. 128, Tianhe Guangzhou Guangdong, 510610, China

Abstract: We should optimize the location of nucleic acid testing institutions, consider the actual situation, and carry out differentiated management for high-risk and low-risk nucleic acid testing according to the needs of epidemic prevention and control. Based on the coverage model, the optimization model of nucleic acid detection and location of nucleic acid collection points is established with the objective of minimizing the social cost of emergency, and lingo and CPLEX tools are used to optimize the solution. The feasibility of the model is verified by an example.

Keywords: Location optimization; Coverage model; Emergency facilities; Lingo

1. INTRODUCTION

The sudden epidemic broke the original peaceful life and caused great economic losses. Sudden major emergencies, the emergency medical system is a huge test. Major emergencies are defined as events with low frequency and great impact. Once they happen, they need a lot of first-time emergency rescue and assistance from local governments and (or) countries, such as natural disasters, public health events, major production safety accidents and terrorist attacks. When COVID-19 occurs, once a city has a case, it will often carry out nationwide nucleic acid detection. How to quickly carry out the layout of emergency medical facilities is particularly important. It requires not only the maximum coverage in a short time, but also the minimum social cost. Once the emergency medical facilities are established, they will operate for a long time. It not only affects the direct cost of operation, but also has a great impact on the coverage, work efficiency and management level., It is very important to choose the right deployment location in order to reduce social costs and improve work efficiency.

2. LITERATURE REVIEW

The location problem is to use a series of methods to find the optimal location of the target facilities, so that the target facilities can meet the demand, cost, policy, environment and other requirements. The location of emergency facilities can be divided into many types according to different standards ^[1]. Permanent emergency facilities refer to the long-term existence of emergency facilities in a fixed place; temporary emergency facilities refer to the short-term existence in a fixed place, which is used to deal with the deployment of emergency facilities for large-scale emergencies. Emergency facilities can be divided into emergency medical facilities, emergency logistics facilities and disaster prevention and mitigation facilities in the view of usage. Disaster prevention and mitigation facilities refer to the disaster reduction and avoidance places that need to be

* Corresponding author. Email: csjpw@scut.edu.cn (Jiangping Wan), 1339765449@qq.com (Xuejian Li)

established when major disasters occur. Today, the research of emergency facility location model mainly includes basic location problem, dynamic location problem, stochastic location problem, robust optimization location problem and other location problems. The basic location problem is mainly studied from the coverage model, p-median model and p-center model; the dynamic location model considers the time parameters, usually considering different locations in different time periods; the stochastic location problem considers the location problem when the parameters obey the probability distribution; the robust optimization location problem considers the location problem when the parameter distribution is uncertain location, through robust optimization model, to find the optimal results under random fluctuations. The model considered in this paper is a deterministic mixed integer programming location problem. Many scholars at home and abroad have proposed a variety of models and algorithms in the corresponding fields: Yu Dongmei et al.^[2] considered the location problem of emergency facilities in the situation of facility interruption, proposed to establish an optimization model of reliability emergency facility selection with sharing uncertain demand and limited service capacity by extending MCLP model, and improved Gray Wolf algorithm to solve the model; Based on the theory of bounded rationality, Chen Gang et al.^[3] consider the individual behavior of the affected residents in the perspective of survival probability of the wounded. This paper proposes the maximum robust location model for the survival probability of the wounded under uncertain demand from the tactical level; Sun huali^[4] and others propose the minimum delivery time and material delivery time model in the perspective of material delivery time, and verify the effectiveness of the model with several examples; Sun jianping and others^[5] study the location problem of high-speed rail emergency resources. This paper establishes the location optimization model of high-speed railway emergency resource reserve point, and uses the sequential selection genetic algorithm to solve the model, and obtains a good optimization result. Chen zhizong and others^[6] consider the fairness and efficiency of emergency rescue facilities, integrate the maximum coverage model, p-center model and p-median model, and put forward a multi-objective location model, which is proved to be effective by an example.

In the above studies, the main focus is on the location of earthquake, anti-terrorism, shelters and so on. There are few studies on the location of epidemic situation, and most of the studies on the location of emergency facilities do not consider the capacity constraints and economic issues. In the epidemic situation, nucleic acid testing for the whole people should not only consider the full coverage, but also the risk of epidemic transmission caused by large-scale population flow. It is particularly important to quickly allocate the population for orderly testing and consider the social cost in the optimal solution. In this paper, based on the coverage model, an optimization model of emergency facility location with limited service capacity is established to reduce potential risks and provide scientific and reasonable decision-making basis for emergency management departments in the perspective of residents and society.

3. PROBLEM DESCRIPTION AND MODEL CONSTRUCTION

3.1 Problem description

Based on the mathematical programming method, the location of nucleic acid detection points and nucleic acid collection points in key areas, the number of detection reagents provided to them and the population they should serve are determined. Area network graph can be represented by directed graph $G=(V, A)$, where V is a group of nodes and A is a group of arcs. The set V can be divided into $\{V_0, V_1, V_2, V_3\}$, where V_0 is the distribution center set, v_0 is the location of the supply point, and $v_0 \in V_0$ is the set of population gathering points, V_2 is the set of detection points to be selected, and V_3 is the set of collection points to be selected. The number of population at the node $v_i \in V_1$ is q_i , W_i is defined as the set of detection points to be selected within the coverage radius r_1 km of $v_i \in V_1$, $W_i \in V_2$, L_i is defined as the set of detection points at the node v_i . The collection of points to be selected within the coverage radius r_2 km of $v_i \in V_1$, $L_i \in V_3$. The route of

transport vehicle from v_0 to $v_j \in V_2$ is represented by arc (v_0, v_j) , the route of crowd v_i from $v_i \in V_1$ to $v_j \in V_2$ is represented by arc (v_i, v_j) , the route of collection point v_k from $v_k \in V_3$ to $v_j \in V_2$ is represented by arc (v_k, v_j) , and the route of crowd v_i from $v_i \in V_1$ to $v_k \in V_3$ is represented by arc (v_i, v_k) .

For the convenience of analysis, assumptions are made in the following:

- 1) It is assumed that the test is carried out in the way of one person and one reagent;
- 2) Suppose that the samples collected from the collection point are sent to the nearest detection point for detection according to the shortest distance principle;
- 3) One person only tested once;
- 4) Assume that all nucleic acids are detected in the region;
- 5) It is assumed that the sample detection in severe epidemic areas is provided with additional detection capability, that is, the samples sent to the corresponding detection points are processed separately and are not constrained by the original detection capability.

3.2 Model building

3.2.1 Symbol description

1) Decision variables

x_{ij} : The proportion of v_j people going to v_j for testing;

x_{ik} : v_i the proportion of people who need to go to the collection point to receive nucleic acid collection;

y_j : 0-1 variable, when j detection point is selected, $y_j = 1$, otherwise $y_j = 0$;

g_k : 0-1 variable, when k detection point is selected, $g_k = 1$, otherwise $g_k = 0$.

2) Parameters

q_i : The number of population at node $v_i \in V_1$;

α_{ij} : Time cost and transportation cost of unit citizen inspection;

ξ_{ik} : Time cost and transportation cost of unit citizens going to collection points;

β_j : From the distribution center to each testing point j, the unit cost of testing supplies;

η_k : The transportation cost of transporting the collected samples from the collection point K to the nearest detection point for testing;

θ_k : From the distribution center to each collection point K, the cost of sample processing per unit of reagent collection;

γ_j : The fixed cost of testing point includes the salary of testing personnel and the cost of setting up testing point;

σ_k : The fixed cost of collection point includes the salary of collection personnel and the cost of setting collection point;

d_{ij}^* : Distance of residents from community v_i to detection point v_j ;

v : Average travel speed of residents;

t_1 : Average time spent collecting nucleic acid;

c_1 : Unit travel cost of users;

c_2 : Unit time cost lost by users going to test;

c_3 : Unit transportation cost from reagent production site v_0 to detection site v_j ;

c_4 : Unit transportation cost from v_k to v_j ;

c_5 : Unit collection cost of v_k ;

c_6 : Fixed cost of v_j ;

c_7 : Daily salary of the inspector at v_j ;

c_8 : Construction cost of acquisition point v_k ;

c_9 : Daily salary of collector in v_k collection point;

C_{10} : Unit cost of testing supplies;

C_{11} : Unit cost of sampling supplies;

d_{oj}^r : Distance between reagent distribution center and v_j ;

d_{ok}^r : Distance between distribution center v_0 and detection point v_k ;

z_j : The number of testing personnel needed for v_j of each testing point;

r_k : The number of acquisition personnel required for v_k acquisition point;

n_1 : Single person daily testing ability of testing personnel;

n_2 : Single person daily collection capacity of collection point;

d_{kj}^g : Distance between v_k acquisition point and v_j detection point;

M : Infinity;

F_j : Maximum detection capability of v_j of detection point;

F_k : Maximum acquisition capacity of acquisition point v_k .

3.2.2 Modeling

(1) Objective function

Considering the complexity and particularity of nucleic acid detection location in reality, this paper takes minimizing the total social cost in the distribution process as the optimization objective to measure the economic benefits of the model. The total social cost includes two aspects: the user perspective cost and the government public cost. User cost mainly considers user time cost and transportation cost $\sum_{i \in V_1} \sum_{j \in W_i} \alpha_{ij} q_i x_{ij}$ and $\sum_{i \in V_1} \sum_{k \in L_i} \xi_{ik} x_{ik} q_i$; government public cost includes: reagent transportation cost $\sum_{i \in V_1} \sum_{j \in W_i} \beta_j q_i x_{ij}$, collection point transportation cost $\sum_{i \in V_1} \sum_{k \in L_i} \eta_k q_i x_{ik}$, collection reagent transportation cost $\sum_{i \in V_1} \sum_{k \in L_i} \theta_k q_i x_{ik}$, detection point investment cost $\sum_{j \in V_2} \gamma_j y_j$, collection point investment cost $\sum_{k \in L_i} \sigma_k g_k$. As for the user's time cost, because time and economy are two different concepts and cannot be calculated directly, it is necessary to convert the time cost into the value cost that the user can create in that period of time. Income method and production method are used for conversion for different types of users.

The income method is used for conversion for those who do not have regular work. The conversion formula is as follows: the cost of income reduction due to nucleic acid detection is reduced for those who do not have regular work. The production method is adopted for conversion for the personnel with fixed work, and the loss that caused by the personnel can not create greater value due to the loss of this period of time. The conversion formula is as follows:

$$TC_1 = \frac{IC}{WC} \quad TC_2 = \frac{GDP}{WN * WT}$$

Where, TC_1 is the unit time cost converted by income method, where IC is the annual per capita income and WT is the average annual working hours. Where, TC_2 is the unit time cost converted by production method, where GDP is GDP and WN is the annual fixed number of workers.

Assuming that the proportion of people without fixed jobs in a region is ε , then the proportion of people with fixed jobs is $1 - \varepsilon$, then the average unit time cost of the region is $C_2 = \varepsilon TC_1 + (1 - \varepsilon) TC_2$.

To sum up, the final objective function is as follows:

$$\min \sum_{i \in V_1} \sum_{j \in W_i} \alpha_{ij} q_i x_{ij} + \sum_{i \in V_1} \sum_{k \in L_i} \xi_{ik} x_{ik} q_i + \sum_{i \in V_1} \sum_{j \in W_i} \beta_j q_i x_{ij} + \sum_{i \in V_1} \sum_{k \in L_i} \eta_k q_i x_{ik} + \sum_{i \in V_1} \sum_{k \in L_i} \theta_k q_i x_{ik} + \sum_{j \in V_2} \gamma_j y_j + \sum_{k \in L_i} \sigma_k g_k$$

Among them, the travel cost per unit citizen is composed of transportation cost and the time cost of participating in the detection $\alpha_{ij} = 2 * d_{ij}^* C_1 + \left(\frac{2 * d_{ij}^*}{v} + t_1\right) C_2$, the cost of nucleic acid Collection $\xi_{ik} =$

$\left(\frac{2 * d_{ik}^*}{v} + t_1\right) C_2$, the cost of testing supplies per unit $\beta_j = C_3 d_{oj}^r$, the cost of transporting samples nearby per

unit $\eta_k = C_4 \min\left[(1 - y_j)M + d_{kj}^g\right]$, the cost of sampling supplies per unit $\theta_k = C_5 d_{ok}^r$, the investment cost

of testing points $\gamma_j = C_6 + z_j C_7$, and the investment cost of sampling points $\sigma_k = C_8 + r_k C_9$.

(2) Constraints

All people will be detected for constraints.

$$\sum_{j \in W_i} x_{ij} + \sum_{k \in L_i} x_{ik} = 1, \quad i \in V_i$$

Ensure that only the selected checkpoint is open for detection.

$$x_{ij} \leq y_j, \quad i \in V_1, j \in W_i$$

Ensure that only the selected detection points are open for collection.

$$x_{ik} \leq g_k, \quad i \in V_1, k \in L_i$$

Detection capability constraints of checkpoints.

$$\sum_{i \in V_i} q_i x_{ij} \leq F_j, \quad j \in W_i$$

Acquisition capacity constraints of acquisition points.

$$\sum_{i \in V_i} q_i x_{ik} \leq F_k, \quad k \in L_i$$

Calculation formula of detection capability.

$$F_j = z_j * n_1$$

Calculation formula of acquisition capacity.

$$F_k = z_k * n_2$$

Variable nonnegative constraint

$$x_{ij}, x_{ik} \geq 0,$$

0-1 constraint, when j detection point is selected, y_j is 1, otherwise it is 0; when k acquisition point is selected, g_k is 1, otherwise it is 0.

$$y_j \in \{0,1\}, g_k \in \{0,1\}$$

4. EXAMPLE TEST

4.1 Parameter setting

It is modified in order to verify the effectiveness of the model based on the medium-sized random example designed in literature. Suppose that 30 cells are randomly distributed in a 70km × 80km area, one cell is randomly selected as the high-risk area (the cell No. 25 is selected in this paper), and 10 test points are randomly distributed. Four sampling points are generated within 3km around the high-risk cell 25, and the reagent distribution center is at (60, 10), as shown in the Figure 1. It is required to complete the full detection within three days. In this example, Euclidean distance is used to calculate the distance between points. The coordinates and population of the plot are shown in the Table 1, and the coordinates and detection capacity of the selected points of the detection point are shown in the Table 2. The detection point within 25km from the cell can be used as the alternative detection point of the cell. In addition, t_1 takes 30 minutes, c_1 takes 1.5 yuan / km, c_2 takes 0.2 yuan / km, c_3 takes 0.2 yuan / km, c_4 takes 0.2 yuan / km, c_5 takes 2 yuan / person, c_7 takes 200 yuan / person, c_9 takes 200 yuan / person, c_{10} takes 50 yuan / piece, c_{11} takes 10 yuan / piece.

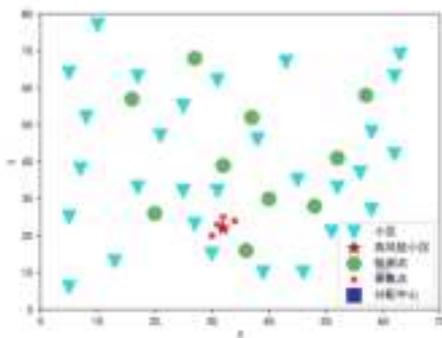


Figure 1 Distribution of residential area, detection point and acquisition point

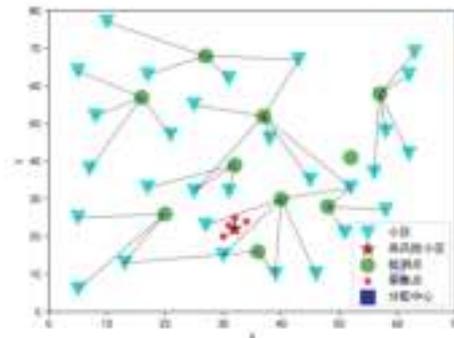


Figure 2 Site selection results

Table 1. Parameters of crowd gathering point

District number	coordinate /KM	population /10000 people	District number	coordinate /KM	population /10000 people
1	(31,32)	1.4	16	(52,33)	0.4
2	(38,46)	1.1	17	(58,27)	1.3
3	(31,62)	1.7	18	(5,6)	0.7
4	(25,55)	0.7	19	(13,13)	1.2
5	(17,63)	0.6	20	(62,42)	0.3
6	(25,32)	1.2	21	(63,69)	1.4
7	(5,64)	0.3	22	(21,47)	1.0
8	(8,52)	1.3	23	(17,33)	0.4
9	(5,25)	0.6	24	(7,38)	1.3
10	(10,77)	0.5	25	(32,22)	0.7
11	(45,35)	0.8	26	(58,48)	0.4
12	(27,23)	0.7	27	(51,21)	1.1
13	(56,37)	0.3	28	(62,63)	0.4
14	(39,10)	1.5	29	(30,15)	1.2
15	(46,10)	0.2	30	(43,67)	0.5

Table 2. Parameters of alternative test points

number	coordinate /KM	Daily detection capability	Three day detection capability	Construction cost / 10000 yuan	Number of personnel
1	(16,57)	1.3	3.9	130	13
2	(20,26)	0.6	1.8	60	6
3	(27,68)	1.2	3.6	120	12
4	(32,39)	0.9	2.7	90	9
5	(36,16)	0.4	1.2	40	4
6	(40,30)	1.1	3.3	110	11
7	(37,52)	1.1	3.3	110	11
8	(48,28)	0.8	2.4	80	8
9	(52,41)	1.5	4.5	150	15
10	(57,58)	1.1	3.3	110	11

Table 3. Parameters of alternative sampling points

number	coordinate /KM	Daily detection capability	Number of personnel	Construction cost / 10000 yuan
1	(32,25)	4200	6	2.1
2	(34,24)	3800	5	1.9
3	(30,20)	3600	5	1.8
4	(31,23)	2000	4	1.0

4.2 Result analysis

Through lingo programming, this kind of problem is mixed integer programming, and the minimum social cost is 3.970685×10^7 yuan. The results showed that the alternative test point 9 was not selected, and the other points were selected to open. Table 4 shows the results of population proportion from each population gathering point to each open test point and from high-risk areas to each open collection point. Table 5 shows the

unit transportation cost between the acquisition point and the nearest open detection point. It can be seen from Table 4 that the selected candidate can make the demand of each small area covered and met, and there is a small amount of excess detection capacity in No. 3 and No. 10 detection points, which can be arranged as the emergency response capacity to deal with emergencies, and the overall utilization rate of the candidate points is 97.25%; acquisition points 1 and 3 are selected and open, and point 1 operates at full load, and point 3 has a surplus of 800 person times. The utilization rate was 77.8%. It can be seen from Table 5 that the samples collected from collection point 1 are sent to the sixth detection point for nucleic acid detection, while the samples collected from collection point 3 are sent to detection point 5 for nucleic acid detection. From Figure 2, we can see the corresponding cell situation of each detection point.

Table 4. The results of the ratio of alternative test points to population

number	Inspection point									Collection point	
	1	2	3	4	5	6	7	8	10	1	3
1				1.00							
2							1.00				
3			1.00								
4							1.00				
5			1.00								
6				0.71			0.29				
7	1.00										
8	1.00										
9		1.00									
10			1.00								
11							1.00				
12						1.00					
13									1.00		
14					0.31	0.69					
15						1.00					
16						0.60	0.33	0.07			
17								1.00			
18		1.00									
19		0.39			0.61						
20									1.00		
21									1.00		
22	1.00										
23				1.00							
24	1.00										
25										0.6	0.4
26									1.00		
27								1.00			
28									1.00		
29						1.00					
30			0.39				0.61				
Total	3.9	1.8	3.6	2.7	1.2	3.3	3.3	2.4	3.3	0.42	0.28
Surplus	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.1	0	0.08

Table 5. Unit cost of transportation near the collection point and results of corresponding detection point

η_k	1	2	3	4
Unit freight	1.88	1.64	1.44	1.72
Corresponding detection point	6	5	5	5

5. CONCLUSIONS

Aiming at the location problem of nucleic acid detection institutions in emergency management of sudden epidemic situation, differentiated management is carried out for high-risk and medium low-risk areas combined with practical problems. In this paper, a distribution model based on coverage model is constructed with the goal of comprehensive coverage of residential areas and minimum social cost. A specific example is taken to optimize the solution. The commercial solution software lingo and CPLEX are used to solve the problem. Don't solve, get the final optimization result. Thus, the feasibility of the model is verified, which provides ideas and methods for the planning and layout of nucleic acid testing institutions of urban emergency management departments, and provides reference for the decision-making of relevant departments to deal with the epidemic situation.

ACKNOWLEDGEMENT

This research was supported by Guangdong Tobacco Monopoly Bureau (Company) science and technology projects (Contract No.201944000200036).

REFERENCES

- [1] Toregas C.ReVelle C. (1972). Optimal location under time or distance constraints. Papers of the Regional Science Association, 28:133-143.
- [2] Yu Dongmei, Gao Leifu, Zhao Shijie. (2020). Research on reliability location model considering interruption risk and defense of emergency facilities. Operations research and management, 29 (9): 53-61.
- [3] Chen Gang, Fu Jiangyue, he Meiling.(2019). Study on the location of emergency shelters considering residents' choice behavior. Operations research and management, 28 (9): 6-14.
- [4] Sun Huali, Xiang Meikang, Xue Yaofeng. (2019). Robust optimization of emergency facility location and path under uncertain information [J]. Journal of systems management, 28 (6): 1126-1133.
- [5] Sun Jianping, Gong Suping, Tang Zhaoping, Liu Huan. (2019). Site selection optimization of high-speed rail emergency resource reserve based on set coverage model. China Science and technology paper, 12 (19): 2213-2218.
- [6] Chen zhizong, you Jianxin. (2006). Multi objective decision making model for site selection of emergency rescue facilities for major emergencies. Management science, 52(4): 10-14.

Full Research Paper**Research on the Competitiveness and Trade Potential of China-India****Pharmaceutical Trade under the One Belt and One Road Initiative***Lin Ni^{1*}, Shihui Shuai², Wei Li¹*¹School of Economics and Management, China University of Geosciences, Wuhan, 430074, China²School of English and International Studies, Beijing Foreign Studies University, Beijing, 100089, China

Abstract: The implementation of "One Belt And One Road" initiative brings both opportunities and challenges for China to strengthen the pharmaceutical trade cooperation with India. Based on the data of pharmaceutical products trade from 2001 to 2018, this paper examines the pharmaceutical trade competition between China and India, establishes an expanded trade gravity model between China and countries along the "One Belt And One Road" Initiative, and measures the export potential of Chinese pharmaceutical products to India. The results show that China's pharmaceutical trade volume to India is positively affected by GDP of the two countries, Indian population and language, and negatively affected by distance. China's pharmaceutical export trade potential to India as a whole belongs to "potential reshape" type. It is urgent need for the two countries to further strengthen trade cooperation in pharmaceutical products in the context of the global spread of Covid-19 pandemic. Finally, the countermeasures and suggestions to promote the development of China-India pharmaceutical trade are put forward.

Keywords: India, pharmaceutical products, gravity model, co-integration test, trade potential

1. INTRODUCTION

In recent years, the "One Belt And One Road" health cooperation and efforts to create the "Health Silk Road" have strongly promoted the development of the pharmaceutical industry in relevant countries. With the worldwide pandemic of the CoviD-19 epidemic, the international attention to public health and hygiene and the development of pharmaceutical trade will be promoted. China and India are both emerging economies in Asia. The pharmaceutical industry of the two countries has entered the ranks of world powers in terms of manufacturing capacity, and the pharmaceutical trade between the two countries is particularly prominent.

In recent years, domestic and foreign scholars have made many achievements in the study of pharmaceutical trade and Cross-border E-commerce. Relevant literatures are summarized as follows: Regarding China-India pharmaceutical trade, Mai Liyi (2018) compared the differences between China and India in the internationalization mode of generic drugs industry, and summarized the enlightenment of the Indian pharmaceutical internationalization mode to the internationalization of Chinese generic drugs^[1]. Liu Shu (2018) pointed out that the promotion of China-India pharmaceutical industry cooperation is inseparable from the joint efforts of government departments, industrial chambers of commerce and pharmaceutical enterprises, and the key is to select the right cooperation field and path^[2]. Musadiq et al. (2018) pointed out that Indian traditional medicine natural plant products have a high trade status at the global national and regional level through the study of Indian traditional medicine market^[3]. There has been a wealth of research on the impact of reviews presented on e-commerce platforms on consumers' cognitive and behavioral decisions^[4-5]. Scholars have basically reached a consensus on the positive promoting effect of perceived security on consumers' purchasing behavior^[6]. Meskaran et al(2014). summarized the factors affecting consumers' perception of security in e-commerce transactions, and pointed out that e-commerce platform elements would have an impact on buying

* Corresponding author. Email: 972697846@qq.com(Lin Ni), shuaibin@cug.edu.cn(Shihui Shuai).

intention through consumers' perception of security^[7]. Zhang Xiaheng(2017) believed that with cross-border e-commerce spreading globally, India's cross-border e-commerce market is emerging. The development of cross-border e-commerce in India is characterized by huge development potential, rapid growth rate, strong mobile trend, and preference for working hours shopping and payment methods of cash on delivery^[8]. Li Zongwei et al.(2017) regarded that the transaction volume information of e-commerce platforms can convey signals of group purchase and product identification to consumers, which improves consumers' perceived value of products sold on e-commerce platforms^[9]. Zhang Xiaodong(2018)found that logistics capability is an important clue for consumers to identify cross-border e-commerce brands, and can help consumers form brand preferences for cross-border e-commerce^[10]. Existing research results show that there are few research results on quantitative measurement of the trade competitiveness of Chinese and Indian pharmaceutical products, and the analysis on the trade potential of Chinese pharmaceutical products exported to India is also relatively lacking. Therefore, In this paper, theoretical analysis and empirical test are combined to analyze the competitiveness of China-India pharmaceutical trade. Trade gravity model is used to analyze the influencing factors of China's pharmaceutical exports to India based on panel data and to measure the trade potential. The research results are helpful to clarify the influence mechanism of China-India pharmaceutical trade, enrich the theoretical analysis framework of related research, and have positive significance for the construction of human health community. At the same time, they can also provide empirical reference for promoting the development of China-India pharmaceutical trade.

2. COMPETITIVENESS ANALYSIS OF CHINA-INDIA PHARMACEUTICAL TRADE

As a major country in the world in terms of the manufacturing capacity of the pharmaceutical industry, the analysis of the competition between China and India in pharmaceutical trade can reveal the trade potential of pharmaceutical products between China and India.

2.1 Classification of medicinal products

According to the characteristics of pharmaceutical products, based on the 4-digit HS code of pharmaceutical products [International Convention for Harmonized Commodity Description and Coding System (Harmonized System)], pharmaceutical products with the code of 30 are divided into the following six 4-digit pharmaceutical products according to HS96, as shown in Table 1.

Table 1. Classification of pharmaceutical products under HS96 standard

The HS code	Category Product Name
3001	Extracts and products from human and animal tissues
3002	Biological products
3003	Western medicine raw materials
3004	Western medicine preparations
3005	Medical dressings
3006	Special medical supplies (including sterile suture materials, X-ray contrast agents, medical boxes, etc.)

2.2 Competitiveness analysis of China-India pharmaceutical trade

The Revealed Comparative Advantage Index (RCA index) is defined as the ratio between the share of the export volume of a certain commodity of a country in its total export volume and the share of the export volume of such commodity in the total world export volume. The calculation formula is as follows:

$$RCA_{ik} = (X_{ik} / X_i) / (X_{wk} / X_w) \quad (1)$$

In the formula, RCA_{ik} represents the explicit comparative advantage index of Class k goods in Country i, X_{ik} represents the export volume of Class k products in Country i, and X_i represents the total export volume of all products in Country i. X_{wk} represents the global exports of Class k products, and X_w represents the total exports of all products in the world. If RCA_{ik} is less than 0.8, it means that country i has weak international competitiveness in category k products. If $0.8 < RCA_{ik} < 1.25$, it means that Country i has certain international competitiveness in category k products. If $1.25 < RCA_{ik} < 2.5$, it means that Country i has a strong international competitiveness in category k products. If $RCA_{ik} > 2.5$, it means that Country i has a strong international competitiveness in category k products [11].

The sample industries selected by China and India are shown in Table 2 and Table 3 for their displayed comparative advantage index under the four-digit classification standard of International Trade Classification (HS).

Table 2. RCA index of six pharmaceutical products in China from 2001 to 2018

	3001	3002	3003	3004	3005	3006
2001	22.97	0.86	1.31	0.47	18.86	0.99
2002	26.29	8.07	0.88	0.44	22.52	1.76
2003	34.57	6.20	0.88	0.44	24.98	2.26
2004	40.13	3.57	0.82	0.44	27.93	2.63
2005	37.13	3.13	0.66	0.45	26.70	2.22
2006	30.13	4.73	1.02	0.46	28.18	3.48
2007	33.32	4.97	1.3	0.49	27.13	3.56
2008	32.34	4.07	1.46	0.47	25.55	3.69
2009	43.04	4.07	1.82	0.42	19.63	4.32
2010	37.40	3.96	3.47	0.38	17.11	3.55
2011	20.16	3.12	3.19	0.55	17.45	2.92
2012	19.48	4.15	3.81	0.61	16.92	3.28
2013	19.55	6.28	4.01	0.64	16.56	3.78
2014	19.67	8.51	3.84	0.67	16.25	4.25
2015	22.74	9.46	4.46	0.67	15.36	4.87
2016	23.89	17.77	4.56	0.67	14.63	5.07
2017	27.76	15.78	3.72	0.73	14.74	4.32
2018	25.84	13.69	3.06	0.79	13.36	3.60

It can be seen from Table 2 that from 2001 to 2018, China has strong international competitiveness in 3001, 3002, 3005 and 3006 pharmaceutical products. In 2010, the RCA_{ik} of category 3003 pharmaceutical products exceeded 2.5, and it gradually became a strong international competitiveness. There are only a few of the 3,004 pharmaceutical products in China that have strong international competitiveness. Pharmaceutical products are generally manufactured in two steps: the drug substance and the drug preparation. China has become the largest producer and exporter of API in the world. In 2018, 80% of China's pharmaceutical exports were APIs, and the largest buyer was India, where China acts as an upstream supplier.

Table 3. RCA index of six pharmaceutical products in India from 2001 to 2018

	3001	3002	3003	3004	3005	3006
2001	1.09	7.72	4.59	0.94	0.49	0.29
2002	1.08	9.87	2.71	0.97	0.48	0.32
2003	1.21	10.18	3.03	0.97	0.51	0.32
2004	1.12	10.21	2.86	1.01	0.42	0.40
2005	1.16	6.28	1.51	1.12	0.57	0.29
2006	0.26	11.95	1.92	1.07	0.55	0.25
2007	1.66	9.32	2.43	1.08	0.54	0.37
2008	0.70	3.77	2.57	1.10	0.45	0.32
2009	0.53	8.75	2.81	1.07	0.45	0.40
2010	0.20	3.45	4.37	1.14	0.59	0.38
2011	0.21	4.33	3.36	1.18	0.53	0.41
2012	0.39	4.42	2.72	1.24	0.27	0.45
2013	0.41	6.00	2.19	1.27	0.35	0.54
2014	0.44	7.16	1.62	1.31	0.31	0.73
2015	0.51	8.30	1.20	1.33	0.27	0.56
2016	0.60	10.69	1.26	1.34	0.26	0.56
2017	0.43	10.73	1.08	1.42	0.28	0.59
2018	0.42	10.10	0.84	1.47	0.29	0.52

The pharmaceutical industry is a pillar of India's economic development. India has helped upgrade the industry with legal protections for generic drugs, one-stop shopping sites for drugs, and advanced business models adopted by pharmaceutical retail chains. For example, some pharmaceutical retail chains combine basic diagnostic functions of clinics with drug-testing services and even a service to transport patients, illustrating the different cultures of pharmaceutical consumption in India and China. It can be seen from Table 3 that India has strong international competitiveness in category 3002 pharmaceutical products, but there is not a big difference in the explicit comparative advantage index between China and India in category 3002 pharmaceutical products. India's international competitiveness in category 3003 pharmaceutical products has declined every year since 2011. In terms of category 3004 pharmaceutical products, the WTO gave India a transitional period of 10 years to "protect pharmaceutical products" after its entry into the WTO. Indian pharmaceutical companies forced to overtake by virtue of their institutional advantages. Up to now, India is undertaking preparations that China is not qualified for, which is conducive to attracting foreign investment and exploring the international market. To sum up, China and India should further strengthen pharmaceutical trade cooperation, scale production of pharmaceutical products with strong international competitiveness, and increase the export of pharmaceutical products with strong international competitiveness. To be specific, China should increase the export of 3001, 3003, 3005 and 3006 pharmaceutical products to India; At the same time, India should step up its export of 3004 categories of pharmaceutical products to China.

3. ANALYSIS ON THE TRADE POTENTIAL OF CHINESE PHARMACEUTICAL PRODUCTS EXPORTED TO INDIA

In this paper, the gravity model is used to calculate and analyze the trade potential of Chinese pharmaceutical products exported to India.

3.1 Analysis on the influencing factors of Chinese pharmaceutical products export to India

3.1.1 Sample selection and data sources

In this paper, the selected 13 "One Belt And One Road" countries includes India, Egypt, Philippines, Thailand, Vietnam, Pakistan, Turkey, Singapore, Poland, Russia, Malaysia, Bangladesh, Indonesia etc. The reasons are as follows: first, China and these countries larger pharmaceutical trade, secondly, these countries basically cover the major cooperative countries of ASEAN, West Asia, South Asia, CIS, Central and Eastern Europe and North Africa along the "One Belt And One Road" route. Their economic, cultural, social systems and living standards are different, which can better reflect the overall characteristics along the "One Belt And One Road". Therefore, the selection of the above countries is very convincing for the estimation and analysis of the trade potential of China's pharmaceutical exports to India. The export data of 30 categories of medicines encoded in HS96 version is from Trade map database, the GDP and total population data of each country are from World Development Indicator database of World Bank, and the geographical distance between China and each country is from French Center for International Economic Research (CEPII).

3.1.2 Model setting and variable selection

Gravity model is an effective tool to study international trade problems. Combined with the actual research content and the availability of data, this paper will use the logarithmic form of gravity model to explore the influencing factors of Chinese pharmaceutical products export to India under the background of "One Belt And One Road". The specific equation is as follows:

$$\ln T_{ijt} = \beta_0 + \beta_1 \ln GDP_{it} + \beta_2 \ln GDP_{jt} + \beta_3 \ln POP_j + \beta_4 \ln DIS_{ij} + \beta_5 FTA_{ij} + \beta_6 LAN_{ij} + \mu \quad (2)$$

In the formula, T_{ijt} the explained variable, refers to the pharmaceutical product export volume between China i and the trading country (or region) j , β_0 represents the constant term, β_k ($k = 1, 2, \dots, 5$) is the regression coefficient of explanatory variables, μ represents the random error term, and the explanatory variables are as follows:

(1) GDP_{it} stands for China's gross domestic product. It reflects China's export supply capacity and also indicates the economic development scale of a country or region. The larger the economic scale, the greater the trade flow.

(2) GDP_{jt} is the gross domestic product of importing country j . It reflects the demand capacity of the importing country. The greater the GDP of the importing country, the greater the trade flow.

(3) POP_j is the population of country j of the importing country. The greater the population, the greater the need for imports.

(4) DIS_{ij} represents the spherical distance between China and the capital of country j . The greater the distance, the higher the transport costs, the less trade flows.

(5) FTA_{ij} indicates whether the two countries have a free trade agreement in force. Free trade agreements reflect the level of regional integration and are represented by dummy variables. 1 indicates that they have entered into force, otherwise 0 is set. The preferential policies of free trade agreements will reduce bilateral trade barriers to some extent and increase trade flows.

(6) LAN_{ij} expresses whether the two countries share a common language. Language affects communication cost, which is represented by dummy variable. 1 means common language, otherwise 0. A smooth language can eliminate more trade barriers, and a common language will promote the development of bilateral trade.

3.1.3 Unit root test

In order to verify the stationarity of the sequence, unit root test is carried out on each variable selected in this paper. In this paper, LLC, IPS, ADF-Fisher and PP-Fisher tests were used to conduct unit root test for the first-order difference of the original data of the explained variables and the main explanatory variables in the model. According to the analysis of the test results, the original data of $\ln GDP_{it}$, $\ln GDP_{jt}$, $\ln POP_i$, $\ln DIS_{ij}$, FTA_{ij} and LAN_{ij} did

not reject the null hypothesis of "unit root existence", which indicated that the original data was non-stationary. Therefore, the unit root test of first-order difference was needed for $\ln GDP_{it}$, $\ln GDP_{jt}$, $\ln POP_i$, $\ln DIS_{ij}$, FTA_{ij} and LAN_{ij} . According to the test results, $\ln GDP_{it}$, $\ln GDP_{jt}$, $\ln POP_i$ passed 1% significance level test, $\ln DIS_{ij}$ passed the significance level test of 5% and 10%. Namely, both the first-order difference reject the null hypothesis and the data are stable, but FTA_{ij} variable raw data and the first order difference are not rejected "unit root" null hypothesis, indicates that the variable is not through stationarity test, thus eliminating the variable.

Table 4. Unit root test results

variable	LLC Inspection	IPS Inspection	ADF-Fisher Inspection	PP-Fisher Inspection
$\ln GDP_{it}$	-2.8767***	-0.3221	15.9231	30.8596*
D ($\ln GDP_{it}$)	-7.9931***	-3.6172***	40.9143***	71.9053***
$\ln GDP_{jt}$	-3.0708***	0.1227	13.8291	29.8140***
D($\ln GDP_{jt}$)	-2.5208***	-3.2634***	21.6166***	39.1392***
$\ln POP_j$	-3.7526***	-0.5987	24.3259	14.2423
D($\ln POP_j$)	-15.5064***	-2.4173***	43.8724***	25.2629***
LAN_{ij}	-4.3182***	-1.1277	21.1132*	21.4461*
D(LAN_{ij})	-7.0317***	-1.8665***	48.4475***	57.6142***
$\ln DIS_{ij}$	-2.7301**	-0.2033	18.7535	25.1509**
D($\ln DIS_{ij}$)	-3.7336***	-1.4077*	22.8182*	25.1563**
FTA_{ij}	0.0813	0.7081	5.1176	30.9358
D(FTA_{ij})	-1.5708*	-0.7275	17.7621	52.0237***

Note: D in the table represents first-order difference. *, ** and *** indicate significant at the level of 10%, 5% and 1%, respectively.

3.1.4 Panel co-integration test

The co-integration test methods can be divided into two categories, one is Engle and Granger two-step test including Pedroni test and Kao test, and the other is Johansen co-integration test. This paper mainly adopts Kao test, and the test results are as follows:

Table 5. Co-integration test results

Inspection methods	the null hypothesis	statistic	t-Statistic	Prob.	conclusion
Kao test	no co-integration relationship	ADF	-3.3695***	0.0000	reject

3.1.5 Empirical results

After passing the stationarity test above, the variables that failed the test were deleted in this paper to get a new gravity model. The regression results are shown in Table 6.

Table 6. Regression results of gravity model

Variable	Coefficient	T Statistics	P value
C	-11.25987	-14.49707	0.0000
$\ln GDP_{it}$	0.942061	14.66546	0.0000
$\ln GDP_{jt}$	0.570741	8.509772	0.0000
$\ln POP_i$	0.125905	2.600167	0.0099
$\ln DIS_{ij}$	-0.677407	-4.575687	0.0000
LAN_{ij}	0.312351	5.921334	0.0000

$R^2=0.791128$ Adjusted $R^2=0.785607$

$F=143.2981$ $p=0.000000$ $DW=2.420551$

Data source: collated according to Eviews output

The regression results show that the determination coefficient of the regression equation R^2 is 0.791128 and the adjusted R^2 is 0.785607 at the 1% confidence level. The model has good goodness-of-fit. The F statistic is 143.2981, and the corresponding p value is 0, indicating that the regression equation has passed the test on the whole, and the independent variable has a high degree of explanation to the dependent variable.

In terms of the practical significance of the regression results of the model, firstly, the coefficient sign of the variable GDP_{it} is positive, indicating that the change of China's GDP will have an important positive impact on the export of pharmaceutical products. Other things being equal, if China's GDP increases by 1%, China's pharmaceutical exports will increase by 0.94%. The bigger China's economy means more pharmaceutical products are produced in China, which boosts exports. Second, the coefficient symbol of the variable GDP_{jt} and is positive, which has a positive impact on the export volume of China's pharmaceutical products. That is, if other conditions remain unchanged, the GDP of the importing country increases by 1%, the export volume of China's pharmaceutical products will increase by 0.57%, indicating that the expansion of the economic scale of the importing country will improve the demand structure of the importing country and increase the import demand for China's pharmaceutical products. In terms of coefficient, China's GDP has a greater impact on the export volume of China's pharmaceutical products than the GDP of importing countries, and China's economic scale and supply capacity are more decisive to a certain extent. Thirdly, the coefficient of variable POP_j is 0.12, indicating that the demographic dividend of importing countries still has a positive impact on China's pharmaceutical export trade. As a developing country with a large population in the world, India has a strong potential demand for pharmaceutical products, and the export trade of pharmaceutical products from China to India is also full of opportunities. Fourth, the coefficient of the variable DIS_{ij} is negative, in line with expectations. Geographic distance impedes China's export trade in pharmaceutical products. In other words, if other conditions remain unchanged, every 1% decrease in the distance between China and its partner countries will increase China's pharmaceutical exports by 0.67%. Therefore, the establishment of better infrastructure and the minimization of transportation costs will provide a boost to pharmaceutical exports. At last. The dummy variable LAN_{ij} has a positive promoting effect on China's pharmaceutical export trade. English is very influential in India, and English is one of the official languages of India. The language advantage of Indian human resources makes the culture of India closer to multinational companies, and reduces the language and cultural barriers in India's international cooperation. At the same time, many senior executives of Indian pharmaceutical companies have overseas working and management background and professional managers familiar with international business cooperation, which forms the comparative advantage of Indian pharmaceutical trade. Especially for cross-border e-commerce, the most difficult problem is communication and payment. As a cross-border e-commerce platform for the first batch of Indian drugs, the operation model of "India's First Pharmacy" is reflected in the presence of a separate operation team for customers. For example, all the customer service staff of the shopping mall for Chinese customers are Chinese nationals or proficient in Chinese. This method effectively avoids the problem of communication and payment. Although the operation cost has increased, the user feedback is the best. China and India should strengthen cultural and linguistic exchanges and promote the development of medical trade.

3.2 Estimation of trade potential of Chinese pharmaceutical products exported to India

For the measurement of export trade potential, In this paper, the method of Liu Qingfeng and Jiang Shuzhu (2002) is adopted to bring the data of explanatory variables over the years into the extended trade gravity model, so as to obtain the simulated value of the export volume in the theoretical state. Then, the actual export volume is divided by the simulated export volume, and the ratio obtained is the trade potential value of the export volume. If the ratio is greater than or equal to 1.20, it is a "potential reshape ", indicating that the existing trade

potential has been exhausted; if the ratio is between 0.80 and 1.20, it is a "potential development" type, indicating that the potential of both sides of trade has not been fully exploited, and there is still room for further expansion of bilateral trade relations. If the ratio is less than or equal to 0.80, it is a "huge potential", indicating that there is a huge trade potential, and the possible factors that hinder the growth of export trade need to be excluded. Accordingly, the parameter data of China and India from 2001 to 2018 were substituted into Equation (2) to obtain the simulated value of China's pharmaceutical products export to India from 2001 to 2018, and then the potential value of export trade was obtained. As shown in Figure 1, 2001-2018 China exports to India's exports is greater than the actual simulation pharmaceutical products, and the ratio of the two is greater than or equal to 1.20, the overall belongs to "potential reshape".

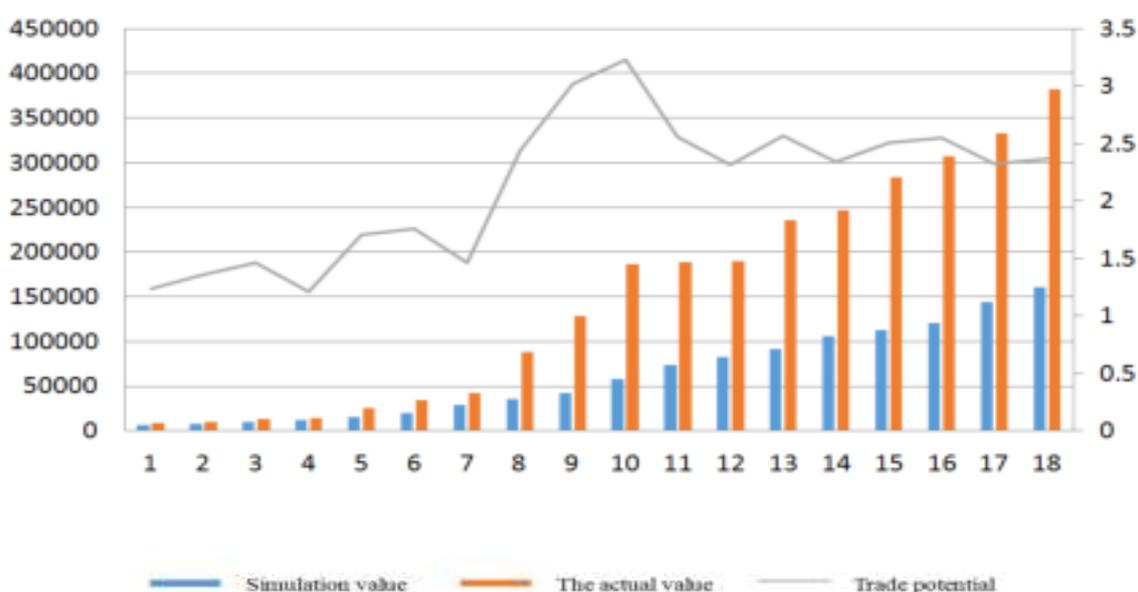


Figure 1. Analog value, actual value and ratio chart of China's pharmaceutical exports to India

Data source: collated according to Eviews output

4.CONCLUSIONS AND SUGGESTION

4.1 Conclusion

First, through the empirical study on the explicit comparative advantage index (RCA) of trade between China and India, it can be concluded that China has strong international competitiveness in 3001, 3002, 3005 and 3006 pharmaceutical products. India, on the other hand, has strong international competitiveness in 3004 and other pharmaceutical products. The differences between China and India further consolidate the foundation of pharmaceutical trade cooperation, and the two countries urgently need to further strengthen economic and trade cooperation and expand the trade scale.

Second, extended Trade Gravity Model was used to empirically test the significance of China's influence factors on India's pharmaceutical exports. The regression analysis shows that the GDP of China and India, the population and language of India promote the export of Chinese pharmaceutical products to India, while the transportation cost represented by the bilateral distance hinders the development of the export trade of Chinese pharmaceutical products to India.

Third, from the results of export trade potential analysis, China's export trade of pharmaceutical products to India as a whole belongs to the "potential reshape" type. This indicates that China needs to develop a multi-level and diversified new trade pattern while maintaining the existing development pattern of pharmaceutical trade.

4.2 Suggestion

In the external environment of rising protectionism, sluggish world economy and shrinking global market, uncertainties and destabilizing factors in global trade have increased significantly, which has made India's diplomatic and military routes increasingly warlike and led to increased trade frictions between China and India.

However, in the face of the most serious pandemic of infectious diseases in the past century, the fundamentals of China-India trade in pharmaceutical products remain sound in the long run. It is more urgent to accelerate the cultivation of new competitive advantages in China-India pharmaceutical trade. Under the current new development pattern of domestic cycle as the main body and domestic and international double cycle mutual promotion, in order to promote the innovative development of pharmaceutical trade and promote the joint construction of high-quality development of "One Belt And One Road", this paper puts forward the following development suggestions:

4.2.1 For government

First, Governments should strengthen exchanges and cooperation. The government should take the "One Belt And One Road" initiative as an opportunity to explore trade liberalization of important medical materials, actively and steadily promote the construction of a "green channel" for goods, maintain the safe and smooth operation of the global industrial chain and supply chain. We will promote high-quality "One Belt and One Road" cooperation and work together to overcome the challenge of the epidemic. Secondly, the government should increase the investment in the pharmaceutical industry, and improve the international competitiveness of China's pharmaceutical products through scientific and technological innovation and digital transformation. The government can fully rely on the Asian Infrastructure Investment Bank to expand China's infrastructure investment in the "One Belt And One Road" countries such as India, promote the construction of the Bangladesh-China-India-Myanmar Economic Corridor, strengthen the construction of digital infrastructure, actively reduce the negative impact of distance factors on the trade of pharmaceutical products between the two countries, and stimulate greater trade potential. Last but not least, we should promote medical culture and people-to-people exchanges between China and India. We should actively cultivate talents proficient in the languages, customs and economic and trade laws of both sides, further improve the quality of our language services, improve the international universality of Chinese, use English as an effective communication tool, and appropriately use Hindi to enhance feelings and cultural identity, facilitate orderly exchanges of personnel, and reduce trade risks.

4.2.2 For companies

First, Chinese and Indian pharmaceutical companies should strengthen cooperation. Cooperation in the field of medicine and health is an important part of the "One Belt and One Road" construction. Chinese pharmaceutical enterprises should make full use of various preferential policies and international resources under the "One Belt and One Road" initiative to improve the quality and growth of China-India pharmaceutical cooperation, diversify the product structure of China-India pharmaceutical trade, and reduce the impact of competition of homogenized pharmaceutical products. At the same time, Chinese pharmaceutical companies should fully understand Indian customs, medical demand preferences and local laws and regulations, customize sales in terms of pharmaceutical product grade, price, trademark and curative effect to expand the scale of pharmaceutical trade. Secondly, enterprises should enhance the modernization level of the industrial chain and supply chain, strengthen the innovation of pharmaceutical products. At the same time, in the production process of pharmaceutical products, enterprises should strengthen supervision to ensure product quality. Finally, Indian cross-border e-commerce enterprises should further reduce the cost of cross-border trade, improve the logistics and distribution system of pharmaceutical products, improve the safety of cross-border pharmaceutical products, conduct new media marketing and social media interaction on the clustering websites of consumers in their

respective market segments, and create first-class consumption experience.

ACKNOWLEDGEMENT

This research was supported by the 2019 General Teaching Research Project of China University of Geosciences (Wuhan), "Optimization and Reconstruction of Teaching Content of International Finance from the Perspective of Curriculum Ideology and Politics" (Project No.2019A25).

REFERENCES

- [1] Mai Liyi, Chen Xin, An Jimeng, Jiang Jie, Yang Guodong.(2018). Comparison of the Generic Internationalization Way between China and India and Its Inspiration to Chinese Pharmaceutical Industry. *Chinese Journal of Pharmaceuticals*, 49(5):690-697(in chinese)
- [2] Liu Shu.(2018). The Belt and Road Initiative and Sino — India Cooperation in the Pharmaceuticals Industry: Prospects, Challenges and Paths Forward. *South Asian Studies*, (1):50-67+158(in chinese)
- [3] Musadiq H. Bhat, Ashok K. Jain, Mufida Fayaz.(2018) . Indian Herbal Drug Industry. Challenges and Future Prospects. *Plant and Human Health*, (1):657-6733
- [4] SIMONSON I.(2016). Imperfect Progress: An Objective Quality Assessment of the Role of User Reviews in Consumer Decision Making, A Commentary on deLanghe, Fernbach, and Lichtenstein. *Journal of Consumer Research*, 42(6):840-845
- [5] MOORE S G.(2015). Attitude Predictability and Helpfulness in Online Reviews: The Role of Explained Actions and Reactions. *Journal of Consumer Research* ,42(1):30-44
- [6] Kim C., Galliers R. D., Shin N. (2012). Factors influencing Internet shopping value and customer repurchase intention. *Electronic Commerce Research and Applications*, 11(4): 374-387.
- [7] Meskaran F., Shanmugamm B., Ismail Z.(2014). Factors affecting on security perception in online purchase intention.*Advanced Science Letters*, 20(10): 2004-2008.
- [8] Zhang Xiaoheng(2017) .The Opportunity for China and India Cross-border E-commerce Market under "The Belt and Road" Strategy. *Journal of Yunnan Open University*,19(1):50-53+58(in chinese)
- [9] Li Zongwei, Zhang Yanhui, Luan Dongqing(2017) . What Factors Influence Consumers' Online Purchasing Decisions? —Customer Perceived Value Drivers. *Management Review*,29(8):136-146(in chinese)
- [10] Zhang Xiao-dong, HE Pan(2018). Study on the Influence Mechanism of Consumer's Brand Preference in Cross-border E-commerce. *Consumer Economics*,34(4):82-89(in chinese)
- [11] Yang Changjian.(2020). Study on Influencing Factors and Potential of China-India Bilateral Trade.Nanchang: Jiangxi University of Finance and Economics(in chinese)

Short Research Papers

Analysis on Public Opinion Sentiment Evolution of COVID-19 Based on Weibo Data

Jiangping Wan*, Xu Liu, Yihang Zuo, Jianfeng Luo

School of Business Administration, South China University of Technology, Guangzhou, 510000, China

Abstract: In this paper, we collected more than 60,000 weibo comments data from 2020 January 20 to December 28, by Python crawler. Subsequently, we used the SnowNLP model based on the naive Bayes algorithm to classify the text corpus with sentiment orientation, analyzed the evolution of epidemic-related topics, and visualized the display from the two dimensions of time and space. On temporal dimension, the emotional attitudes of netizens experienced an anxiety fluctuation period, a stable transition period, and a period of deterioration in public opinion during the beginning of the outbreak (January 20-April 28). The overall emotional attitude of netizens showed negative characteristics. netizens' sentiment experienced a period of rising volatility and steady improvement, with positive sentiment dominating during the normalization phase (May 1st-December 28th). On spatial dimension, we found that there were significant differences in the emotional state and attention of users in different administrative regions with geographic statistical analysis. The more severe the epidemic situation, the higher the topic participation of weibo users and the lower the emotional index. This research provides theoretical reference and event significance for targeted public opinion guidance at the macro level.

Keywords: COVID-19 pandemic, sentiment analysis, evolution of emotional situation, SnowNLP model, Naive Bayes Algorithm

1. INTRODUCTION

Today, social media led by Sina Weibo has gradually become the main venue for the spread of public opinion on hot events and reflecting social conditions and public opinions. Its short text, high interaction and diversified characteristics make it convenient for users to participate in the discussion of hot topics. Taking the weibo topic #Following COVID-19 as an example, by January 20, 2021, the topic had been read 22.87 billion times and discussed 1.99 million times. Netizens' discussions on the COVID-19 have lasted for a long time and were highly popular. At the same time, weibo has accumulated a large amount of user behavior data.. It is conducive to the research of public opinion propagation mechanism, emotional situation evolution, hot topic mining and other aspects by taking the high sequence text as the information base^[1]. Based on weibo data, this paper explores the development stages and evolution process of public opinion with Natural Language Processing (NLP) and sentiment analysis technology, and reveals the evolution rule of public opinion of COVID-19, so as to provide reference for relevant departments to monitor and control online public opinion.

2. LITERATURE REVIEW

Text sentiment analysis plays an important role in revealing the evolution of online public opinion. In recent years, domestic and foreign scholars have successively carried out research on user emotion based on social media. PA et al^[2] collected Twitter platform corpus data and constructed a classifier to classify positive, negative and neutral emotions. Chen Xingshu et al^[3] obtained the hot topics concerned by users at the early stage of the epidemic through clustering analysis of weibo comment data. Han et al.^[4] based on random forest and LDA algorithm to classify the topic of COVID-19 weibo data and explored evolution rules, and proposed early

* Corresponding author. Email: csjpw@scut.edu.cn(Jiangping Wan) , v-xuliu6@Microsoft.com (Xu Liu)

governance strategy. Zhao Yang et al^[5] based on the improved cnn-svm algorithm to perform sentiment analysis and satisfaction scoring on user comments of Haitao APP and provide user feedback and operation strategies for app operators. Ren Zhongjie et al^[6] proposed a weibo public opinion evolution analysis model for emergencies, which divided the development stages of public opinion by defining emotional heat. Huang Faliang^[7] and others integrated emoticons and user features into the topic model LDA, and proposed a weibo topic emotion mining model TSMMF that integrates multi-source features to realize the synchronous analysis of topics and emotions.

Most of the existing researches aim to improve the performance of sentiment classification algorithm, and focus on the evolution process of emotion tendency, which can deepen the research on the development law of public opinion. This paper used Weibo comments as the data source to conduct a temporal and spatial analysis of the evolution of public opinion on COVID-19, and provided references for public opinion supervision departments on a practical level.

3. Data Sources And Research Methods

3.1 Data sources

The comment text of microblog reflects users' subjective views and emotional attitudes towards hot events, and the emotional orientation extraction of comment text is helpful to the research of public opinion trend^[8], the participation of microblog users in hot topics can produce greater public opinion influence than the blog itself^[9], aggravate the spread of events and affect the trend of public opinion. Therefore, we scraped comments on the weibo published daily by the People's Daily and CCTV News about the national epidemic. Taking a weibo content as a target unit and three days as a time node, a total of 62,189 comments were crawled as the basic data of the corpus. Among them, since there were no new local confirmed cases from September to mid October, no data was collected during this period. The content captured by each weibo includes commentator ID, personal homepage URL, comment content, number of likes, number of replies, and comment timestamp, and then crawled the regional information according to the personal homepage URL. The crawled data needs to be preprocessed. First, the duplicate data is removed according to the time dimension, and secondly, empty data and irrelevant data were removed according to the comment content dimension. There were 58,910 valid data after processing.

3.2 Research methods

We constructed a weibo emotional posture evolution model based on the three dimensions of emotional evolution, hot topics, and geographic distribution. Firstly, the comment data of the two official weibo of people's daily and CCTV news are used as the information base, which was stored after cleaning and preprocessing. Secondly, we annotated the text emotion category data and trained the SnowNLP model to complete the classification calculation of emotion orientation. Then, we divided the development stages of public opinion by fusion of emotional tendencies and evolution characteristics, and explored the evolution characteristics of public opinion on the COVID-19 epidemic. Finally, we performed a visual display based on word frequency and spatial dimensions to reveal the evolution of the COVID-19.

3.3 Emotion analysis method

The SnowNLP sentiment analysis model is constructed based on the Naive Bayes algorithm, and its built-in corpus is based on the text of shopping reviews, which has obvious lag. Therefore, it is necessary to replace the positive text pos.txt and negative text neg.txt with the text manually labeled with sentiment orientation. Then train the model and calculate the sentiment tendency value for each comment. The process of emotion judgment is as follows: Calculate the prior probabilities $P(\text{pos})$ and $P(\text{neg})$ of positive and negative emotions through Bayes' theorem, perform text segmentation on the comments to be sentimentally calculated,

and then calculate the posterior probability of each word. $P(\text{Word}|\text{neg})$ and $P(\text{word}|\text{pos})$ are compared to select categories with greater probability. Take the comment "I hope everyone will be safe and healthy, Wuhan, come on" as an example, and mark the positive emotion as P and the negative emotion as N, then:

$$P(P | \text{"hope", "safe", "healthy", "Wuhan", "come on"}) = P(P) * P(\text{"hope", "safe", "healthy", "Wuhan", "come on"}) / P(\text{"hope", "safe", "healthy", "Wuhan", "come on"})$$

$$P(N | \text{"hope", "safe", "healthy", "Wuhan", "come on"}) = P(N) * P(\text{"hope", "safe", "healthy", "Wuhan", "come on"}) / P(\text{"hope", "safe", "healthy", "Wuhan", "come on"})$$

According to the naive Bayes algorithm, the premise is that each feature remains independent. Therefore, assuming that each word is independent of each other, then:

$$P(\text{"hope", "safe", "healthy", "Wuhan", "come on"} | P) = P(\text{"hope"} | s) * P(\text{"safe"} | s) * P(\text{"healthy"} | s) * P(\text{"Wuhan"} | s) * P(\text{"come on"} | s)$$

Finally, the probability $P(P | x)$ of each comment belonging to positive emotion is calculated to measure the emotional tendency of the text according to the snownlp model. The closer its value is to 1, the closer the comment is to positive emotion, and the closer its value to 0, the closer the comment is to negative emotion.

4. Result analysis

4.1 Emotional trends of weibo users

4.1.1 Time series analysis based on emotional mean

The development of the epidemic has entered a normalization stage since April 28 according to the "China's Behavior in Fighting against the COVID-19 Epidemic" white paper which was issued by the Information Office of the State Council. Therefore, we define January 20-April 28 as the initial stage of the outbreak, and April 28-December 28 as the normalization stage of the epidemic. SnowNLP calculates the emotional value and the result is between 0 and 1. It can intuitively show the evolution of the emotional situation throughout the period with calculating the average daily emotional tendency value and drawing a trend graph. It is illustrated in Figure 1 that the evolution of the daily average sentiment value over time in the whole period.

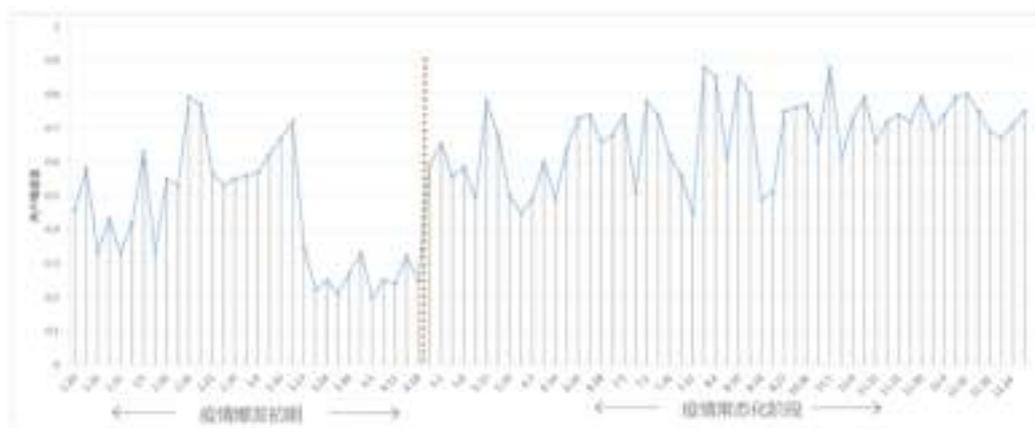


Figure 1. The evolution trend of public sentiment.

It can be seen from the trend graph that in the early stage of the outbreak (January 20 to April 28), the user's emotion experienced a fluctuating process from negative to positive, and then to worsening. The negative emotion concentrated in the two stages of January 20 to February 13 and March 17 to April 21, especially in the latter period. The extreme value of emotion value dropped to 0.2, and lasted for one month. The user's sentiment showed a stable and positive state, with the average sentiment value maintained at a relatively high level and

fluctuations were small after entering the normalization stage of the epidemic (April 28 - December 28).

We found that the reasons for the concentration of negative emotions in the first stage were as follows (January 20 to February 13) based on the top 20 hot comments and the development of the epidemic situation: First, at the beginning of the epidemic, authoritative experts announced that the unknown epidemic was “preventable and controllable”, as well as the wrong punishment of eight medical workers for spreading “rumors”, which led to the large-scale spread of negative emotions in the early stage of the epidemic. Second, the rapid growth of confirmed cases has triggered fear and fear among netizens. COVID-19 has become the most important challenge of our country, and the macro environment has also created a severe and tense atmosphere coupled with the far-reaching impact of the lockdown of Wuhan. Third, netizens questioned the inefficient allocation of materials of Wuhan Red Cross Society, which led to public doubts about the government, government officials and even the system. The above three aspects were the main reasons for the negative emotions in the first stage of the epidemic. The main reason for the continued negativity (April 28 - December 28) of the second stage was the local-related cases caused by imported cases. Related comments accounted for 15 of the TOP20 hot reviews. However, compared with the number of likes and replies in the first stage, we have found that the popularity of public opinion on the epidemic situation in the second stage was much lower than that in the first stage, which was also related to the gradual improvement of the epidemic situation at the macro level and the continuous resumption of work and production. Although there were more negative emotions, the popularity of public opinion has also declined.

4.1.2 Stage division of public opinion based on emotional tendency

(1) At the beginning of the outbreak

The sentiment value greater than 0.6 is defined as positive emotion, and the emotion value less than 0.1 is defined as extreme negative emotion. It was illustrated in Figure 2 the trend chart of the proportion of positive sentiment and extreme negative sentiment at the beginning of the outbreak. The abscissa represents the timeline of the outbreak. The abscissa represents the time line of the outbreak, and the ordinate represents the proportion of emotion. From Figure 2, it can be seen that users' emotions fluctuated greatly in the early stage of the outbreak, and the proportion of positive emotions and extreme negative emotions changed alternately, showing the characteristics of ups and downs.

We have divided the evolution of public opinion into three stages through the comprehensive analysis of the evolution of daily average emotional value, the evolution of public opinion tendency and the discrete coefficient of each period, combined with the life cycle theory of public opinion evolution,. The first stage was the anxiety fluctuation period from January 20 to February 7. During this period, with the outbreak of the early epidemic and the nationwide attention, netizens' emotions show the characteristics of instability. The second stage was the smooth transition period from February 10 to March 13. The existing cases in many provinces have been cleared, and the public opinion situation has also appeared positive emotions occupy an absolute proportion with the unblocking of Wuhan and the gradual improvement of the epidemic situation, the number of newly diagnosed patients has dropped sharply. The third stage was the worsening period of public opinion from March 17 to April 21. This stage did not continue the positive situation of the previous stage. The proportion of extreme negative sentiment was too high, and the average value exceeded 50% during this period. Netizens have illustrated concentrated and persistent dissatisfaction. Since the epidemic has been erupting for more than two months, this period was relatively easy to cause fluctuations in public opinion. The lack of effective official guidance was also an important reason for the deterioration of public opinion.

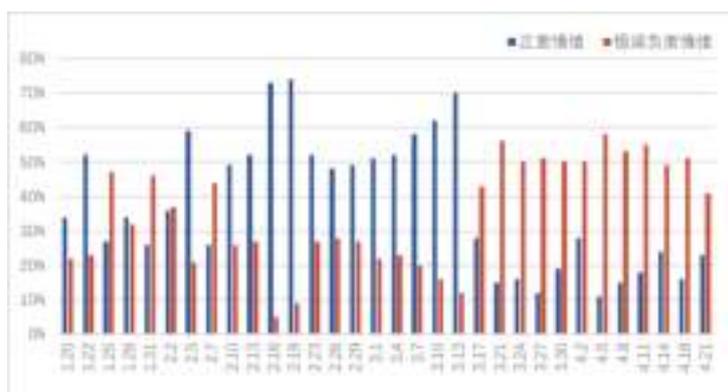


Figure 2. The evolution trend of public sentiment.

(2) Epidemic normalization period

It could be found that the overall sentiment of netizens was significantly higher than that of the initial stage of the outbreak, with small fluctuations and no extreme negatives period by drawing a scatter plot of the daily average sentiment value (see Figure 3 below) and comparing the initial stage of the outbreak. Since there were no new local confirmed cases from September to mid-October, considering the emotional evolution and the development of the epidemic, we have divided the development of public opinion during the normalization period into two stages. The first stage was the rising period of volatility from May 1st to August 27th. The sporadic outbreaks of regional local cases increased in various places, resulting in fluctuations in public opinion during this period. It could be seen in Figure 3 that negative emotions were concentrated in mid-June and mid-to-late July, corresponding to the new outbreak in Beijing and Urumqi in Xinjiang, respectively, but the spread of the epidemic was far weaker than that in Wuhan, so there was no extreme negative Emotional period. The second stage is the stable and positive period from October 13th to December 28th. There was only sporadic local increase in the country and no new regional diagnoses during this period. Therefore, both the number of cases and the impact scope were very small. The sentiment value also reached an unprecedented high and remained stable, with a small degree of dispersion. It illustrated that netizens' acceptance of the epidemic and their emotional value have increased significantly.

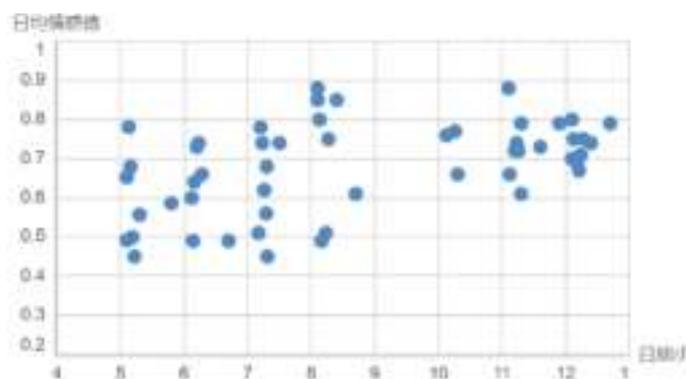


Figure 3. Distribution of daily average sentiment value during the normalization period

4.2 Topics Concerned by Weibo Users

4.2.1 Word frequency statistics

This paper used word frequency statistics to construct a bar chart of relevant vocabulary in the Weibo field from January 20 to December 28 in order to further analyze the hot topics that netizens pay attention to during the epidemic, and explore the causes of the various stages of public opinion.

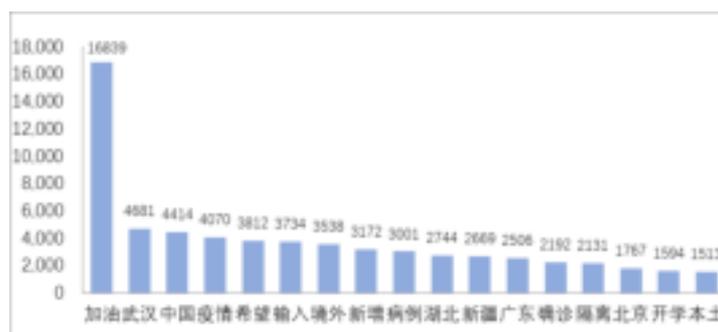


Figure 4. The evolution trend of public sentiment.

As shown in Figure 4, the keyword with the highest frequency was "jiayou", which appears more frequently than other words and represents the overall emotion of netizens. In addition, six geographic information-related words also appeared in the high-frequency words. The COVID-19 spread from Wuhan to the whole country in the early stage. After entering the period of normalization of the epidemic, local and severe epidemics had occurred in Beijing and Xinjiang. Therefore, the above-mentioned areas had attracted widespread attention from netizens. In addition, the vocabulary reflecting the hot spots of netizens' concerns are "yiqing", "shuru", "jingwai", "xinzeng", and "geli". These words reflected netizens' attention to the development of the epidemic and its response measures. Generally speaking, the topics that netizens paid attention to be positive, and there were no negative words in the high-frequency words.

4.2.2 Word frequency evolution

(1) At the beginning of the outbreak

The font size of the word cloud image reflects the number of word frequency. It can be found that in the early stage, the hot topics and the development direction of the epidemic have also changed. At the regional level, the focus of netizens has shifted from Wuhan to Hubei area and then to Guangdong area, the migration law is consistent with the geographical law at the beginning of the epidemic. At the level of the anti-epidemic focus, initially gathering the power of the whole country to support the Hubei region, and gradually turning to preventing the flow of foreign imports from causing a second outbreak of domestic epidemics. The domestic epidemic has basically subsided, but the risks bring by overseas imports have greatly increased. The relevant pictures are illustrated below. We found that in Phase II, the two keywords "jieshu" and "shengli" appeared 450 and 390 times respectively, and neither of these two words appeared in the TOP100 vocabulary of stage I with the number of word frequencies and hot comments. It illustrated that netizens were full of confidence in epidemic control with the stabilization of the national epidemic situation in Phase II. It should be noted that the three keywords of "import", "overseas" appeared 743 times, 565 times, and respectively. It indicated that there were already imported cases in stage II when the domestic epidemic had greatly eased.



Figure 5-1. Phase I word cloud.



Figure 5-2. Phase II word cloud.



Figure 5-3. Phase III word cloud.

(2) Epidemic normalization period

At the regional level, the most serious areas of stage IV epidemics are Xinjiang, Beijing, Jilin and other

regions. With the control of the epidemic, there were no specific areas in the cloud map of stage V, which also reflects that the epidemic has gradually subsided and improved, showing a process from divergence to convergence. In addition, compared with stage III, high-frequency words such as "jingwai" and "shuru" have disappeared in stage IV, indicating that China had successfully controlled the local related cases caused by overseas imports. The high-frequency words appearing in stage V were "fanghu", "xiwang" and "jiayou", reflecting that netizens still attach importance to the epidemic situation at this stage, and the protection problem had become a hot topic of users' attention at this stage. The word cloud picture of the normalization period were shown in Figure 6-1 and Figure 6-2.



Figure 6-1 . Phase IV word cloud.



Figure 6-2 . Phase V word cloud.

4.3 Geographical statistical analysis of the epidemic

We have drawn a nationwide cumulative diagnosis map and a nationwide emotional index map for each administrative region, and the average emotional value of each province was used as the province's Emotional index in order to more intuitively reveal the emotional situation distribution of the COVID-19 from a geographical level with the python-based pyecharts tool. Users in Guangdong, Hubei, and Beijing accounted for 12.43%, 9.51%, and 9.17%, respectively. These areas had more confirmed cases of this epidemic and greater pressure for prevention and control. Secondly, the regions accounted for more than 5% include Heilongjiang, Zhejiang, Shandong, Jiangsu and Shanghai. These regions users had a higher degree of participation in topics. In addition, we also conducted statistics on the sentiment index of various regions across the country and found that Hubei and Guangdong have the lowest sentiment index, 0.553 and 0.579, respectively. These two regions correspond to the two regions with the most local confirmed cases and the greatest impact from overseas imports. The four regions of Tibet, Guizhou, Qinghai, and Ningxia had the highest emotional values. We had found that there is a significant negative correlation between the emotional index of each region and the cumulative number of confirmed cases with geographic statistical analysis, That was, the more severe the epidemic, the more negative the emotional tendency of netizens and the lower the emotional index. Sentiment analysis based on geographic dimensions provides a reference for relevant departments to conduct targeted regional public opinion guidance. The distribution of cumulative number of confirmed cases of COVID-19 and the distribution of the emotional index of netizens in various regions were illustrated in Figure 7 and Figure 8.

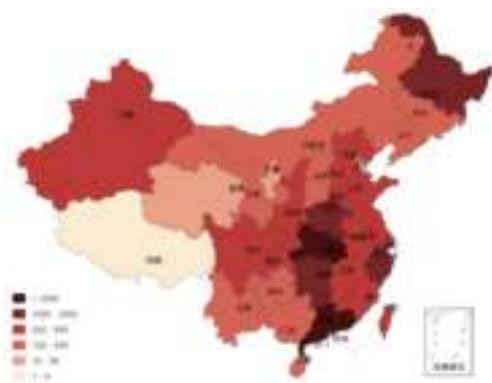


Figure 7 .Cumulative number of confirmed cases

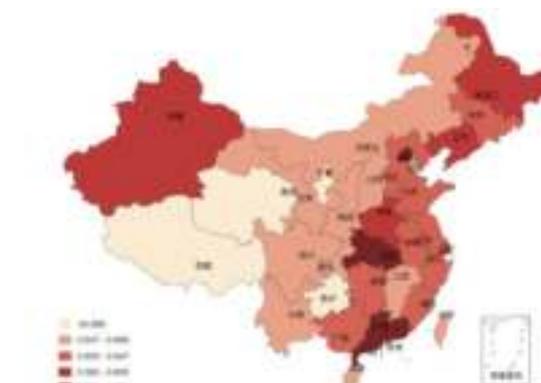


Figure 8 .Sentiment index of all administrative regions

5. CONCLUSIONS

This paper crawled the weibo comment data of People's Daily and CCTV News from 2020 January 20th to December 28th. We combined with the average daily sentiment evolution, sentiment tendency evolution and dispersion coefficient to divided the stages of public opinion development and revealed the evolution process of weibo users' sentiment situation from two dimensions of time and space with SnowNLP. The findings were in the following: (1) During the beginning of the outbreak, user emotions experienced a period of anxiety fluctuations (January 20-February 7), a smooth transition period (February 10-March 13), and a period of deteriorating public sentiment (March 13 -April 21), the overall mood was negative and fluctuating. While entering the period of normalization of the epidemic, user sentiment experienced a period of rising volatility (May 1-August 27), a period of steady improvement (October 13-December 28), and the overall sentiment was more positive and more positive smooth. (2) During the early stage of the outbreak of the epidemic, the region of Internet users' attention showed the characteristics of convergence to divergence, the region of Internet users' attention changed from divergence to convergence during in the normalization stage of the epidemic,. (3) The participation and emotional index of users in each province on the topic of the epidemic were directly related to the degree of development of the epidemic in this province. The more severe the epidemic situation, the higher the user's attention and the lower the emotional index.

ACKNOWLEDGEMENT

This research was supported by Student Research Project (key technology of enterprise knowledge map and its intelligent application) South China University of Technology projects (Granted No. X202010561159).

REFERENCES

- [1] Chen Zi, Gao Tao, Luo nianxue, et al.(2017). Discussion on the effectiveness of social media reflecting the spatial and temporal distribution of natural disasters. *Surveying and Mapping Science*, 42 (8): 44-48129.(in Chinese)
- [2] Pak A, Paroubek P. (2010). Twitter as a corpus for sentiment analysis and opinion mining[C]// International Conference on Language Resources and Evaluation, Valletta, Malta. DBLP: 1320-1326.
- [3] Chen X S et al. (2020). Spatial and temporal analysis on public opinion evolution of epidemic situation about novel coronavirus pneumonia.*Journal of Sichuan University (Natural Science Edition)*, 57(2):409416. (in Chinese)
- [4] Han X,Wang J,Zhang M,et al. (2020). Using social media to mine and analyze public opinion related to COVID-19 in China[J]. *International Journal of Enviromental Research and Public Health*,17(8):2788.
- [5] Zhao Yang, Li Qiqi, Chen Yuhan, et al. (2018). Examining Consumer Reviews of Overseas Shopping APP with Sentiment Analysis[J],*Data Analysis and Knowledge Discovery*, 2(11):19-27. (in Chinese)
- [6] Ren Zhongjie et al. (2019).Analysis of emotional situation evolution of Emergencies Based on microblog data mining -- Taking Tianjin 8 · 12 accident as an example [J]. *Intelligence journal*, 38 (02): 140-148. (in Chinese)
- [7] Huang FL,et al. (2017). Mining topic sentiment in microblogging based on multi-feature fusion[J].*Chinese Journal of Computers*, 40(4):872-888.] . (in Chinese)
- [8] Cao Yanbo. (2018). Analysis on the temporal and spatial characteristics of public opinion of Tonghai earthquake with ms5.0 in Yunnan Province Based on Sina Weibo [J], *earthquake research*, 41 (4): 525-533. (in Chinese)
- [9] Chen Zhiqun, Ju ting. (2020). Analysis of microblog comment tendency based on bet and bidirectional LSTM. *Intelligence theory and practice*: 43 (8): 173-177. (in Chinese)

Full Research Paper

What Motivates People to Share Online Rumors? Deconstructing the Ambiguity of Rumors from a Perspective of Digital Storytelling

Xiangcheng Yang¹, Yi Jiang^{1*}, Wenqian Xu¹, Siqu Guo¹

¹ School of Economics and Management, China University of Geosciences, China

Abstract: With the proliferation of social networks and the development of digital technology, the content structure and propagation mode of rumors have become more complicated with ambiguity, which has greatly influenced people's behaviors when facing digitalized rumors. Based on the digital storytelling theory, this study takes an early initiative by deconstructing and identifying the basic components of online rumors and revealing the conditions under which people's sharing behaviors in a social environment. A data set of health-related rumors related to Covid-19 was used to test the research hypotheses. The results indicated that causality explicitness, element integrality and source explicitness have different influences on rumor sharing behavior. And rumor vividness plays a negative moderating effect during the sharing process. This research offers insight to viewers and website authorities on ways to monitor and debunk online rumors.

Keywords: Online rumors, Digital storytelling theory, Sharing behavior, Covid-19

1. Introduction

With the dramatic change in the media landscape in recent years, rumors, mixed with other real information, can be easily and quickly disseminated on social platforms. People might receive and notice rumors when they use the mobile phone and browse certain-scene information on social networks. The emergence of convenient interaction channels also transformed people from information recipients to spreaders of information. Especially in a health crisis, people with anxiety are more likely to share rumors ^[1], which causes unnecessary panic and confusion. For example, in the period of COVID-19, the rumor "taking Isatis Root Granules, a kind of medicine for clearing heat, would play a role in reducing and eliminating the virus" caused a big amount of discussion on the internet, and then caused people to hoard related medicine for speculation.

Compared with the traditional media, such as newspapers and television, rumors on the internet, which are digitalized, have a more complex content structure and communication modality. Specifically, online rumors are often disseminated with a characteristic of *ambiguity*, by concealing the actual information sources and the total news clue ^[2]. It is one of the core essences of distinguishing rumors from true news. However, this distinction might be unobserved in social networks in that internet users have been used to fragmented online content ^[3]. And then it comes to an opportunity for "rumors" to hide and tamper with part of real information through digital media, which is not easily detected by internet users.

On the other hand, digital technology can reshape and sensationalize the original information by the combination of text, picture and short video. And prior research has indicated that content that uses drama-supporting elements, taking digital media as a carrier, may positively affect engagement and the arousal of emotions, and then create viewers' sharing behavior ^[4]. Thus, rumors, with pictures and videos, may increase the perceived importance and validity of viewers when they face such a package of content. For instance, a video of people snapping up in supermarkets on promotion day was uploaded and shared on social networks during the period of COVID-19, which makes people misunderstand that there may be a food shortage. Obviously, under the

* Corresponding author. Email: wuhanjoey@163.com

help of video, rumors may get more powerful impetus than that with pure text.

Thus, when rumors meet digital technology, the possible results not only show that the speed of rumor spreading will become faster, but also the embedding of digital technology may easily make people lose focus of information by creating stronger ambiguity and concealing the real content.

Previous studies on rumor have argued that the main factors of rumor sharing include the type of rumors, and the personal perception, such as dread or wish rumors, perceived uncertainty, and trust^[5]. Even with the emergence of social media, the research on online rumors just regarded digital technology as an external environment of rumor transmission to check the role of the above factors on online behaviors, such as verification on searching websites, sharing on social platform, and debunking rumors. Meanwhile, much scholarly attention from the information system field has been proved that digital technology, with the label of modularity and combination, reshape the traditional information and content^[6]. The lack of understanding the nature of rumor-digitized makes it difficult to clarify how digital technology can influence the spread of rumors. Obviously, the prior research on the relationship between rumors and sharing behavior will not enough explain this new complex context of digitalized rumors in digital platforms. Hence, this paper seeks to deconstruct the digital rumor to identify different patterns of creating ambiguity by digital technology, and then to explore how they interactively influence the sharing behavior.

To address the above research question, the target is divided into the two-fold objective: the identification of online rumors' structure and the empirical analysis of such components' effect on sharing behavior. Theoretically, the digital storytelling theory (DST) holds that the creation of a new method of weaving story and narrative processing through digital technology could influence the way people think and their decision-making^[7]. This theory identified the three key components of digital content: foundational component (such as thematic characteristics), structural component (such as antecedents and consequences), and presentational component (combined use of video, pictures, and text)^[8]. This study argues that the spread of rumors, in essence, is to cause confusion by reducing the integrity of foundational and structural components of information and combining vivid digital media. Thus, by applying this theory to the rumor context, we identify three types of features of digital rumors: (1) element integrality and source explicitness, (2) causality explicitness, and (3) vividness of digital media as the key influential variables of viewers' sharing behaviors. Further empirically, to test how the factors interactively operate, we collected and analyzed the rumor data sets in Chinese social platforms during the period of COVID-19.

This study extends prior research in several key ways. First, we expand the rumor research, by deconstructing and identifying the basic components of digital rumors, which uncover their interrelated relationship. Secondly, we specially explored the characteristics of the content ambiguity and digital media in comparison with prior research on textual rumor, to reflect effectively the changing principle by the embedded digital tools, revealing the conditions under which people's sharing behaviors in a social environment. Third, our study is the first study to apply digital storytelling theory in the context of rumor research. It also contributes to the research on other digital contents such as live chat, by dissecting the microscopic and potential structure behind the contents. Finally, we provide practical insights and algorithm ideas related to AI-enabled rumor recognition for social platform managers to reduce the possibility of sharing unverified information.

2. Literature Review and Theoretical Development

2.1 Online rumors

The rumor was defined as a widely spread proposition for the belief of topical reference disseminated without official verifications. Compared with traditional rumors, in addition to the same essence, the dissemination medium of online rumors has converted from the oral and printed transmission into the digital media of social

networks. The anonymity and openness of the internet provide viewers with opportunities to share in an unrestricted way, which leads to the faster and wider spread of rumors and more multiple forms of presentation. Especially when encountering social crisis, inevitably accompanied by collective anxiety, improvisational group behavior and adaptive collaboration among the public, online rumors are characterized by high importance and ambiguity^[9]. In this case, people often share rumors on social media to dispel fear, seek confirmation or make up for vague information, which aggravates the harm of rumors.

Prior research has explored the key factors inducing sharing rumors from the perspective of rumor source, content and viewer's characteristics. First, previous studies have pointed out that the credibility of rumor disseminators will influence the sharing behavior^[10], and information with professional news labels is more likely to be shared by others than information with ambiguous labels or without a label. Secondly, in terms of the content of rumors, the importance of rumors, as well as rumor types such as whether rumors are dread-based or wish-based^[2], will also have an impact on viewer's behavior. For example, Chua and Banerjee explored the influence of personal involvement, the rumor type, text length, overhead information and the presence of counter-rumors on rumor sharing decisions^[5]. Lastly, the personality characteristics, such as gender, personality traits, educational background and individual cognitive beliefs, have been proved to be related to rumors sharing and viewers with different cognitive characteristics may show differences in their behaviors of dealing with rumors.

The pervasiveness of digital technologies in almost every industry has led to new forms of social interaction, control and change^[5]. In recent years, rumors have been endowed with new forms of communication, content carriers and features by digital technology. Through various forms of information exchange, rumors can be spread with high traffic on social media websites^[11]. The packaging of digital technologies which affects viewers' perception as never before combined with the influence of the internet complicates the spread and impact of rumors. How online rumors influence viewers' behavior in such situations has converted in an unknown and comprehensive direction.

2.2 Digital storytelling theory

Digital storytelling, through the positive combination of storytelling and digital technology, has become a more creative way of story editing and narrative processing, and a propaganda tool that influences the way people think and behave. The concept of digital storytelling theory can be summarized as expressing traditional stories as digital works by music, videos and animation, which is used extensively in pedagogy, marketing, art, software development and other areas. Based on the digital storytelling theory, social media covers combinations of text, images and videos, and users communicate their actions and ideas through storytelling and generate self-representation^[11]. As an emerging narrative form, digital storytelling can induce people's emotional emergence, prompting them to make a series of corresponding behavioral responses. Researchers believe that in the case of limited personal cognition, storytelling plays a key role in people's decision-making^[12].

Pavlik (2017) identified the three key features of digital narration in the digital media environment: foundational components (such as thematic characteristics, emotional characteristics), structural components (first-person narratives or segmented, or number quotes) and presentational components (combined use of video, pictures, and text)^[13]. Among them, the foundational components is the basic structure of the content, that is, the basic information and main idea that the content conveys to viewers, which can help viewers quickly and comprehensively understand the text content. Structural components demonstrate the narrative structure of content, that is "How content is laid out", which is especially important to assess the trust-building potential of text arguments^[14]. The presentational components focus on "how content is conveyed to consumers". With the gradual development of digital multimedia technology, content creators can use emerging technologies to enhance the display effect of content and convey information to consumers in a more intuitive way.

In social networks, people can participate through like, comment, share, etc., which makes rumors rife.

Sharing has become an essential component to measure rumor propagation. In a broad sense, rumor sharing doesn't mean that people are sharing the rumor on social networking sites, it simply represents the intention of viewers to share news which might lead to rumor propagation. Previous researches failed to deal with the complex situation under the influence of the internet and digital technology to systematically study the narrative characteristics of digital rumors and deconstruct them. Packaging in digital technology, online rumors possess digital storytelling features and take social media as a storytelling space, which gives the digital storytelling theory the ability to make up the gap. In this study, the spread of online rumors is regarded as a process of digital storytelling, the foundational components of online rumors such as story content, story structure and digital presentation form are directly or indirectly narrated through digital stories. We deconstruct the content features of the rumors into three levels: foundational components, structural components, and presentational components based on the perspective of digital storytelling theory.

This study suggests that the ambiguity of online rumors results from the partial absence of the three components under the embedding of digital technology. And it may also influence the perception and next response to the rumors. Thus in the next section, we develop the hypotheses to explore the relationship between the content features of online rumors and rumor sharing behavior.

The foundational components are represented by an underlying message that the content is delivering to the information receiver^[13]. Since one of the foundational characteristics of rumors is that they provide people with information that describes the event, we include element integrality as part of the foundational components. Besides, considering the source of the news conveys to people the credibility of the information, source explicitness is introduced as the other one of the foundational components. Rumor can be considered as a piece of news; it inspires us that we can use the five elements of journalism to model the element carried in the rumor. More precisely, to conclude the overall characteristics of rumors, element integrality is defined as whether there is time, place, character, event, and cause of the rumor. Five elements of rumors can sketch the contours of the object it describes and influence people's relevant sharing behaviors by influencing people's thinking. Rumor is usually modified or omitted for journalism element ambiguity, on the one hand, a rumor which contained fewer journalism elements will be better able to hide the content of falsity, on the other hand, this kind of rumor is more concise in appearance, and its emotional transmission is stronger, viewers in social networks are also more accustomed to this fragmented way of information transmission and more willing to go to share this kind of rumor. Thus, we propose that:

H1. The element integrality of online rumors has a negative impact on rumor sharing behavior.

The source explicitness of rumors is defined as whether the source is explicitly mentioned in the content of the rumor, ignoring whether the source is true or not. Rumor itself is a kind of persuasive information, so the source explicitness of online rumors to a large extent determines the spread effect of the rumor. Information with credible sources contributes to rumor control and high information quality^[15], the source of news is closely related to its credibility, some rumors increase the credibility of their content by making up and showing the source of the news, it helps eliminate the viewer's doubt and uncertainty. When a social crisis occurs, viewers have very limited information to accept and will make decisions based on the behavior of others. When a rumor has an explicit source mark, viewers will have the illusion that it is a piece of real news verified by others viewers are more willing to believe and have the intention to share rumors with explicit sources^[16]. More recent studies have shown that whether the source is in the title will influence the sharing of rumors^[17]. Hence, we propose that:

H2. The source explicitness of online rumors has a positive impact on rumor sharing behavior.

Structural components refer to the narrative approach of the content to trigger consumer feedback^[13]. Rumors convey the causal logic described by the whole rumor information through the narrative mode of its content, which is closely related to the persuasiveness of information. The causality explicitness of rumors is defined as whether

the causal claim has been clarified clearly in rumors. The previous study has shown that the causality explicitness of news affects people's decisions to some extent^[18]. Most health-related news stories make a prominent causal claim include in either the headline or first two sentences. It is these headlines and main claims that are most eye-catching, most shared and that also frame the rest of a story. Due to the ambiguity and arbitrariness of the fabrication of online rumors, the causality of their content often cannot stand scrutiny. Rumors without clearly causal logic could be easily denied by information viewers^[19], and thus not shared. Those well-designed rumors have better causal logic and are more likely to be believed and spread by the viewers. Hence, we propose that:

H3. The causality explicitness of online rumors has a positive impact on rumor sharing behavior.

2.3 Media Richness theory

Media richness theory is proposed by Richard L. Daft and Robert H. Lengel, two organization theorists, to explain how media choice affects the interaction and decision-making between organizations and individuals. With the development of digital technology, information media has shown a trend of diversification. Therefore, media richness theory is also widely used in various new scenes in information system fields, such as virtual teams, online learning, digital deception, and so on. According to the existing literature, in the digital environment, the carriers with different media richness (such as plain text, pictures and videos) will cause the difference in users' perception of information in the process of disseminating knowledge and information. The core idea is that different media divide the richness according to their information-carrying capacity, and the interaction degree between media richness and information content determines the behavior results of viewers.

In this study, the media richness of digital rumor refers to the potential information-carrying capacity of rumor media. The media that can overcome different knowledge backgrounds and make it easier for the information receiver to understand the information are considered to be rich, such as picture and text, while the media that need a long time to understand or provide less information, such as plain text, are considered to be low richness media. High richness media is conducive to the dissemination of complex and ambiguous information, while low richness media is suitable for conventional and easy to understand information. After the deconstruction of digital rumor, the flexible combination of its components will present different content complexity. Under the media interaction of different richness, the effect of each component on user sharing behavior may present different results.

Combined with the vivid description of digital story theory, the media with high richness represents stands for a kind of concept of “telepresence”. In contrast to an unmediated environment, a person in a mediated environment cannot directly feel, touch, taste, or smell an object. Instead, they rely on the medium to create telepresence that represents a direct experience. Vividness refers to the extent to which a mediated environment can simulate a direct sensory experience. content publishers can convey more vivid and real information to consumers utilizing text, pictures, and videos. This diversified media can create a more vivid information dissemination environment, which can help consumers understand the information conveyed by the content more conveniently in the context of media^[14]. Thus, the vividness of rumors is defined as the diversity of presentation forms of in-process propagations that might stimulate the perception of viewers. Vivid information may produce an effect on judgments of a message's general persuasiveness, suggesting that people might infer persuasion based on interest and attention. Moreover, previous research has proved that the vividness of information presentation can interact with variables such as interactivity to influence the decision-making and behavior of viewers^[20].

Previous studies generally believe that high richness media is conducive to complex and high ambiguity information dissemination, while low richness media is suitable for conventional and easy to understand information. However, some studies have found that the influence of media richness on deception behavior and social network information presentation shows an inhibitory mechanism. On the one hand, the media with high media richness may contain redundant information, to distract users from focusing on the core information; on

the other hand, studies have shown that people tend to be overconfident in the expectation of information authenticity in the context of digital deception, and the richer and more diverse the information carried by media will increase the pressure of users' trust, thus inhibiting users' sharing behavior. Therefore, from this perspective, media richness may have an antagonistic moderating effect on users' behavior of sharing digital rumors, and we hypothesize:

H4a. The vividness of online rumors weakens the effect of element integrity on rumor sharing behavior.

H4b. The vividness of online rumors weakens the effect of source explicitness on rumor sharing behavior.

H4c. The vividness of online rumors weakens the effect of causality explicitness on rumor sharing behavior.

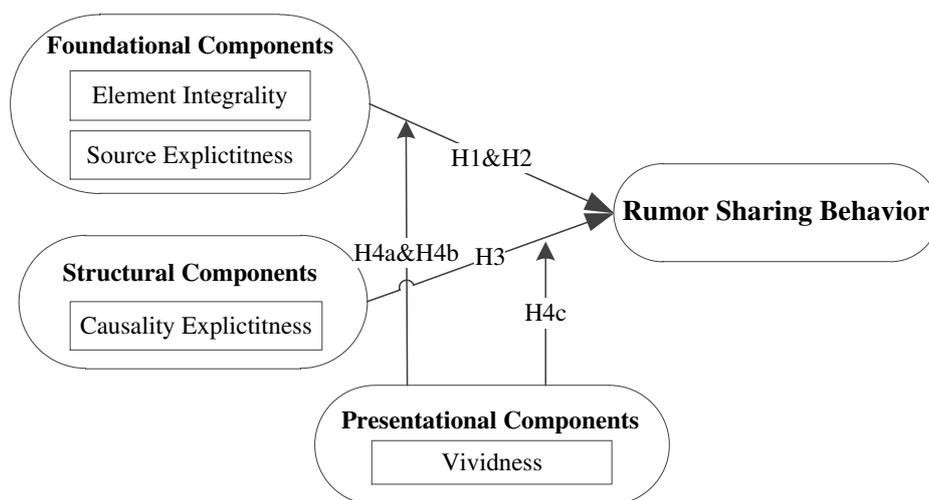


Figure 1. Research model

Based on the above hypotheses, we proposed the research model (see Figure 1), to describe the effects of foundational components and structural components on rumor sharing with the moderation of presentational components.

3. Methodology

3.1 Data collection

Including the title and the refutation date, rumor refutation message about COVID-19 posted on fact.qq.com (a leading news verification website in China) between January 18 and March 15 in the year 2020 were collected. To excavate the characteristics of the rumor, the original text of the rumor is searched with the title of rumors on Weibo, in which users can access various mobile terminals to realize instant information sharing, dissemination, and interaction in the form of multimedia such as text, pictures, and videos.

Information index launched by Baidu refers to the weighted sum of netizens' behaviors such as reading, commenting, forwarding, like and dislike, concerning words. Based on the definition of rumor sharing, the information index of keywords of rumor can reflect the intensity of rumor sharing. Therefore, the information index is taken as the quantitative index to measure the intensity of rumor sharing. The principle of selecting keywords is the keyword that can best represent the event, which means that the information index of the keyword can reflect the intensity of rumor sharing to the greatest extent. The work of keyword extraction was mainly completed by 5 undergraduates, with each of whom independently selected several keywords for each rumor. Moreover, only when all undergraduates choose it, can the selected word be used as the keywords of rumors to ensure the accuracy of keyword extraction.

3.2 Content encoding

Content analysis is a popular technique in feedback research by transforming the meaning of text comments

into objective data using systematic procedures to ensure the objectivity, reproducibility, and reliability of the data analysis. We set up a five-person coding team to encode variables, the encoding process takes the following steps. Firstly, the coding group made judgments according to the encoding rules, and each person was responsible for the coding part of the rumor data. After the preliminary coding was completed, each person checked the codes of other groups and coded for the second time. Finally, the research team further verified the rumor that the verification result was inconsistent with the first encoding result, and the team members voted to decide the result. The encoding rules for each variable are as follows:

Element integrality. We define the elements of rumor as to whether the news includes time, place, character, event, and cause of the rumor. Whether five elements are in the rumor context which we collected from the Weibo platform is checked, encoding result is the count of the elements that appear in the context. For example, if the time, place, and person are stated in the text, the encoding result is 3.

Source explicitness. To encode the rumor source, the original context in the rumor propagation is found in the social platform. If a clear source can be found in the context, it is coded 1, otherwise, it is 0.

Causality explicitness. Causality explicitness of rumor refers to the clarity of the causal logic conveyed by rumors. Common ways to blur causality explicitness in online rumors include the deliberate vagueness of the details of the said object, or reversal of causality. We encode the causality of rumors, with rumors with causal logic coded as 1, and rumors without causal logic coded as 0.

Vividness. The vividness of rumors is used as a moderating variable to influence the sharing of rumors. At present, there are three main forms of news presentation on the Internet, namely text, picture, and video. The vividness in the process of rumor propagation is mainly reflected in these three ways. Thus, the vividness of rumors is measured by the count of the presentation forms used in the rumor propagation.

3.3 Data analysis and results

To understand the influence mechanism of rumors' characteristics on internet viewers' behaviors of sharing rumors, linear regressions were used to estimate the study variables, independent variables and moderating variables in the model are centralized during the regression of moderating effects.

Table 1. Descriptive statistics and Correlation matrix

Variable	V1	V2	V3	V4	V5	SD	Mean
V1 Causality explicitness	1.0000					0.4821	0.3618
V2 Element integrality	0.3312	1.0000				1.5155	2.6447
V3 Source explicitness	0.1673	0.5102	1.0000			0.4781	0.3487
V4 Vividness	0.0071	0.3966	0.3689	1.0000		0.5224	1.3224
V5 Sharing behavior	0.2497	-0.0331	0.1508	-0.0988	1.0000	1.4725	0.2916

152 rumor samples were used for regression analysis, descriptive statistics for variables used in linear regression and their correlation coefficient matrices are shown in Table 1. As shown in Table 1, when rumors were published, many netizens chose to spread rumors, the information index of keywords related to rumors increased by an average of 29.16% month-on-month. In terms of the characteristics of rumors, the causality and source of COVID-19 rumors are generally ambiguous, with average values of 0.3618(SD=0.4821) and 0.3487(SD=0.4781), respectively. Interestingly, the forms of presentation of rumors are not as extensive as expected to spread simpler and more straightforward content, with an average of 1.3224(SD=0.5224) for the vividness of rumors. It can be seen from the table that the rumor sharing behaviors of netizens are correlated to all independent variables to a certain extent, and the correlation coefficients of independent variables are all less than 0.6, so the multicollinearity

problem will not be caused during regression. In order to test whether there is heteroskedasticity in each model, BP test results reject the null hypothesis of homoscedasticity and consider that there is heteroscedasticity ($P < 0.01$). Therefore, OLS with robust standard errors is used to test the model in this paper.

An overview of linear regression results is shown in Table 2, summarizing the main findings of the three models tested, Model 1 only considers the main effect, Model 2 add the moderator to Model 1, Model 3 add the interaction term between the moderator and the independent variable to Model 2. To guarantee the statistical correctness of each model, variance inflation factor (VIF) test is carried out for each model, multicollinearity problems is not a problem, as the maximum VIF index calculated was 1.71 for the research variables, it is far less than the standard value of 10.

Tabel 2. Results of linear regression

Variables	Model 1	Model 2	Model 3
Causality explicitness	0.8956(2.76)**	0.8486 (2.78)**	0.8448(2.84)**
Element integrality	-0.2392(-2.68)**	-0.2040(-2.60)*	-0.1967(-2.68)**
Source explicitness	0.7003(2.14)*	0.7768(2.17)*	0.8627(2.31)*
Vividness		-0.3115(-1.51)	-0.4202(-2.33)*
Causality explicitness × Vividness			-1.101(-2.69)**
Element integrality × Vividness			0.2485(2.25)*
Source explicitness × Vividness			-0.7841(-1.65)
_cons	0.3561(1.88)	0.3099(2.15)*	0.2879(2.49)*
R ²	0.1157	0.1253	0.1737
Mean VIF	1.32	1.45	1.45

Notes: t-statistics in parentheses. ** $p < 0.01$; * $p < 0.05$

The explanatory power (R^2) of Model 1 is 0.1157. In terms of the results of the estimated coefficients in Model 1, as far as foundational component ambiguity is concerned, the number of journalism elements of rumor has a significant negative impact on rumor propagation ($\beta = -0.2392$, $P < 0.01$). Our results prove that the absence of journalism elements will instead promote the spread of rumors, supporting hypothesis H1. On the other hand of foundational component ambiguity, the presence or absence of news sources in rumor content has a significant impact on rumor propagation, and rumors with news sources in rumor content are more likely to be shared ($\beta = 0.7003$, $P < 0.05$). Thus, hypothesis H2 is also accepted. As far as the structural component is concerned, the Causality of rumors has a statistically significant positive impact on rumor sharing behavior and the influence effect is the most obvious ($\beta = 0.8956$, $P < 0.01$), which means that rumors with clearer causality will encourage viewers to carry out rumor sharing, and the hypothesis H3 is supported.

We then examined the moderating effect of vividness through model 3. And Model 3 with moderating effect has a stronger explanatory power and better model fitting effect than Model 1 ($R^2 = 0.1737$). The results demonstrate that rumor vividness exerts a significant moderating impact on the influence of causality explicitness ($\beta = -1.101$, $P < 0.01$) and element integrality ($\beta = 0.2485$, $P < 0.05$) on rumor sharing, while the moderating impact of rumor vividness is stronger for element explicitness. As depicted in Figure 2 and Figure 3, at low levels of rumor vividness, online rumors sharing behavior decreases rapidly when element integrality increases, online rumors sharing behavior rises more rapidly as causality explicitness rises. This means that the vividness of online rumors can weaken the influence of element integrality and causality explicitness on online rumors sharing behavior, H4a and H4c are supported. And the moderating effect of online rumors' vividness on the influence of source explicitness is not significant ($\beta = -0.7841$, $P > 0.05$), H4b is not supported.

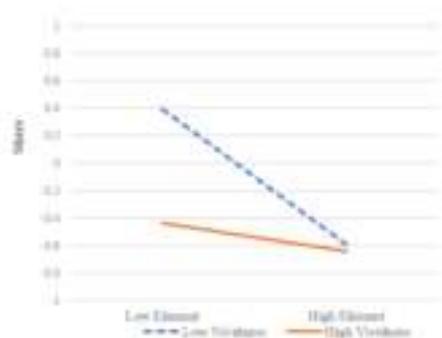


Figure 2. The moderation effect of Vividness on Element

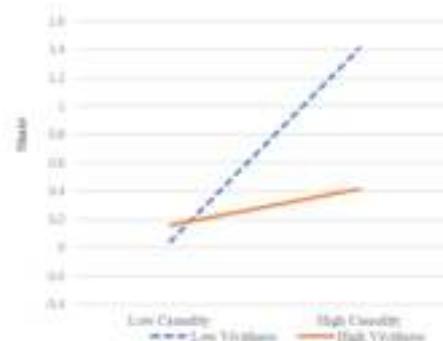


Figure 3. The moderation effect of Vividness on

Causality

4. Discussion and Implications

4.1 Discussion

Drawing on the integration of digital storytelling theory and media richness theory, this study identified the basic components of digital rumors and empirically analyze their interactive relationship on sharing behavior. The results support most of the hypotheses proposed.

First, the results demonstrate that causality explicitness, element integrity and source explicitness exerts great influence on rumor sharing behavior. Among which causality explicitness and source explicitness have positive effects on rumor sharing behavior, while element integrity has negative effects on rumor sharing behavior. Rumors tend to blur the message at different levels, the degree of ambiguity of rumors can influence the sharing behavior of rumors. Mindful individuals are often likely to evaluate the effectiveness of social media content before sharing it. When people are exposed to information that may be a rumor, they tend to assess the information's credibility by assessing its causality ambiguity and source ambiguity. If the causality and source of the rumor are explicit, people are more willing to share the rumor. However, when the element integrity of rumor is high, people are reluctant to spread rumors. There is often a phenomenon that rumor journalism elements are unconsciously transformed and gradually decrease with the propagation. Rumors with complete news elements do not have stronger propagation power. In the information fragmentation era, people prefer a concise message, rumors with concrete and easy to understand are the easiest to become popular, while complex opinions are unlikely to be widely spread.

Secondly, this study sheds light on the moderating effect of rumor vividness in online rumor sharing behavior. Rumor vividness was tested and validated as a negative moderator between element, causality and online rumor sharing behavior. As technology evolves, rumors "recreate facts" in the form of videos or pictures. The iterative updating of media technology means is very rapid, news production and communication channels have been very different from the traditional media era. Simple, inexpensive techniques can make spliced fake videos and images more realistic; facts are "reshaped" so that the audience cannot tell the difference between true and false information, and rumors are no longer just words, but more lurid, frightening, and believable videos or images.

In the digital context, the rich media affect the viewer's perception of the causal logic and element integrity of the content of rumors, weakening the traditional influence path from content ambiguity to communication behavior. Compared with rumors lacking vividness, rumors with high vividness greatly change people's attitude towards rumors^[20]. The vividness of rumor greatly attracts the attention of information seekers and reduces people's thinking about the authenticity of information. The diverse forms of presentation of rumors weaken the influence of the characteristics of rumor content on the rumor sharing behavior.

4.2 Theoretical Implication

With the development of social networks and digital technology, rumors are spreading more rapidly and in more diverse forms. However, previous studies have not considered the interaction between digital technologies and rumors in social networks, which is insufficient to explain the new complex situation of digital rumors in social networks. Our job offers several theoretical implications to the literature.

First, this study provides a new theoretical perspective to investigate the antecedents of online rumors sharing in the digital world. Previous studies on rumors have discussed the propagation of rumors in terms of the types, ambiguities and personality characteristics of the receivers, etc. This study contributes to the literature by further identifying the intrinsic component of online rumors and unfolding how these components affect sharing online rumors. Specifically, the composition of the digital rumor is deconstructed into three parts based on the digital storytelling theory, namely, the foundational components, the structural components, and presentational components, and we found the effects of these attributes on rumor sharing. Our work can be conducive to find the essential fragments of digital rumors which cause the sharing behaviors. Researchers who are interested in detecting the clues of online rumors may refer to this study to assess the specific attributes in the spread of information.

Second, our work extends the application sector of digitalization research by uncovering the traces of digitization in rumor spreading. Yoo et al. (2010) proposed the layered modular architecture to describe one of the critical attributes of a digital system. However, the progression of rumors' digitalization is not precisely reflected in the existing research on online rumors. Related research usually regards digital technology or digital platform as an environment to explore the process of rumor spreading in the digital context^[5]. This study focused on the most three typical changes of rumors' design after rumors are digitalized and shown on the social network. Based on the digital storytelling theory, this study further remodels the relationship between components of digital rumors and rumors sharing. Thus, our empirical findings contribute to alleviating the gap between the practice of rumor digitalization and the existing rumor research.

Third, our findings contribute new insights to the literature by applying the media richness theory into the context of digital rumors. Previous research has indicated that media richness may play different roles in different situations, as most of them argued the positive effects of media richness on performance. Interestingly, the empirical results showed that media richness of rumor spreading could negatively moderate the effects of the element integrality and causality explicitness on online rumor sharing. In other words, the variety of rumors' presentation may restrain the intention of sharing rumors, to a certain extent. In fact, the explanation for the findings in this study has some support in related research, which indicated that media richness may show the opposite effect in the context of digital deception. This study attempts to supplement the research of media richness with new findings and insights by discussing media richness's impact on sharing digital rumors.

4.3 Practical Implications

In the information fragmentation era, the content of social media is more scattered and the process of information deduction is greatly simplified. At the same time, rumor makers are also increasingly good at confusing viewers' judgment through diversified rumor carriers such as pictures and videos. This paper has two practical implications for website administrators, information seekers, and governmental agencies.

First, our empirical findings suggest that short and fragmented news may create ambiguity due to the lack of causality. To make clear the whole story, information seekers need to be vigilant when facing fragmented news and ensure to independently verify sensitive and radical information before sharing. In fact, the openness of the internet makes it easy to trace the source of rumor information. In contact with complex and sensitive social information and communication, the source of information should be actively traced to verify whether the source is fabricated and whether the content is true or false with a rational attitude.

Second, this study provides new ideas for designing the method of rumor detection. Currently, the algorithms for rumor monitoring are based on the evaluation methods such as message content or its sentiment score. Our findings can instruct practitioners of social networking sites to consider adding the new attributes discussed in this study (e.g. length of content, information integrity, media richness) in detection design. Managers should pay attention to verifying the source and authenticity of information and timely release rumor refutation message. Besides, often when the public has a demand for information and the official channels do not release it in time, the rumor will spread quickly. We suggest that government agencies should make information open and transparent, release information of public concern promptly, and squeeze the space of rumor propagation with real information to the greatest extent.

5. Limitation and Further Research

This study has also several limitations. First, all data used in this study are collected in Chinese social networks. However, internet users with different cultural backgrounds or in different stages of COVID-19 may take a different attitude toward rumors. Thus, future research could consider expanding the samples to cover more regions.

Second, considering the interference of rumors' types, most data in this study, which were collected during the COVID-19, are health-related. The influence mechanism of online rumors' characteristics on netizens' sharing behaviors may not be applicable to other types of public security emergencies. In future studies, the sample types and sources could be expanded to further test the universality of this study.

Finally, our study does not explore the intrinsic psychological mechanism of internet users' willingness to share rumors on social network. In future studies, more psychological variables will be introduced to deconstruct the black box of motivation in online rumors sharing.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant No. 71702176 and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (Grant No. 1810491C07 and No. 1910491C08).

REFERENCE

- [1] Oh H J, Lee H. (2019). When Do People Verify and Share Health Rumors on Social Media? The Effects of Message Importance, Health Anxiety, and Health Literacy. *Journal of Health Communication*.
- [2] Bordia P, DiFonzo N. (2004). Problem solving in social interactions on the Internet: Rumor as social cognition. *Social Psychology Quarterly*, 67(1): 33-49.
- [3] Silverman C. (2015). Lies, damn lies, and viral content: How news websites spread (and Debunk) online rumors, unverified claims and misinformation. *Tow Center for Digital Journalism*, 168(4): 134-140.
- [4] Tellis G J, MacInnis D J, Tirunillai S, et al. (2019). What Drives Virality (Sharing) of Online Digital Content? The Critical Role of Information, Emotion, and Brand Prominence. *Journal of Marketing*, 83(4): 1-20.
- [5] Chua A Y K, Banerjee S. (2018). Intentions to trust and share online health rumors: An experiment with medical professionals. *Computers in Human Behavior*, 87: 1-9.
- [6] Yoo Y J, Henfridsson O, Lyytinen K. (2010). The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, 21(4): 724-735.
- [7] Herskovitz S, Crystal M. (2010). The essential brand persona: storytelling and branding. *Journal of business strategy*.
- [8] Bassano C, Barile S, Piciocchi P, et al. (2019). Storytelling about places: Tourism marketing in the digital age. *Cities*, 87: 10-20.

- [9] Oh O, Agrawal M, Rao H R. (2013). Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises. *MIS Quarterly* 37(2): 407-426.
- [10] Liu F, Burton-Jones A, Xu D. (2014). Rumors on social media in disasters: Extending transmission to retransmission. *Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014*.
- [11] Lund N F, Cohen S A, Scarles C. (2018). The power of social media storytelling in destination branding. *Journal of Destination Marketing & Management*, 8: 271-280.
- [12] Herskovitz S, Crystal M. (2010). The essential brand persona: Storytelling and branding. *Journal of Business Strategy* 31(3): 21-28.
- [13] Pavlik J V, Pavlik J O, Understanding Quality in Digital Storytelling: A Theoretically Based Analysis of the Interactive Documentary, in *Digital Transformation in Journalism and News Media: Media Management, Media Convergence and Globalization*, M. Friedrichsen and Y. Kamalipour, Editors. 2017, Springer International Publishing: Cham. p. 381-396.
- [14] Pavlou P A, Dimoka A. (2006). The nature and role of feedback text comments in online marketplaces: implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4): 392-414.
- [15] Oh O, Kwon K, Rao R. An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010. 2010. 231.
- [16] Wathen C N, Burkell J. (2002). Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2): 134-144.
- [17] Kim A, Dennis A R. (2019). Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media. *Mis Quarterly*, 43(3): 1025-1039.
- [18] Adams R, Challenger A, Bratton L, et al. (2019). Claims of causality in health news: A randomised trial. *BMC Medicine*, 17: 91.
- [19] Zubiaga A, Liakata M, Procter R, et al. (2016). Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads. *Plos One*, 11(3).
- [20] Liu J, Li C, Ji Y G, et al. (2017). Like it or not: The Fortune 500's Facebook strategies to generate users' electronic word-of-mouth. *Computers in Human Behavior*, 73: 605-613.

Full Research Paper

An Empirical Study on Inter-firm Knowledge Governance, Knowledge Transfer and Innovation Performance

Xiang Yang¹, Zhao Li^{2*}

¹School of Management, Wuhan Textile University, China

²School of Management, China University of Mining and Technology, China

Abstract: In order to comprehend the impact of knowledge governance upon knowledge transfers and the capacity of amphibious innovation, this paper presents a framework based on emerging Chinese strategic industries. This study has been based on a sample of 323 research staffs and middle-top managers from 107 emerging strategic firms. Structural equation modeling and the hierarchical regression method were used to test our hypothesis. The results show that inter-firm knowledge governance has a significant positive effect on knowledge transfers, and network structures have a negative moderating effect on the relationship between social mechanisms and knowledge transfers; exploitative innovation and explorative innovation have a significant positive effect on market performance and technical performance, and exploitative innovation and explorative innovations have a negative moderating effect on the relationship between knowledge transfers and market performance. The results provide a theoretical and methodological contribution to the study of knowledge governance and can thus improve managerial practice in an area of knowledge-based companies.

Keywords: knowledge governance; knowledge transfer; innovation performance; network structure; amphibious innovation

1. INTRODUCTION

Knowledge governance is choosing the appropriate governance structure and coordination mechanisms to influence the process of knowledge exchange, knowledge utilization, knowledge transfer, knowledge sharing and knowledge creation of inter-firm and intra-firm^[1]. As a new research field which systematically study the intersection of knowledge and organization, it gradually develops the knowledge governance view and becomes the popular issues in the field of knowledge management by degrees. Innovation mode evolution like the emerging of open innovation and scientific and technological progress such as the widely use of the internet and mobile internet embed inter-firm's knowledge activities in a larger and more complex knowledge network, meaning that the participants of knowledge activities are presented as diversified and networked, in which situation institution innovation and mechanism innovation seem particularly crucial, while trying to explore institution innovation and mechanism innovation in the knowledge innovation process between enterprises is just what knowledge governance about. The large industrial documentary films the Heavy Machine of Great Nations records the successful experience of eighteen industry leaders' independent innovation and transformation of development mode in the equipment manufacturing industry of China. Through reviewing enterprises' transformation and upgrading and interviewing the key persons in the process of innovation, it highlights the close relationship between the institution innovation, mechanism innovation and knowledge innovation. Enterprises generally face three main problems in the process of knowledge innovation. The first is how to transfer knowledge resource effectively between firms. Further, how to use knowledge resource effectively to carry on knowledge innovation. The last one is how to produce the synergistic effect of knowledge activities of inter-firm and intra-firm. This study specifically attempt to put forward some suggestions of institution innovation and mechanism innovation for knowledge transfer between firms effectively and some reference of practice for efficient utilizing knowledge resource in firms so as to improve innovation performance,

and in the meantime provide theoretical support for the practice of transforming the economic development mode, optimizing industrial structure and developing strategic emerging industries in China, as well as enrich and extend the knowledge governance view in theory, through the research on knowledge governance, knowledge transfer and innovation performance based on the moderating effect of network structure and amphibious innovation.

2. LITERATURE REVIEW AND RESEARCH HYPOTHESES

2.1 Knowledge governance

The concept of knowledge governance comes from Grandori who considers knowledge governance is the governance specifically to knowledge exchange, transfer and sharing of inter-firm and intra-firm. In recent years, scholars enrich and extend the content of knowledge governance^[2]. Michailova & Foss regarded knowledge governance as choosing the appropriate governance structure and coordination mechanism to influence the process of knowledge exchange, utilization, transfer, sharing and creation of inter-firm and intra-firm, and clearly put forward what knowledge governance study, how to carry on the study and why they research on it. Foss et al considered that knowledge governance is to meet the organizations' expectation by using formal or informal organizational mechanism, structure and so on to influence the process of knowledge transfer, sharing, integration, utilization, and creation. The literature on knowledge governance theme above means a lot to form the knowledge governance view, however, it is a pity that these literature basically adopt the way of literature review and theoretical deduction. There is no doubt that it can be more persuasive with further empirical examine. Therefore, this study tries to make some contributions to the empirical aspects. Research category of enterprise knowledge governance mainly can be divided into two levels, namely inter-firm knowledge and intra-firm knowledge governance^[3]. The main research object of inter-firm knowledge governance mechanism is the diversified participants in the enterprise knowledge network. As to intra-firm knowledge governance mechanism, the main research object is teams and individuals, and the knowledge governance in the team level is basically informal governance, for example, the team atmosphere and team identity, based on the social and cultural factors in Chinese situation.

The research in the field of knowledge management is usually based on the knowledge-based view, and yet knowledge governance view is the new scope in knowledge management field and mainly focus on the function mechanism of organization institution, mechanism, incentive, structure and other elements to enterprise knowledge activities. Knowledge governance view may have more important meaning and value as the participants of knowledge activities in firms present as diversified and networked.

2.2 Inter-firm knowledge governance and knowledge transfer

The integration and design of inter-firm knowledge governance mechanisms are mostly studied from three dimensions: market mechanism, social mechanism and hierarchical mechanism. The studies of market mechanism mainly refer to the method, channel, trade and contract that enterprises gain knowledge resource^[4]. Researches on social mechanism largely cover cooperation intention, communication, commitment, trust, network relationship^[5]. In the case of hierarchical mechanism, it concerns network management method and the attitude of the top managers primarily^[6]. The literatures on inter-firm knowledge governance are numerous, however, each piece of them only involves part of market mechanism, social mechanism and hierarchical mechanism, and insufficient research has been conducted regarding the systematically integration and design of inter-firm knowledge governance mechanism. Therefore, the objective of this study is trying to enrich the related research. This paper adopt the definition of knowledge transfer which means taking advantage of other organizations' knowledge, abilities and skills and emphasize the knowledge transfer to the enterprise from each participant in the knowledge network between enterprises^{[7][8]} This definition includes a lot of activities between

enterprises like the mining and association of knowledge resource, and this fits the characteristics of this research. The existing literature specific to inter-firm knowledge governance mechanism mostly adopts the method of literature review and theoretical deduction. Hence, this study seems extraordinary valuable on the basis of integration. For the above-mentioned analysis, we have the following hypotheses:

H1a: Market mechanism has a positive impact on knowledge transfer.

H1b: Social mechanism has a positive impact on knowledge transfer.

H1c: Hierarchical mechanism has positive impact on knowledge transfer.

2.3 Knowledge governance and knowledge transfer: the moderating effect of network structure

Phelps et al. insisted that knowledge network mainly has the following characteristic elements: structural properties, relational properties, nodal properties and knowledge properties. The existing related study of the inter-firm mainly include: (1) the structural properties, including enterprise network position, ego network position and whole network position^[9]; (2) relational properties, including the impact of relational strength, competition, technology and geographical proximity on knowledge activities^[10]; (3) nodal properties, including the absorption ability, conversion ability, cooperation ability on the enterprise level and the diversity and depth of cooperators' knowledge^[10]. The related literature on knowledge network above-mentioned explain the content and domain of knowledge network primarily and yet network structure is a significant part of knowledge network and provides the theoretical support for this study to investigate the moderation effect in the knowledge situation. Network structure refers to the content of the aspects such as network scale and network centrality. The stronger the network structure, the more the object for enterprises to gain knowledge resource, thereby the greater the impact market mechanism on knowledge transfer could be. The stronger the network structure is, the more distant the relationships between participants and enterprises could be, therefore the smaller the impact social mechanism on knowledge transfer could be. The stronger the network structure is, the more scattered management's focus will be, thus the smaller the impact hierarchical mechanism on knowledge transfer could be. Accordingly, we have the following hypotheses:

H2a: Network structure positively moderates the relationships of market mechanism and knowledge transfer.

H2b: Network structure negatively moderates the relationships of social mechanism and knowledge transfer.

H2c: Network structure negatively moderates the relationships of hierarchical mechanism and knowledge transfer.

2.4 Knowledge transfer and innovation performance: the moderating effect of amphibious innovation

The study on knowledge transfer and innovation performance between enterprises is relatively mature and presents a nearly consistent result that knowledge transfer has a positive effect on innovation performance. This study focuses on the moderating effect of amphibious innovation, that is, exploitative innovation and explorative innovation, on the relationship between knowledge transfer and innovation performance. On the one hand, with the copy, improve, generalize and promote feature, exploitative innovation is designed to satisfy the existing customers and the current market and turns up as the improvement of product or service, the promotion of process efficiency and so on^[11]. On the other, explorative innovation is characterized by searching, breakthrough, experiment, creation and adventure. It is born to meet the new customers and the new demand, and appears in the form of new product, new service, new flow, etc^[12]. The content of amphibious innovation mentioned above covers the definition the existing related research on amphibious innovation in principle and is fairly considerate. Although the study on the relationships of amphibious innovation and innovation performance is pretty much, the study based on the subdivision of market performance and technical performance remains further explored. Besides, there is a big difference between market performance and technical performance, and it deserves a

deep-going inquiry. Therefore, we have the following hypotheses:

H3a: Exploitative innovation has a positive impact on market performance.

H3b: Exploitative innovation has a positive impact on technical performance.

H3c: Explorative innovation has a positive impact on market performance.

H3d: Explorative innovation has a positive impact on technical performance.

Wang et al. studied the equilibrium effect among exploitative innovation and explorative innovation, however, their interaction with knowledge transfer was not tested^[13]. Knowledge transfer is the tie of knowledge activities between enterprises and has an important influence on enterprise's innovation activities; while amphibious innovation is the crucial part of organizational learning and knowledge activities in firms and plays a vital role for enterprise's innovation activities as well. At the present stage, how is the synergistic effect between knowledge transfer and amphibious innovation? It needs exploratory analysis, since knowledge transfer belongs to the knowledge activities of inter-firm and jet amphibious innovation is a kind of knowledge activities of intra-firm. So the question that whether there is a synergistic effect in the knowledge activities of inter-firm and intra-firm comes out and the existing literature has little attention on it, so we attempt to put forward the following assumptions:

H4a: Exploitative innovation significantly moderates the relationship between knowledge transfer and market performance.

H4b: Exploitative innovation significantly moderates the relationship between knowledge transfer and technical performance.

H4c: Explorative innovation significantly moderates the relationship between knowledge transfer and market performance.

H4d: Explorative innovation significantly moderates the relationship between knowledge transfer and technical performance.

Based on the theoretical analysis and research hypotheses above, the research framework of this study is shown in Figure 1.

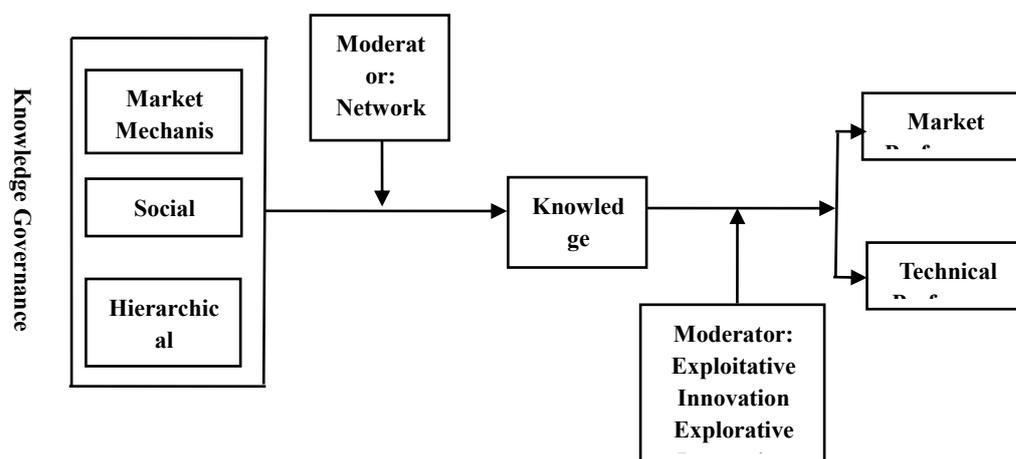


Figure 1. Research framework

3. METHODS

3.1 Variables measurement

Nine latent variables including market mechanism, social mechanism, hierarchical mechanism, knowledge transfer, network structure, exploitative innovation, explorative innovation, market performance and technical performance need to be measured in this study. We used Likert-7 scale for the measures among which 1

represents strongly disagree and on the contrary 7 represents strongly agree.

Market mechanism, social mechanism and hierarchical mechanism in our research were adapted from Gooderham's market relationship, social relationship and hierarchical relationship^[14]. Knowledge transfer was adapted from the work of Maurer et al^[15]. Network structure was developed by us according to its content. Exploitative innovation and explorative innovation were adapted from Lubatkin et al.^[16]. Items of market performance and technical performance were measured based on the integration of related literature. Furthermore, enterprise age, enterprise scale and industry were included in our study as control variables.

To ensure the content validity and fitness of the questionnaire, we invited two management professors and six management postgraduates to check all the measurement items of the questionnaire, and modified it in accordance with their opinions. After that, we chose 20 related persons distribution in 5 firms to conduct the pre-test and in-depth interviews to complete the questionnaire. Finally, we conducted an official survey among the chosen sample.

3.2 Data collection and sample characteristics

We conducted a questionnaire survey on 485 R & D personnel and senior and middle managers from 107 strategic emerging firms in Wuhan city circle and Pearl River Delta regions, which are national innovation demonstration zone in China. The 107 firms have certain industry representativeness since they widely distributed in seven strategic emerging industries including new-generation information technology, high-end equipment manufacturing, advanced materials, alternative-fuel cars, energy-saving and environmental protection, alternative energy and biotechnology, which are selected by the State Council of China. The reason this study chose the strategic emerging firms in Wuhan city circle and Pearl River Delta regions as the sample is that: (1) Strategic emerging industry is suitable for the requirements of this study because of its characteristics of knowledge intensive. (2) Strategic emerging industry is the leading force to transform the pattern of economic development and realize industrial upgrading and has the important role of demonstration which promotes the practical significance of this study. Besides, the strategic emerging industry in Wuhan city circle and Pearl River Delta regions exhibit a robust momentum of development. (3) This study team has always focused on the growth of strategic emerging industry among which there are abundant social resources and especially the powerful alumnus resources in Wuhan city circle and Pearl River Delta regions make this study feasible. Taken together, the sample of this study is not only available to the requirements of the study but also representative. To ensure that the data can comprehensively reflect the enterprise conditions objectively, about five R & D personnel and senior and middle managers in each firm were investigated based on the full communication with the firms. We sent out 485 questionnaires in total and received 374 responses of which 323 responses were valid, and the valid rate was 66.598%. Table 1 displays the descriptive information of the data set. According to Table 1, the majority of the subjects were male and the proportion is as high as 72.755 %. The working years were almost more than five. The vast majority of the respondents were senior and middle managers. The enterprise ages exceeded five years largely, and more than 50% had a scale above 300 persons. The enterprises distributed in seven strategic emerging industries uniformly, and there was little difference between samples of Wuhan city circle and Pearl River Delta regions. Based on these comparisons, the characteristics of the sample lived up to the requirements of this study.

Table 1. Basic characteristics statistics of the sample (N=323)

Variables	Item	Count	%
Gender	Male	235	72.755
	Female	88	27.245

Firm Age (years)	≤5	31	9.597
	>5 and ≤10	58	17.957
	>10 and ≤15	115	35.604
	>15	119	36.842
Firm Industry	energy conservation and environmental protection	44	13.622
	new materials	55	17.028
	biological medicine	61	18.885
	high-end equipment manufacturing	48	14.861
	new energy	40	12.384
	new generation of information technology	55	17.028
Working years	≤3	44	13.622
	>3 and ≤6	68	21.053
	>6 and ≤10	137	42.415
	>10	74	22.910
Firm Scale (persons)	≤100	11	3.406
	>100 and ≤300	92	28.483
	>300 and ≤500	119	36.842
	>500	101	31.269
Occupation	Top managers	30	9.288
	Middle managers	208	64.396
	Ordinary staff	85	26.316
Area	Wuhan city circle	168	52.012
	Pearl River Delta regions	155	47.988

3.3 Discriminant validity test

Table 2. Descriptive statistics and discriminant validity test

Variables	1	2	3	4	5	6	7	8	9	10	11	12
Firm Ages	1											
Firm Scale	0.483*	1										
Industry	0.027	0.036	1									
Market Mechanism	-0.140+	-0.074	0.071	1								
Social	-0.171+	-0.158*	0.137+	0.478*	1							
Hierarchical	-0.115+	-0.098	0.068	0.330*	0.505*	1						
Knowledge	-0.060	-0.116+	0.059	0.388*	0.500*	0.501*	1					
Network	-0.060	0.000	0.029	0.195*	0.301*	0.352*	0.390*	1				
Exploitive	-0.136+	-0.072	0.025	0.435*	0.470*	0.447*	0.562*	0.399*	1			
Explorative	-0.102	-0.024	0.086	0.402*	0.443*	0.486*	0.455*	0.296*	0.570*	1		
Market	0.126+	0.010	-0.053	0.239*	0.174*	0.192*	0.352*	0.315*	0.430*	0.443*	1	
Technical	0.028	0.038	-0.027	0.278*	0.288*	0.313*	0.348*	0.291*	0.449*	0.502*	0.705*	1
Mean	2.997	2.960	3.712	5.221	5.397	5.322	5.435	5.179	5.500	5.492	5.319	5.560
Square	0.967	0.857	1.840	0.728	0.767	0.667	0.697	0.705	0.707	0.728	0.643	0.703

Note: * means significant at the $p < 0.001$ level; + stands for significant at the $p < 0.05$ level.

4. DATA ANALYSIS AND HYPOTHESIS TESTING

4.1 Reliability and validity

Firstly, we tested the reliability of the nine variables using SPSS 17.0 and it turned out that the Cronbach's α of each variable is greater than 0.7, which means the questionnaire were reliable. Secondly, we employed AMOS 7.0 to carry on the CFA. The standardized coefficient of each factor was above 0.5, which was significant at the 0.001 level, indicating that the questionnaire had a good validity. Table 3 depicts the result of reliability and validity testing.

Table 3. Test of reliability and validity

Variables	Item	Standardized Coefficient	α	Construct
Market Mechanism (MM)	MM1	0.685	0.743	
	MM2	0.732		
	MM3	0.577		
	MM4	0.595		
Social Mechanism (SM)	SM1	0.780	0.819	Gooderham et al.
	SM2	0.755		
	SM3	0.700		
	SM4	0.683		
Hierarchical Mechanism (HM)	HM1	0.763	0.815	
	HM2	0.773		
	HM3	0.646		
	HM4	0.715		
Knowledge Transfer (KT)	KT1	0.755	0.829	Maurer et al.
	KT2	0.774		
	KT3	0.696		
	KT4	0.739		
Network Structure (NS)	NS1	0.836	0.851	Self-developed
	NS2	0.735		
	NS3	0.865		
Exploitive Innovation (ELI)	ELI1	0.764	0.874	
	ELI2	0.839		
	ELI3	0.761		
	ELI4	0.818		
Explorative Innovation (EXI)	EXI1	0.822	0.878	Lubatkin et al.
	EXI2	0.779		
	EXI3	0.787		
	EXI4	0.818		
Market Performance (MP)	MP1	0.770	0.818	
	MP2	0.840		
	MP3	0.721		
Technical Performance (TP)	TP1	0.825	0.792	Self-developed
	TP2	0.696		
	TP3	0.726		
Interaction of ELI and KT (IELI)	IEL1:ELI1×KT1	0.658	0.844	Measurement of Interaction
	IEL2:ELI2×KT2	0.834		

	IEL3:ELI3×KT3	0.815	
	IEL4:ELI4×KT4	0.744	
	IEX1:EXI1×KT1	0.655	
Interaction of EXI and KT (IEXI)	IEX2:EXI2×KT2	0.801	0.819
	IEX3:EXI3×KT3	0.738	
	IEX4:EXI4×KT4	0.736	

Note: the Standardized Coefficients were significant at the 0.001 level; specific content of each variable item in Table 3 is shown in appendix.

4.2 Hierarchical regression analysis

This paper adopted the hierarchical regression analysis to examine the moderating effect of network structure. Before the hierarchical regression, the variables were centralized and the impact of control variables on knowledge transfer was checked through null model. The R2 and adjusted R2 were tiny and the F-value was non-significant, which indicated that the control variables were not the main factors that influence knowledge transfer. Therefore, the control variables were not included in the hierarchical regression model. The hierarchical regression result was exhibited in Table 4.

Table 4. Result of hierarchical regression analysis

	Knowledge Transfer			
	Model 1	Model 2	Model 3	Model 4
Constant	-	-	-	-
Independent Variables				
MM	0.157**	0.322***		
SM	0.265***		0.408***	
HM	0.315***			0.412***
NS		0.329***	0.292***	0.246***
Interaction				
MM*NS		-0.010		
SM*NS			-0.112**	
HM*NS				-0.022
R2	0.351	0.254	0.325	0.304
adjusted R2	0.345	0.246	0.319	0.297
F-value	57.609***	36.112***	51.166***	46.429***

Note: *** means significant at the 0.001 level and ** significant at the 0.050 level.

The result of hierarchical regression demonstrated that: (1) Market mechanism, social mechanism and hierarchical mechanism had a positive impact on knowledge transfer, that is to say, the H1a、H1b、H1c were supported. (2) The moderating effect of Network structure on the relationships of market mechanism and knowledge transfer was not significant and H2a was not supported. Network structure negatively moderated the relationships of social mechanism and knowledge transfer and the H2b was confirmed. The moderating effect of Network structure on the relationships of hierarchical mechanism and knowledge transfer was not significant and H2c was not supported.

4.3 Structural equation model

We employed the structural equation model to examine the moderating effect of exploitative innovation

and explorative innovation. According to the data handling advices on variables' interaction of Wen et al. (2003), the original data were dealt through centralized processing. Then the effect of interaction of exploitative innovation and knowledge transfer, and the interaction of explorative innovation and knowledge transfer on innovation performance were examined, through the construct of interaction based on the cross product of the two indicators. The analysis results of structural equation model were presented in Table 5.

Table 5. Analysis results of structural equation model

	Path	Standardized Coefficient	T value	Goodness of fit					
				χ^2	df	χ^2/df	RMSEA	GFI	CFI
Model A	KS→MP	0.118	1.208						
	ELI→MP	0.437***	4.069	136.751	82	1.668	0.046	0.947	0.976
	IEL→MP	-0.156*	-1.848						
Model B	KS→TP	0.094	0.852						
	ELI→TP	0.605***	4.862	138.655	82	1.691	0.046	0.947	0.975
	IEL→TP	0.039	0.413						
Model C	KS→MP	0.187**	2.541						
	EXI→MP	0.329***	4.780	135.607	84	1.614	0.044	0.949	0.976
	IEX→MP	-0.321***	-3.944						
Model D	KS→TP	0.158*	1.918						
	EXI→TP	0.500***	6.427	127.118	83	1.532	0.041	0.951	0.979
	IEX→TP	-0.089	-1.032						

Note: *** means significant at the 0.001 level, ** significant at the 0.050 level and * significant at the 0.100 level.

The result of the structural equation model showed that: (1) Exploitative innovation had a positive impact on market performance; Exploitative innovation had a positive impact on technical performance; Explorative innovation had a positive impact on market performance; Explorative innovation had a positive impact on technical performance. Thus, H3a、H3b、H3c and H3d were supported. (2) Exploitative innovation negatively moderated the relationship between knowledge transfer and market performance; Explorative innovation negatively moderated the relationship between knowledge transfer and market performance. Thus, H4a、H4c、were supported while H4b and H4d was not.

5. DISCUSSION

(1) Inter-firm knowledge governance mechanism and knowledge transfer

The three dimensions of inter-firm knowledge governance mechanism, including market mechanism, social mechanism and hierarchical mechanism, all had significant influences on knowledge transfer. This study and result responded to the research of Foss et al. in the form of an empirical study on knowledge governance and knowledge transfer. In the meantime, it verified and expanded the opinions of Gooderham et al. that knowledge governance mechanism helped to promote knowledge transfer between different firms. This study focused on the inter-firm knowledge governance of strategic emerging industry enterprises. It not only extended the research scope but also enriched the research contents, which was full of theoretical value and practical guidance to some extent. It also provided useful steps for inter-firms' knowledge governance practice.

(2) The moderating effect of network structure

The network structure negatively moderated the relationships between social mechanism and knowledge transfer. This suggested that the stronger the network structure, the more distracted the spot firms had on social

resources. It may lead to the predicament that went against the exertion of synergetic effects of social mechanism. This made enterprises in a dilemma. On the one hand, the more powerful the network structure was, the more outstanding the function of market mechanism and hierarchical mechanism were. On the other hand, the social mechanism had a negative moderating effect when the network structure was powerful. In this situation, firms should build a strong network structure to improve the utility of market mechanism and hierarchical mechanism. Meanwhile, as to the aspect of social resources, firms should be targeted and distinguish between primary and secondary. Focus on the advantages of social resources was needed so that the negatively moderating effect of social mechanism could be avoid.

(3) Amphibious innovation and innovation performance

Exploitative innovation had a positive impact on market performance; Exploitative innovation had a positive impact on technical performance; Explorative innovation had a positive impact on market performance; Explorative innovation had a positive impact on technical performance. This made the positive effect exploitative innovation and explorative innovation had on market performance and technical performance stand out. Either exploitative innovation or explorative innovation was the basic organizational learning behavior, from which we can see the effect of learning curve on business performance was still unshakable. The study made it clear that firms should gain knowledge through knowledge governance of inter-firm as well as pay attention to whether they could take advantage of it effectively. Only the two branches achieved the synergies could firms maximize the effectiveness of knowledge activities.

(4)The moderating effect of amphibious innovation

Exploitative innovation negatively moderated the relationship between knowledge transfer and market performance; Explorative innovation negatively moderated the relationship between knowledge transfer and market performance. That was an interesting result which seemed against the common sense. But through careful analysis, we found that it reflected the inharmonious of the knowledge transfer of inter-firm and the knowledge creation of intra-firm. The knowledge gained from knowledge transfer of inter-firm was not translated well into firm's innovation performance, especially the market performance. Maybe it's just because the knowledge enterprise acquired was not what they required and the value to enterprise innovation was little. The reason may also be the unsound knowledge management mechanism of enterprise that did not give full play to the utility of knowledge resources. In addition, the lack of good communication between R&D activities and marketing activities in firms may be the cause as well.

6. CONCLUSION

This study adopts a questionnaire survey for 485 research staffs and middle-top managers from 107 strategic emerging firms in China, follows which the moderating effect model of network structure on the relationship of knowledge governance and knowledge transfer and moderating effect model of amphibious innovation on the relationship of knowledge transfer and innovation performance are constructed. Based on these, we employ the structural equation model and hierarchical regression analysis to conduct an empirical research. The conclusions of this paper include: (1) Inter-firm knowledge governance (including market mechanisms, social mechanisms and hierarchical mechanisms) has a significant positive effect on knowledge transfer. (2) Network structure has a negative moderating effect on the relationship of social mechanisms and knowledge transfer. (3) Exploitative innovation and explorative innovation have significant positive effect on market performance and technical performance. (4) Exploitative innovation and explorative innovations have a negative moderating effect on the relationship of knowledge transfer and market performance.

This study enriches and expands the existing related research, and the conclusions have a significant guidance on inter-firm knowledge governance practice not only in strategic emerging industries but also

knowledge-intensive firms. Besides, the conclusions have important reference value to improve innovation performance through the exploitative innovative and explorative innovation within the enterprise.

But this study remains to be further improved. Firstly, the data collection was conducted through questionnaire, a subjective way. Although we have taken a lot of measures to ensure the quality of the subjective data, it is less convictive than the objective data. It's better to collect data with the combination of a variety of channels, such as the combination of depth interview and questionnaire-based survey and the combination of case study and the open information channel like Microblog, Wechat and professional forums. Secondly, the utility of knowledge governance is not just presented as knowledge transfer, and it may have an impact on knowledge exploitation and knowledge creation. This study simplified the process in this aspect, and the follow-up studies can conduct a comprehensive inspection. Thirdly, innovation performance is affected by not only the effect of knowledge transfer between enterprises, but also the learning curve inside enterprise and knowledge creation. Follow-up studies can pay attention to the cross layer research of inter-firm and intra-firm to reflect the impact and synergistic effect of knowledge activities of inter-firm and intra-firm more fully. Fourthly, we employed hierarchical regression model and structural equation model in this study. Future research may consider the hierarchical linear model (HLM) to carry on the comparative analysis between different industries and the cross layer analysis of inter-firm and intra-firm.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 72072139.

REFERENCES

- [1] Foss, N.J., Husted, K. and Michailova, S. (2010), Governing knowledge sharing in organizations: levels of analysis, governance mechanisms and research directions, *Journal of Management Studies*, 47(3):455-482.
- [2] Kerkhoff, V.L. and Pilbeam, V. (2017). Understanding socio-cultural dimensions of environmental decision-making: a knowledge governance approach. *Environmental Science & Policy*, 73 (7):29-37.
- [3] Cao, Y. and Xiang, Y. (2012). The impact of knowledge governance on knowledge sharing. *Management Decision*, 50(4):591-610.
- [4] Alexy, O., George, G. and Salter, A. (2013). Cui bono? the selective revealing of knowledge and its implications for innovative activity. *Academy of Management Review*, 38 (2):270-291.
- [5] Lavie, D., Haunschild, P.R. and Khanna, P. (2012). Organizational differences, relational mechanisms, and alliance performance. *Strategic Management Journal*, 33 (13):1453–1479.
- [6] Milagres, R., and Burcharth, A. (2019). Knowledge transfer in interorganizational partnerships: what do we know?. *Business Process Management Journal*, 25 (1):27-68.
- [7] Cong S , Lingshuang K , Francesco C , , Ulf H(2020). Demand and willingness for knowledge transfer in springboard subsidiaries of Chinese multinationals. *Journal of Business Research* 109:297–309.
- [8] X. Wang, J. Wang and R. Zhang(2019) .The optimal feasible knowledge transfer path in a knowledge creation driven team. *Data & Knowledge Engineering* 119: 105–122.
- [9] Guan, J.C. and Liu, N. (2016). Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological field of nano-energy. *Research Policy*, 45 (1):97-112.
- [10] L. Wang, S.Li,Z.You(2020). The effects of knowledge transfer on innovation capability: A moderated mediation model of absorptive capability and network reliance. *Journal of High Technology Management Research* 31:1-13.
- [11] Yan, Y. and Guan, J.C. (2017), Social capital, exploitative and exploratory innovations: The mediating roles of ego-network dynamics. *Technological Forecasting & Social Change*, 126: 244-258.
- [12] Enkel, E. Heil, S. Hengstler, M. and Wirth, H. (2017). Exploratory and exploitative innovation: To what extent do the

- dimensions of individual level absorptive capacity contribute?. *Technovation*, 60:29-38.
- [13] Wang, F.B., Chen, J.X. and Yang, Y. (2012).An analysis of the effect of the explorative and exploitative innovations and their balance. *Management World*, 28(3):96-112.
- [14] Gooderham, P., Minbaeva, D.B. and Pedersen, T. (2011).Governance mechanisms for the promotion of social capital for knowledge transfer in multinational corporations.*Journal of Management Studies*, 48 (1):123-150.
- [15] Maurer, I., Bartsch, V. and Ebers, M. (2011).The value of intra-organizational social capital: how it fosters knowledge transfer innovation performance, and growth. *Organization Studies*, 32 (2):157-185.
- [16] Lubatkin, M.H., Zeki, S., Ling, Y. and Veiga, J.F. (2006).Ambidexterity and performance in small-to- medium-sized firms: the pivotal role of top management team behavioral integration.*Journal of Management*, 32 (5): 646-672.

Full Research Paper

Meta-analysis of COVID-19 Clinical Symptoms, Prevalence of Comorbidities and Influencing Factors of Severity and Mortality Cases

Yuting yang¹, Liyi Zhang^{2*}

^{1,2}School of Information Management, Wuhan University, PR China

Abstract: As the Corona Virus Disease 2019 (COVID-19) continues to spread globally, it is necessary to systematically understand the characteristics of the disease and cases. The purpose of this meta-analysis was to evaluate the clinical characteristics of patients infected with COVID-19 and the prevalence of comorbidities, as well as the risk of potential diseases in severe patients (or deceased patients) compared with non-severe patients (or survivors). As of July 31, 2020, we had used several databases for literature search. A random effects model was used to summarize the prevalence, odds ratio (OR) and 95% confidence interval (95% CI). The final meta-analysis included 79 studies, analyzing a total of 330,464 infected patients from countries in Asia, Europe and North America. The meta-analysis results showed that the average age of the patients was 55.46 years old, and there were more male patients than females. The most common clinical symptoms were fever, followed by cough, dyspnea, fatigue, sore throat, headache and diarrhea. Hypertension, diabetes and cardiovascular disease were the most common complications. In addition, chronic kidney disease, respiratory disease, cancer, chronic liver disease and HIV often appeared. We also observed that the prevalence of some clinical symptoms and complications varies from region to region. When compared with non-severe patients, the combined odd ratios of some comorbidities were all higher in severe patients. A comparison between deceased and surviving patients found that there was also a higher risk of comorbidities in deceased patients. Our research showed that comorbidities may be a risk factor that affects the development of COVID-19 into severe illness or death. We recommended that high-risk patients adopt more targeted infection prevention and treatment strategies to reduce the risk of future COVID-19 diseases.

Keywords: COVID-19, comorbidities, clinical features, epidemiology, meta-analysis

1. BACKGROUND

The novel coronavirus (SARS-CoV-2) is a pathogen not previously found in humans. It was reported for the first time in the bronchoalveolar lavage fluid samples of three patients in Wuhan Jinyintan Hospital on January 24, 2020, and it was confirmed as the cause of COVID-19 [1]. After in-depth research on the full-length genome, it was discovered that the virus is a new type of β -coronavirus that has not been detected in humans or animals. The new type of coronavirus pneumonia it caused was named COVID-19 by the World Health Organization (WHO). It was also named SARS-CoV-2 by the International Committee on Taxonomy of Viruses, and belongs to the same coronavirus family as severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) [2]. Since its outbreak in China at the end of 2019, SARS-CoV-2 has spread all over the world with its extremely contagious nature. On January 30, 2020, it was declared by the WHO that this was a public health emergency of international concern, and was later declared as a global pandemic on March 11, 2020.

The continuous outbreak of COVID-19 infection has seriously threatened international health and the

* Corresponding author. Email: lyzhang@whu.edu.cn

economy. As the number of confirmed cases continues to increase, clinical research on patients and antiviral drugs is accelerating. Many countries have already started vaccine research. China, the United Kingdom, the United States and other countries developed related vaccines, but have not yet put them into large-scale use. At present, it is still useful to take preventive measures to reduce the movement and gathering of people. Simultaneously, more and more retrospective studies have reported the clinical manifestations and information of patients with COVID-19. The research of Alfonso et al. [3] shows that some underlying diseases may be risk factors for aggravation and death of patients. In previous studies on SARS and MERS, clinical and risk factors were reported. Badawi and Ryoo reported that the prevalence of chronic diseases is higher in MERS-CoV patients [4]. With the increase in the number of published studies, there are certain differences in the impact of comorbidities on the severity of COVID-19 and the outcome of death. Some studies have reported an association between the severity of COVID-19 and the outcome of previous diseases, while some other reports are not related. In order to further explore the clinical characteristics of COVID-19 and the prevalence of comorbidities and its relationship with worse outcome (i.e., severe disease), and to obtain more convincing results, we searched related literature and adopted meta-analysis methods, to assess the demographic characteristics, clinical symptoms, and prevalence of comorbidities of patients, including subgroup analysis by country or region. The risk of underlying diseases in severe patients (or deceased patients) compared with non-severe patients (or surviving patients) is also assessed. The findings of this article may be beneficial to patient management in the clinical stage, help identify patients who are susceptible to severe diseases, and facilitate tailor-made treatment plans for patients. It may also help clinicians to take corresponding treatment measures at an earlier stage.

2. METHOD

2.1 Search strategy and inclusion criteria

This research was based on the recommendations of the Preferred Reporting Project (PRISMA) of the meta-analysis. First, a systematic literature search was conducted on the research published in PubMed, Scopus, Embase, Web of Science, and the preprinted literature database bioRxiv. For related literature published before August 31, 2020, the following search terms were used: "Corona Virus Disease-2019", "2019 novel coronavirus", "COVID-19", "comorbidities", "clinical characteristics", "epidemiological", combined with Boolean logic operators (AND, OR). We also conducted a manual search and checked the reference list of included studies to identify missing studies. The obtained documentation results were managed by EndNote X9.0 software to eliminate duplicates.

First filter the initial search results based on the title and abstract, then read the full text of articles that meet the theme to check whether the article contains the information required. The inclusion criteria of the article are as follows: the study population is patients diagnosed with COVID-19, the research design is a case study, the outcome indicators include epidemiological, clinical features, and relevant descriptions of comorbidities in infected patients. Articles without clinical features, clinical experience, and case reports with a sample size of less than 10 are not included in the selection. We excluded studies that specifically targeted pregnant women, children, and elderly patients. The steps of document retrieval are shown in Figure 1.

2.2 Data collection and extraction

Extract data basic data such as author, study time, country, continent, age, gender, and number of cases from the literature. We also extracted the number and proportion of cases with clinical symptoms. The number and proportion of comorbidities were also extracted. Besides, if there is a study involving severe cases and deaths, the comorbidities in severe and deceased patients will be also within the statistical range.

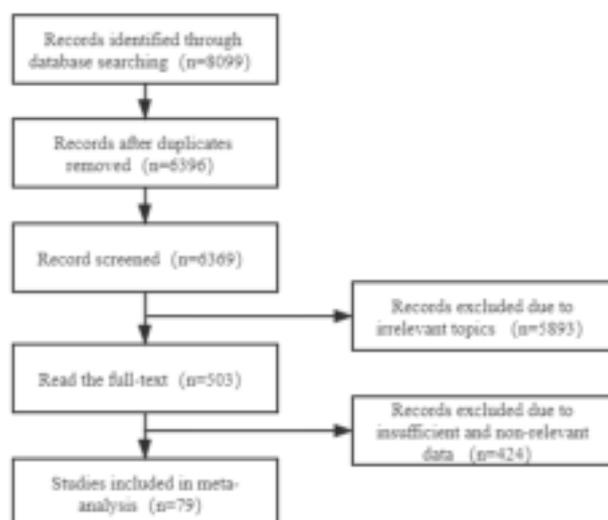


Figure 1. Study selection flow diagram

3. STATISTICAL METHODS

This article adopted the statistical method of meta-analysis, which aims to combine the scientific results of multiple comparable studies or experiments, and obtain aggregated estimates by summarizing relevant information, thereby improving statistical capabilities. In this study, the mean \pm standard deviation was calculated to describe the distribution of continuous variables (such as age). For continuous data using median and quartile or number of samples, the estimation method proposed by Wan^[5] was used to obtain the sample mean and standard deviation. The combined prevalence and its 95% confidence interval (95% CI) were used to summarize the weighted effect size of each study grouping variable. This study used odds ratio (OR) to describe the risk of comorbidities in severe and deceased patients.

The measures of heterogeneity were evaluated and reported, including Cochran's Q statistic and I^2 statistic. This study used a fixed effects model or a random effects model to assess the heterogeneity of the research. When the $P < 0.10$ of Cochran's Q statistic^[6], there is heterogeneity and it needs to be estimated using a random effects model. Higgin and Thompson's I^2 statistic expresses the percentage of effect size variability not caused by sampling error. When the I^2 statistic is medium (50%-74%) or high (75%), the random effects model will used to calculate the combined estimated value of the effect size^[7]. For the group with high heterogeneity, subgroup analysis was conducted according to research region.

The publication bias was assessed by visual inspection of the funnel plot asymmetry and Egger regression test^[8]. If the significance level P is greater than 0.05, it is considered to be unbiased. When evidence of publication bias is found, the trim-and-fill method is used to estimate and adjust the potential missing studies, and the effect size is recalculated accordingly^[9]. All calculations are carried out by Comprehensive Meta Analysis 3.0 software.

4. RESULT ANALYSIS

4.1 Research selection and feature analysis

Initially, 8099 articles were retrieved using the search strategy. After deleting duplicate articles, 6396 articles were obtained. The abstract and title were further evaluated, and 503 articles were selected for full-text evaluation. Among them, 424 were excluded due to lack of relevant required information, and 79 articles reached the predetermined inclusion and exclusion criteria for quantitative meta-analysis. This study included 79 studies published as of June 28, 2020, and contains cases from 21 countries including the United States, China, Italy, Spain, India, the United Kingdom, South Korea, Kuwait, and Iran. A total of 330,464 patients were reported in all

studies.

4.2 Meta-analysis results

4.2.1 Prevalence of clinical symptoms and comorbidities

The variables of the meta-analysis research were shown in Table 1. The average age of the subjects studied was 55.46 (95% CI: 52.53-58.39) years old, and there were more male patients than females (59.96%, 95% CI: 57.46-62.40%, $I^2=99.26\%$). The results of meta-analysis showed that the most common clinical symptom was fever, and the combined prevalence rate was 73.49% (95%CI: 67.07-79.05%, $I^2=97.75\%$), followed by cough (62.91%, 95%CI: 57.15- 68.32%, $I^2=97.38\%$), dyspnea (43.26%, 95%CI: 37.44–49.29%, $I^2=96.79\%$) and fatigue (29.32%, 95%CI: 22.99–36.58%, $I^2=97.15\%$). Other symptoms such as sore throat (14.87%, 95% CI: 12.01–18.26%, $I^2=84.77\%$), headache (13.38%, 95% CI: 10.04–17.61%, $I^2=91.65\%$) and diarrhea (12.43%, 95% CI: 9.67–15.83%, $I^2=94.34\%$) also appeared from time to time.

Common comorbidities included hypertension, diabetes, respiratory disease, cardiovascular disease, chronic liver disease, chronic kidney disease, cancer and HIV. Among them, the most common comorbidities were hypertension (32.82%, 95%CI: 28.80–37.10%, $I^2=99.75\%$), diabetes (19.77%, 95%CI: 17.21-22.61%, $I^2=99.57\%$) and cardiovascular disease (11.10%, 95%CI: 8.86-13.81%, $I^2=99.61\%$). A small proportion of patients developed chronic kidney disease (8.41%, 95%CI: 5.72-12.20%, $I^2=99.84\%$), respiratory system diseases (7.53%, 95%CI: 5.62-10.02%, $I^2=99.72\%$), cancer (5.44%, 95%CI: 4.69–6.30%, $I^2=96.82\%$), chronic liver disease (2.28%, 95%CI: 1.82–2.84%, $I^2=98.24\%$), and HIV (1.03%, 95%CI: 0.58-1.82%, $I^2=97.23$). In the evaluation of clinical characteristics, I^2 was above 90%, showing statistical heterogeneity ($P<0.001$), as shown in Table 1. Through the analysis of the prevalence of comorbidities, it was observed that the Q statistic $P<0.1$, and the I^2 of the comorbidities are all above 90%, showing significant heterogeneity.

Table 1. Meta-analysis results

	N	Prevalence ^a	Lower limit	Upper limit	Q	P	I^2
Male	73	59.96%	57.46%	62.40%	9778.31	0.00	99.26
Fever	47	73.49%	67.07%	79.05%	2048.59	0.00	97.75
Cough	51	62.91%	57.15%	68.32%	1909.09	0.00	97.38
Fatigue	31	29.32%	22.99%	36.58%	1052.31	0.00	97.15
Dyspnea	43	43.26%	37.44%	49.29%	1306.92	0.00	96.79
Diarrhea	33	12.43%	9.67%	15.83%	565.69	0.00	94.34
Headache	31	13.38%	10.04%	17.61%	359.22	0.00	91.65
Sore Throat	23	14.87%	12.01%	18.26%	144.47	0.00	84.77
Hypertension	58	32.82%	28.80%	37.10%	22669.63	0.00	99.75
Diabetes	66	19.77%	17.21%	22.61%	14983.19	0.00	99.57
Respiratory System Disease	64	7.53%	5.62%	10.02%	22583.26	0.00	99.72
Cardiovascular Diseases	63	11.10%	8.86%	13.81%	15767.18	0.00	99.61
Chronic Liver Disease	32	2.28%	1.82%	2.84%	1761.60	0.00	98.24
Chronic Kidney Disease	55	8.41%	5.72%	12.20%	33793.63	0.00	99.84
Cancer	41	5.44%	4.69%	6.30%	1257.86	0.00	96.82
HIV	13	1.03%	0.58%	1.82%	433.11	0.00	97.23

^a Meta-analysis for the prevalence was calculated from binary random-effects model analysis

4.2.2 Subgroup analysis

A subgroup analysis was conducted on the prevalence of clinical symptoms and comorbidities. The

subgroups were divided into Asian countries, European countries and North American countries according to regions. There were 32 studies from Asian countries, 15 of which were from China, 23 of European countries, 9 of which were from Italy, 6 from Spain, and 24 from North America, most of which were from the United States. The average age of patients from Asian countries was 48.23 (95% CI: 45.14–51.32), the proportion of male patients was 63.26% (95% CI: 55.42–70.46%). The average age of European patients was 65.23 (95% CI: 59.92–70.54), the proportion of male patients was 58.21% (95% CI: 54.48–61.85%). In North America, the average age of patients was 58.08 (95% CI: 55.09–60.07) years, 57.08% (95% CI: 53.07–61.01%) were male patients. We had observed that the combined prevalence of some clinical symptoms and complications varies from region to region.

In terms of clinical symptoms, fever and cough were the most prevalent manifestations in Asia, Europe and North America, and dyspnea and fatigue were also common symptoms. The prevalence of cough in North American patients exceeded fever, and was the most common clinical symptom. At the same time, the prevalence of dyspnea ($Q=20.77$, $P<0.001$) and diarrhea ($Q=30.13$, $P<0.001$) in North American were obviously higher than that in Asia and Europe.

As for comorbidities, the most common were hypertension and diabetes, and cardiovascular disease and respiratory system diseases also occurred from time to time. There were significant differences in the prevalence of some comorbidities in different regions. The prevalence of hypertension ($Q=43.33$, $P<0.001$), cardiovascular disease ($Q=11.56$, $P<0.001$) and cancer ($Q=10.99$, $P<0.001$) in European and North American patients was significantly higher than that in Asia. The prevalence of diabetes ($Q=47.93$, $P<0.001$), chronic kidney disease ($Q=12.36$, $P<0.001$) and HIV ($Q=11.91$, $P<0.001$) among North American patients was significantly higher than that of the other two regions.

Table 2. Subgroup analysis results

		N	Prevalence	Lower limit	Upper limit	Q	P	I ²
Male	Asia	31	63.26%	55.42%	70.46%	2468.36	0.00	98.78
	Europe	22	58.21%	54.48%	61.85%	1950.19	0.00	98.92
	North America	20	57.08%	53.07%	61.01%	2276.45	0.00	99.17
Fever	Asia	21	73.73%	64.13%	81.50%	368.62	0.00	94.57
	Europe	13	76.13%	65.46%	84.30%	768.69	0.00	98.44
	North America	13	68.89%	57.45%	78.41%	256.54	0.00	95.32
Cough	Asia	22	59.88%	50.18%	68.86%	554.22	0.00	96.21
	Europe	15	56.92%	47.26%	66.09%	743.66	0.00	98.12
	North America	14	73.42%	62.25%	82.22%	285.36	0.00	95.44
Fatigue	Asia	19	25.12%	15.39%	38.22%	657.97	0.00	97.26
	Europe	5	31.83%	21.05%	44.99%	131.58	0.00	96.96
	North America	7	39.42%	22.52%	59.30%	114.56	0.00	94.76
Dyspnea	Asia	17	26.19%	16.20%	39.44%	421.90	0.00	96.21
	Europe	15	46.38%	39.83%	53.07%	356.17	0.00	96.07
	North America	11	64.78%	55.08%	73.39%	100.37	0.00	90.04
Diarrhea	Asia	14	6.05%	3.86%	9.35%	86.56	0.00	84.98
	Europe	9	15.09%	11.15%	20.11%	92.76	0.00	91.38
	North America	10	23.89%	18.64%	30.08%	47.39	0.00	81.01
Headache	Asia	17	11.77%	7.62%	17.76%	221.12	0.00	92.76
	Europe	5	15.55%	5.73%	35.79%	93.06	0.00	95.70
	North America	9	15.92%	11.19%	22.13%	34.00	0.00	76.47
Sore Throat	Asia	10	17.98%	13.87%	23.00%	63.10	0.00	85.74
	Europe	3	11.27%	4.01%	27.86%	19.04	0.00	89.49
	North America	10	12.60%	8.24%	18.79%	51.68	0.00	82.59

		N	Prevalence	Lower limit	Upper limit	Q	P	I ²
Hypertension	Asia	22	16.70%	12.45%	22.04%	960.67	0.00	97.81
	Europe	18	42.82%	35.27%	50.71%	4152.18	0.00	99.59
	North America	18	46.89%	38.33%	55.64%	9971.74	0.00	99.83
Diabetes	Asia	24	12.80%	9.94%	16.35%	576.24	0.00	96.01
	Europe	20	18.72%	14.89%	23.26%	2713.35	0.00	99.30
	North America	22	30.06%	26.92%	33.39%	1607.09	0.00	98.69
Respiratory System Disease	Asia	22	3.83%	1.33%	10.57%	4634.93	0.00	99.55
	Europe	21	10.64%	6.89%	16.08%	6588.69	0.00	99.70
	North America	21	9.13%	6.23%	13.19%	3350.17	0.00	99.40
Cardiovascular Diseases	Asia	23	6.91%	4.89%	9.67%	409.65	0.00	94.63
	Europe	21	13.90%	8.24%	22.50%	9556.89	0.00	99.79
	North America	19	14.37%	10.78%	18.90%	3096.77	0.00	99.42
Chronic Liver Disease	Asia	12	2.67%	1.10%	6.30%	298.66	0.00	96.32
	Europe	10	2.00%	1.01%	3.92%	751.82	0.00	98.80
	North America	10	2.08%	1.14%	3.75%	482.40	0.00	98.13
Chronic Kidney Disease	Asia	15	3.12%	1.72%	5.60%	335.00	0.00	95.82
	Europe	19	7.57%	5.90%	9.65%	897.45	0.00	97.99
	North America	21	17.09%	7.55%	34.20%	23953.92	0.00	99.92
Cancer	Asia	14	2.49%	1.22%	4.99%	369.16	0.00	96.48
	Europe	16	8.20%	6.60%	10.14%	781.64	0.00	98.08
	North America	11	6.71%	5.93%	7.59%	25.40	0.00	60.63
HIV	Asia	3	0.58%	0.13%	2.56%	10.68	0.00	81.27
	Europe	3	0.47%	0.23%	0.95%	45.52	0.00	95.61
	North America	7	1.70%	1.25%	2.32%	17.38	0.01	65.47

4.2.3 Non-severe and severe cases

A comparison between severe and non-severe patients found that the average age of severe patients was 62.17 (95% CI: 59.29–65.06), and the average age of non-severe patients was 53.02 (95% CI: 47.58–58.45). Compared with non-severe patients, severe patients presented a higher risk of complications. The ORs of cardiovascular disease, diabetes, hypertension, chronic liver disease, were 2.32 (95%CI: 1.31–4.10, I²=80.73%), 2.02 (95%CI: 1.41–2.90, I²=76.30%), 1.90 (95%CI: 1.31–2.74, I²=83.10%), and 1.89 (95%CI: 1.42–2.52, I²=0.24%) respectively. Respiratory system diseases (OR=1.79, 95%CI: 1.03–3.11, I²=64.60%), chronic kidney disease (OR=1.54, 95%CI: 0.95–2.49, I²=62.04%), cancer (OR=1.50, 95%CI: 0.94–2.39, I²=30.66%) and HIV (OR=1.34, 95%CI: 0.30–5.93, I²=0.00%) were also closely associated with severe disease.

4.2.4 Surviving and non-surviving cases

A comparison between the survivors and deceased patients found that the average age of the dead patients was 71.47 (95% CI: 67.56–75.38), and the average age of the survivors was 60.48 (95% CI: 57.61–63.35). Compared with surviving patients, deceased patients presented a higher risk of complications. The ORs for cardiovascular disease, chronic kidney disease, hypertension, cancer, were 4.05 (95%CI: 2.63–6.24, I²=83.30%), 3.69 (95%CI: 2.13–6.41, I²=92.02%), 2.46 (95%CI: 1.85–3.29, I²=81.11%), and 2.30 (95%CI: 1.76–3.01, I²=54.92%) respectively. Chronic liver disease (OR=2.26, 95%CI: 1.32–3.87, I²=32.68%), respiratory system diseases (OR=2.24, 95%CI: 1.77–2.84, I²=52.59%), diabetes (OR=2.04, 95%CI: 1.7–2.65, I²=79.16%), and HIV (OR=0.74, 95%CI: 0.43–1.27, I²=0.00%) were also closely associated with severe diseases.

4.2.5 Publication bias and sensitivity analysis

The publication bias was evaluated by visually inspecting the symmetry of the funnel plot and Egger's

regression. Asymmetry was observed in the funnel plot of the prevalence of fatigue, diarrhea, and chronic liver disease, with Egger's P of 0.024, 0.0035, and 0.019 respectively. In the analysis of severe cases, the funnel plot of chronic liver disease showed asymmetry (Egger's $P=0.0078$), indicating the existence of publication bias. However, after using the trim-and-fill method to process it, and adjusting the size of the asymmetry of the funnel plot, the adjusted effect was still the same as the original effect, indicating that the result was reliable.

By omitting each study one by one to conduct sensitivity tests to assess the stability of the combined results, no obvious significant deviations were found. The one study removed method was used to exclude one document at a time, and the influence of each document on the results of the meta-analysis was studied. The results of the sensitivity analysis showed that the combined effect value of each study does not change much, suggesting that the results of this meta-analysis were relatively robust and credible.

5. DISCUSSION

This study was based on 79 laboratory-confirmed COVID-19 research data. The results found that the proportion of male patients is greater than that of females (59.96%, 95% CI: 57.46-62.40%), which is in consistent with the conclusions of previous studies. In the previous two outbreaks of coronavirus respiratory system diseases, namely severe acute respiratory disease (SARS) and Middle East respiratory disease (MERS), it was also observed that the infection rate of men was significantly higher than that of women [4]. In Mohammad's study, it was also found that the hospital mortality rate of men among COVID-19 patients was significantly higher than that of women (OR 3.4, 95% CI: 1.2-9.1, $P = 0.01$) [10].

Many previous studies had shown that women are less likely to be infected by a variety of bacteria and viruses than men, partly because women have stronger innate and adaptive immune responses. The severity of the disease may be related to the immunomodulatory effects of hormones. Estrogen and testosterone sex hormones have the effect of regulating the immune response. In general, estrogen is conducive to enhancing immunity, and testosterone can suppress immunity, which may also be the reason why women are not susceptible to virus infection [11]. Some scholars had also raised the influence of smoking. According to research, angiotensin converting enzyme 2 (ACE2) is the receptor of SARS-CoV-2 in the lower respiratory tracts, and the expression of ACE2 in the lower respiratory tract of smokers is higher, which indicates that smokers have higher risk of COVID-19 illness [12]. The smoking prevalence of men in China is much higher than that of women (57.6% vs. 6.7%). In other countries, the smoking rate of men is also usually higher than that of women. These findings may reveal the reason for the higher risk of men in COVID-19. Although approximately 70% of health and care workers in the world are women, and they are more exposed to the risk of infection, most studies currently showed that men with COVID-19 have a higher overall infection and mortality rate [13]. More research is needed to explore gender differences during the COVID-19 pandemic.

We extracted data on the prevalence of the most common symptoms of COVID-19. These most common clinical manifestations were also similar to previous reports. Fever and cough were the most common clinical symptoms in COVID-19 patients. The prevalence of fever symptoms in COVID-19 patients was similar to that of SARS and MERS, but the prevalence of cough in SARS and COVID-19 was higher than that of MERS. (<50%) [14]. In North America, the prevalence of cough symptoms (73.42%) exceeded fever (68.89%), and became the most common clinical symptom. The analysis of subgroups divided by regions found that the prevalence of dyspnea ($Q=20.77$, $P<0.001$) and diarrhea ($Q=30.13$, $P<0.001$) in North American patients was significantly higher than that in Asia and Europe. The reason for this difference remained to be further investigated.

According to the results of our meta-analysis, the most common comorbidities were hypertension, diabetes, cardiovascular disease and so on. The analysis of regions divided into different subgroups showed that the prevalence of complications such as hypertension, cardiovascular disease, cancer, diabetes, chronic kidney disease,

and HIV also differs significantly in different regions. This may be caused by differences in the patients' environment, social conditions, and medical conditions, thus further research is needed to determine the reason for the difference. This was similar to the conclusion obtained by Sunny Goel et al.^[15]. Their study also reported that the prevalence of respiratory system diseases (asthma and chronic obstructive pulmonary disease) and liver disease in American and European patients was significantly higher than that in Asia. However, no significant differences were observed in our study (respiratory system diseases: $Q=3.27$, $P=0.19$; chronic liver disease: $Q=0.29$, $P=0.87$).

Compared with non-severe patients, the risk of comorbidities was higher in severe people. Similarly, compared with survivors, non-survivors also had a higher risk of comorbidities. However, no higher risk of death has been observed in HIV patients (OR: 0.74, 95% CI: 0.43–1.27), which may be due to the fact that fewer relevant literatures were included. There is not much information reported on HIV comorbidities, and further research is needed to explain the relationship and interaction between HIV and COVID-19 pathogenesis.

According to age-based analysis, the average age of patients was 55.46 years, which was close to the number reported in previous studies^[14]. In fact, it had been shown that elderly patients are more likely to have more serious complications and further deaths, which may be related to the relatively weak immune system of the elderly or the higher prevalence of comorbidities. Comorbidities and susceptibility diseases may damage the function of macrophage and lymphocyte function, thereby reducing immunity, which is related to the pathogenesis of COVID-19^[16]. Previous studies had also pointed out that diabetes and heart disease are also significantly related to MERS-CoV illness. The meta-analysis of Yang et al.^[17] pointed out that age and comorbidities are highly correlated in COVID-19 patients. Another study also confirmed that the elderly and patients with comorbidities (including diabetes, hypertension, cardiovascular disease, liver disease, malignant tumors) are more likely to develop severe diseases (62.1%, 25.0%, $p < 0.001$).

Hypertension is the most common comorbidity among COVID-19 patients. Previous studies had also explored the relationship between hypertension and the severity of COVID-19. A study from Zekavat et al. found that patients with hypertension patients have a higher risk of acute respiratory disease and chronic lower respiratory disease. A systematic review from Chen et al. showed that hypertension is one of the main comorbidities leading to the death of COVID-19.

Diabetes increases metabolic disorders and induces inflammatory infections, which makes diabetic patients more susceptible to SARS-CoV-2. At the same time, the blood glucose level of diabetic patients may fluctuate sharply due to viral infection, which is not conducive to the patient's recovery. Considering that diabetes can weaken the complexity of the human immune system, it is important to take preventive measures to combat the high incidence of COVID-19 in this vulnerable group. Such patients must be reminded to pay special attention to blood glucose control, blood glucose monitoring and timely treatment to avoid serious complications of COVID-19.

Heart disease patients with SARS-CoV-2 may develop serious diseases, which indicates that SARS-CoV-2 may cause more infections in cardiovascular disease patients and make their condition worse. Gaurav's research showed that cardiovascular disease is directly related to the severity of COVID-19 patients, and they found that cardiovascular disease is associated with an approximately three-fold increase in the probability of severe COVID-19 infection^[18].

Chronic kidney disease usually causes various metabolic and electrolyte abnormalities, and can lead to serious consequences due to acute kidney injury. A report on COVID-19 indicated that 30% of cases developed moderate to severe kidney damage. According to research, its pathogenesis may be that kidney tissue is one of the main binding sites for COVID-19, where the expression of ACE2 increases^[19].

Respiratory system diseases such as patients with chronic obstructive pulmonary disease have less resistance to viruses and are prone to acute respiratory distress syndrome. Similar to other respiratory infections, COVID-

19 mainly affects the lungs and invades pulmonary alveolar epithelial cells. Patients who have suffered from respiratory system diseases may have hypoxia at baseline, so they are more likely to be at risk of impaired lung function after infection [20]. Furthermore, according to studies by Imai et al., the expression of ACE2 (the main receptor of COVID-19) is higher in people with acute respiratory distress syndrome and acute respiratory injury. This may partly explain the relationship between patients with lung disease and severe COVID-19.

Liang et al. released a study of 1590 COVID-19 patients, including 18 cancer patients, which is higher than the cancer incidence rate of the entire Chinese population. Another cross-sectional study published by Jing et al. analyzed 1524 cancer patients in a hospital in Wuhan. It reported 12 cancer patients diagnosed with COVID-19 and concluded that the COVID-19 infection rate among cancer patients was higher than the cumulative incidence reported in Wuhan during the same period (0.79% vs 0.37%).

A study by Zhang et al. showed that the expression of ACE2 in the liver makes it easy to be injured by SARS-CoV-2 infection. Other drugs used to treat COVID-19, such as lopinavir/ritonavir may also adversely affect the liver, causing further liver damage.

For immunodeficiency and HIV patients, the risk of serious diseases associated with COVID-19 infection will greatly increase. Compared with SARS-CoV-2 single-infected patients, COVID-19 patients with HIV are also prone to increase the prevalence of other comorbidities, which may further lead to aggravation of the disease. However, in this study, it was found that there was no significantly higher risk of HIV comorbidities among patients who died from COVID-19, and the relationship between the two in terms of pathogenesis and therapeutic impact needs to be further explored.

Based on the results of our research, the following prevention strategy was proposed, that was, individuals with higher risk should receive immunization first. People with high blood pressure, diabetes, cardiovascular disease, chronic kidney disease, respiratory disease, cancer and chronic liver disease are at higher risk of worse outcome from COVID-19, so they should be given priority when allocating vaccines. Lau et al. investigated the influenza vaccination of 91,605 diabetic patients and found that influenza vaccine can significantly reduce the incidence of influenza and pneumonia in diabetic patients. The study of Remschmidt et al. also confirmed that in people with chronic comorbidities, influenza vaccination significantly reduced morbidity and mortality. In the season when respiratory system diseases caused by influenza, respiratory syncytial virus and other respiratory viruses are highly prevalent, influenza vaccination is still necessary. On the one hand, it can still be used to prevent influenza. On the other hand, it can help reduce the possibility of infection with COVID-19 confused. Therefore, patients with chronic diseases such as hypertension, diabetes, respiratory system diseases and cardiovascular diseases should be included in the influenza and future COVID-19 vaccination recommendations. In view of the limited level of evidence in this study, scholars in this field can conduct more adequate research to prove this association in the future. The prevalence of chronic diseases is increasing year by year, and targeted public health vaccination interventions must be taken to better protect people with chronic diseases from COVID-19 and other respiratory viruses.

From a clinical point of view, this research was very relevant to the practice of patient care. Our research found that there is a certain correlation between comorbidities and the severity of COVID-19 disease and the outcome of death, which may help medical and health departments to guide susceptibility population prevention and assessment of the risk of patient deterioration. It can also contribute to guiding prevention and treatment strategies. Since the clinical results of COVID-19 are more related to the elderly, men, and comorbidities, more active hospital care measures may be needed for such patients. Abnormal examinations of the lungs, liver, and kidneys on admission can be used as predictors of disease severity. Susceptible patients with certain known risk factors may need to take additional preventive measures to help prevent SARS-CoV-2 infection. High-risk patients who are infected can be encouraged to seek medical attention in the early stages of the disease. In the case of

hospitalization, doctors should consider that the medical management of such patients may be more difficult because they are at higher risk of serious illness and death.

6. CONCLUSIONS

COVID-19 is a new type of virus. Its outbreak and epidemic are threatening the safety of all countries and the entire ecosystem in the world. In the face of this new clinical disease, it is necessary for all countries and departments to make joint efforts in the fields of epidemiology, diagnosis, treatment and prevention during the epidemic period in order to contain the further spread of the disease. In the context of the COVID-19 outbreak, the clinical symptoms, laboratory results, and imaging analysis of the disease should be systematically studied and evaluated to correctly understand the virus itself and the characteristics of the disease caused by it. This study analyzed 330,464 patients through a systematic review and random effects meta-analysis. Through the combined prevalence and 95% confidence interval estimation, the following summary was obtained: the most common clinical manifestation in COVID-19 patients was fever, symptoms such as cough, dyspnea and fatigue, sore throat, headache, diarrhea were common; the elderly, men, and the presence of comorbidities were some of the risk factors; among patients with comorbidities, the most common were hypertension, diabetes, cardiovascular disease, chronic kidney disease, respiratory disease, cancer, chronic liver disease and HIV also often appeared; through the analysis of the comorbidities of severe and non-severe patients (deceased and surviving patients), it was found that there was a higher risk of comorbidities in severe and deceased patients.

The symptoms, prevalence, and epidemiological characteristics of COVID-19 patients are an important topic in the research community, because more data can be obtained and more specific and stable discoveries can be obtained. Our study conducted a meta-analysis of patient characteristics and compared the performance of patient characteristics in different regions. The results of this study can help medical and public health professionals and the public better understand the symptoms associated with COVID-19. This study confirmed that patients with comorbidities have an increased risk of developing serious or death, which was consistent with the conclusions of other previous studies. Considering this higher risk, we suggested that people with one or more potential comorbidities need to pay more attention, especially those in areas with high infection rates, should be protected and avoid close contact with community members. In addition, with the development of effective antiviral treatments and vaccines, we must vigorously consider adopting key intervention measures to protect this vulnerable group.

However, there were still certain limitations in this study. First, the lack of reports on asymptomatic infections and mildly symptomatic patients may cause this study to overestimate the severity of the disease. Secondly, most of the studies included in this meta-analysis were not randomized controlled trials. In addition, the meta-analysis results found that there was a high heterogeneity statistical data, which may be related to the large difference in the sample size between the study design and the study. Some studies provided incomplete or insufficient patient clinical data, and missing information may lead to biased results. Moreover, patients had great heterogeneity, because the age, race, and economic status of patients in the study were not same, and different treatment methods and drug use during the treatment process may also affect statistics. In this study, we only considered the region as the basis for subgroup classification. Furthermore, the literature included in this study was not sufficient. As the pandemic spreads globally, future studies can obtain more regional data and collect more comprehensive data, higher-quality clinical trials for more reliable conclusions.

So far, whether it is a case report or a cross-field study, the clinical findings are basically the same. However, more data is needed to further determine the epidemiology, pathogenesis, duration of virus shedding, and clinical scope of the COVID-19 virus infection. Clinically, these data will help guide countries that are currently undergoing epidemics to conduct adequate clinical surveillance. Additionally, this study provided assistance in

the screening and prevention of COVID-19 in clinical practice, and the risk stratification and management of patients. Meanwhile, it was useful for decision makers, clinicians and researchers to take supportive interventions and implement infection control measures to prevent further spread.

REFERENCES

- [1] Huang C, Wang Y, Li X, et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223): 497-506.
- [2] Gorbalenya A E, Baker S C, Baric R S, et al. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 5(4): 536-544.
- [3] Rodriguez-morales A J, Cardona-ospina J A, Gutierrez-ocampo E, et al. (2020). Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. *Travel Med Infect Dis*, 34: 101623.
- [4] Badawi A, Ryoo S G. (2016). Prevalence of comorbidities in the Middle East respiratory syndrome coronavirus (MERS-CoV): a systematic review and meta-analysis. *Int J Infect Dis*, 49: 129-133.
- [5] Wan X, Wang W, Liu J, et al. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*, 14: 135.
- [6] Cochran W G. (1954). The Combination of Estimates from Different Experiments. *Biometrics*, 10(1).
- [7] Higgins J P, Thompson S G, Deeks J J, et al. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414): 557-560.
- [8] Egger M, Davey Smith G, Schneider M, et al. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109): 629-634.
- [9] Duval S, Tweedie R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2): 455-463.
- [10] Nasiri M J, Haddadi S, Tahvildari A, et al. (2020). COVID-19 Clinical Characteristics, and Sex-Specific Risk of Mortality: Systematic Review and Meta-Analysis. *Front Med (Lausanne)*, 7: 459.
- [11] Taneja V. (2018). Sex Hormones Determine Immune Response. *Frontiers in Immunology*, 9.
- [12] Leung J M, Yang C X, Tam A, et al. (2020). ACE-2 expression in the small airway epithelia of smokers and COPD patients: implications for COVID-19. *European Respiratory Journal*, 55(5).
- [13] Gebhard C, Regitz-zagrosek V, Neuhauser H K, et al. (2020). Impact of sex and gender on COVID-19 outcomes in Europe. *Biology of Sex Differences*, 11(1).
- [14] Yin Y, Wunderink R G. (2018). MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*, 23(2): 130-137.
- [15] Goel S, Jain T, Hooda A, et al. (2020). Clinical Characteristics and In-Hospital Mortality for COVID-19 Across The Globe. *Cardiology and Therapy*, 9(2): 553-559.
- [16] Landi F, Barillaro C, Bellieni A, et al. (2020). The New Challenge of Geriatrics: Saving Frail Older People from the SARS-COV-2 Pandemic Infection. *The journal of nutrition, health & aging*, 24(5): 466-470.
- [17] Yang J, Zheng Y, Gou X, et al. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases*, 94: 91-95.
- [18] Aggarwal G, Cheruiyot I, Aggarwal S, et al. (2020). Association of Cardiovascular Disease With Coronavirus Disease 2019 (COVID-19) Severity: A Meta-Analysis. *Current Problems in Cardiology*, 45(8).
- [19] Batlle D, Soler M J, Sparks M A, et al. (2020). Acute Kidney Injury in COVID-19: Emerging Evidence of a Distinct Pathophysiology. *Journal of the American Society of Nephrology*, 31(7): 1380-1383.
- [20] Gold M S, Sehayek D, Gabrielli S, et al. (2020). COVID-19 and comorbidities: a systematic review and meta-analysis. *Postgraduate Medicine*: 1-7.

Full Research Paper**Mediating Effect of ICT Impacts on ICT and Health Goals of MDGs:****A Cross-national Panel Analysis***Jing Fan¹, Zishan Huang², Mingxing Shao^{3*}, Hongquan Li⁴, Yan Wan⁵*¹International Business School, Beijing Foreign Studies University, China² School of Economics and Management, Beijing University of Posts and Telecommunications, China

Abstract: This work contributes to the discussion on the relationship between ICT factors, ICT impact and national health outcomes. The method used is a cross-national study of 148 countries using secondary data from World Economic Forum, World Bank, the United Nations and other international organizations. This study applies a panel data set from 2012 to 2016 and structures a mediating effect model to explore dynamic relationships between all the constructs, among which ICT factors are independent variables, ICT impacts are mediating variables and health indicators selected from Goal 4 and Goal 5 of UN Millennium Development Goals (MDGs) are dependent variables. The results show that there are significant negative associations between all ICT factors and health indicators, while only some of the partial mediating effects are proved. Economic impact has a significant mediating effect on the association between ICT infrastructure and both of the two health indicators. Social impact only mediates the influence of ICT infrastructure on under-five mortality rate, and ICT skills on maternal mortality ratio as well.

Keywords: Information and communication technology (ICT), health outcomes, cross-national research, mediating effect

1. INTRODUCTION

Information and communications technology (ICT) is an inclusive term, covering all communication equipment or application software: for example, radio, television, mobile phone, computer, network hardware and software, satellite system, etc., as well as various services and application software related to it, such as video conference and distance learning. The importance of ICT is not the technology as such, but its enabling function in facilitating enhanced access to information and communication across large distances.

This study aims to explore two research questions: first, the association between ICT factors and health outcomes are the basis; second, ICT impacts are introduced as mediation variables. Answering these questions can help global policy makers to formulate health resource allocation and investment strategies, especially in the fields of health care and technology in developing countries, so as to achieve the best interests of people worldwide.

The method used is a cross-national study of 148 countries using data from World Economic Forum, World Bank, the United Nations and other relevant international organizations. This study applies a panel data set from 2012-2016 and structures a classic mediating effect model to explore dynamic relationships between all the constructs.

2. RESULTS

Our results indicate that overall ICT factors substantially facilitate national public health delivery. However, ICT factors have different acting mechanism on health outcomes. Some use economic impact as a mediation, some use social impact, and others seem to totally skip a mediating process. From the economic perspective, ICT infrastructure is the only ICT factor that shows mediation function. We can clearly see the necessity of ICT infrastructure in facilitating economic growth. Thus, the mediation of economic impact is inevitable.

* Corresponding author. Email: shaomingxing@bfsu.edu.cn(Mingxing Shao) , fanjing@bfsu.edu.cn(Jing Fan)

From the social perspective, only with ICT infrastructure and skills can social impact jointly influence the two health indicators respectively. NRI measures social impact by impact of ICTs on access to basic services, Internet access in schools, ICT use and government efficiency and E-Participation Index. It is quite obvious that social impact mainly emphasizes on a wider scope of a citizen's daily life, including access to education, government and Internet, etc. Thus, we can draw a possible path of ICT infrastructure laying a solid ground for the wide spreading of social applications including newborn care education, government support of childbirth, etc.

Likewise, given that ICT skills obviously highlight education level, it is connected even more tightly with social impact. A reasonable assumption is that maternal health requires higher concern than newborn health, especially in some developing countries where newborn is seen as more important than mothers and women may be treated as fertility machine in the birth giving process. Therefore, ICT skills, which is closely linked to education level, may influence maternal health through the channel of social impact.

Affordability and skills, on the other hand, fail to influence economic in our study. Affordability also fails in social case. Although not in line with our hypotheses, we can accept that the mediating effect is relatively weak in our research paradigm.

3. CONTRIBUTIONS

The contributions of this study can be categorized into theoretical and practical contributions.

Theoretically, Previous studies on ICT and health development are mostly conceptual, or remain micro level, focusing on the actual operation of technology. This study raises the theoretical altitude to a macro level, introducing a wider sample of countries from the United Nations when a large number of fellow studies only look at EU, other regional organizations or even just single countries. Second, on top of ICT factors, we bring in ICT impacts as mediations in our model, which is a huge innovation in the area. We hope it can serve as a good beginning for researchers to try other theoretical models and make it more and more sufficient and accurate. Finally, the study aims to review the realization of MDGs (2000-2015) and provide inspiration for SDGs (Sustainable Development Goals, 2015-2030).

Practically, this study aims to help global policy makers formulate health resource allocation and investment strategies, especially in the fields of healthcare and technology in developing countries. As we provide a new perspective to see how ICT factors associate with its impact, developing countries can use it as a guide to get financial assistance from or cooperate with other countries or organizations to improve health deliveries from their weakest areas with the most urgent needs according to their own conditions.

REFERENCES

- [1] Raghupathi, W. Raghupathi. (2013). Exploring the relationship between ICTs and public health at country level: a health analytics approach, *Int. J. Health. Inf. Syst. Inf.*, 8(3): 1-22.
- [2] Tavares, A. I. (2018). eHealth, ICT and its relationship with self-reported health outcomes in the EU countries. *International Journal of Medical Informatics*, 112: 104-113.
- [3] Seddon, J. J. J. M., & Currie, W. L. (2017). Healthcare financialization and the digital divide in the European Union: Narrative and numbers. *Information & Management*, 54(8): 1084-1096.
- [4] Si, Y. , Su, M. , Dong, W. , Yang, Z. , & Chen, X. . (2018). Importance of medical information for health-care utilization in china: a cross-sectional study. *The Lancet*, 392, S49.
- [5] Dieleman, J. L. , Krycia, C. , Agyepong, I. A. , Sarah, A. , Bollyky, T. J. , & Bump, J. B. , et al. (2019). The g20 and development assistance for health: historical trends and crucial questions to inform a new era. *Lancet (London, England)*, 394(10193): 173-183.

Full Research Paper

Understanding the Status of Reemployment of Elderly Talents in Digital Society: Evidence from Aged Job Search Website in China

Junjie Zhou*, Guoxin Wang, Huiying Wei, Chengcheng Cai,

Xiaomin Zhan, Xinling Lao, Wenhao Wang

School of Business, Shantou University, Shantou, 515063, China

Abstract: Retirees reemployment has been an effective way to fight the severe aging population and shortage of talented workers in many countries. However, the current situation of willingness and behaviors of reemployment for elderly talents, and the enterprises' hiring requirements for the elderly are still unclear. In order to address this question, a total of 1045 pieces of occupational position and 70636 pieces of elderly talents' information were collected from *Retiree Talent Website* by self-designed crawler. Based on prior studies and Holland's person-environment fit theory, content analysis, data mining, and statistical methods were used to explore the collected data. The results of descriptive analysis indicate that the reemployment demand of the elderly cannot be satisfied in most areas, especially in coastal areas; the number of jobs in labor-intensive industries and knowledge intensive industries is the largest, which is consistent with the industry that most elderly talents expect to engage in; enterprises tend to provide realistic positions and enterprising positions in recruitment, which is also consistent with the expectations of elderly talents. The results of exploratory analysis show that requirements of hard skills and soft skills in enterprises' recruitment are tightly associated with work experience; different types of industries and jobs have different qualification requirements for the elderly. Based on the results of data analysis, this study provides insights into the current supply and demand situation of elderly reemployment market in the digital society, and proposes corresponding policy recommendations for relevant departments.

Keywords: Elderly talents, Reemployment, Aging population, Information and communication technologies (ICTs)

1. INTRODUCTION

It is now nearly axiomatic that the severe aging population has led to considerably various social and economic issue^[1]. In recent years, aged workforces and shortage of young workers caused by the aging population are increasingly becoming a worldwide problem^[2], especially in many less-developed countries. Generally, effective labor supply is the foundation of further industrial upgrades, and the demographic dividend is an important factor for many developing countries to sustain long-term rapid development^[3]. However, the aging of the population, especially the aging of the working population, has led to a decline in the proportion of the workforce of the right age, which makes many less-developed countries lose their advantage in labor resources. Meanwhile, entering a society with an aging population will inevitably be accompanied by a large number of laborers with increasing age and weakening of their ability to innovate, adapt, and update knowledge, restricting the development of emerging industries and the upgrade of industrial structure. In a word, the aging population and shortage of talented workforces make the retirees reemployment on the agenda^[4].

Online recruitment services are now rapidly changing the landscape of the hiring traditions on the job market by providing an online platform to link both the job-seeker and employer^[5]. To some extent, the emergence of the online job search market has some positive effects on job seekers' outcome in the labor market^[6]. Because of its convenience and less age discrimination in real workplaces, more and more older adults, especially those

Junjie Zhou* (Corresponding author), E-mail: jjzhou@stu.edu.cn; Guoxin Wang, E-mail: 18gxwang@stu.edu.cn; Huiying Wei, E-mail: 18hywei@stu.edu.cn; Chengcheng Cai, E-mail: 18cccai@stu.edu.cn; Xiaomin Zhan, E-mail: 18xmzhan@stu.edu.cn; Xinling Lao, E-mail: 18xllao@stu.edu.cn; Wenhao Wang, E-mail: 20whwang@stu.edu.cn.

elderly talents, begin to search for jobs online after retirement. Prior studies found that retirees with lower retirement age, longer education and better health are more likely to continue working after retirement^[7]. Examples have also shown that most middle-aged and older employees are willing to adapt to the changing employment and re-enter the job market^[8]. Hence, an increasing number of older adults choose delayed retirement or bridge employment, which means that they could fully utilize their accumulated experience, knowledge, skills and social networks at work.

To identify the aging workers' employment status and then effectively utilize these retirees talents resource have become a crucial concern today^[4]. As more and more senior citizens start using internet and other ICTs to go shopping and communicating with their families and search for a job^{[9][10][11]}, online information has become an important channel for researchers to explore the employment status of elderly talents. Hence, it is relatively easier to reach elderly talents' employment demand through their UGC in online platforms, like aged support forums, job-seeking website, social media and other counterparts. Due to the current situation of willingness and behaviors of reemployment for elderly talents, and the enterprises' hiring requirements for the elderly are still unclear, we propose the research questions as follows:

Q1: What are the current status of elderly job-seekers and enterprise recruiters in reemployment?

Q2: What are the main requirements and qualifications of enterprises when recruiting elderly talents?

To answer these questions, we designed a python program and crawled data from *Retired talent website*, one of the most popular job search website for the elderly in China. Furthermore, data coding by four coders and Statistical methods were then conducted to analyze the data available. In the end, we organize the rest of this paper as follows: section 2 presents the literature review; section 3 describes the methodology, data source, and way of coding. The data analysis section outlines the results of our main findings. The final section is the conclusion as well as some contributions and limitations.

2. LITERATURE REVIEW

The questions about elderly reemployment related to some fields of labour economy, human resource management, and so on^[12]. In this study, we focus on the reemployment of elderly workers and retirees in digital society, then review two research streams of relevant studies. Specifically, we review some studies about elderly reemployment and research relative to the theory of Holland's RIASEC that we used to code the data later.

2.1 Elderly reemployment

Earlier in this study, we mentioned several semantically different words relative to older adults, like elderly worker, older worker, aged workforce, retiree, and so on. As the main goal of this study is to figure out the situation of demand and supply of retirees resource with expertise, we choose the word *elderly talent*, which is defined as older workers with professional knowledge, high skill or other expertise as being age 55 or over according to China labor policy. Then, we go over relevant studies that involve the possible factors underlying the phenomenon of reemployment; the antecedents that affect retirees' willingness to re-enter job markets; and potential outcome of the elderly used in digital society.

The aging of population is a worldwide issue. In China, the number of aged workers is increasing, with people generally living in a higher standards of life and advanced healthcare system^[13]. Due to the aging of population, many scholars have agreed that people can work longer^[14]. Actually, most retirees are healthy, energetic and have rich working experience and resources. Hence, it is unwise that older workers have to leave their job position at mandatory age, which absolutely means gratuitous depletion of highly talented resources. In order to sustain the economic developmen affected by the reduction of young labor force and fully utilize the elderly human resource, many countries have introduced reemployment policies to bring back retired employees, such as delayed retirement, bridge employment, pension benefits and so on^{[15][16]}.

However, older adults usually face more barriers when attaining a satisfactory job^[17]. Many elderly face a prolonged period in life in which they are relatively healthy and vigorous but lack a recognized role in the economic and social life of society. For example, age discrimination in hiring usually makes it difficult for older adults to re-enter the job market^[18]. In addition to these discrimination and barriers that affect the reemployment of the elderly, prior studies show that individual attributes, psychological factors, family factors and retirement planning also affect the reemployment and retirement decision-making of the elderly^[19]. Besides, from some perspectives of macroeconomics and social psychology, economic situation, rates of job loss, national retirement and social security policies will also affect the reemployment of elderly^[7] ^[20].

With the popularization of ICTs, it is worth noting that an increasing number of older adults begin to adopt and continue to use the internet in daily life^[10]. Compared with being introduced by relatives or friends, participating in social job fairs, registering with intermediary agencies, job advertisements, finding jobs based on Internet platforms (e.g., elderly peer support groups, online job markets, virtual communities) is becoming more and more popular for the elderly^[17]. In terms of the favor occupational position that retired elderly people are engaged in, prior studies also found that they are mainly engaged in factories, service industries, equipment operators. In general, most elderly people are used to choosing a similar job as before when they are reemployed.

2.2 Holland's person-environment fit theory of occupational choice

The theory of occupational choice, or another more famous title, the Holland's RIASEC model, which is one of the most popular professional career development theories^[21]^[22]. According to Holland's theory, personality is a critical factor in determining an individual's career choice^[22]. In general, people gravitate towards work environments that correspond to their occupational interests, and make vocational choices based on their perceptions of the degree to which work environments fit their interests^[22]^[23]. Based on Holland's viewpoints, he classified the work types into six categories, including realistic, investigative, artistic, social, enterprising, conventional (namely, RIASEC model).

Even though it has been decades since Hollands' theory was put forward, it is still widely used in practice today^[24]. Holland's theory of occupational choice focuses on the classification of occupations, and divides workers into different types according to their personality and career selection tendency, which provides a basis for workers to find their proper jobs, and makes it easier and more efficient for enterprises to recruit employees with different personalities according to special positions in the recruitment. Over the years, existing studies have shown the effects of Holland's theory on job performance, leadership behavior, vocational interests and choices, job attitudes^[23] ^[24] ^[25]. Therefore, our subsequent content analysis and vocational classification are based on this classic theory, so as to explore its adaptability and potential managerial implications generated by new jobs and new work environment in the digital age.

2.3 Summary

According to previous studies, few researches aiming to identify the situation of reemployment of elderly in China^[4], not to mention exploring the retirees' cognitive and behavioral under the context of digital society. In addition, less attention is paid to the online information produced by the elderly, which directly reflect the demand and supply side of enterprises and elderly workers in the job market. As a result, we plan to solve our questions through objective data, which can overcome the limitation of questionnaire conducted normally by national statistic department and some second-hand data widely used in most relevant studies.

3. METHODOLOGY

3.1 Research methods

We employed a qualitative-quantitative mixed approach to solve the research questions above. Specially, content analysis was first used to code the data; descriptive statistics were then adopted to analyse coded dataset

for answering both question 1 and 2; association rule mining (ARM) and Chi-square test were finally used to explore the data for answering question 2. More concretely, We encoded and classified the data according to the abovementioned theories and used software like SPSS Modeler 18 to conduct data analysis and visualization. In the end, we concluded our conclusions and carried out some implications for further research and managerial practice.

3.2 Data collection and coding

3.2.1 Data collection

The data of our study were collected from *Retiree Talent Website* (<http://www.ltxjob.com/>), one of the most popular job searching websites for elderly people in China. The website officially started to run on December 5, 2009, which is a service platform for retired talents to provide re-employment. Its purpose is to provide employment information to all kinds of retired talents and enterprises publicly. Until now, the website has more than 10,000 visitors per day, more than 10,000 recruitment units have actively registered on the website, and more than 10,000 retired talents have registered and released their job search information^[26].

We used a self-design program to crawl the recruitment data of enterprises and job-hunting data of retired talents online. The recruitment information released by enterprises (including job title, industry, workplace, responsibilities, qualification requirements, etc.) and the job-hunting information released by retired talents (including expected position, self-evaluation, working experience, expected working place and industry, etc.) were captured according to the regions. After cleaning the data available, we finally obtained 1045 pieces of enterprise data and 70636 pieces of retired talents' data.

3.2.2 Data coding

We encoded the processed data according to Holland's theory of occupational choice (RIASEC model) and the classification of industries by western economists in the early years. Based on Holland's occupational theory and industrial classification standards, we coded the job types from six categories: realistic, conventional, investigative, enterprising, artistic and social, and coded the industry types from resource-intensive, labor-intensive, knowledge-intensive and capital-intensive. The coding process includes three stages: firstly, four coders started to code 200 pieces of data for the positions provided by the enterprises and their industries, the expected positions of the elderly and the industries where they work based on the theory; secondly, we exchanged the problems in the coding process, such as the inability to determine the categories, then discussed and unified the definitions of the problems; finally, we continued to encoding the remaining data.

3.2.3 Variables

After cleaning and coding the data, we extracted variables (e.g., Job title, Industry, Job type, Workplace, Qualification Requirements and so on) from coded dataset for analysis further. Since the qualification is text data, we split the qualification into 11 new variables (i.e., age, gender, physical condition, registered residence, relevant certificates, educational background, work experience, hard skill, soft skill, job attitude, and personality traits) for subsequent analysis. The specific table of variables is following:

Table 1. Description of variables

Variables	Definition	Allowed values
Job Title	The concrete name of the job	e.g. Workshop Director
Industry	The type of industries that a enterprise is belonged to	1: Resource-intensive Industry 2: Labor-intensive Industry 3: Knowledge-intensive Industry 4: Capital-intensive Industrie
Job Type	Type of job based on the Holland's RIASEC model	1: Realistic Position 2: Conventional Position 3: Investigative Position 4: Enterprising Position

Variables	Definition	Allowed values
		5: Artistic Position 6: Social Position
Workplace	Province of workplace	e.g.ZheJiang Province
Age	Age requirements of enterprise in recruitment	
Gender	Gender requirements of enterprise in recruitment	
Physical condition	physical condition requirements of enterprise in recruitment	
Registered residence	Registered residence requirements of enterprise in recruitment	
Relevant Certificates	Relevant certificates requirements of enterprise in recruitment	
Educational background	Educational background requirements of enterprise in recruitment	
Work experience	Work experience requirements of enterprise in recruitment	
Hard skill	Hard skill requirements of enterprise in recruitment	
Soft skill	Soft skill requirements of enterprise in recruitment	
Job attitude	Job attitude requirements of enterprise in recruitment	
Personality traits	personality traits requirements of enterprise in recruitment	

4. DATA ANALYSIS

In order to answer the question 1, we employed a descriptive statistic approach to analyse the coded data to understand the current status of elderly job-seekers and enterprise recruiters in employment from four perspectives: workplace, job type, industry and qualification requirement (e.g., age, gender, physical condition, registered residence, educational background, hard skill, soft skill, job attitude and personality traits and so on). In these four perspectives, the situation of both demand side(elderly job-seekers) and supply side (recruiters enterprises) would be discussed.

4.1 Distribution of workplace about recruiters and elderly job-seekers

As shown in Figure 1, most of the enterprises providing job positions online are located in the southeast coastal provinces. Among all the provinces, Zhejiang ranked first (38.85%), followed by Shanghai (16.27%) and Guangdong (10.24%). Additionally, the number of elderly job-seekers shows a decreasing trend from coastal provinces to inland provinces. Among them, Guangdong (12.90%), Shanghai (12.36%) and Beijing (7.86%) are the most popular areas for the elderly, while demand in inland areas such as Qinghai (0.13%) and Tibet (0.01%) is less.

It is obvious that the employment demand for elderly is in short supply in all regions, that the gap between supply and demand in coastal areas is particularly prominent, which means the reemployment demand of the most elderly cannot be satisfied.

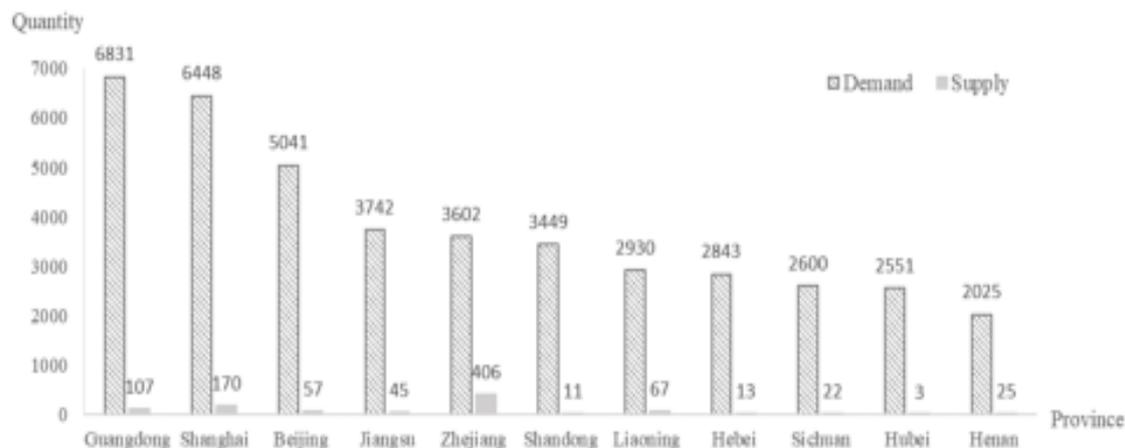


Figure 1. Comparison of supply and demand of the elderly in different provinces

4.2 The distribution of industry types that recruiters belong to and elderly job-seekers are expected

In terms of the industry type of recruitment enterprises, the results show that labor-intensive industries have the largest number(60%), followed by knowledge-intensive industries (34%), resource-intensive industries(3%), and capital-intensive industries(3%). Among the labor-intensive industries, the largest demands of the elderly are mainly machinery, construction, and food processing; While in knowledge-intensive industries, the largest demands focused on electronics, electricity, communications, professional services and cultural education.

For the expected industries, about 28% of the elderly have no specific requirements regarding the industries they want to engage in. Among the four main types of industries, labor-intensive industries occupy the highest proportion(44%), followed by knowledge-intensive industries (23%), while capital-intensive industries (1%) and resource-intensive industries(1%). Among labor-intensive industries, the jobs that the elderly expect to be engaged in are those that rely more on labor and less on technology and equipment, such as construction, machinery manufacturing and transportation industries; while for knowledge-intensive industries, the elderly prefer to do some work that is highly dependent on knowledge, experience and skills, such as finance, investment, and education.

4.3 The distribution of job types that recruiters are providing and elderly job-seekers are expected

For the recruiters, there are three main types of recruitment positions: skilled position (47%), management position (21%) and office position (18%). Further, enterprises tend to recruit electrical engineers, mechanical engineers, constructors and other positions with high operation skills in skilled positions; in office positions, enterprises tend to recruit accounting, cashiers, and other financial related jobs; while in management positions, enterprises tend to recruit sales, project, finance, production manager, or supervisor.

In terms of job seekers' expected positions, there are three main types of expected positions for retired elderly people, which are skilled position (37%), office position (24%) and management position (27%). It can be seen that the retired elderly are more inclined to work in positions related to skills, affairs and economic management.

4.4 Analysis on the requirements of employment qualifications for the elderly

In supply side, we conducted content analysis on the qualification requirements issued by enterprises. As shown in Table 2, among the 1045 enterprises, the most mentioned are work experience (58%), hard skills (47%) and education background (34%). For work experience, most enterprises require candidates to have 2-3 years of work experience; For hard skills, enterprises repeatedly mentioned that they want candidates to master drawing software and office software; For education background, college degree or above is preferred.

Table 2. Detail of qualification requirements

Variables	<i>N(Per.)</i>	Specific Requirements
Age	156(15%)	18-55 years old (71%); Over 55 years old (17%)
Gender	75(7%)	Male (52%); female (35%)
Physical Condition	117(11%)	Health (83%); No infectious and congenital diseases (17%)
Registered Residence	32(3%)	Zhejiang (28%); Shanghai (19%)
Educational Background	354(34%)	College degree or above (47%); Bachelor's degree or above (23%)
Relevant Certificates	115(11%)	Certificate of Engineer (17%); Certificate of electrician and welder (17%)
Work Experience	607(58%)	3 years working experience or above (19%); 2 years working experience or above (14%)
Hard Skills	490(47%)	Master CAD and other drawing software (19%); Master office software (17%)
Soft Skills	301(29%)	Communication skills (55%); Team spirit (31%)
Job Attitude	289(28%)	Responsible (36%)
Personality Traits	162(16%)	Cheerful personality (10%)

For demand side, through the content analysis of the self-evaluation of the elderly, we found that most of them held more positive self-evaluation on their job attitude and soft skills. Additionally, a few also mentioned the rich life experience, work experience and client relationships in their fields. In terms of hard skills, some have many years of design experience, and are familiar with auto, Pro/E and other software. In terms of education standard, only 0.02% of them had a junior college degree, and about 0.03% of them had a bachelor degree or above.

4.5 Other exploratory analysis

In order to answer the question 2, we further explore enterprises' data available by taking some exploratory analysis on qualification requirements, job types and industries.

4.5.1 Association rule mining: Qualification Requirement

we first carried out association rule mining (ARM) based on the Apriori algorithm to explore potential indicators of qualification requirement. Specially, SPSS modeler 18 was used to employ ARM and the two basic indicators were set appropriately (i.e., support was set as 10% and confidence was set as 80%).

Table 3. The results of the ARM

Consequent	Antecedent	Total	<i>Sup. %</i>	<i>Conf.%</i>	<i>Rule-support %</i>
Work	Soft Skills & Educational	149	14.258	91.275	13.014
Experience	Background				
Work	Soft Skills & Educational	118	11.292	90.678	10.239
Experience	Background & Hard skills				
Work	Job Attitude & Soft Skills	115	11.005	84.348	9.282
Experience	&Hard skills				
Work	Educational Background &	230	22.010	83.043	18.278
Experience	Hard skills				
Work	Soft Skills & Hard skills	213	20.383	81.690	16.651
Experience					

As shown in Table 3, work experience as the consequent item is associated with soft skills and educational background(support, 14.258%; confidence, 91.275), soft skills, educational background and hard skills(support, 11.292; confidence, 90.678%), job attitude, soft skills and hard skills(support, 11.005%, confidence, 84.348%), educational background and hard skills(support, 22.010%, confidence, 83.043%), soft skills and hard

skills(support, 20.383%, confidence, 81.690%). Through these five rules, we can draw the conclusion that when companies put forward the requirements of hard skills and soft skills in recruitment, they often require candidates to have certain work experience at the same time.

4.5.2 χ^2 analysis: qualification requirements and job type

Chi-square test (χ^2) was used to examine the relationship among job types and 11 qualification requirements. The results indicate that different jobs have no significant effect on age, gender, physical condition, educational background, hard skills and relevant certificates. However, there were significant differences in registered residence, work experience, soft skills, network resources, work attitude and personality traits. ($p < 0.01$) (see Table 4).

In order to understand the specific differences, we analyzed the crosstab. As shown in Table 5, Conventional position has higher requirements for registered residence (7%); Enterprising position has higher requirements for work experience (72%); Investigative Position has lower requirements for soft skills (16%); Social position has higher requirements for network resources (26%); Investigative position has lower requirements for work attitude (8%); Investigative position has lower requirements for personality traits (0%).

Table 4. The result of χ^2 analysis

Variable	χ^2	df	<i>p</i>
Registered Residence	18.934	5	0.002
Work Experience	30.983	5	0.000
Soft Skills	42.741	5	0.000
Network Resources	73.537	5	0.000
Job Attitude	16.412	5	0.006
Personality Traits	19.771	5	0.001

Table 5. The result of significant cross (chi square) analysis

		Job Type					
		Realistic	Conventional	Investigative	Enterprising	Artistic	Social
Registered	<i>N.</i>	9	14	0	7	2	0
Residence	<i>Per.</i>	1.81	7.33	0.00	3.23	5.88	0.00
Work	<i>N.</i>	251	115	14	157	20	50
Experience	<i>Per.</i>	50.60	60.21	56.00	72.35	58.82	60.98
Soft Skills	<i>N.</i>	104	56	4	94	14	29
	<i>Per.</i>	20.97	29.32	16.00	43.32	41.18	35.37
Network	<i>N.</i>	5	5	1	18	0	21
Resources	<i>Per.</i>	1.01	2.62	4.00	8.29	0.00	25.61
Job	<i>N.</i>	134	67	2	62	10	14
Attitude	<i>Per.</i>	27.02	35.08	8.00	28.57	29.41	17.07
Personality	<i>N.</i>	62	35	0	48	5	12
Traits	<i>Per.</i>	12.50	18.32	0.00	22.12	14.71	14.63

4.5.3 χ^2 analysis : qualification requirements and industry

Chi square test was used to study the different relationship between industries and 11 qualification requirements. Our results indicate that different industries have no significant effect on gender, physical condition, registered residence, education background, hard skills, soft skill, work experience, network resources, relevant certificates, work attitude and personality traits. However, there were significant differences on age and physical

condition ($p < 0.01$) (see Table 6). Further analysis of the values in the crosstab (Table 7) indicated that knowledge-intensive industries have a greater correlation with age, while labor-intensive industries have a lower requirement for age; resource-intensive industries have a greater correlation with physical condition.

Table 6. the result of χ^2 test

Variable	χ^2	df	p
Age	24.404	3	0.000
Physical Condition	17.998	3	0.000

Table 7. The result of significant cross (chi square) analysis results

		Resource-intensive Industry	Labor-intensive Industry	Knowledge-intensive Industry	Capital-intensive Industry
Age	<i>N.</i>	10	69	74	3
	<i>Per.</i>	35.71	11.06	20.67	8.57
Physical Condition	<i>N.</i>	11	58	45	3
	<i>Per.</i>	39.29	9.29	12.57	8.57

5. CONCLUSION AND DISCUSSION

On the basis of prior studies and the methods of content analysis and descriptive statistics, this study discussed the status of elderly job seekers and enterprise recruiters in employment from 4 dimensions: workplace, job type, industry and qualification requirement. It is found that there is a high imbalance between supply of job positions and demand of job seekers. The results of descriptive comparative analysis indicate that the reemployment demand of the elderly cannot be satisfied in most areas, especially in coastal areas; the number of jobs in labor-intensive industries and knowledge intensive industries is the largest, which is consistent with the industry that most elderly talents expect to engage in; enterprises tend to provide realistic positions and enterprising positions in recruitment, which is also consistent with the expectations of elderly talents. Additionally, we also examined enterprise's requirements for elderly human resources through association rule mining and Chi-square tests. The results of ARM show that requirements of hard skills and soft skills in enterprises' recruitment are closely associated with work experience. Besides, different types of industries and jobs have diverse qualification requirements for the elderly. Namely, Knowledge-intensive industries have higher requirements for the age of the elderly, while labor-intensive industries have lower; and the requirements for the physical condition of the elderly in resource-intensive are higher.

This study makes several contributions. First, the main theoretical contribution is that we examined the status of supply and demand of elderly employment in digital society from multiple perspectives, and put forward some insightful policy recommendations. Second, we integrated the classical theory of Holland's RIASEC model with numerous data collected online, expanded the theoretical boundary and enriched relevant theories of occupational choice. Third, based on the results of data analysis, this study provided insights into the current supply and demand situation of the elderly reemployment market, and offered some direction to future work.

The findings of our study also have some practical implications. For the elderly, it is reasonable for them to choose suitable strategies to match their qualifications to with proper position (realistic, conventional, investigative, enterprising, artistic, or social position) and industry type (resource-intensive, labor-intensive, knowledge-intensive and capital-intensive). As the demand for the elderly is far greater than the job supply that will cause a huge waste of elderly human resources, thus, more attention should be paid to the development and utilization of elderly human resources. Based on the international experience and findings of this study, we could put forward

several countermeasures that the authorities should take into consideration: (1) Implementing a flexible retirement system to effectively protect the right of the elderly to work. (2) Providing a platform for the elderly's learning and training. (3) Adjusting the industrial structure to encourage labor-intensive and knowledge intensive industries to absorb more elderly people. (4) Enhancing the legislation for the employment of the elderly.

This study has useful insights but also has several limitations. Firstly, we only studied the reemployment of the elderly in the Chinese context, and did not pay attention to diverse races. Secondly, we only used data from a widely used job search website, and the generalizability of the conclusions and suggestions may be undermined.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71501062; the Key Projects of Philosophy and Social Sciences Research of Chinese Ministry of Education under Grant 19JZD021; the Guangdong Provincial Science and Technology Research Project under Grant 2019A101002110; and STU Scientific Research Initiation Grant under Grant STF18011.

REFERENCES

- [1] Li, N. (2017). Is the problem of population aging real? an answer obtained from the labor force participation of the g7 countries. *China Population and Development Studies*, 1(2), 1-15.
- [2] Taylor M A, Bisson J B. (2020). Changes in cognitive function: Practical and theoretical considerations for training the aging workforce. *Human Resource Management Review*, 30(2): 100684.
- [3] Barsukov, V. N. (2019). From the Demographic Dividend to Population Ageing: World Trends in the System-Wide Transition. *Economic and Social Changes: Facts, Trends, Forecast*, 12(4), 167-182.
- [4] Chen S, Chen G. (2017). Do Employers Discriminate Against the Aging Workers when Hiring Employees in China?. *Journal of International Development*, 29(6): 805-824.
- [5] Yan, R., Le, R., Song, Y., Zhang, T., Zhang, X., & Zhao, D. (2019). Interview choice reveals your preference on the market: To improve job-resume matching through profiling memories. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 914-922.
- [6] Gürtzgen N, né Nolte A D, Pohlman L, et al. (2018). Do digital information technologies help unemployed job seekers find a job? Evidence from the broadband internet expansion in Germany. *European Economic Review*, 103657.
- [7] Wang, M., Zhan, Y., Liu, S., & Shultz, K. S. (2008). Antecedents of bridge employment: A longitudinal investigation. *Journal of applied Psychology*, 93(4), 818.
- [8] Lin S T, Sun J H, Chen C J. (2020). Re-enter the job market: job satisfaction and career transition competency among middle-aged and older adults. *Educational Gerontology*, 46(12): 774-784.
- [9] Vroman, K. G., Arthanat, S., & Lysack, C. (2015). "Who over 65 is online?" Older adults' dispositions toward information communication technology. *Computers in Human Behavior*, 43, 156-166.
- [10] Ghasemaghaei M, Hassanein K, Benbasat I. (2019). Assessing the design choices for online recommendation agents for older adults: Older does not always mean simpler information technology. *MIS Quarterly*, 43(1): 329-346.
- [11] Sun K, Zhou J. Understanding the Impacts of Internet Use on Senior Citizens' Social Participation in China: Evidence from Longitudinal Panel Data. *Telematics and Informatics*, 101566.
- [12] Harris, K., Krygsman, S., Waschenko, J., & Laliberte Rudman, D. (2018). Ageism and the older worker: A scoping review. *The Gerontologist*, 58(2), e1-e14.
- [13] Gong, C. H., Kendig, H., & He, X. (2016). Factors predicting health services use among older people in China: An analysis of the China Health and Retirement Longitudinal Study 2013. *BMC health services research*, 16(1), 1-16.
- [14] Morrow-Howell N, Wang Y. (2013). Productive engagement of older adults: elements of a cross-cultural research agenda. *Ageing International*, 38(2): 159-170.

- [15] Beehr, T. A., & Bennett, M. M. (2015). Working after retirement: Features of bridge employment and research directions. *Work, Aging and Retirement*, 1(1), 112-128.
- [16] Lei, L., & Heng, X. (2019). Research on the Effect of China's Delayed Retirement Age Policy on Young People Employment. *Management Review*, 31(1), 255.
- [17] Dordoni, P., & Argentero, P. (2015). When age stereotypes are employment barriers: a conceptual analysis and a literature review on older workers stereotypes. *Ageing International*, 40(4), 393-412.
- [18] Neumark, D., Burn, I., Button, P., & Chehras, N. (2019). Do state laws protecting older workers from discrimination reduce age discrimination in hiring? Evidence from a field experiment. *The Journal of Law and Economics*, 62(2), 373-402.
- [19] Wang M, Shultz K S. (2010). Employee retirement: A review and recommendations for future investigation. *Journal of Management*, 36(1): 172-206.
- [20] Cahill K E ,Giandrea M D, Quinn J F. (2013).Bridge employment ,in M Wang(ed.),*The Oxford Handbook of Retirement*, Oxford: Oxford University Press, 293-310
- [21] Holland J L. (1959).A theory of vocational choice. *Journal of counseling psychology*, 6(1): 35.
- [22] Holland J L.(1997).*Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Psychological Assessment Resources.
- [23] Woods, S. A., Edmonds, G. W., Hampson, S. E., & Lievens, F. (2020). How our work influences who we are: Testing a theory of vocational and personality development over fifty years. *Journal of Research in Personality*, 85, 103930.
- [24] Sheldon, K. M., Holliday, G., Titova, L., & Benson, C. (2020). Comparing Holland and Self-Determination Theory measures of career preference as predictors of career choice. *Journal of Career Assessment*, 28(1), 28-42.
- [25] Al-Ajlouni, M. M. (2018). Human Resources and their Tendency to Information Security Crimes Based on Holland Theory. *Information Resources Management Journal (IRMJ)*, 31(4), 44-58.
- [26] Retiree Talent Website.(2020.01.28). <http://www.baimin.com/top/97672.html>

Full Research Paper**Verifying Online Health Rumors on Social Media: An Empirical Research****Based on the Stimulus-Organism-Response Framework***Ying Liu¹, Yi Jiang^{1*}, Shuang Zhang¹, Zhenkun Wei¹*¹ School of Economics and Management, China University of Geosciences, China

Abstract: Many health rumors often appear on social media. The widespread spread of these rumors not only damages the reputation of social media platform, but also has a serious impact on individuals and society. Therefore, the research on the spread of health rumors has attracted academic attention. Based on the stimulus-organism-response framework, this study investigates the mediating role of people's perception for verifying rumors, especially adding the variables of social consistency to explore the interacting effect on the intention to verify. An experiment was conducted to test the research model and the empirical result indicated the positive effect of perceived authenticity and perceived importance and the negative effect of perceived trust on verifying rumors. Besides, social consistency will moderate positively the effect of perceived trust and moderate negatively the effect of perceived importance. Compared to the wish rumors, the dread rumors will produce higher perceived importance and perceived trust. This study provides theoretical implications and practical guidance for online rumor research and practices.

Keywords: Online rumors, Verifying behavior, Covid-19, the S-O-R framework

1. INTRODUCTION

The internet provides convenience for users to quickly obtain rich health information and has become the main choice for people to search for health knowledge. However, hidden dangers also arise, such as the proliferation of online health rumors. All kinds of articles about health and disease often appear on social media. The fuzziness of the source and content of these articles often makes it difficult for the information receiver to select the truth from them^[1]. When health-related rumors spread wantonly and become a "virus", it will cause users unnecessary anxiety and bring trouble to their lives. Some alarmist health rumors will arouse the public's psychological panic and destroy the social trust system^[2]. For example, in the period of COVID-19, the rumor "taking Isatis Root Granules, a kind of medicine for clearing heat, would play a role in reducing and eliminating the virus" caused a big amount of discussion on the internet, and then caused people to hoard related medicine for speculation. Therefore, the research on the spread of health rumors has attracted the attention of the academic community.

Rumor studies have argued that the main motive of rumor sharing is fact-checking^[3]. When people perceive the rumor as important, they share it with others in an attempt to verify its truthfulness. On the Internet, people verify facts not only by sharing information but also by seeking additional information on social media^[4] or by searching websites^[5]. Thus, rumor verification and rumor sharing should be treated separately. However, prior research mainly focused on people's sharing behaviors from the effect of message label and message valence and relatively there is a lack of research on rumor verification. Thus in this study, we try to unfold the process of verifying rumors and clarify the main motivations and their interactive relationships.

Compared with the traditional media, such as newspapers and television, rumors on the social network have

* Corresponding author. Email: wuhanjoey@163.com

a more complex content structure and communication modality, especially during health crises. First, online rumors are often disseminated with a characteristic of ambiguity, by concealing the actual information sources and the total news clue [6]. It is one of the core essences of distinguishing rumors from true news. However, this distinction might be unobserved in social networks in that internet users have been used to fragmented online content [7]. And then it comes to an opportunity for “rumors” to hide and tamper with part of real information through digital media, which is not easily detected by internet users.

Second, especially during health crises, individuals with an excessive health concern hold an obsessive belief that they will have a serious illness soon. Research has shown a link between health anxiety and a need for health information. As online health information is often incomplete and ambiguous, they may need to spend an excess of time confirming the validity of the information [8].

Third, with the popularity of social media, it enlarges consumer sovereignty and gives users full opportunities to voice. The evaluation of news on social networking sites, WeChat circle of friends, microblog, blog, and other media have had a great impact on people's concept and cognition. When people face unverified information regarding the practice of medicine and healthcare, online reviews may affect people's confirmation behavior. Previous studies on rumor have argued that the main factors of rumor creation. Even with the emergence of social media, the research on online rumors just regarded social network as an external environment of rumor transmission to check the role of the above factors on online behaviors.

In all, the lack of understanding the above three natures of online health rumors on the social network makes it difficult to clarify what motivates people to verify the rumors. Hence, this paper seeks to explore how the ambiguity of rumors take effect on the verifying behavior on the social network by influencing people's perception.

To address the above research question, the target is divided into the two-fold objective: the construction of an integrated model demonstrating the process of rumor verifying and the empirical analysis of components' interactive effect during this process. Theoretically, the need to understand how people's perceptions elicited by the health rumors affect the verifying behavior leads us to the Stimulus-Organism-Response (S-O-R) framework. Although the S-O-R model takes a rather broad approach, it provides an adequate theoretical framework for this study. According to the S-O-R framework, environmental prompts are the starting point of an intentional or unintentional psychological or behavioral process. Specifically, stimuli from environmental prompts affect individuals' internal cognitive and affective states, including their perceptions, experiences, and evaluations, which in turn affect their behavior. Given that rumors' natures work as an important social interactivity environmental stimulus that provides viewers unique information perception, it is logical to apply the S-O-R model as the theoretical framework for studying this phenomenon. Further empirically, to test how the factors interactively operate, we collected the data sets through an experiment with a questionnaire survey during the period of COVID-19.

This study extends prior research in several key ways. First, we expand the rumor research, by identifying the mediating role of people's perception, which uncover their interrelated relationship during the verifying rumors. Secondly, we specially explored the characteristics of the content ambiguity in comparison with prior research on rumors' type, to reflect effectively the changing principle by the embedded digital tools, revealing the conditions under which people's verifying behaviors in a social environment. Third, our study integrated the S-O-R model and social network research into the context of rumor. which contributes to the rumor research by dissecting the microscopic and potential structure behind the contents. Finally, we provide practical insights and algorithm ideas related to AI-enabled rumor recognition for social platform managers to reduce the possibility of sharing unverified information.

2. LITERATURE REVIEW AND THEORETICAL DEVELOPMENT

2.1 Online Health Rumors and Emergency

Nowadays, with the development of new media and social platforms, the speed of information dissemination has been accelerated and the scope of information dissemination has been expanded. Based on the theory of Peterson and Gist, a rumor is an unproven explanation or reason for something of public concern. Roger, an American scholar, defined health communication as transforming medical research results into common medical knowledge known to the public and improving the life quality and health level of a country or region through changing attitudes and behaviors. At present, the academic research on rumors mainly focuses on the generation stage of rumors and the influence of the characteristics of rumors on user behavior, but there is no detailed understanding of the inner role and perception of users.

By far, the academic research on rumor mainly focuses on the characteristics of rumor itself. The dimension of rumor stimulus includes two variables: content fuzziness and content source, both of which are characteristics of rumor content.

The fuzziness of rumors plays an important role in the public's perceived uncertainty, perceived importance and perceived trust, while the internal process of the public plays a crucial role in the cognition and dissemination of rumors. Yi et al. studied the inner experience of users when crisis events occurred through the humor model in his research, using perceived sincerity and perceived severity as mediators of the research in his article [9]. Therefore, rumor fuzziness has a significant impact on users' inner perception, and the research on this is of great significance to the understanding of users' reactions in the process of rumor spreading.

In the researches on different types of rumors, most of them are classified according to the internal reaction of the audience after receiving the rumor stimulation. They play a key role in the chain of rumor propagation, and their internal reaction has an impact on the communication intention. Individual participation forces people to trust and share [10]. Fear-type rumors lead to people's willingness to trust and share more than hope. Bordia proposed a categorization of claims (rumor interactive analysis system) and found the difference between horror rumors and hope rumors in content categories related to anxiety [6]. Scholars have shown that fear-based rumors and hope-based rumors have different effects on users' behaviors.

However, existing studies have not yet solved the influence of the psychological process on user behavior or even its important role in the whole rumor propagation process. The current studies have not yet explained it from the perspective of the rumor propagation process. At the same time, the current research on user psychology has not been carried out in the context of emergencies and medical and health events. Therefore, it is necessary to conduct a study on user psychology to have a deeper understanding of the rumor spreading process, which needs to be refined.

In addition, users' social networks are complex and diverse, and their social properties also have an impact on their inner perception and behavior. Heshan Sun believes that in the process of technology acceptance, users often imitate others in social contact and degrade the information and cognition they have acquired so as to maintain consistency with their social partners in social contact [11]. This is the way that users tend to conform to social behavior due to herd behavior (herd mentality). In this case, the individual behavior characteristics, the evolution of views and the process of information transmission in social networks have become the research spots in the interdisciplinary field. Therefore, it is necessary to use a psychological research model to study user perception, to have a deeper understanding of the whole rumor propagation process.

2.2 The S-O-R Framework

In order to study the influence of users' inner perception on user behavior and rumor propagation, the S-O-R framework is adopted, which provides a sufficient theoretical framework for the purpose of this study. The S-O-R framework was first proposed by environmental psychologists Mehrabian and Russell [12]. It is based on the

influence of different environmental stimuli on people's cognition, emotion and behavior. This model has been widely used in the study of social behavior by focusing on the influence of people's inner perception and emotion on people's behavior. Victor R. Prybutok, using the S-O-R framework, proposed that the sense of immersion and perceived interest are two important mediating variables, and explained the theoretical mechanism of how social existence, as an environmental and social factor, affects audience loyalty [13]. However, the SOR model has not been used to explain the rumor propagation process.

In this study, the S-O-R model is applied to study the spreading process of network health rumors in the context of public health emergencies. Stimulation is a rumor generating mechanism. For rumor stimulation of different types or fuzziness, users will feel different values in their hearts, so as to influence user behavior and communication process. Meanwhile, social consistency, as an important environmental stimulus, will trigger the psychological process of individual response. Considering that user's psychological perception is an important part of rumor propagation, it is reasonable to take the S-O-R model as a theoretical framework to study this phenomenon. In summary, the SOR model provides a structured guide to building a comprehensive model that opens up the black box of how Internet rumors affect the public and drive them to react differently.

Based on the S-O-R framework, this study proposed the research model as shown in Figure 1 and describe how rumor type and ambiguity influence the intention to verify through the perceptions. In the next section, we develop the hypotheses to explore the relationship between the features of online rumors and rumor verifying behavior.

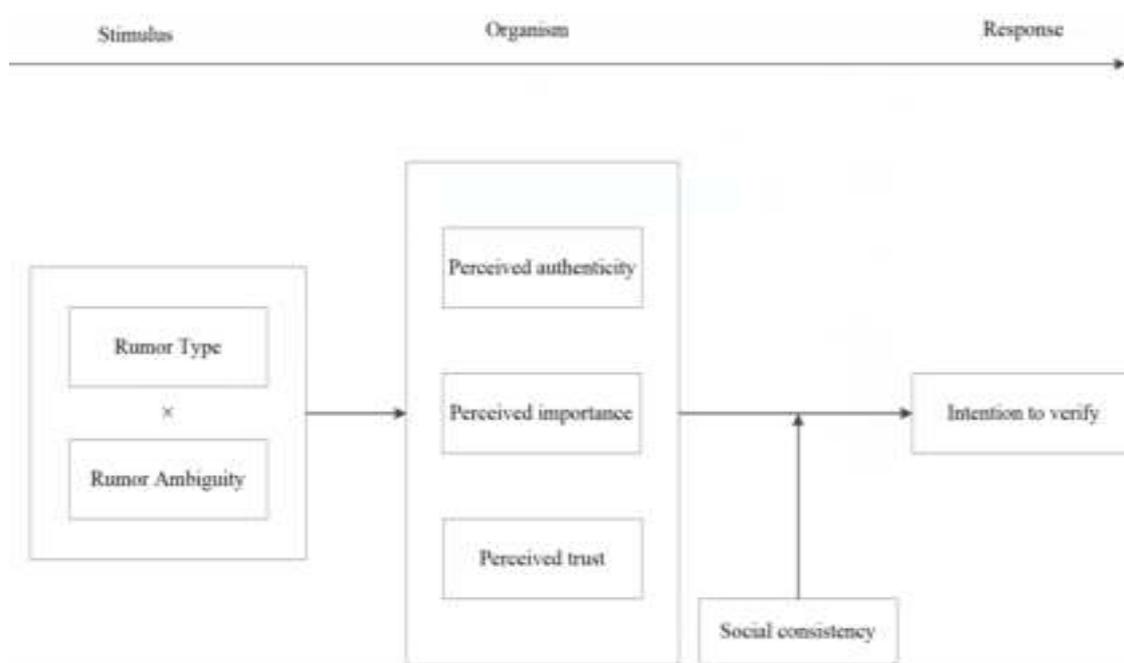


Figure 1. Research framework

The individual's confirmed intention emphasizes a trust tendency, and individuals who have a low degree of trust in uncertain information will have a stronger confirmed intention. Perceived authenticity measures the relevant knowledge an individual possesses, including scientific knowledge, experience, reliable information, etc. For example, for cancer-related rumors, ordinary people may have a stronger intention to confirm than professional doctors, or people who have had similar experiences will easily judge the authenticity of uncertain information, and the corresponding verification intention may not be very strong. As such, we hypothesize:

H1: The individual confirmed intention is positively related to the perceived authenticity.

Oh and Lee proposed that perceived message importance was associated with intentions to verify and share. People often pay more attention to things that are of interest to them or that are important to them and take corresponding measures and actions to achieve their expectations^[14]. The same is true for the role of rumors, such as uncertain information. For the rumors in the COVID-19 epidemic used in research, the unique environment has brought a certain degree of anxiety to individuals, and the manifestation of importance will also be strengthened. Collectively, we hypothesize that:

H2: The individual confirmed intention is positively related to the perceived importance.

Perceived trust reflects the individual's acceptance of rumors. Individuals with low levels of trust in rumors are more likely to ignore the rumor, and individuals with high levels of trust in uncertain information are more willing to produce certain confirmation intentions to satisfy themselves. Verification requirements. However, Chua shows that most of the experimental participants with medical background did not show the intention to confirm the rumor information^[10]. Based on these reasons, we propose:

H3: The individual confirmed intention is negatively related to the perceived trust.

Perceived authenticity means people's perceptual evaluation of the authenticity of rumors. The arrival of the information explosion era has greatly reduced the importance of information authenticity, and in the context of the COVID-19 epidemic, sharing information can play a role in alleviating anxiety. But even so, as a person in a social network, the authenticity of the information shared still has a certain social significance for him^[15].

The individual's perception of importance not only affects the confirmation of intention but also stimulates a strong intention to share. People tend to share information that they think is important to friends or relatives and other people they care about. The important information here is often not only important to yourself but also to people around you and the surrounding environment.

People tend to have more confidence in what they believe in to support their ideas and tend to convey their opinions through sharing. This means that trust can promote the individual's intention to share uncertain information to a certain extent, thereby further affecting the spread of rumors. Therefore, for rumors, such specific uncertain information, the perceived trust of individuals will promote the sharing intention and show a certain positive correlation trend, then we propose:

H4a: Social consistency moderates positively the effect of perceived trust on the intention of verify.

H4b: Social consistency moderates negatively the effect of perceived importance on the intention of verify.

H4c: Social consistency moderates positively the effect of perceived authenticity on the intention of verify.

Rosnow et al suggest that under the influence of importance or belief, dread rumors are transmitted more often than wish rumors even with controls for other characteristics^[16]. Prashant studied the influence of dread rumors and wish rumors on anxiety and find their differences between anxiety-related content categories^[6]. Under normal circumstances, we tend to believe that dread rumors can more easily resonate with the audience. The tense atmosphere brought by dread rumors makes people believe in the authenticity of the event, and can also attract the attention of individuals^[17]. In addition, people have a mentality of believing what they have rather than what they don't. Therefore, dread rumors can also make individuals feel more trustworthy. Consistent with the consensus reached in prior literature, we expect that:

H5: Compared to the wish rumors, the dread rumors will produce higher perceived authenticity, perceived importance and perceived trust.

When people accept a dread rumor, their first response is panic and suspicion. At this time, people tend to think more rationally about the stimuli brought by external information. In this case, the cause and effect of rumors

is more important than the source. Vague causal logic can give individuals a higher sense of trust and dependence than information with source ambiguity. As Michael A. Kamins mentioned in their research, bad rumors are more trustworthy when the logical expression is uncertain and have a greater impact on the spread of rumors in social media ^[18]. Hence, we propose:

H6: When the rumors are of the dread type, the causal ambiguity of the rumors (compared to the source ambiguity) can produce higher perceived trust.

The individual's perceived authenticity is embodied in the individual's degree of certainty about the content of the rumors. It is hoped that wish rumors can give people a pleasant belief than dread rumors, but at this time individuals will also pay more attention to the source of information. Compared with the causal ambiguity of rumors, individuals may be more sensitive to rumors of uncertain sources, because the causal logic of the source is easier to judge in a short time. Unrestricted information sources can avoid the individual's suspicion to a certain extent, otherwise, the individual will pay too much attention to the source of rumors and affect the spread of rumors. Therefore, the uncertain source of information can make the individual have higher perceptual certainty. Then we propose the following hypothesis:

H7: When the rumors are of the wish type, the source ambiguity of the rumors (compared to the causal ambiguity) can produce higher perceived authenticity.

3. METHODOLOGY

3.1 Study design

To test our hypotheses and explore the impact of different types and ambiguity of rumors on users' confirmed and sharing intention under different variables, an experimental survey was conducted. To design the rumor content in the experimental questionnaire, we first selected real-time rumor refutation information from January 18, 2020 to March 15, 2020 on the "Authentic Platform" which is a national platform dedicated to news verification in China, and then adopted the web crawler technology to collect relevant data. Finally, the main field contents include title, date, verifier, type of verifier, conclusion, rumor type and URL.

A total of 494 rumor data were collected and coded independently by 5 members in 0-1 coding mode to determine whether the information was wish or dread and whether it was source ambiguity or causal ambiguity. Only the rumors that were coded consistently by 5 members were used as the experimental design. In the end, there were 5 rumors of Causal ambiguity×Wish type, 5 rumors of Causal ambiguity×Dread type, 7 rumors of Source ambiguity×Wish type, and 8 rumors of Source ambiguity×Dread type. To ensure the consistency of the number of rumors in each group, three of them were selected for the experimental design. Finally, the 12 rumors after screening were designed into four questionnaires with 3 rumors for participants to browse according to the classification of rumors stimulation.

3.2 Stimuli

We employed a 2 (ambiguity: source vs. causal)×2 (type: wish vs. dread) factorial design. A total of 4 situation questionnaires were formed: Causal ambiguity×Wish type、Causal ambiguity×Dread type、Source ambiguity×Wish type、Source ambiguity×Dread type. As shown in Tabel 1, a total of 12 rumors during the outbreak of COVID-19 epidemic were used as questionnaire questions. Five-point scale (strongly disagree, disagree, general, agree, strongly agree) developed by Likert was used to measure the parameters.

Table 1. Rumors selected as the experimental stimuli

Rumors stimulate	Rumor content
Causal ambiguity × Wish type	Drinking brown sugar, ginger, shallot and garlic boil water, you will not be infected with the COVID-19 epidemic.
	Since February 3, 3 million masks have been put into 1000 pharmacies in Beijing every day.
	Chengdu citizens can make an appointment to receive masks on the Tianfu Citizen Cloud APP on February 4.
Causal ambiguity × Dread type	Someone in the car from Xiantao, Hubei Province, escaped after being detected high temperature.
	Five people traveling from Wuhan to Shanghai stayed in Jianguo Hotel and all were infected.
	Sichuan highway has been closed.
Source ambiguity × Wish type	The COVID-19 epidemic can be protected by making use of diapers.
	The COVID-19 epidemic can be effectively prevented and treated with antibiotics.
	The COVID-19 epidemic can be prevented by taking vitamin C effervescent tablets.
Source ambiguity × Dread type	The COVID-19 epidemic will outbreak seasonally like the influenza virus.
	The tap water is disinfected with the upper limit standard and chlorinated and should be left standing for at least two hours.
	The COVID-19 epidemic will spread through seafood.

3.3 Participants and procedure

A total of 210 participants were recruited for this experiment, and the network questionnaire was used to conduct the experiment. Table 2 shows detailed descriptive information about participants. We deemed college students represent a large portion of Web users and the sample used in our study is feasible. Of the 210 participants, 43 percent were men. Most of the participants (76 percent) were between 19 and 35 years old.

We set up a virtual social network environment, obtained 12 rumors after screening through the pre-coding, then searched for relevant information on Weibo and other social platforms, finally collected the main dissemination content. Taking into account the differences in information comments and user attention from different sources, each piece of rumor information was manually commented on, and screenshots were processed by imitating social platform design information as the content viewed by participants in the formal experiment.

Before the start of the experiment, we informed the participants that they will receive a small gift at the end of the experiment to encourage them to take part in the experiment seriously. At the beginning of the experiment, participants were randomly assigned to one of four rumor scenarios, they were instructed to browse 3 rumors in the specific rumor scenario and answered relevant questions respectively. We changed the questionnaire title to "Research on the Influencing Factors of Online Information Dissemination", and gave simulated scenarios to better reflect the individual's psychological reaction process during the severe period of the new coronary pneumonia epidemic, and try to avoid the impact of time delay on the experimental results. The specific content of the scenario simulation in the questionnaire is: "Looking back in time, assuming that you are an ordinary citizen who is forced to stay at home due to a severe epidemic, and you pay attention to the trend of the epidemic every day. You need to be rational about the credibility of the following Weibo information, think about how this information will affect your life, and make response and decision accordingly." Participants need to imagine that they are in such an environment, trace the time back to the more serious stage of the epidemic, and answer basic personal information such as gender, age, and education level.

During the experiment, we will not tell the participants that the contents of the questionnaire items are "rumors" or "unsubstantiated information", so as not to affect the results of the research due to the ambiguity of the subjects. The entire experimental survey was supervised and directed by five research assistants. In addition, the questionnaire items used in the experiment were modified after the preliminary survey.

3.4 Measurement

Perceived authenticity is to measure whether participants have the relevant background knowledge to help identify rumors. When there is a certain knowledge background or relevant experience, individuals can often make rational judgments about uncertain information. Refer to the scale of RL Rosnow^[16], we measured the perceived uncertainty by "I have relevant knowledge (such as: scientific knowledge, experience, reliable information, etc.) to judge the authenticity of this information".

Perceived importance represents the importance of an individual to uncertain information, which is divided into three dimensions, we measured the perceived importance by "I think this information is important to me" and "The information is important to people around me".

Perceived trust refers to the tendency and dependence of an individual on rumors. It is measured by two items of "I believe this information" and "I will rely on this information to make subsequent decisions" with reference to the Chua, A&Banerjee's scale^[10].

Social consistency means that people have the same tendency towards a certain comment or event in social networks. It is measured by two items of "When I read the comments of netizens on the information, I think the netizens' views on the authenticity of the information are the same as my own." and "When I read the comments of netizens on the information, I think the netizens' attitude toward the information is consistent with my attitude".

Intention to verify is measured by three items: "I will browse related websites to judge the authenticity of this information", "I will pay attention to the latest development of this information" and "I will go to other websites to find useful information".

4. DATA ANALYSIS AND RESULTS

We evaluated the convergence effect of the items based on the discriminant validity and convergence validity of the research tool. The convergence validity is tested in two steps. First, Cronbach's alpha values of intention to verify (ITV), perceived trust (PT), social consistency (SC) and perceived importance (PI) are 0.896, 0.849, 0.937 and 0.891, respectively, greater than 0.8. Second, all project factor loads exceeded 0.63 (Table 2).

Table 2. Item loadings (KMO=0.825)

	Cronbach alpha	Intention to Verify	Perceived Trust	Social Consistency	Perceived importance	Perceived authenticity
ITV1	0.896	0.931	0.138	-0.001	0.121	0.079
ITV2		0.906	0.112	-0.027	0.117	0.207
ITV3		0.736	0.149	0.027	0.558	-0.033
PT1	0.849	0.111	0.870	0.330	0.132	0.010
PT2		0.150	0.835	0.263	0.261	0.015
SC1	0.937	0.020	0.254	0.939	0.060	0.005
SC2		-0.044	0.281	0.917	0.147	-0.010
PI1	0.891	0.294	0.538	0.238	0.661	0.116
PI2		0.402	0.519	0.190	0.636	0.077
PA1	-	0.174	0.028	-0.005	0.052	0.979

196 data were used for regression analysis, and an overview of linear regression results is shown in Table 3, summarizing the main findings of the three models tested, Model 1 only considers the main effect, Model 2 add the moderator to Model 1, Model 3 add the interaction term between the moderator and the independent variable to Model 2. To guarantee the statistical correctness of each model, variance inflation factor(VIF) test is carried

out for each model, multicollinearity problems is not a problem, as the maximum VIF index calculated was 1.71 for the research variables, it is far less than the standard value of 10.

Table 3. Results of linear regression

Variables	Model 1	Model 2	Model 3
Perceived authenticity	0.196(2.98)**	0.193 (2.99)**	0.166(2.39)**
Perceived importance	0.882(8.61)***	0.914(9.07)***	0.878(8.68)***
Perceived Trust	-0.409(-4.05)***	-0.286(-2.68)***	-0.246(-2.23)*
Social Consistency		-0.254(-3.11)**	-0.184(-2.12)*
Perceived authenticity × Social Consistency			-0.046(-0.53)
Perceived importance × Social Consistency			-0.367(-2.39)**
Perceived Trust × Social Consistency			0.367(2.64)**
_cons	1.265(4.47)***	1.585(5.37)***	1.448(4.90)***
R ²	0.375	0.405	0.43
Mean VIF	1.78	1.79	2.08

Notes: t-statistics in parentheses. ***p<0.001; **p<0.01; *p<0.05

The explanatory power (R²) of Model 1 is 0.375. In terms of the results of the estimated coefficients in Model 1, as far as people's perceptions are concerned, the perceived authenticity has a significant positive impact on the intention to verify rumors ($\beta = 0.196$, $P < 0.01$), which supported hypothesis H1. Similarly, the perceived importance has a positive impact on rumor verifying ($\beta = 0.882$, $P < 0.001$). Thus, hypothesis H2 is also accepted. Differently, the perceived trust has a statistically significant negative impact on rumor verifying behavior ($\beta = -0.409$, $P < 0.001$), which means the hypothesis H3 is supported.

We then examined the moderating effect of social consistency through model 3. And Model 3 with moderating effect has a stronger explanatory power and better model fitting effect than Model 1 (R²= 0.43). The results demonstrate that social consistency exerts a significant moderating impact on the influence of perceived importance ($\beta = -0.367$, $P < 0.01$) and perceived trust ($\beta = 0.367$, $P < 0.01$) on rumor verifying. As depicted in Figure 2 and Figure 3, at high levels of social consistency, the effect of perceived importance on verifying rumors will decrease while the perceived importance's effect will increase. This means that the social consistency can take different moderations on the process of rumor verifying behavior, H4a and H4b are supported. And the moderating effect on the influence of the perceived authenticity is not significant ($\beta = -0.046$, $P > 0.05$), H4c is not supported.

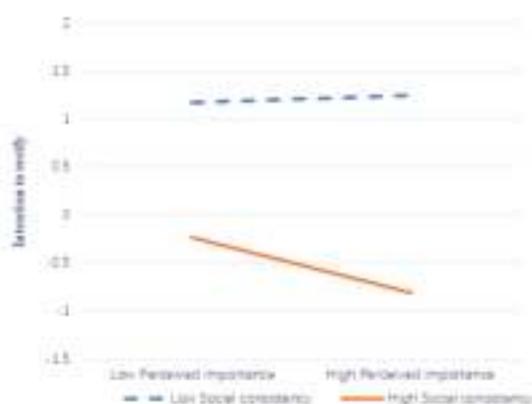


Figure 2. The moderation effect on the perceived importance

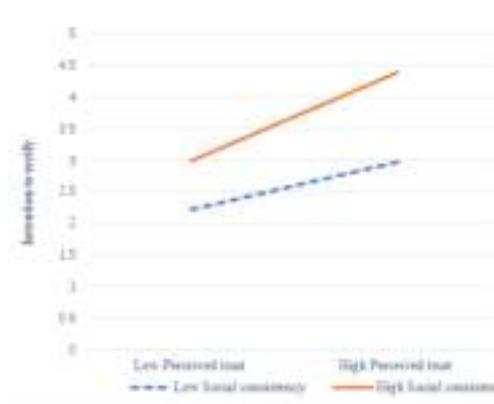


Figure 3. The moderation effect on perceived trust

We used an independent sample t-test to verify H5. The results indicated that the dread rumors were generally perceived as conveying a more increasing in the perceived authenticity ($M_{\text{dread}} = 3.33$, $M_{\text{wish}} = 3.17$, $p > 0.05$), perceived importance ($M_{\text{dread}} = 3.20$, $M_{\text{wish}} = 2.75$, $p < 0.001$) and perceived trust ($M_{\text{dread}} = 2.85$, $M_{\text{wish}} = 2.53$, $p < 0.001$). Except for the perceived authenticity, the other two perceptions will increase when people face dread rumors and H5 can be partially supported. Similarly, to test H6 and H7, we used the independent sample t-test in the dread rumor group and the wish rumor group respectively, in order to check the different effects of the source ambiguity and the casual ambiguity of rumors. The results indicated that when the rumors are of the dread type, the causal ambiguity of the rumors (compared to the source ambiguity) can produce higher perceived trust ($M_{\text{source}} = 2.68$, $M_{\text{causal}} = 3.03$, $p = 0.008$). When the rumors are of the wish type, the source ambiguity of the rumors (compared to the causal ambiguity) can produce higher perceived authenticity ($M_{\text{source}} = 3.34$, $M_{\text{causal}} = 3.01$, $p = 0.02$). H6 and H7 are both supported.

5. DISCUSSION AND IMPLICATIONS

The starting point of this paper is to understand: 1) How people's behavior of verifying online health rumors is triggered through the perception of rumors with social consistency. 2) The impact of different rumor type and ambiguity on personal perception. According to experimental data, most of the proposed hypotheses were supported. The experimental results show that people's intention to verify the online rumors is related to perceived authenticity and perceived importance positively, and perceived trust influences the verifying behavior negatively. This is because perceived trust reflects the individual's acceptance of rumors. Individuals with low levels of trust in rumors are more likely to ignore the rumor, and individuals with high levels of trust in uncertain information are more willing to produce certain confirmation intentions to satisfy themselves.

The experimental results also show that the social consistency exerts a significant moderating impact on the influence of perceived importance and perceived trust on rumor verifying. In other words, at high levels of social consistency, the effect of perceived importance on verifying rumors will decrease while the perceived importance's effect will increase. Besides, dread rumors may create more psychological perceptions, and when the rumors are of the dread type, the causal ambiguity of the rumors (compared to the source ambiguity) can produce higher perceived trust. When the rumors are of the wish type, the source ambiguity of the rumors (compared to the causal ambiguity) can produce higher perceived authenticity.

5.1 Theoretical implication

The theoretical contributions of this study mainly include the following three aspects: first, there is less research on online health rumors during the crisis. Although previous literature has confirmed that the credibility of rumors' type significantly affects the spread of rumors, the ambiguity of rumors has no impact on the spread intention. This study divides the ambiguity of rumors into sources ambiguity and causal ambiguity and confirms that they have a different impact on the perceptions with different types of rumors, which provides a reference for the follow-up research variables of online health rumors.

Second, based on the S-O-R framework, this study considers the mediating role of perceptions on the intention to verify rumors, and confirms that people's willingness to verify information from different psychological sources. This suggests a potential mechanism that might explain the inconsistent findings on the relationship between rumor natures and people's responses in prior studies.

Third, about the interactive role of social influence in shaping user verifying, this study confirms that social consistency has a direct effect on the verifying intention of health rumors, and also have a moderating effect of the perceived importance and the perceived trust on the intention to verify. Therefore, it reveals the important role of social effects in the spread of health rumors and enriches the research in the field of information processing and

health rumor spread.

5.2 Practical implications

First of all, this paper confirms that users are more likely to verify rumors which contain ambiguities. Therefore, online health community managers should pay attention to the rumors including sources and content, establish a "gatekeeper mechanism", and verify the authenticity of health information sources in time, so as to prevent users from passing off authority or encouraging the masses to lead to the spread of rumors. Medical professionals are urged to actively participate in correcting the wrong health information on the Internet. Ordinary users should also seek multiple sources to verify the correctness of the information, and should not easily believe the health information on the Internet.

Secondly, the results show that fear rumor is easier to influence the people's perception than hope rumor, and users are more willing to spread fear rumor from quantitative sources and hope rumor regardless of source. Therefore, managers should pay more attention to the negative and anxious fear rumors, especially the fear rumors with large reading volume, praise volume and forwarding volume, and verify their authenticity in time, so as to avoid the fear rumors causing panic and anxiety of information receivers. In addition, it is necessary to verify the authenticity of the source of hope rumors with expert authority certification or widely spread, so as to avoid someone pretending to be a professional to confuse the masses.

Thirdly, this study proves that individual perceptions have a significant impact on rumor verifying. Users with low knowledge beliefs are more likely to spread health rumors, especially those with quantitative sources; users with high knowledge beliefs are more likely to spread health rumors with qualitative sources. Therefore, for managers, they should strive to promote the dialogue between authoritative people and laymen on the Internet, and make use of the professional knowledge of medical authoritative personnel to help ordinary users improve their health knowledge level and cultivate strong cognitive beliefs. For ordinary users, when receiving online health information, they should make logical judgments combined with their medical knowledge, and at the same time, they should check with the medical staff.

6. LIMITATION AND FURTHER RESEARCH

There are still some limitations in this study: first, most of the subjects come from universities and their surrounding communities, which may be limited by age and education level, and the future research objects should be selected from different geographical groups; second, this paper considers the emotional types of rumors, and there is other rumors classification, such as long rumors and short rumors, text rumors and picture rumors, etc We can consider the comparative study of multiple rumor classifications at the same time. Third, the spread of online health rumors will also be intervened by the relevant government departments and health departments. For example, the correct health knowledge publicity by the relevant departments through social media and critical education for malicious rumor makers may improve the spread of online health rumors. Future research can be further explored from the third-party intervention in addition to the source, object, and content of the spread ^[19].

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant No. 71702176 and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (Grant No. 1810491C07 and No. 1910491C08).

REFERENCES

- [1] Allcott H, Gentzkow M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2): 211-36.

- [2] Chua A Y, Banerjee S. (2017). To share or not to share: The role of epistemic belief in online health rumors. *International journal of medical informatics*, 108: 36-41.
- [3] DiFonzo N, Bordia P. (2007). Rumor, gossip and urban legends. *Diogenes*, 54(1): 19-35.
- [4] Bode L, Vraga E K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4): 619-638.
- [5] Vraga E K, Bode L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, 39(5): 621-645.
- [6] Bordia P, DiFonzo N. (2004). Problem solving in social interactions on the Internet: Rumor as social cognition. *Social Psychology Quarterly*, 67(1): 33-49.
- [7] Silverman C. (2015). Lies, damn lies, and viral content: How news websites spread (and Debunk) online rumors, unverified claims and misinformation. *Tow Center for Digital Journalism*, 168(4): 134-140.
- [8] Starcevic V, Berle D. (2013). Cyberchondria: towards a better understanding of excessive health-related Internet use. *Expert Review of Neurotherapeutics*, 13(2): 205-213.
- [9] Xiao Y, Cauberghe V, Hudders L. (2018). Humour as a double-edged sword in response to crises versus rumours: The effectiveness of humorously framed crisis response messages on social media. *Journal of Contingencies and Crisis Management*, 26(2): 247-260.
- [10] Chua A Y, Banerjee S. (2018). Intentions to trust and share online health rumors: An experiment with medical professionals. *Computers in Human Behavior*, 87: 1-9.
- [11] Sun H. (2013). A longitudinal study of herd behavior in the adoption and continued use of technology. *Mis Quarterly*: 1013-1041.
- [12] Mehrabian A, Russell J A. *An approach to environmental psychology*. the MIT Press, 1974.
- [13] Fang J, Chen L, Wen C, et al. (2018). Co-viewing experience in video websites: The effect of social presence on e-Loyalty. *International Journal of Electronic Commerce*, 22(3): 446-476.
- [14] Oh H J, Lee H. (2019). When do people verify and share health rumors on social media? The effects of message importance, health anxiety, and health literacy. *Journal of health communication*, 24(11): 837-847.
- [15] Dong X B, Li H B. (2018). Does online media sequence matter in product marketing? *Electronic Commerce Research and Applications*, 28: 44-53.
- [16] Rosnow R L, Esposito J L, Gibney L. (1988). Factors influencing rumor spreading: Replication and extension. *Language & Communication*.
- [17] Liwei X, Jiangnan Q, Wenjing G, et al. (2020). The Dynamic Effects of Perceptions of Dread Risk and Unknown Risk on SNS Sharing Behavior During EID Events: Do Crisis Stages Matter? *Journal of the Association for Information Systems*, 21(3): 545-573.
- [18] Kamins M A, Folkes V S, Perner L. (1997). Consumer responses to rumors: Good news, bad news. *Journal of consumer psychology*, 6(2): 165-187.
- [19] Pal A, Chua A Y K, Goh D H L. (2019). Debunking rumors on social media: The use of denials. *Computers in Human Behavior*, 96: 110-122.

Task Decomposition for Knowledge-intensive Crowdsourcing: Managing Dependency and Structural Complexity

Xiaojie SHI^{1*}

¹School of Information Management, Wuhan University, Wuhan, 430072, China

Abstract: Knowledge-intensive crowdsourcing (KIC) is expected to provide flexible access to knowledge and expertise. In this research, we proposed a task decomposition method to support the design decisions for KIC tasks decomposition and investigated how the level of granularity affects the crowdsourcing performance. To address the structural complexity, we employed the business process lens and coordination theory to describe the components of KIC tasks and their structural relationships. Afterwards, seven arithmetic tasks with different levels of structural complexity were designed and decomposed into subtasks with different levels of granularity by the proposed decomposition method. A laboratory experiment including 1960 groups of tests to simulate a real crowdsourcing environment was conducted to explore the relationship between different levels of granularity and the crowdsourcing performance. The results suggest that moderate decomposition helps to reduce the completion time and improve the quality of outcomes. A critical point of the level of granularity at which the completion time achieves the minimum is identified.

Keywords: knowledge-intensive crowdsourcing, business process lens, task decomposition, coordination theory

1. INTRODUCTION

Knowledge-intensive crowdsourcing refers to accomplishing knowledge-intensive tasks through the collaborative creation of knowledge content and requires special skills obtained by the means of crowdsourcing. In the dynamic and globalized environment, an increasing number of companies are attempting to make use of KIC to access knowledge and expertise. Despite its great potential, KIC often encounters a high risk of failure such as timeouts and poor quality of outcomes^[1,2]. Task decomposition is one of the classical approaches to resisting project failure risks, and it has been explored and verified in many domains, such as service systems, cloud manufacturing and product development^[3].

However, research on the KIC task decomposition is still in its infancy by far. Existing research on the management of KIC projects often focuses on task assignments or collaboration between crowd workers, e.g. . Task decomposition is often regarded as a design decision in the preparation of a crowdsourcing project with limited attention and is conducted by task providers or crowd workers based on their experience in practice, leaving the method of task decomposition in ambiguity. A key issue in task decomposition is identifying the proper level of granularity, which describes the grain size of the subtasks^[4]. To the best of our knowledge, an in-depth investigation of KIC tasks decomposition to address the level of granularity and its impact on crowdsourcing performance has not been reported in KIC literature. A generic task decomposition method that could address the complex interdependency between subtasks and achieve maximum benefits from the decomposition is desired. To address this knowledge gap, this study proposes a KIC task decomposition method and aims to reveal how level of granularity affects the performance of crowdsourcing in terms of the completion time and quality.

To manage the structural complexity of KIC tasks, a business process lens and the coordination theory have been employed. Through a business process perspective, the decomposition of KIC tasks is an undertaking in which interdependent subtasks are constructed with a logical sequence. It allows the transformation of a complex

* Corresponding author. Email: sprinashi@whu.edu.cn

task decomposition problem into a business process coordination problem that aims to manage the dependencies between subtasks. The dependencies of the subtasks are clarified by the coordination theory in this study. The proposed task decomposition method was demonstrated through a laboratory experiment that consisted of 1960 groups of tests. The results indicate that a moderate decomposition of KIC tasks results in less completion time and higher-quality crowdsourcing. With regard to the level of granularity, a critical point at which the completion time of the crowdsourcing reaches the minimum was discovered from the experiments. The results of our research give decomposition guidance, which will help in identifying an appropriate level of granularity for supporting the design decision for KIC task decomposition and improving the crowdsourcing performance.

The rest of this paper is structured as follows. In Section 2, we introduce the business process lens and the coordination theory which serve as the theory background of our study. In Section 3, seven types of tasks are designed according to the coordination theory. Based on this, a task decomposition method is proposed to decompose these tasks into suitably sized subtasks. Experiments exploring the relationship between the level of granularity and the performance of crowdsourcing are presented in Section 4. Finally, in Section 5 we discuss the results of our study to provide both the theoretical and practical implications. Finally, we conclude the paper by proposing the limitation and the future research direction in Section 6.

2. BACKGROUND AND RELATED WORKS

2.1 Business process lens and coordination theory

Decomposing a KIC task into a set of subtasks has to address the interdependencies among them. Dependencies arise when actions taken by one referent affect the actions or outcomes of another referent [5]. To analyze and manage these dependencies, a business process lens has been adopted in this study. A business process, according to van der Aalst and Hee [6], can be defined as a combination of individual activities and a workflow describing their logical order. The business process lens allows investigating the independent crowdsourcing elements and linking them into an integrated process.

The business process lens views the subtasks and their logical relationships within a knowledge-intensive task as a set of independent activities to be coordinated. Furthermore, the business process lens provides an organic view through which to understand the elements of a KIC task and their structures. Analyzing a KIC task from the business process lens provides insight into how its structure impacts its execution. In this way, comparing different business process designs enables the comparison and analysis of different decomposition solutions.

To understand and manage the dependencies within knowledge-intensive tasks, this research employed the coordination theory. The coordination theory, proposed by Malone and Crowston [7], involves how activities can be coordinated to work together harmoniously. According to the coordination theory, there are three kinds of basic dependence relationships between activities that bring the coordination problems (Figure 1): flow, share and fit. The relationship of flow means that the output of one activity is the exclusive input of another activity. Share means that two different activities require the same limited resources, such as human resources and money. The fit relationship indicates that the former two activities jointly form the latter activity. These three basic relationships and their combinations constitute dependencies in knowledge-intensive activities.



Figure 1. Three basic dependencies in processes [7]

The above three basic relationships and their combinations can enable the design of different processes that represent a variety of knowledge-intensive tasks. If these three kinds of dependencies and their combinations are

well-managed within the task, the task decomposition can be addressed. Therefore, investigating and analyzing common dependencies and their related coordination mechanisms facilitate the development of a generic decomposition method for different KIC tasks.

2.2 Task decomposition

While there is limited research on the decomposition of KIC tasks, task decomposition has been studied in many other domains, such as robot systems, cloud manufacturing and product development. Relative studies of task decomposition contain decomposition approaches from different aspects. As a widely adopted methodology, task modular decomposition^[8] divides a task into multiple subtask modules with certain associations among those subtasks. It attaches great importance to the interdependencies among subtasks. Quantitative criteria such as the degree of coupling^[3,9] and the degree of equilibrium^[10] are often used in task modular decomposition to evaluate the decomposition outcomes.

The degree of coupling measures the amount of information interactions between subtasks [11]. The greater the amount of information interaction between subtasks, the higher is the degree of coupling. Frequent information interactions between subtasks demand collaboration, which will increase the difficulties of crowdsourcing these subtasks to different crowd workers. The greater the amount of the information interactions between the subtasks, the more dependencies there are between them. It is vital to enable a suitably low coupling degree between two subtasks.

The dependency structure matrix (DSM) is a common method used to depict the correlation between two subtasks^[12]. It divides a task into n subtasks and distributes it into an $n \times n$ matrix. As shown in Eq. (1)

, the design structure matrix is an n -th-order square matrix that stands for all of the subtasks, and n represents the number of subtasks. a_{ij} in the DSM represents the information interaction between the subtasks in the row and the corresponding subtasks in the column. The value of a_{ij} is shown in Eq.(). If there is a correlation between Subtask i and Subtask j , then the combination between them is marked as 1 in the matrix. Otherwise, the value of their combination is 0. By using the DSM, the correlation between two subtasks can be clearly reflected.

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \quad (1)$$

$$a_{ij} = \begin{cases} \text{cond}(t_i, t_j) & , i \neq j \\ 0 & , i = j \end{cases} \quad (2)$$

$$R = \begin{cases} \sum_{s, t \in T} \frac{\text{cond}(t_1, t_2)}{|T| \times |T| - 1}, & |T| > 1 \\ 0, & |T| \leq 1 \end{cases} \quad (3)$$

The formula used to calculate the coupling coefficient is given in Eq.(3)^[13]. In this formula, R represents the coupling coefficient. t_i and t_j are two independent subtasks, $\sum_{t_1, t_2 \in T} \text{cond}(s, t)$ stands for the amount of information interaction between the subtasks, and $|T|$ presents the number of subtasks.

The degree of equilibrium reflects a subtask's uniformity in size^[10]. An appropriate equilibrium degree of subtasks helps balance the completion time of the subtasks and consequently improve the overall completion time of the crowdsourcing project^[14]. In the context of crowdsourcing, the standard deviation of a subtask execution time is used to quantify its degree of equilibrium^[10]. The equilibrium degree of a subtask can be evaluated by Eq.(4). D represents the value of the standard deviation of the task execution time, and n is the number of subtasks. \bar{T} is the average of the task execution time.

$$D = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (t_i - \bar{t})^2} \quad (4)$$

2.3 Related works addressing the task decomposition of KIC

Knowledge-intensive tasks are often large in scale and high in structural complexity and cannot be accomplished by an individual worker. The key to performing knowledge-intensive tasks is to decompose them into several subtasks and coordinate among these subtasks^[15]. In related studies, there are two main research

perspectives: technical and business perspectives. The technical perspective mainly focuses on how to supply crowdsourcing using technologies and emphasizes technical methods, techniques, and frameworks for solving problems in crowdsourcing . The business perspective focuses on how to complete certain business goals efficiently and effectively^[16], involving crowdsourcing outcomes in terms of time, quality or cost.

Though existing studies have promoted the development of KIC, neither technical nor the business perspective studies well address the following two issues. First, current studies on decomposition limit their vision to specific types of tasks, such as article writing^[17] and proofreading^[18], leading to a low generalizability of the conclusions. Generic methods that can guide decompositions of different types of KIC tasks are still lacking. Second, how to achieve maximum benefits from the decomposition of KIC tasks with a proper level of granularity is still unclear. Many KIC studies do not focus on the decomposition of tasks, leaving this undertaking up to the task providers or crowd workers. Inappropriate levels of granularity will have an adverse effect on the crowdsourcing performance and bring problems such as low completion quality, high cost in terms of time and money or even failure of the crowdsourcing project^[19]. We are therefore motivated to investigate a generic task decomposition method with predictable effects on the crowdsourcing performance.

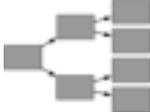
3. THE DECOMPOSITION METHOD FOR KIC TASKS

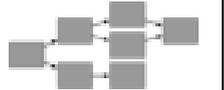
3.1 Task structural complexity

To study the structural complexity of KIC tasks, in this section, seven types of business processes with different levels of structural complexity are designed. As the business process lens helps transfer a KIC decomposition problem into a business process problem, the design of KIC tasks is actually the design of business processes. As defined in Section 2.2, flow, fit and share are the three basic dependencies. Table 1 demonstrates how single or multiple types of dependencies constitute tasks with different levels of structural complexity. In the figures listed in the second column, the rectangle represents a basic activity that is a minimum-sized building block constituting a KIC task, and the arrow represents the sequence relationships between the activities. Based on the three types of dependencies, seven knowledge-intensive crowdsourcing tasks with different levels of structural complexity are obtained.

Type 1 is made up of the flow relationships, which indicate that there are only flow relationships inside the

Table 1. Task structural complexity demonstrations

Type	Demonstration	Level of structural complexity	Constitution
1		low	flow
2		low	fit
3		low	share
4		medium	flow, fit
5		medium	flow, share
6		medium	share, fit

7		high	flow, fit, share
---	---	------	------------------

business process. The category is regarded as the business processes with the least structural complexity. Type 2 includes business processes that only possess fit relationships. Similarly, Task 3 contains only share relationships. Types 1, 2 and 3 of business processes represent KIC tasks with a low level of structural complexity. Combining any two types of dependencies could produce tasks with a higher level of structural complexity. We therefore have the flow and fit relationships with which to form Type 4, the flow and share relationships to form Type 5, and the share and fit relationships to form Type 6. These 3 types of business processes are regarded as tasks with a medium level of structural complexity. Type 7 contains all of the three basic relationships—fit, share, and flow—and represents KIC tasks with a high level of structural complexity.

3.2 The task decomposition method

In previous studies, the decomposition is dominated by crowd workers or task providers based on their own experience^[1,15]. Decomposition solutions differ when deduced by different people. In addition, if the sizes of the subtasks are not balanced, this will lead to a large difference between the completion times of the subtasks, which will affect the overall completion progress of the whole task^[10]. In consideration of the above two conditions, a task decomposition method that helps control the decomposition process by utilizing the concepts of the task coupling degree and task equilibrium degree is proposed in this section. The method can be divided into 3 steps, and it is introduced as follows.

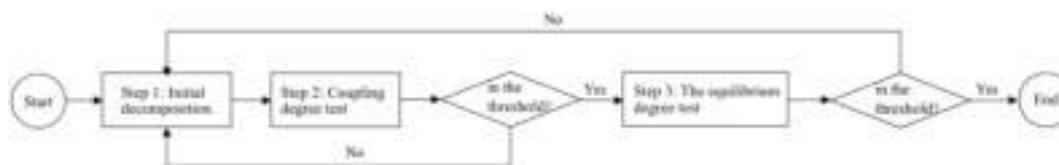


Figure 2. The steps in task decomposition

Step 1: Initial decomposition. In this step, a preliminary decomposition based on the structure of the KIC task is conducted. Tasks with a low level of structural complexity that consist of a single kind of relationship, such as Types 1, 2, and 3, can be evenly divided. For task types with higher structural complexity, we try to divide a task in as balanced a way as possible during the decomposition. This means that the number of flow, fit, and share relationships in each subtask should be approximately the same. At the end of this step, the temporary subtask Set T is obtained.

Step 2: Coupling degree test. This step aims to ensure the independency of subtasks after the initial decomposition. The coupling degree is controlled within a reasonable range, which might vary for different types of tasks and decomposition goals. To maintain the proper coupling degree of subtasks, it is necessary to set a reasonable decomposition threshold in advance. The appropriate decomposition threshold can be determined according to the actual decomposition experience or repeated experiments. In this step, if the degree of coupling of the task meets the requirements of the threshold, the subtasks in Set T are independent of each other, and then the decomposition can be continued with Step 3. If not, there are too many information interactions between the subtasks. In this case, a return to step 1 is needed. The re-decomposition requires merging subtasks with a higher degree of coupling.

Step 3: The equilibrium degree test. This step uses the completion time of a subtask to measure the scale of a subtask. A balanced completion time means that the subtasks are of balanced sizes to avoid the situation in which a few subtasks take much more time than do the others, resulting in an unnecessary long project duration. The completion time of each subtask is evaluated by experts in this step. Similar to the case for Step 2, a reasonable

threshold of the equilibrium degree is given in advance. If the equilibrium degree of the subtasks is within the threshold, that the subtasks are balanced and the decomposition is effective. Then, an ultimate decomposition result Set U is obtained. If the equilibrium degree of the subtasks exceeds the threshold, the scale of the subtasks is unbalanced, and need to go back to Step 1 for re-decomposition. The subtasks that have an overly long completion time need to be split, and subtasks with a completion time that is too short should be merged.

4. EXPERIMENT

In this chapter, a knowledge-intensive task is designed and decomposed into a number of subtasks according to the decomposition method proposed in Section 3.2. After the decomposition, a crowdsourcing experiment was conducted to explore how the level of granularity of the tasks will affect the performance of the crowdsourcing. In this study, conducting a laboratory experiment is chosen rather than conducting experiments in a real-world crowdsourcing environment. This is because there are many factors that could impact the performance of crowdsourcing in a real crowdsourcing environment. For example, crowd workers may be interrupted by personal affairs. In such a case, the performance of the crowdsourcing would be greatly affected and could not well reflect the effect of the task decomposition. What's more, since the background and the ability of the crowd workers various, the participants of the laboratory experiment are randomly selected. Therefore, this study uses a laboratory experiment to explore the effects of the task decomposition to avoid influences from other factors.

The experiments were conducted in February 2019. The participants of the experiments were undergraduate students of our university in various majors. Recruiting college students to conduct lab experiments for crowdsourcing or process design research is a strategy has been adopted by previous studies, and its validity has been verified [20].

4.1 Design of the knowledge-intensive tasks

The performance of crowdsourcing can be measured in many ways, such as according to time, price, and quality. In the experiment, we judge the performance of crowdsourcing by two parameters: the completion time of the task and the quality of the results. Existing KIC studies often employ experiments to investigate KIC tasks. The experiments simulate the design and execution of KIC tasks including arithmetic tasks, article editing, article writing, software developing, etc. This research chose arithmetic tasks as the knowledge-intensive tasks for the following two reasons. The first is that the quality of the outcomes can be easily judged by the accuracy of the calculation results. The other reason is that arithmetic tasks are relatively short in terms of completion time and are low in cost, therefore allowing the feasibility of conducting a large number of crowdsourcing tasks in the experiment.

Seven different kinds of KIC tasks of arithmetic were designed. For example, Task 7 (Table 2) is an arithmetic task that consists of 11 activities. There are several dependencies between these subtasks. The result of subtask f relies on the calculation output of subtask c. The calculation processes of subtask b and subtasks c depend on the

Table 2. The task of Type 7

activity	Please calculate the value of k.
activity 1	$a=12+4-6$
activity 2	$b=a+9+7$
activity 3	$c=a-15+7$
activity 4	$d=b-5+1$
activity 5	$e=b-5+8$
activity 6	$f=c+1+4$
activity 7	$g=d-e+2$
activity 8	$h=f-3+5$

activity 9	$i=f+1+3$
activity 10	$j=h+i-6$
activity 11	$k=g+j-3$

result of subtask a. Subtask g depends on the results of d and e. According to the concept of the relationships of fit, share and flow, the relationship between subtask f and subtask c is a flow relationship. Subtask a, subtask b and subtask c are in a share relationship. Subtasks d, e and g are in a fit relationship. Likewise, the arithmetic task can be translated into a business process problem, as shown in Figure 3. In a similar way, the other tasks can be translated into business processes.

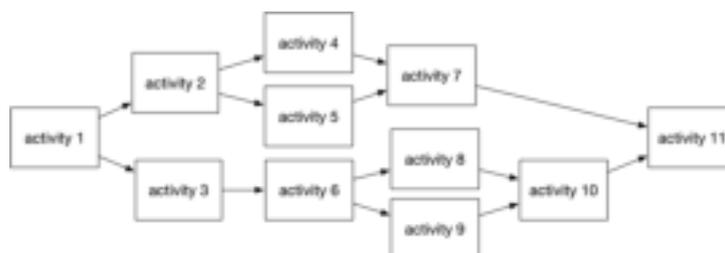


Figure 3. The task of Type 7 as expressed through the business process lens

4.2 Decomposition of the tasks

The decomposition method presented in Section 3.2 is applied in this section to decompose the knowledge-intensive tasks into subtasks to below a certain level of granularity. Decomposing the Type 7 tasks into 5 subtasks serves as an example demonstrating the decomposition process.

In Step 1, a preliminary decomposition of the task according to the structure is conducted. Overall, the flow, share, and fit relationships in each subtask are evenly distributed. The decomposition results, as well as the dependency types and activities, are presented in Table 3. They meet the requirements of Step 1, and the decomposition continues with Step 2.

Table 3. The task decomposition for Type 7

Subtasks	Demonstration	Dependency types
Subtask 1		share, flow
Subtask 2		share, fit
Subtask 3		fit
Subtask 4		share
Subtask 5		fit

In the second step, we need to judge and analyze the coupling degree of the decomposition results from Step 1. Based on experience, we set the task coupling degree threshold to 0.6. According to Eq.(2) and Eq.(3), we can get the task information association matrix and the degree of coupling of the task is within the defined threshold. The decomposition proceeds to the next step. In Step 3, the job is to analyze and control the degree of equilibrium of the subtasks. Based on experience, we set the threshold value to 4 s. We evaluate the time required for each subtask based on previous crowdsourcing experiments. According to Eq.(4), the calculated equilibrium degree of the subtasks is 2.76 s, which is less than the threshold of 4 s and therefore meets the conditions. At the end of the process, the task decomposition solution is valid.

4.3 Data collection and analysis

There are 7 kinds of knowledge-intensive tasks, which are of different levels of structural complexity. They are decomposed into 7 seven different levels of granularity, respectively. Thus, 196 different kinds of subtasks are obtained. Each kind of subtask requires a participant to complete it independently in our laboratory room under monitoring. To eliminate or reduce the impact of different calculation speeds of individuals on the experiment, the test for each subtask was performed repeatedly by different participants 10 times. Furthermore, to avoid the impact of different academic background on the calculation speed, we averagely assigned participants in different majors to each repeating group. At the end, in total, 1960 sets of results were collected in the experiment. As the performance of crowdsourcing is reflected by the completion time and accuracy, the results and completion times of the subtasks are recorded in the experiment. The experimental platform is powered by www.wjx.cn, which can assign the tasks randomly to the participants and record the execution times of the subtasks.

By adding the completion times of each group of subtasks, the overall completion time of the seven complex types of crowdsourcing tasks with different levels of granularity is obtained. The results are presented in Figure 4. We find that at the same level of granularity, the more complex is the knowledge-intensive task, the longer is the time it takes during crowdsourcing. By breaking down these knowledge-intensive tasks, the crowdsourcing time can be reduced. To further analyze the changes of the completion time with different levels of granularity, we calculate the rate of time change by $\Delta t/t$. We assume that when the value of $|\Delta t/t|$ is below 0.06, the time change can be neglected.

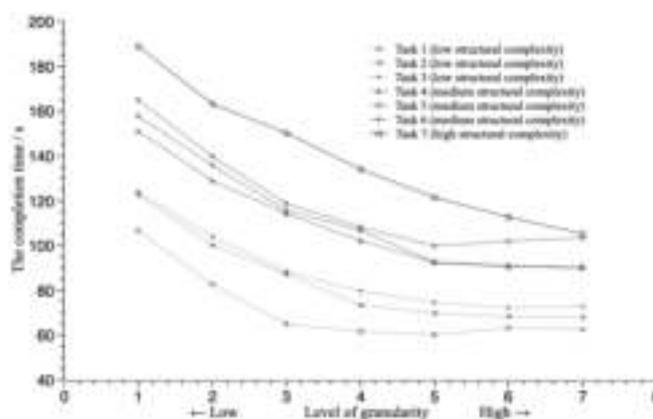


Figure 4. The completion times of the tasks with different levels of granularity for different types of tasks

For Type 1, the values of $|(t_3 - t_4)/t_3|$, $|(t_4 - t_5)/t_4|$, $|(t_5 - t_6)/t_5|$ and $|(t_6 - t_7)/t_6|$ are below 0.06. This indicates that when the level of granularity is less than 3, the crowdsourcing time decreases as the decomposition increases. However, when the task is decomposed into more than 3 subtasks, the overall completion time no longer reduces as the decomposition proceeds further. In contrast, the completion time stays the same and even increases. Therefore, we define the level of granularity of 3 as a critical point for Type 1, at which point the overall completion time reaches its minimum. For tasks with low levels of structural complexity, the critical point occurs when the level of granularity is 4. For tasks with medium levels of structural complexity, such a point occurs when

the level of granularity is 5. For Type 7, the structural complexity of which is the highest, such a point still does not appear when the task is decomposed to a level of granularity of 7. Therefore, where the critical point appears is related to the level of structural complexity of the tasks. The more complex is the task, the higher is the level of granularity at which the critical point appears. By calculating the accuracy of each task, we observe the relationship between the level of granularity and the accuracy of tasks with different levels of structural complexity. Figure 5 presents the accuracies of the calculation results at different levels of granularity for the 7 types of tasks.

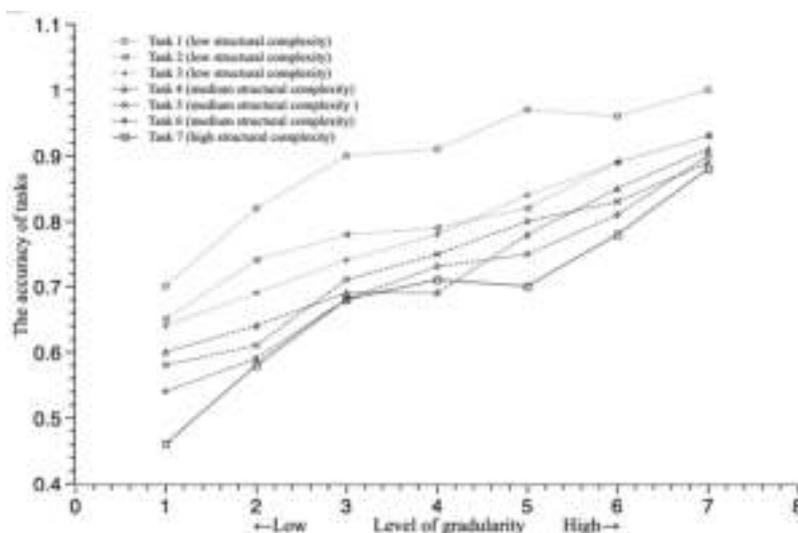


Figure 5. The accuracy of the results at different levels of granularity for each type of task

Based on the results, we found that the accuracy of each task stays the lowest when the task is not decomposed. The most complex task 7 has an accuracy of only 0.45, while task 1, which has a low level of structural complexity, has an accuracy of 0.7. For the same KIC task, as the level of level of granularity increases, the accuracy increases gradually. To compare the accuracies of tasks with different levels of structural complexity, we averaged the accuracies of tasks 1-3 and 4-6, respectively, and obtained the average accuracy for different levels of structural complexity. By comparing the accuracies of the tasks with these three different levels of structural complexity, we find that at the same level of granularity, the less complex is the KIC task, the higher-quality will be the crowdsourcing outcome.

For each decomposition solution, 10 repeated experiments were performed to ensure the credibility of the data. To examine the degree of dispersion of the completion times of each of these 10 groups of tasks, the standard deviation of the completion times for different types of tasks at different levels of granularity are calculated. The findings show that for each type of task, as the level of granularity increases, the standard deviation of the completion time decreases, indicating that a high level of granularity degree helps improve the predictability of the completion time of the tasks. The structural complexity of knowledge-intensive tasks brings instability to the completion time of the crowdsourcing. The results also indicate that as the level of granularity increases, the deviation often reduces gradually. This means that higher level of granularity results in less uncertainty in the completion time.

5. DISCUSSION

The theoretical model (Figure 6) summarizes the key findings of the experiment. This study investigates how the level of granularity effects KIC performance by the two dimensions of quality and time, where the level of granularity is an independent variable and KIC performance is a dependent variable. Structural complexity exerts an influence on the relationship between the level of granularity and KIC performance, and thus serves as the moderate variable in the research. In order to avoid the influence of other unrelated variables, we control the

external disturbances that may impact the results of the experiment, such as the interruption from personal affairs and their different calculation speeds.

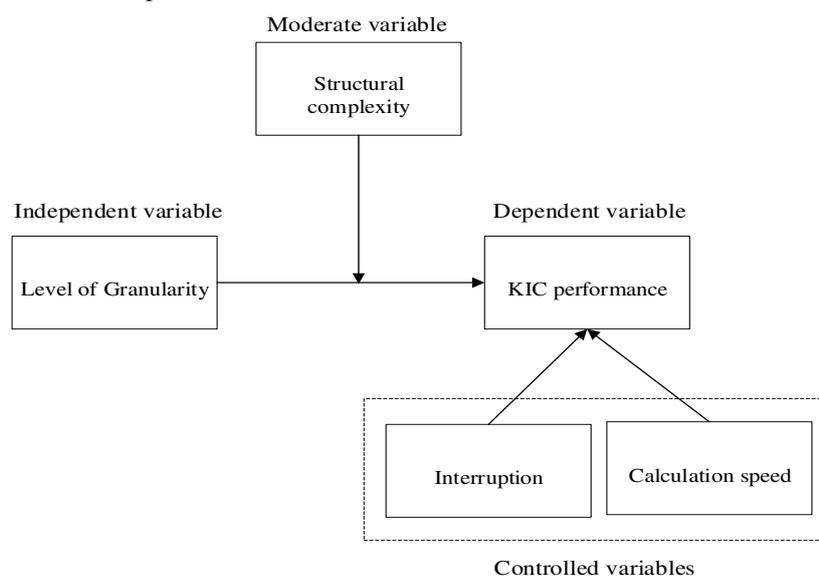


Figure 6. The theoretical model

The research discussed in this paper can have potential impacts on KIC research. Previous research regarded task decomposition as a given preparation stage of KIC projects^[15]. In contrast, our study investigates and explains the relationships between the level of granularity and the completion time and quality of KIC tasks. The results in this study allow KIC researchers to include and control task decomposition in their research design.

In addition, this study indicates a new approach to the research of KIC task decomposition by applying business process methods and the coordination theory. We use the three basic dependence relationships to model the relationships between subtasks. Thus, an abstract decomposition problem is transferred into a concrete business process modeling problem. Business process methods are applied as the coordination mechanism to enable task decomposition. According to the coordination theory, dependencies and the mechanisms used for managing them are generic. Although further investigation of the different aspects influencing crowdsourcing performance is required to develop a more comprehensive method, this approach is promising in terms of its generalizability. As the business process can serve as a template of for creating multiple, real-life instances of the same process, it enables the development of generic process patterns to decompose KIC tasks, removing the different types and features of the tasks.

The results of our research also provide useful practical insights for use in the task decomposition of KIC. The results demonstrate the benefits of task decomposition in crowdsourcing. The structural complexity of the tasks influences the relationship between the level of granularity and the crowdsourcing performance. Thus, KIC tasks with more complex structures should be decomposed into subtasks with a smaller size. An important finding of this research is the existence of a critical point in the relationship between the level of granularity and the completion time. As the task is reaches a higher level of granularity, the completion time of the crowdsourcing continually declines. However, once the critical point has been reached, the completion time stays the same or even inclines when the level of granularity becomes higher. Thus, to complete KIC tasks within a minimum amount of time, it is vital to identify the critical point. Furthermore, the decrease of the completion time becomes slow when the level of granularity nears the critical point. Employing a low level of granularity could result in the reduction of the completion time of task decomposition. The decision of whether to continue with a higher level of granularity to reach the critical point is a tradeoff between the cost of task decomposition and the savings from further reducing the completion time of the crowdsourcing task. In addition, increasing the level of granularity

could make the completion time of a task more predictable. These findings allow for reasonable decision-making with regard to KIC task decomposition.

6. CONCLUSIONS

To fill the knowledge gap in managing KIC task decomposition, we start from the dependency relationships of KIC subtasks and investigate them through a business process lens. In the research, seven types of KIC tasks with different levels of structural complexity were designed according to the coordination theory. A decomposition based on the structure of the tasks was conducted to obtain 196 groups of subtasks that represent different level of granularity. To explore how the level of granularity will affect the performance of a KIC, a series of laboratory experiments was conducted. The results indicate that the performance of the KIC is closely related to the level of granularity. Task decomposition helps reduce the completion time of the crowdsourcing and improve the predictability of the completion time, which further helps reduce the timeout risk. The quality of the crowdsourcing also improves as a result of task decomposition. A critical point exists at which the level of granularity enables the minimum completion time of the crowdsourcing. The proposed decomposition method and the findings of the experiment provide guidelines for supporting the design decision for effective KIC task decomposition which could improve the crowdsourcing performance at the end.

In this study, there are several limitations that should be viewed as the starting points for further research. First, our analysis is based on experiments using arithmetic tasks. More kinds of knowledge-intensive tasks should be explored to further test the generalizability of the conclusion. Second, to constrain the scale of the KIC tasks in the experiment, tasks with the different structural complexities consist of the same number of subtasks. Whether and how the scale of a KIC task influences the crowdsourcing performance remains to be investigated. Third, a critical point at which the task completion time will reach its minimum has been identified in the research. Knowing the position of this point can enable more precise decision-making in preparing crowdsourcing projects. However, currently, this point can only be obtained through a large number of crowdsourcing experiments and is time- and labor-consuming. Whether there is a method that can be used conveniently to locate or estimate this point is worth exploring.

REFERENCES

- [1] Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, 16(3), 417-434.
- [2] Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., & Allahbakhsh, M. (2018). Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *Acm Computing Surveys*, 51(1), 7.
- [3] Zhang, X., Yang, Y., & Bao, B. (2016). Task decomposition and grouping for customer collaboration in product development. *Journal of Intelligent Systems*, 25(3), 361-375.
- [4] Chiriac, N., Hölttä-Otto, K., Lysy, D., & Suh, E. S. (2011). Level of modularity and different levels of system granularity. *Journal of Mechanical Design*, 133(10), 101-117.
- [5] Mccann, J. E., & Ferry, D. L. (1979). An Approach for Assessing and Managing Inter-Unit Interdependence. *Academy of Management Review*, 4(1), 113-119.
- [6] Van Der Aalst, W., & Van Hee, K. (2004). *Workflow management: models, methods and systems*. Boston: MIT Press.
- [7] Malone, T. W., & Crowston, K. (1994). The Interdisciplinary Study of Coordination. *Acm Computing Surveys*, 26(1), 87-119.
- [8] Habib, M., & Paul, C. (2010). A survey of the algorithmic aspects of modular decomposition. *Computer Science Review*, 4(1), 41-59.
- [9] Liu, W., Ji, J., Yang, Y., & Zhang, L. (2018). Capability-Based Design Task Decomposition in Heavy Military Vehicle Collaborative Development Process. *The International Academy for Production Engineering*, 70(1), 13-18.

- [10] Bao, B., Yang, Y., Li, F., & Xue, C. (2014). Decomposition model in product customization collaborative development task. *Computer Intergrated Manufacturing Systems*, 20(7), 1537-1545.
- [11] Smith, R. P., & Eppinger, S. D. (1997). A predictive model of sequential iteration in engineering design. *Management Science*, 43(8), 1104-1120.
- [12] Maheswari, J. U., & Varghese, K. (2005). Project scheduling using dependency structure matrix. *International Journal of Project Management*, 23(3), 223-230.
- [13] Pang, H., & Fang, Z. (2008). Task decomposition strategy and granularity design in networked collaborative environment. *Computer Intergrated Manufacturing Systems*, 14(3), 425.
- [14] Mizusawa, K., Tajima, K., Matsubara, M., Amagasa, T., & Morishima, A. (2018). Efficient pipeline processing of crowdsourcing workflows. Paper presented at the Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, October 22 - 26, 2018
- [15] Jiang, J., An, B., Jiang, Y., Lin, D., Bu, Z., Cao, J., et al. (2018). Understanding crowdsourcing systems from a multiagent perspective and approach. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 13(2), 8.
- [16] Luz, N., Silva, N., & Novais, P. (2015). A survey of task-oriented crowdsourcing. *Artificial Intelligence Review*, 44(2), 187-213.
- [17] Kittur, A. (2011). CrowdForge: crowdsourcing complex work. Paper presented at the Chi 11 Extended Abstracts on Human Factors in Computing Systems, Santa Barbara, USA, October 16 - 19, 2011
- [18] Jiang, H., & Matsubara, S. (2014). Efficient task decomposition in crowdsourcing. Paper presented at the International Conference on Principles and Practice of Multi-Agent Systems, Gold Coast, Queensland, Australia, December 1-5, 2014
- [19] Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., & Dustdar, S. (2013). Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, 17(2), 76-81.
- [20] Thuan, N. H., Antunes, P., & Johnstone, D. (2018). A Decision Tool for Business Process Crowdsourcing: Ontology, Design, and Evaluation. *Group Decision and Negotiation*, 27(2), 285-312.

Full Research Paper

E-commerce Platform Resources, Sequences of Digital Strategic Actions and Competitive Advantage: An Empirical Study from Online Tourism Industry

Lewei Hu^{1,2*}, Jing Zhao³, Yi Jiang³

¹School of Business, Jiangnan University

²Institute of Intelligent Decision-making, Jiangnan University

³Research Center for Digital Business Management, School of Economics and Management, China University of Geosciences

Abstract: This study examines the effects of digital strategic actions sequences on business value in long-term competitive dynamics. Based on the competitive dynamics theory and configurational perspective, we first use fuzzy-set qualitative comparative analysis to analyze the news data of Chinese online tourism industry, and find out the corresponding configurations of digital strategic actions. Then, we use panel regression to explore the relationship between the unpredictability and complexity of digital strategic actions sequences and competitive performance in the continuous competitive interaction, and find the inverted U-shaped relationship between them. Finally, we discuss implications for theory and practice.

Keywords: Business Value of IT, Competitive Dynamics, Sequences of Digital Strategic Actions, Complexity, Unpredictability.

1. INTRODUCTION

Facing with the current fierce competition, e-commerce platform needs to continuously initiate *digital strategic actions (DSAs)* to provide differentiated products and services for creating and accumulating competitive advantages. However, the application of digital technology has increased the intensity of confrontation among e-commerce platforms, which creates more uncertainty for platform managers^[1]. The similarity of the resources owned by e-commerce platforms causes competitors to imitate and respond quickly. Thus, it is difficult for a single competitive action to win the competitive advantage for e-commerce platform. Our motivation is to explore the value creation of e-commerce platform from a perspective of actions sequences in competitive dynamics^[2]. Specifically, this study focuses on the impact of dynamic characteristics (including unpredictability and complexity) of the DSAs sequences on the advantage of e-commerce platform^[3, 4].

2. METHODS

Based on the paradigm of empirical research, our study has extracted data from news as the basis of constructing variables, and then established econometric model, and verified hypotheses. Specifically, first of all, the configurations of the DSAs are obtained based on structure content analysis and fuzzy-set qualitative comparative analysis^[5]. Then, the dynamic characteristics of the DSAs sequences are analyzed by using Markov chain and Shannon index. Finally, the feasibly generalized least squares (FGLS) is used to explore the rules of sequences of the DSAs to create business value.

* Corresponding author. Email: pedagogical@foxmail.com, pedagogical@jhun.edu.cn

3. RESULTS

With the FGLS analysis, our research has testified econometric model. Table 1 summarizes the regression results of FGLS estimation method testing hypotheses. In the table, UEAS and CPAS represent the unpredictability and complexity of the DSAs sequence respectively. As a dependent variable, PVCR represents competitive performance, which is calculated on the basis of page views.

Table 1. Regression results

	Dependent variable: PVCR							
	Linear model		Quadratic model					
UEAS	0.011***	(0.001)	0.040***	(0.002)				
UEAS ²			-0.035***	(0.001)				
CPAS		0.016***	(0.001)	0.011***	(0.002)			
CPAS ²				-0.009***	(0.002)			
FA	-0.001	(0.001)	-0.000	(0.001)	0.000	(0.001)	-0.000	(0.001)
Intercept	0.031	(0.040)	0.012	(0.043)	0.019	(0.037)	0.046	(0.054)
Wald Chi2	101.19***		241.39***		945.24***		55.82***	

Notes: ***, **, and * denote statistical significance at 1 percent, 5 percent, and 10 percent, respectively. Standard deviation in brackets.

The empirical results support our two hypotheses, that are, the unpredictability and complexity of the DSAs sequences initiated by the e-commerce platform will exhibit an inverted U-shaped relationship with the competitive performance.

4. CONTRIBUTIONS

This research has several contributes to research on the business value of IT. Firstly, in view of the characteristics of the complex interaction of digital technologies and partner resources in e-commerce platform, this study explores the relationship between the dynamic combinations of two types of resources and different types of the DSAs based on the configurational approach, and clarifies that in competitive dynamics, the value of e-commerce platform comes from the dynamic combinations of advantage resources. Secondly, our study found the inverted u-shaped relationship between the dynamic characteristics of the DSAs sequences and the competitive performance of e-commerce platforms and revealed the non-linear change trend of business value caused by the continuous competitive dynamics among e-commerce platforms. Thirdly, Managers of e-commerce platforms should understand that when orchestrating and formulating DSAs sequences, they cannot blindly pursue their suddenness and complexity at the cost of exceeding their control and management capabilities.

REFERENCES

- [1] Mithas, S., A. Tafti, and W. Mitchell, How a Firm's Competitive Environment and Digital Strategic Posture Influence Digital Business Strategy. *Mis Quarterly*, 2013. 37(2): p. 511-536.
- [2] Ndofor, H.A., D.G. Sirmon, and X.M. He, Firm Resources, Competitive Actions and Performance: Investigating a Mediated Model with Evidence from the IN-VITRO Diagnostics Industry. *Strategic Management Journal*, 2011. 32(6): p. 640-657.
- [3] Connelly, B.L., et al., Competitive Repertoire Complexity: Governance Antecedents and Performance Outcomes. *Strategic Management Journal*, 2017. 38(5): p. 1151-1173.
- [4] Rindova, V., W.J. Ferrier, and R. Wiltbank, Value from Gestalt: How Sequences of Competitive Actions Create Advantage for Firms in Nascent Markets. *Strategic Management Journal*, 2010. 31(13): p. 1474-1497.
- [5] Campbell, J.T., D.G. Sirmon, and M. Schijven, Fuzzy Logic and the Market: A Configurational Approach to Investor Perceptions of Acquisition Announcements. *Academy of Management Journal*, 2016. 59(1): p. 163-187.

Full Research Paper**The Mechanism and strategy of Digital Logistics Park Cross Supply Chain****Collaboration based on Evolutionary Game***Zhou Xingjian*^{1,2*}, *Cai Lihua*³, *Xu Gechen*^{1†}, *Huang Yingpei*¹¹School of management, Wuhan Textile University, China²Post Doctoral Research Station of Management Science and Engineering, Nanchang University, China³School of environment and Bioengineering, Wuhan Technology and Business University, China

Abstract: In order to overcome the disadvantages of "go-it-alone" in logistics parks, it is necessary to select appropriate horizontal collaboration strategies to promote cooperation among logistics parks. Considering a kind of horizontal collaboration cross supply chain between logistics parks under digital Cloud platform, there is a three-party composed of one Cloud platform and two logistics parks. According to the benefit function of each party, a dynamic replication equation is constructed to analyze the stability strategy of the three-party Evolutionary Game, and then a Jacobian matrix is constructed to analyze the stability of the equilibrium point of the three-party Evolutionary Game of the resource sharing strategy between logistics parks. Moreover, two gradually stable strategies (0,0,0) and (1,1,1) are formed, and the strategy is affected by the initial sharing proportion of resources between logistics parks, the risk of customer order loss and the proportion of Cloud platform subsidy. The simulation analysis shows that the higher the initial sharing proportion of resources among logistics parks, the more likely the logistics parks are to choose the horizontal collaboration strategy, and the more likely the Cloud platform are to choose the subsidy strategy; the higher the subsidy proportion, the more willing the logistics parks are to choose the horizontal collaboration strategy. Combined with the actual operation of the logistics park, it is suggested to strengthen the guiding effect of the core logistics park, strengthen the supervision and guiding ability of the Cloud platform, emphasize the risk sharing and revenue sharing among the logistics parks, and jointly promote the collaborative operation and development among the logistics parks.

Keywords: digital logistics park, horizontal cooperation, cross supply chain, Evolutionary Game

1. INTRODUCTION

With the development of digital transformation, the logistics park will become the key leverage fulcrum and the primary breakthrough of the digital transformation and intelligent transformation of logistics industry. Especially with the support of Cloud computing, big data, Internet of Things and other information technologies, the industry began to explore the construction of digital logistics park alliance through Cloud platform. For example, "China Logistics Park map" created by CFLP, interconnects more than five hundred logistics parks to form a public welfare national digital logistics park platform. During the period of COVID-19, it coordinated with government department to make a central dispatch, forming a synergy between five logistics parks in Wuhan, Ezhou and Xiangyang, which makes the transportation and delivery the emergency materials into Hubei Province more fast and efficient. The mode of sharing logistics resources through the Cloud platform provides the feasibility for the horizontal cooperation between logistics parks. However, the cooperation between logistics parks will lead to the loss of some customer orders and the leakage of trade secrets, and there are revenue damage and security risks. How do the Cloud platform and logistics parks choose the cooperation strategy and maximize the benefits of all parties? It is an urgent problem to be solved.

Considering that the cooperation strategy between logistics parks is evolved and developed from multi-party

* Corresponding author. Email: wuliuwtu@163.com(Zhou Xingjian)

† Corresponding author. Email: 526392430@qq.com(Xu Gechen)

competition and cooperation Game^[9], which is not only affected by the "bounded rationality" of logistics parks, but also constrained by the Cloud platform, the paper uses the Evolutionary Game method, takes the Cloud platform as an external effective constraint, constructs a three-party cooperation decision model, and analyzes the cross supply chain horizontal cooperation mechanism between logistics parks.

2. CONCLUSIONS

Through the strategy selection and simulation analysis of tripartite Evolutionary Game between Cloud platform and logistics parks, the main conclusions are as follows:

(1) There are two strategies for cross supply chain horizontal cooperation among logistics parks:①all logistics parks choose not to share resources and the Cloud platform party chooses not to subsidize strategy;② all logistics parks choose to share resources and the Cloud platform party chooses to implement subsidy strategy.

(2) The cross supply chain horizontal cooperation strategy among logistics parks is related to the initial share ratio of resources. At the same time, when the initial sharing ratio of resources among logistics parks gradually increases, the rate of Cloud platform's subsidy ratio converging to 1 will be significantly accelerated, and the Cloud platform finally chooses to implement the subsidy strategy.

(3) The cross supply chain horizontal cooperation strategy among logistics parks is promoted by the subsidy proportion of Cloud platform. For the sake of maximizing the interests, the willingness of the logistics park to choose the horizontal cooperation strategy will also increase.

Combined with the actual operation of logistics parks, the following suggestions are put forward. We should strengthen the guiding effect of core logistics parks, strengthen the supervision and guidance ability of Cloud platform, and emphasize risks sharing and revenues sharing among logistics parks.

Based on the Cloud platform there are other factors that will affect the Game process of resource sharing between logistics parks, such as the coordination of interests between logistics parks. In addition, considering the risk of resource sharing between logistics parks, whether the joint risk governance model with Cloud platform subsidies will have different effects on the resource sharing strategy between logistics parks is a feasible direction of future research.

3. ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China under Grant 20FJL146, National Natural Science Foundation of China under Grant 72062019, Postdoctoral Research Project of Jiangxi Province under Grant 2019KY13, Social Science Foundation of Hubei Province under Grant 2018044, Social Science Foundation of Wuhan City under Grant 2019005, and Research Project Plan of China Logistics Society and China Federation of Logistics and Purchasing in 2021.

REFERENCES

- [1] Shi N, LI Z P, ZHAO S Li, et al. (2019). Construction of modern logistics system based on interconnection. *Science and Technology Management Research*, (15):191-197.
- [2] Chen Yunzhe. (2020). "The Map of China's Logistics Parks" is the navigation of the emergency logistics transfer station for epidemic prevention in Hubei. *Beijing Business Daily*. <https://finance.ifeng.com/c/7tnGVt7NHjk>. (in Chinese)
- [3] Deng Jianxin, YAN Haijuan, SHI Xianlian. (2020). A research on key information for collaborative business response control of logistics information platforms. *Industrial Engineering Journal*, 23(01):10-17(in Chinese)
- [4] Qi E S, LI T B, LIU L, et al. (2017). The Evolutionary Game analysis of the sharing of manufacturing resource in the environment of Cloud manufacturing. *Operation Research and Management Science*, 26(2): 25-34.
- [5] Friedman D. (2019). Evolutionary Games in economics. *Econometrica: Journal of the Econometric Society*, 637-666.

Short Research Paper

Research on Location Optimization of Waste Transfer Center in Harbin*Jiangping Wan^{1*}, Xuejian Li¹, Pengjie Wu¹, Zhaotao Chen¹, Zhiwei Du¹*

¹South China university of technology, Business administration college, Wushan Road No. 381, Tianhe Guangzhou
Guangdong, 510640, China

Abstract: This paper puts forward the background of the topic, expounds the research status of waste transfer center location at home and abroad, and summarizes and establishes a nonlinear mixed 0-1 planning type of multi transfer center location model in the relevant theory. The model gives the location constraints when the alternative point of the transfer center has been determined. Under the constraints, it mainly considers the total transportation cost of the waste passing through the transfer center, the treatment cost due to the classification, compression and packaging of waste, and the fixed investment cost of the construction of the waste transfer center, so as to minimize the total cost of these four parts. Based on Gaode map, on the basis of establishing the model, the investigation, statistics and analysis data are substituted to solve the optimal location result.

Keywords: Waste; Transfer Center; Site Selection; Waste Treatment Plant

1. INTRODUCTION

The problem of waste treatment has gradually become a problem for urban managers. If the city managers are not scientific enough in the site selection of waste treatment facilities and uneven distribution of facilities, it will not only affect people's production and life, but also affect the efficiency of waste transfer, making the overall cost of waste treatment larger.

2. LITERATURE REVIEW

Foreign countries have a deep research history on the location of transshipment center. As early as 1909, Weber's problem was proposed to study the location of single warehouse. The minimum transportation distance between warehouse and user was explored, and the graphic method was given. Since then, many related problems have evolved from Weber problem, such as location allocation problem, dynamic location problem and so on. With the development of society and the rapid development of economy, the location problem is constantly enriched, resulting in many new theories and methods, and put forward many models and methods to solve all kinds of transfer center location problems, such as Kuehn hamburger model method^[1], flexible allocation method^[2], Baumol wolf method^[3], and p-median method^[4] etc..

As early as around 1970, foreign countries put forward the problem of optimizing the transportation route of urban garbage collection vehicles. Beltrami (1974)^[5] and ulusoi (1985)^[6] found that when the capacity of garbage reached the upper limit, it was necessary to transport the garbage to the transfer center, so they divided the problem into two steps. The first step was to find the itinerant path. The second step was to consider which collection point could reach the upper limit according to the capacity of garbage truck, and then analyzed the problem. The obtained path is decomposed to find a feasible result. Since the 21st century, there have been more and more researches on the garbage transportation path, resulting in many new ideas and methods, and many classical models and methods have emerged in this period. For example, Tung (2004)^[7] used mixed 0-1 planning model to study site selection, and amponsah (2004)^[8] proposed double objective model with two constraints of

* Corresponding author. Email: csjpw@scut.edu.cn (Jiangping Wan), 1339765449@qq.com (Xuejian Li)

environment and cost, ghose (2006)^[9]. The optimization model of garbage collection and transportation route is proposed with combining garbage collection with vehicle loading capacity limitation and using GIS technology. Gruler (2015) 错误!未找到引用源。 designed and proposed a heuristic algorithm: WCP by combining Monte Carlo simulation with heuristic algorithm.

3. LOCATION MODEL OF WASTE TRANSFER CENTER IN HARBIN

3.1 Assumptions and elements of the model

(1) Hypothesis

The location problem studied in this paper is to select a certain number of addresses from the address set of all known candidate points to establish a garbage transfer center minimize the total cost.

In order to facilitate modeling and make the model easy to understand, the assumptions are made in the following:

- ① Garbage transportation is completed at one time;
- ② Select the most suitable transfer center within the scope of the location to be selected
- ③ One transfer center can transport waste to multiple waste treatment plants, and one transfer center can also receive waste from multiple transfer stations. Transfer centers cannot transport waste to each other;
- ④ System transportation includes transportation from waste treatment plant to transfer center and from transfer center to transfer station;
- ⑤ The transportation cost of garbage is directly proportional to the transportation volume and distance;
- ⑥ The unit transportation cost between the waste treatment plant and each transfer center and between the transfer center and each transfer station is known constant;
- ⑦ The average daily output of each transfer station is a known constant;
- ⑧ Fixed investment in establishing and operating transit centers;
- ⑨ The waste disposal cost of the transfer center is known;
- ⑩ The capacity and number of transfer centers are limited.

(2) Elements

① Expenses

The costs considered in this model include: fixed cost of investment in transfer center, transportation cost of transporting waste from transfer station to transfer center, transportation cost of transporting waste to waste treatment plant, packing and compression cost of waste in transfer center.

② Refuse transfer volume

Because the data collection can not be carried out on site, according to the "code for design of domestic waste transfer station"^[10], if there is no actual data, it can be estimated according to the following formula:

$$Q = \delta mw / 1000$$

Where: Q -daily transfer capacity of transfer station (T / D);

M -the actual number of people in the service area;

W -per capita daily output of garbage in service area (kg / person · d), 1.0-1.2 kg / person · D

δ -coefficient of garbage output, δ value can be 1.3 ~ 1.4.

3.2 Establishment of location model

In order to construct the model conveniently, the related variables and parameters are defined as follows:

(1) Model parameters

The parameters of the model are shown in Table 1.

Table 1. Model parameters

Parameter symbol	Symbolic meaning
m	Number of waste treatment plants
n	Number of transfer centers to be selected
i	Number of transfer stations
c_{ji}	Unit distribution cost from the j transfer station to the i Transfer Center
j	Garbage output of the j transfer station
g	Operation days in planning period
f_i	Fixed fee of the i Transfer Center
P	Maximum number onsite centers selected

(2) Model variables

x_{ki} — The transportation volume from the k transfer center to the i waste treatment plant;

y_{ij} —The transportation volume from the i transfer station to the j transfer center each time;

z_i —0-1 integer variable, when the i transit center is selected, $z_i = 1$; when the i transit center is not selected, $z_i = 0$.

(3) Objective function

The total cost in the planning period is divided into four parts: the total transportation cost $\sum_{k=1}^m \sum_{i=1}^n g a_{ki} x_{ki}$ of garbage from the transfer center to the garbage treatment plant, the total transportation cost $\sum_{i=1}^n \sum_{j=1}^l g c_{ij} y_{ij}$ of garbage from the transfer station to the transfer center, the garbage treatment cost of the transfer center (where w_i is the flow of the i transfer center, and the index θ can be 1 / 2), and the fixed cost $\sum_{i=1}^n z_i f_i$ of the transfer center. Then the total cost is the sum of the above four items. It is easy to know that the above cost E should be the minimum, that is:

$$\text{Min } E = \min(\sum_{k=1}^m \sum_{i=1}^n g a_{ki} x_{ki} + \sum_{i=1}^n \sum_{j=1}^l g c_{ij} y_{ij} + \sum_{i=1}^n z_i v_i W_i^\theta + \sum_{i=1}^n z_i f_i)$$

(4) Constraints

I. Restriction of waste transfer

The total amount of garbage discharged from a transfer center to each garbage treatment plant shall not exceed the total amount of garbage received by the transfer center:

$$\sum_{i=1}^n x_{ki} \leq A_k, k = 1, 2, \dots, m$$

II. Waste output constraints

For each transfer, the total amount of garbage transported from each transfer station to the transfer center should be equal to the daily garbage production:

$$\sum_{i=1}^n y_{ij} \geq D_j, j = 1, 2, \dots, l$$

III. Balance constraints

During the planning period, the flow is balanced, that is, the amount of garbage in and out of each transfer center is equal:

IV. Capacity constraints

For each transfer, the total amount of waste received by each transfer center shall not exceed the maximum capacity of the transfer center, that is:

V. Number constraint

The number of transfer centers allowed to be established does not exceed the given value p :

Non negative constraint

The variables in the model must be greater than or equal to zero:

VII. Integer constraint

(5) Model

The location model of transfer center is established based on the above analysis:

$$\begin{aligned}
 \text{Min } E = \min & \left(\sum_{k=1}^m \sum_{i=1}^n g a_{ki} x_{ki} + \sum_{i=1}^n \sum_{j=1}^l g c_{ij} y_{ij} + \sum_{i=1}^n z_i v_i W_i^\theta + \sum_{i=1}^n z_i f_i \right) \\
 & \sum_{i=1}^n x_{ki} \leq A_k, k = 1, 2, \dots, m \\
 & \sum_{i=1}^n y_{ij} \geq D_j, j = 1, 2, \dots, l \\
 & \sum_{k=1}^m x_{ij} = \sum_{j=1}^l y_{ij} = W_i, i = 1, 2, \dots, n \\
 & \sum_{k=1}^m x_{ki} \leq z_i M_i, i = 1, 2, \dots, n \\
 & \sum_{i=1}^n z_i \leq P \\
 & x_{ki} \geq 0, y_{ij} \geq 0, k = 1, 2, \dots, m; i = 1, 2, \dots, n; j = 1, 2, \dots, l \\
 & z_i = \begin{cases} 1, & \text{transfer center} \\ 0, & \text{other} \end{cases}
 \end{aligned}$$

4. MODEL PARAMETER ANALYSIS

4.1 Scale standard of transfer station

The scale of transfer station can be divided into small, medium and large scale. The large scale standard is more than 450t / D, the medium scale standard is 150 ~ 450t / D, and the small scale standard is less than 150t / D^[10]. The existing garbage collection methods in Harbin are mixed, including manual collection vehicles, small vehicles and large vehicles.

4.2 Selected sites and related parameters of municipal solid waste in Harbin

The existing population and the area of each district in Harbin can be found on the website of Harbin government. According to the estimation formula $Q = \delta mw/1000$ ($\delta = 1.3, w = 1$), the daily garbage output of each district can be calculated, as shown in Table 2.

Table 2. Population and garbage output of Harbin

Index	Songbei District	Daoli District	Nangang District	Daowai District	Xiangfang District	Bungalow area	Hulan District	Total
Population (10000 people)	19.7	69	105	68.6	72.4	16	62.3	413
Area (km ²)	736.3	479.2	182.9	618.6	339.5	98	2185.9	4640.4
Garbage output (T / D)	256.1	897	1365	891.8	941.3	208	809.9	5369.1

The daily output of waste in Harbin is 5369.1t/d, which is 3758t / D when 70% of the waste is transported to the waste treatment plant for treatment. According to the area and garbage output of each district, the number of medium-sized transfer stations in each district is calculated. According to the principle of meeting the demand, a suitable number of suitable types of transfer stations are matched for each district. Table 3 is illustrated (because the population of cottage district is small, it is not suitable to set up a separate transfer station, so the demand is merged into the adjacent Xiangfang District).

Table 3. Quantity demand of transfer stations in Harbin

Transfer station	Songbei District	Daoli District	Nangang District	Daowai District	Xiangfang District	Bungalow area
Medium transit station	1	2	2	2	2	1
Processing capacity of each station (T)	154	269	410	268	345	486

According to the above principles, combined with the actual situation of Harbin, the following medium-sized transfer stations and large-scale transfer center type I and type II sites and waste treatment plants (Harbin domestic waste incineration power plant of Heilongjiang New Century Energy Co., Ltd., Harbin Yifeng ecological environment Co., Ltd. (Xiangyang domestic waste treatment plant), Harbin Shuangqi environmental protection resources utilization Co., Ltd.) are obtained Table 4 is illustrated the longitude and latitude coordinates of the waste treatment plant in Acheng District.

Table 4. Latitude and longitude coordinates of alternative points

model	Songbei District	Daoli District	Nangang District	Daowai District	Xiangfang District	Bungalow area
Type I	P ₁ (126.61,45.83)	0	P ₂ (126.55,45.61)	0	P ₃ (126.81,45.67)	0
Type II	P ₄ (126.47,45.79)	P ₅ (126.49,45.71)	P ₆ (126.58,45.69)	P ₇ (126.74,45.79) P ₈ (126.82,45.86)	P ₉ (126.77,45.74)	0
Medium 1	Y ₁ (126.49,45.83)	Y ₂ (126.60,45.72)	Y ₃ (126.57,45.59)	Y ₄ (126.84,45.80)	Y ₅ (126.70,45.63)	Y ₁₀ (126.88,45.86)
Medium 2	0	Y ₆ (126.43,45.65)	0	Y ₇ (126.77,45.78)	Y ₈ (126.50,45.68)	0
Medium 3	0	0	Y ₉ (126.51,45.66)			0
Garbage factory	Yifeng garbage disposal plant	G ₁ (126.92, 45.78)		Acheng waste treatment plant	G ₂ (127.05,45.57)	
	New energy	G ₃ (126.69,45.74)		Shuangqi	G ₄ (126.82,45.71)	

4.2 Building cost matrix

4.2.1 Constructing distance matrix

The shortest driving distance between two points can be obtained by path optimization of Gaode map, as shown in Table 5 and Table 6

Table 5. GP distance matrix

P	G ₁	G ₂	G ₃	G ₄
P ₁	40.703	62.721	15.999	44.455
P ₂	53.087	54.125	27.583	31.448
P ₃	29.515	37.007	16.802	23.909
P ₄	70.085	71.123	28.648	48.516
P ₅	67.307	68.345	21.721	45.738
P ₆	34.185	52.510	14.430	31.176
P ₇	15.971	40.466	10.406	38.049
P ₈	18.577	48.951	23.812	46.133
P ₉	21.121	39.220	8.489	29.303

Table 6. PY distance matrix

Y	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉
Y ₁	12.784	28.709	37.351	8.757	48.312	25.819	26.502	39.025	33.03
Y ₂	19.743	15.552	21.526	25.889	10.789	4.534	17.554	31.069	17.205
Y ₃	35.169	3.406	26.145	40.414	26.051	1.115	31.043	12.874	26.602
Y ₄	37.672	55.115	25.506	53.115	57.614	31.006	12.985	12.874	19.178
Y ₅	33.146	23.347	19.257	38.446	28.072	16.502	24.155	45.101	19.714
Y ₆	33.003	17.74	37.315	24.476	10.322	17.93	32.065	55.401	34.54
Y ₇	18.045	35.083	21.899	31.794	31.532	21.132	3.737	14.837	6.954
Y ₈	26.522	12.145	29.219	24.194	31.532	11.091	25.584	55.119	29.676
Y ₉	27.433	10.266	30.13	25.105	9.19	12.002	26.495	56.03	30.587
Y ₁₀	32.869	68.798	32.877	48.312	57.125	38.377	17.563	7.741	26.549

4.2.2 Unit transportation cost

The unit transportation cost is 1.5 yuan (t · K) combined with the actual situation. Transportation cost is calculated according to the above data. The construction of Shanghai waste collection and transportation system is complete. Therefore, the cost and processing capacity of the equipment are estimated by referring to the facility data of Shanghai waste transfer station.

Table 7 and Table 8 are illustrated the unit freight between facilities in this paper

Table 7. Unit freight from transfer center to waste treatment plant

Waste treatment plant	Treatment capacity (T)	Unit freight from transfer center to waste treatment plant (100 yuan / ton)								
		P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉
G ₁	1200	61.05	79.63	44.27	105.13	100.96	51.28	23.96	27.87	31.68
G ₂	380	94.08	81.19	55.51	106.68	102.52	78.77	60.70	73.43	58.83
G ₃	200	24.00	41.37	25.20	42.97	32.58	21.65	15.61	35.72	12.73
G ₄	1600	66.68	47.17	35.86	72.77	68.61	46.76	57.07	69.20	43.95

Table 8 Unit freight from transfer station to transfer center

Transfer Center	Fixed cost (10000 yuan)	Capacity (T)	Unit freight from each transfer station to transfer center (yuan / ton)									
			Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀
P ₁	1000	1000	19.18	29.61	52.75	56.51	49.72	49.50	27.07	39.78	41.15	49.30
P ₂	1000	1000	43.06	23.33	5.11	82.67	35.02	26.61	52.62	18.22	15.40	103.20
P ₃	1000	1000	56.03	32.29	39.22	38.26	28.89	55.97	32.85	43.83	45.20	49.32
P ₄	500	500	13.14	38.83	60.62	79.67	57.67	36.71	47.69	36.29	37.66	72.47
P ₅	500	500	72.47	16.18	39.08	86.42	42.11	15.48	47.30	47.30	13.79	85.69
P ₆	500	500	38.73	6.80	1.67	46.51	24.75	26.90	31.70	16.64	18.00	57.57
P ₇	500	500	39.75	26.33	46.56	19.48	36.23	48.10	5.61	38.38	39.74	26.34
P ₈	500	500	58.54	46.60	19.31	19.31	67.65	83.10	22.26	82.68	84.05	11.61
P ₉	500	500	49.55	25.81	39.90	28.77	29.57	51.81	10.43	44.51	45.88	39.82
Garbage output of each transfer station (T)			154	269	410	268	345	269	268	345	410	486

Treatment cost: according to the literature, the treatment cost of 500t / D waste transfer station is 22.02 yuan / T; the treatment cost of 1000t / D waste transfer station is 19.08 yuan / T^[11].

Therefore, the treatment cost of 500t / D waste transfer station is $w_z = 30.25$. 1000t/d waste transfer station, $w_z = 27.31$. The treatment cost of the proposed transfer station is shown in Table 9.

Table 9. Transfer station processing cost

Cost	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉
Processing cost	27	27	27	30	30	30	30	30	30

5. EXAMPLE ANALYSIS

5.1 Example description

Harbin has four waste treatment plants; 10 transfer stations are selected to be distributed in each area; in order to provide transportation efficiency and save cost, the waste transfer center is planned to be established. After on-site investigation, three type I transfer centers P1, P2, P3 and six type II transfer centers P4, P5, P6, P7, P8, P9 are determined. Based on the economic principle, the best goal is to choose the plan with the lowest cost in the planning period of 10 years (336 working days per year).

According to the above calculation, the following data are obtained: the waste treatment cost of each type I transfer center is 27, 27, 27 (unit: yuan / ton), the waste treatment cost of each type II transfer center is 30, 30, 30, 30, 30, 30, $\theta = 1 / 2$; the unit freight from each transfer center to the waste treatment plant is shown in Table 10;

the fixed capacity cost of each transfer center and the waste output of each transfer station And the unit freight between them is shown in Table 11.

5.2 Solving process based on lingo

The above mathematical model and the calculated data are written into lingo software language, and the programming is input into the software for solving. Because the number of P (P₁-P₉) is unknown, the enumeration method is used to test, and the best one is selected.

Table 10. Treatment capacity of each treatment plant and unit freight from transfer center to treatment plant

Waste treatment plant	Treatment capacity (T)	Unit freight rate from transfer center to waste treatment plant (yuan / ton)								
		P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉
G ₁	1200	61.05	79.63	44.27	105.13	100.96	51.28	23.96	27.87	31.68
G ₂	380	94.08	81.19	55.51	106.68	102.52	78.77	60.70	73.43	58.83
G ₃	200	24.00	41.37	25.20	42.97	32.58	21.65	15.61	35.72	12.73
G ₄	1600	66.68	47.17	35.86	72.77	68.61	46.76	57.07	69.20	43.95

Table 11. Capacity of each transfer center, daily garbage output of each transfer station and unit freight between them

Transfer Center	Fixed cost (10000 yuan)	Capacity (T)	Unit freight from each transfer station to transfer center (yuan / ton)									
			Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀
P ₁	1000	1000	19.18	29.61	52.75	56.51	49.72	49.50	27.07	39.78	41.15	49.30
P ₂	1000	1000	43.06	23.33	5.11	82.67	35.02	26.61	52.62	18.22	15.40	103.20
P ₃	1000	1000	56.03	32.29	39.22	38.26	28.89	55.97	32.85	43.83	45.20	49.32
P ₄	500	500	13.14	38.83	60.62	79.67	57.67	36.71	47.69	36.29	37.66	72.47
P ₅	500	500	72.47	16.18	39.08	86.42	42.11	15.48	47.30	47.30	13.79	85.69
P ₆	500	500	38.73	6.80	1.67	46.51	24.75	26.90	31.70	16.64	18.00	57.57
P ₇	500	500	39.75	26.33	46.56	19.48	36.23	48.10	5.61	38.38	39.74	26.34
P ₈	500	500	58.54	46.60	19.31	19.31	67.65	83.10	22.26	82.68	84.05	11.61
P ₉	500	500	49.55	25.81	39.90	28.77	29.57	51.81	10.43	44.51	45.88	39.82
Garbage output of each transfer station (T)			154	269	410	268	345	269	268	345	410	486

After software calculation, only when x = 4, 5, 6, 7, 8 and 9, the model has the optimal solution. The optimal solution is shown in the Table 12 as follow.

Table 12. Site selection results

X value	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	Total cost (yuan)
x=4	√	√	√				√			8.726942×10 ⁸
x=5		√	√			√	√	√		6.324238×10 ⁸
x=6		√	√			√	√	√	√	6.466272×10 ⁸
x=7		√	√		√	√	√	√	√	6.134563×10 ⁸
x=8		√	√	√	√	√	√	√	√	6.260074×10 ⁸
X=9	√	√	√	√	√	√	√	√	√	6.398269×10 ⁸

When x = 7, the results are as follows

Z (P₁) = 0, Z (P₂) = 1, Z (P₃) = 1, Z (P₄) = 0, Z (P₅) = 1, Z (P₆) = 1, Z (P₇) = 1, Z (P₈) = 1, Z (P₉) = 1, which indicates that the transfer center should be set up at the alternative points P₂, P₃, P₅, P₆, P₇, P₈, P₉ to minimize the total cost of 6.134563 × 10⁸ yuan. Table 13 and Table 14 are illustrated the transportation relationship between waste treatment plant , and transfer center and between transfer center and transfer station respectively.

Table 13. Quantity of waste transported to the waste treatment plant by each transfer center (unit: ton)

Waste treatment plant	P ₂	P ₃	P ₅	P ₆	P ₇	P ₈	P ₉
G ₁	0	0	0	0	500	500	200
G ₂	0	224	0	0	0	0	0
G ₃	0	0	200	0	0	0	0
G ₄	1000	100	0	500	0	0	0
Circulation of Transfer Center	1000	324	200	500	500	500	200
Margin	0	676	300	0	0	0	300

Table 14. Quantity of garbage transported from each transfer station to transfer center (unit: ton)

Transfer station	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Total
P ₂	0	0	176	0	0	69	0	345	410	0	1000
P ₃	0	0	0	0	324	0	0	0	0	0	324
P ₅	0	0	0	0	0	200	0	0	0	0	200
P ₆	0	266	234	0	0	0	0	0	0	0	500
P ₇	154	0	0	254	0	0	92	0	0	0	500
P ₈	0	0	0	14	0	0	0	0	0	486	500
P ₉	0	3	0	0	21	0	176	0	0	0	200
Refuse output of transfer station	154	269	410	268	345	269	268	345	410	486	3224

The optimal planning scheme is illustrated in Table in the following: P₂, P₃, P₅, P₆, P₇, P₈, P₉ are selected as the location of the transfer center, and the total cost is expected to be 6.134563×10^8 yuan within the planning period (10 years).

6. CONCLUSION

In this paper, the related theories of garbage transfer center and its location are studied comprehensively. A multi transfer center location model based on nonlinear mixed 0-1 programming is established, and the model is solved by lingo software. By inputting site selection parameters, the site selection results of waste transfer center in Harbin are calculated.

The concept and development process of waste collection and transportation system and transfer center are studied. A location model with the objective function of minimizing the total cost of the system is established under the condition that the selected points of the waste transfer center have been determined and the capacity and number of transfer centers are limited. Through the path planning of Gaode map, the actual driving distance between each point is obtained, which improves the credibility and authenticity of the location results.

ACKNOWLEDGEMENT

This research was supported by Student Research Project (key technology of enterprise knowledge map and its intelligent application) South China University of Technology projects (Granted No. X202010561159)

REFERENCES

- [1] Kuehn A, Hamburger M. (1963). A Heuristic for Location Warehouses. *Management Science*, 9(6): 643 -666
- [2] Edgar M, Hoover.(1937).*Location in Theory and Leather Industries*. Cambridge Mass: Harvard University Press:429-452
- [3] BaumolW, Wolfe P A. (1958). Warehouse Location Problem. *Operation Research*, 6(2): 252 -263
- [4] Alan T M, Ross A G. (1997). Capacitated Service and Regional Constraints in Location-Allocation Modeling. *Location Science*, 5(2): 103-118
- [5] Beltrami E J, Bodin L D. (1974).Networks and vehicle routing for municipal waste collection. *Networks*, 4(1): 65-94
- [6] Ulusoy G. The fleet, size and mix problem for capacitated are routing. *European Journal of Operational Research*, 22(3): 329-337
- [7] Dang V T, Pinnoi A. (1958). Vehicle routing-scheduling for waste collection in Hanoi: *European Journal of Operational Research*, 2000, 125(3): 449-468
- [8] Amponsah S K, Salhi S. (1958). The investigation of a class of capacitated are routing problems: the collection of garbage in developing countries: *Waste Management*, 2004, 24(7): 711-721
- [9] Ghose M K, Dikshit A K, Sharma S K. (1958).A GIS based transportation model for solid waste disposal-A case study on Asansol municipality[J]. *Waste Manag*, 26(11): 1287-1293
- [10] CCJ / T 47-2016.(2019). Technical specification for domestic waste transfer station. Beijing: Ministry of housing and urban rural development of the people's Republic of China
- [11] Zhang Yan, Wang Yue, Li Xiaoxue, Han Chao. (2019). Cost analysis of domestic waste treatment in T City. *Environmental protection technology*, 25 (02): 56-60 + 64

Full Research Paper**Research on Authors' Co-authorship Network in Supply Chain****Finance in China Based on Social Network Analysis***Yang Zhao*^{1,2*}, *Ning Wang*²¹Center for Studies of Information Resources, Wuhan University, China²School of Information Management, Wuhan University, China

Abstract: This paper used social network analysis method to analyze the high frequency author-co-author network in domestic supply chain finance field. In this study, relevant literatures on supply chain finance collected by CNKI from 2005 to 2019 were selected and key information was extracted. This paper used SATI, the literature citations information statistical analysis tool, to build correlation matrix. Social network analysis software UCINET was used to draw the map of co-authored network. The authors' subnet patterns, network density, centrality, cohesive subgroups and structural holes were analyzed, and the scientific collaboration network characteristics in this field were elaborated to promote academic exchange and development in the field of supply chain finance. The analysis results showed that the authors of supply chain finance in China is not connected enough, the overall network connectivity is weak, and there are few core nodes that play a key role. Therefore, scholars in this field should strengthen cooperation appropriately in the future.

Keywords: Social network analysis, Supply chain finance, Cohesive subgroup, Structural holes

1. INTRODUCTION

Supply chain finance is an interdisciplinary discipline in supply chain, finance, logistics and other fields. It is a specific micro category of comprehensive financial activities based on industrial supply chain and targeted at participants of supply chain based on business flow, logistics and information flow in supply chain operation^[1-2]. There are three typical models of supply chain finance, which are logistics enterprise leading model, enterprise group cooperation model and commercial bank service model^[3]. The rapid development of supply chain finance has brought great opportunities and challenges to the current business model innovation and sustainable development of enterprises. At the same time, it has also proposed many new problems and new directions for the research of management theory, method and technology innovation, which has attracted wide attention in the academic circle.

So far, many scholars at home and abroad have conducted in-depth discussion on the research of supply chain finance, analyzed the development process of supply chain finance, elaborated the characteristics and elements of supply chain finance, put forward the research theme framework of supply chain finance, and studied various specific issues under this framework. Zhang et al. empirically tested the relationship between supply chain finance and enterprise performance by studying the relationship between supply chain finance and enterprise performance and bankruptcy risk, and also expanded the knowledge of supply chain finance^[4]. Hugo K.S. Lam et al. studied the impact of supply chain finance firms with unique characteristics (i.e., enterprise characteristics, cooperation mechanism and service types) on the market value of service providers^[5]. Zhou Han starting from the theory and practice experience of supply chain finance is introduced, and emphatically analyzes the risk of supply chain finance, especially in the participation main body based on the analysis, found that supply chain finance from the bank general financing specific risk and the corresponding risk prevention and methods, and on this basis puts forward some countermeasures and Suggestions on the development of supply chain finance^[6]. Xia Yu et al. used

* Corresponding author. Email: yangzhao_0813@whu.edu.cn (Yang Zhao)

the method of literature research and logical analysis to sort out and review the existing research literature with the evolution of supply chain finance theory as the context, and based on this, made an outlook for the future research^[7]. It can be seen that supply chain finance has been fully studied and discussed by scholars in related fields. However, to our knowledge, there is currently no literature using social network analysis to analyze the co-authored network of supply chain authors. In order to fill this literature gap, this study uses the social network analysis method to discuss the characteristics of co-authorship networks in domestic supply chain fields from five aspects: subnet mode, network density, centrality, cohesive subgroup and structural hole based on the papers published in 2005-2019 collected from CNKI.

2. METHODOLOGY AND DATA

2.1 Data sources

Data source selection is a very important step in social network analysis. CNKI (China Academic Literature Network Publishing Database) was initiated by Tsinghua University and Tsinghua Tongfang in June 1996, aiming at realizing the dissemination, sharing and value-added utilization of knowledge resources in the whole society. With its high-quality content resources, leading technology and professional services, CNKI enjoys a high reputation in the industry. Based on this, this paper selected CNKI as the retrieval database, searched with "supply chain finance" as the main topic, and selected the period from 2005 to 2019. A total of 4,697 pieces of data were retrieved, and 4,678 pieces of data were finally obtained after screening 4,697 pieces of data and removing irrelevant data such as essays, notices and advertisements.

2.2 Data processing tools

The data processing tools used in this paper are SATI and UCINET.

SATI (Statistical Analysis Toolkit for Informatics) is a visual software for statistical analysis of bibliography information developed by Liu Qiyuan from Information Resource Management Department of Zhejiang University using C#. This software can deal with common data in Endnote, NoteFirst and NoteExpress formats. Its main functions include three parts: (1) field information extraction. In the "Options" panel, you can select the field information specified in the bibliography (such as subject word, key word, author, citation, institution, publication year, etc.) and you can choose to store it as a text document; (2) Item frequency statistics. According to the field information extracted from the "Options" panel, the frequency of the elements in the entry (such as subject word, keyword, author, citation, organization, publication year, etc.) is counted and sorted in descending order. (3) Construction of co-occurrence matrix. Can decide for themselves in the SATI Shared output rows and columns of the matrix, and the frequency of the corresponding number in the descending order table entry element as a matrix of knowledge unit operations, to build knowledge unit co-occurrence matrix and generate EXCEL format document, which can be based on the matrix import documents UCINET, NETDRAW and visualization analysis software to generate the co-occurrence network knowledge map.

UCINET (University of California at Irvine Network) is a social network analysis software written by Linton Freeman, an authoritative scholar of social network research at University of California at Irvine. It comes bundled with software such as Pajet, Mage and Netdraw. UCINET is capable of processing raw data in matrix format, and the software itself includes data management and transformation tools. UCINET itself does not contain tools for network visualization, but it can be visualized by exporting the results of data processing to software such as NetDraw. The main functional modules of UCINET are: (1) Network density analysis. (2) Network centrality analysis. (3) Analysis of cohesive subgroups.

2.3 Research Methods

This paper used the social network analysis method to discuss the coauthor relationship in the field of supply chain finance in China. Social network analysis is a method in anthropology, psychology, sociology, mathematics

and statistics, and other fields of development used in the measurement and investigation of the social system ("points") the characteristics of each part and the relationship between each other ("connection"), to indicate it in the form of a network, and then analyzes the relationship between the mode and characteristics of the analysis method [8]. The "point" represents the social actor, which can be any social unit or social entity, and the "connection" represents the connection between the actors or the substantive relationship that shows the occurrence. The basic assumption of social network analysis is that the structure and characteristics of social network, the relationship among them, the distribution and location of points, etc. will influence the behavior and attitude of points to some extent. The types of social network analysis include: ① main network analysis. Starting from one point (subject) of social network, the relationship between it and other points (object) is analyzed. ② Analysis of relation. Analyze the time, content, intensity, affinity and direction of the relationship between two points. The overall network analysis. Considering the relationship between all points, the structure of the integrated network is analyzed [9].

In this article, using the method of social network analysis studies the domestic financial sector, the author co-author network of supply chain relationship, the relationship between the individual, "micro" network with the "macro" structure of the massive social system, combining with the software will be the author of the relationship between using network graph representation, and analyze the results.

3. Basic statistical analysis of co-authorship

3.1 Basic statistical analysis

This paper summarized the number of co-authored papers and the total number of published papers in supply chain finance in China from 2005 to 2019 through a preliminary statistical analysis of a data sample composed of 4,678 articles in supply chain finance in China. The statistical results are shown in Figure 1. It is easy to know from Figure 1 that both the total number of published papers and the number of co-authored papers in the field of supply chain finance in China show an obvious growth trend. The overall growth period can be divided into two stages. The first stage is from 2005 to 2011, which is the stage of rapid growth. The second stage is from 2011 to 2015. In this stage, the number of publications and co-authored papers per year does not change much. The third stage is 2015-2016, during this period, the data on the number of publications and co-authored papers increased dramatically. The fourth stage is 2016-2019, which is a period of fluctuation. In this stage, the growth of domestic research literature in the field of supply chain finance fluctuates relatively steadily. The total number of published papers and the number of co-authored papers in this stage both reached the highest value in 2016, which were 765 and 248 respectively. It is easy to know from the analysis of Figure 1 that domestic research in the field of supply chain finance has been relatively mature and stable in the second stage.

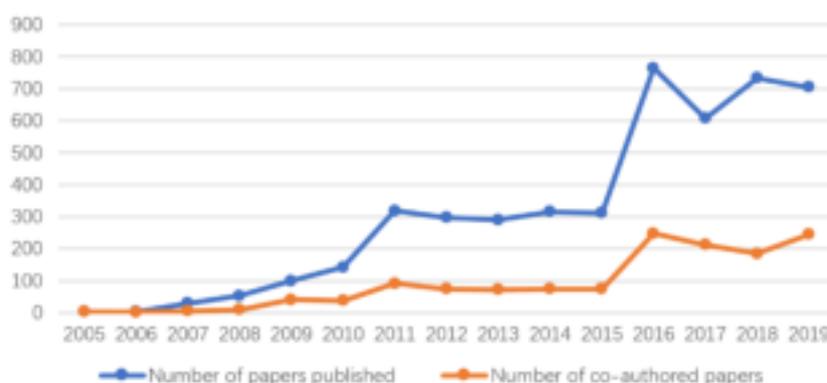


Figure 1. Statistics on the number of papers published and co-authored papers

The top 20 published papers by authors with co-author relationships were statistically analyzed, and the statistical results were shown in Table 1.

Table 1. Ranking of the number of articles published by core authors with co-author relationships

Ranking	Author	Number of Papers	Ranking	Author	Number of Papers
1	Hua Song	40	11	Yuefei Hu	9
2	Peng Xu	15	12	Yong Luo	9
3	Chaofeng Jiang	13	13	Yun Shen	9
4	Xuejian Chu	12	14	Xuehui Yi	9
5	Yongzhi Zhan	12	15	Lei Wang	9
6	Jinlong Chen	12	16	Wei Wang	8
7	Qiang Lu	11	17	Jie Liu	8
8	Juan He	11	18	Taifeng Xia	8
9	Wenli Hu	10	19	Xuxu Gao	8
10	Xuan Yang	10	20	Tao Chen	8

3.2 Subnet pattern analysis

The cooperative network of core authors in the domestic supply chain finance field is shown in Figure 1. Each node in the network represents a researcher, and the lines between nodes represent the cooperative relationship between researchers. The thicker the lines, the closer the connection between researchers is. Unconnected nodes represent independent researchers, which are on the left side of the image.

From the perspective of network mode, there are four subnet modes in the figure. (1) Single point mode. Single-point network refers to that researchers have completed the research work in the field independently. It is easy to know from the observation of the pictures that 21 researchers in the network belong to the single-point subnet. (2) Double karyotype mode. Double karyotype subnet network refers to a two-way network composed of two researchers, that is, two researchers cooperate in publishing papers. In Figure 2, 24 researchers collaborated in pairs to form 12 double karyotype subnets. (3) Bridge mode. A bridge network refers to one researcher who acts as a connecting node, such as a four-person subnet represented by Song Hua, who is the key node connecting the other three people. In the bridge subnet, the contact is the key node. If the contact node breaks, it means the relationship of the whole subnet breaks. There are two bridge-type subnets in the figure. (4) Complete mode. In a complete subnet, any two researcher nodes cooperate, that is, an edge connection. As shown in the subnet composed by Haiqing Hu et al., the four researchers had at least one collaborative experience. Complete subnet is a relatively ideal subnet mode, which represents a wider range of exchange opportunities^[10]. There are six complete subnets in the figure.

Through the further analysis of the network diagram of the core authors in the field of supply chain finance, the following characteristics can be found in the overall network. (1) Poor network connectivity of co-authors. According to the above analysis, in the overall network, most of the subnets are single-point type (21) and double karyotype (12), while the number of bridge type (2) and complete type (6) subnets is small. This indicates that researchers in the field of supply chain finance have strong independence, poor relevance and weak connectivity of the whole network structure. (2) The number of cooperative groups is small, and the scale of subnet is small. In the overall co-authorship network, the largest subnet is composed of Hua Song, Qiang Lu, Sijie Chen and Xuan Yang, and Liang Chen, Haiqing Hu, Lang Zhang and Daohong Zhang. The rest of them are small teams of three or two people. This shows that the research team in the field of supply chain finance in China has not formed a scale. Most of the research work is carried out by independent researchers or small teams, and there is a lack of large-scale and closely connected research team.

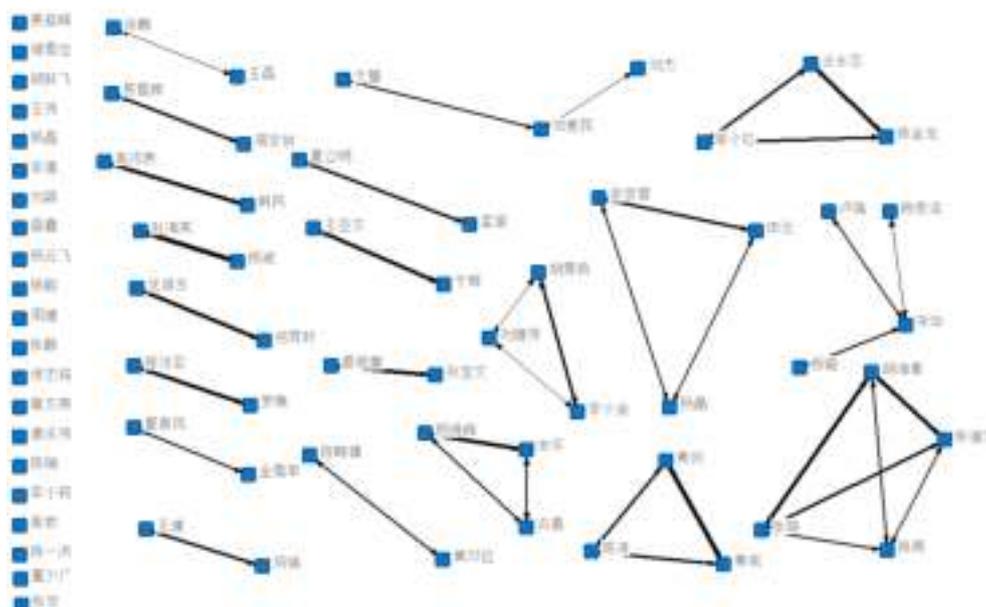


Figure 2. Network diagram of co-authorship relationship of core authors

3.3 Network density analysis

This paper mainly uses two indexes, network density and standard deviation, to measure the degree of coauthor network cohesion in domestic supply chain finance field. Among them, network density refers to the close correlation between researchers in a certain category, which represents a certain number of complex network relationships^[11]. Its calculation formula is as follows:

$$\rho = \frac{2L}{N*(N-1)} \quad (1)$$

Where N represents the number of researchers in the network, and L represents the actual number of connections between researchers in the network.

Standard deviation is the square root of the arithmetic mean of the square difference between the standard value of each unit of the population and its mean. It reflects the degree of dispersion between individuals in a group. Its calculation formula is as follows:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2)$$

Where N is the number of researchers in the network, x_i is the data value attached to the i_{th} researcher in the network, μ is the average value of the data value attached to all subjects in the network.

According to the calculation results of UCINET, the network density and network standard deviation of the domestic co-author network in the field of supply chain finance is 0.0036 and 0.0478, which is significantly higher than the network density value, indicating that the relationship between researchers in this field is relatively loose, with poor cohesion and less cooperation and communication.

3.4 Centrality analysis

Network centrality is an important indicator to measure the whole degree of network centrality, which is used to describe the social status or importance of individuals or organizations in the social network. Network centrality can be divided into three indexes: degree centrality, closeness centrality and between centrality^[12]. Among them, the degree of point-degree centrality is used to measure the activity of a specific node, representing the degree of influence of a certain node in a social network in which nodes are related to each other in a certain scale. The

higher the centrality of the degree, the greater the influence of the node in the social network. Closeness centrality is a centrality represented by distance. The smaller the value, the higher the degree to which a node directly influences other nodes in the network, and the greater its communication influence in the social network. Between centrality refers to the degree to which a node in the network controls the resource information, and represents the degree to which the whole network is centralized, that is, the degree to which the whole network operates around a node or a group of nodes. The size of between centrality is proportional to the amount of resource information controlled. The higher the intermediate centrality of a node is, the greater the degree that the node acts as a medium. Table 2 shows the top 15 researchers in the ranking of point-centricity, near-centricity and intermediate centrality.

Table 2. The centrality ranking of core authors

Degree Centrality			Closeness Centrality				Between Centrality		
Author	Absolute value	Relative value	Proportion	Author	Absolute value	Relative value	Author	Absolute value	Relative value
Haiqing Hu	1.833	1.852	0.051	Hua Song	9603.000	1.031	Hua Song	3.000	0.062
Daohong Zhang	1.758	1.776	0.049	Lang Zhang	9603.000	1.031	Aimin Deng	1.000	0.021
Rui Huang	1.625	1.641	0.046	Haiqing Hu	9603.000	1.031	Peng Xu	0.000	0.000
Jian Huang	1.625	1.641	0.046	Daohong Zhang	9603.000	1.031	Chaofeng Jiang	0.000	0.000
Lang Zhang	1.563	1.578	0.044	Liang Chen	9603.000	1.031	Yongzhi Zhan	0.000	0.000
Tao Chen	1.250	1.263	0.035	Xuan Yang	9605.000	1.031	Jinlong Chen	0.000	0.000
JinLong Chen	1.194	1.206	0.033	Qiang Lu	9605.000	1.031	Xuejian Chu	0.000	0.000
Yongzhi Zhan	1.194	1.206	0.033	Sijie Chen	9605.000	1.031	Juan He	0.000	0.000
Ping Song	1.024	1.034	0.029	Tao Chen	9702.000	1.020	Wenli Hu	0.000	0.000
Haiyan Zhao	1.000	1.010	0.028	Jinlong Chen	9702.000	1.020	Xuan Yang	0.000	0.000
Xiaohong Zou	1.000	1.010	0.028	Yongzhi Zhan	9702.000	1.020	Yuefei Hu	0.000	0.000
Bin Yang	1.000	1.010	0.028	Ping Song	9702.000	1.020	Yong Luo	0.000	0.000
Qifeng Yang	1.000	1.010	0.028	Qifeng Yang	9702.000	1.020	Qiang Lu	0.000	0.000
Yun Shen	0.911	0.920	0.026	Xiaohong Zou	9702.000	1.020	Xuehui Yi	0.000	0.000
Xuxu Gao	0.875	0.884	0.025	Jieping Liu	9702.000	1.020	Lei Wang	0.000	0.000

3.4.1 Degree centrality

According to the above introduction, the higher the degree centrality, the greater the influence of the node in the social network. In combination with Table 2, it is easy to observe that Haiqing Hu with the highest degree centrality with the value of 1.833, and Daohong Zhang with the second degree centrality with the value of 1.758. These two authors have a lot of cooperation with other researchers, which is consistent with the results presented in Figure 1. Table 2 shows that most researchers in the field of supply chain finance have a low degree centrality, which indicates that most researchers conduct independent research and few core nodes play a key role.

3.4.2 Closeness centrality

In combination with the above introduction, the smaller the value of closeness centrality is, the higher the degree to which a node directly influences other nodes in the network, and the greater its communication influence in the social network is. It can be seen from Table 2 that Hua Song, Lang Zhang, Haiqing Hu, Daohong Zhang and Liang Chen have the lowest degree of closeness to the center. They are the core nodes of the collaborative network of integrated supply chain finance and have the greatest communication influence in the collaborative network. According to the analysis in Figure 1, the researchers with the lowest closeness centrality to the center belong to the larger cooperative subnet, which indicates that the establishment of the cooperative subnet is conducive to the scientific researchers to better play the role of information dissemination.

3.4.3 Between centrality

According to the statistical results in Table 2, it is easy to know that the scholars with high between centrality are Hua Song and Aimin Deng. This indicates that these researchers have strong resource control ability and act as an important medium in the whole supply chain financial co-authorship network, which is the key node of the co-authorship network. However, except for the above two authors, between centrality of other authors is 0, which indicates that most researchers in the domestic supply chain finance field do not have the ability to control and connect resources, and the overall connectivity of the co-authorship network is weak.

3.5 Analysis of cohesive subgroups

In social network analysis, cohesive subgroup analysis is also called "small group" analysis. A small group is a subset of actors in which the actors have relatively stable, direct, close, frequent or positive contact [13]. The analysis of cohesive subgroups in social networks is to analyze the relationships and characteristics within and between small groups. By using UCINET's N-Clique analysis, this paper calculates the clique of authors in the field of supply chain finance, so as to analyze the clique characteristics of domestic authors in the field of supply chain finance.

A total of 8 groups were obtained through N-Clique analysis, and the result was consistent with the characteristics presented in Figure 1. (1) Group 1. Group 1 is composed of Hua Song, Qiang Lu, Xua Yangn and Sijie Chen. Except for Qiang Lu from Business School of Beijing Technology and Business University, the other researchers are all from Business School of Renmin University of China. The main research directions of this group are supply chain finance, risk management and financing performance. (2) Group 2. Group 2 consists of three researchers Yongzhi Zhan, Jinlong Chen and Xiaohong Zou, all of whom are from the School of Business Administration of Huaqiao University. The main research directions of the group are network effects, platformer supply chain finance and evolutionary game. (3) Group 3. Group 3 consists of Wenli Hu, Jieping Liu and Xiaojin Li, all of whom are from the School of Finance, Accounting and Finance, Guangdong Academy of Science and Technology. Their research directions are B2B e-commerce, online supply chain finance, and credit model. (4) Group 4. Group 4 consists of three researchers Yun Shen, Jingrong Li and Jing Yang, all from the School of Economics of Sichuan Agricultural University, whose research directions include supply chain finance, poverty reduction effect, multidimensional poverty, and PSM-DID. (5) Group 5. Group 5 is composed of three researchers, Jie Liu, Aimin Deng and Hui Wen, all of whom are from the School of Economics and Trade of Hunan University.

The main research directions of the group are third-party logistics and credit evaluation. (6) Group 6. Group 6 consists of three researchers, Tao Chen, Jian Huang and Rui Huang, all from South China Institute of Finance, Guangdong University of Finance. The three researchers mainly study Internet supply chain finance, organizational form and financial model. (7) Group 7. Group 7 is composed of four scholars: Lang Zhang, Liang Chen, Haiqing Hu and Daohong Zhang. The four researchers are all from the School of Economics and Management, Xi 'an University of Technology. Their main research interests include supply chain finance, credit risk assessment, and support vector machines. (8) Group 8. Group 8 consists of three scholars, Qifeng Yang, Ping Song and Chen Gu, all of whom are from the School of Economics, Wuhan University of Technology. Their main research directions are online supply chain finance and confirmed warehouse financing mode.

3.6 Analysis of structural holes

Structural hole theory is a branch of the school of network analysis. It was proposed by Burt in his book *Structural Hole: Social Structure of Competition* in 1992 [14]. Among them, structural hole refers to the absence of relations between non-redundant contacts, that is, some individuals or individuals in the social network have direct contact with some individuals, but do not have direct contact with other individuals or the relationship is broken. Caves appear from the observation of the whole social network [15].

The four measurement indexes of structural holes are: effective size, efficiency, constraint and hierarchy [16].

Effective scale refers to the degree of non-redundancy in the network. Efficiency is measured in structural holes, the efficiency of a point is equal to the effective size of the point divided by the actual size of the point in the individual network. Constraint represents the ability of the actor to use structural holes in his network, that is, the degree of direct or indirect closeness between a node in the network and other nodes. Hierarchy represents the extent to which the restriction revolves around the actor.

Table 3. Orders of the four indexes of the core author structure hole

Ranking	Author	Effective size	Efficiency	Constraint	Hierarchy
1	Jieping Liu	2.891	0.964	0.937	0.964
2	Liang Chen	2.586	0.646	0.725	0.153
3	Chen Gu	2.451	0.817	0.807	0.389
4	Hua Song	2.092	0.523	0.789	0.528
5	Jing Yang	1.947	0.649	0.841	0.236
6	Lang Zhang	1.654	0.413	0.804	0.139
7	Daohong Zhang	1.588	0.397	0.781	0.124
8	Haiqing Hu	1.577	0.394	0.784	0.131
9	Jingrong Li	1.576	0.525	0.896	0.234
10	Tao Chen	1.542	0.514	0.909	0.007
11	Yun Shen	1.541	0.514	0.857	0.176
12	Xiaohong Zou	1.530	0.510	0.875	0.041
13	Wenli Hu	1.512	0.504	0.986	0.433
14	Aimin Deng	1.486	0.495	0.903	0.536
15	Xiaojin Li	1.360	0.453	0.979	0.433
16	Yongzhi Zhan	1.274	0.425	0.880	0.055
17	Jinlong Chen	1.274	0.425	0.880	0.055
18	Ping Song	1.258	0.419	0.973	0.282
19	Qifeng Yang	1.251	0.417	0.984	0.291
20	Jian Huang	1.179	0.393	0.909	0.044

4. Conclusion

In this paper, using social network analysis method, based on CNKI retrieval library, use SATI, UCINET software for 2005 to 2019 in the field of domestic supply chain finance co-author network systematic analysis and visual display, and from the subnet model, network density, centrality, condensing subgroup, structural holes in the five aspects of domestic co-author network features in-depth in the field of supply chain.

The number of coauthored papers in the field of supply chain finance increased rapidly from 2005 to 2011, and reached a stable period from 2011 to 2015. In 2016, both the total number of published papers and the number of co-authored papers reached the highest value. From 2016 to 2019, the data is in a state of fluctuation. The most published co-authors were Hua Song, Peng Xu and Chaofeng Jiang. In terms of the subnet mode of the co-authorship network, the connectivity of the co-authorship network is poor. In the whole network, most of the subnets are single point mode and double karyotype mode, while the number of bridge mode and complete mode subnets is small. At the same time, the number of cooperative groups is small, and the size of the subnet is small. In terms of network density, according to the calculation results of UCINET, the network standard deviation value is significantly higher than the network density value, which indicates that the relationship between researchers in this field is relatively loose, the cohesion is poor, and the cooperation and communication are less. In terms of centrality, most researchers conduct independent research and do not have the ability to control and connect resources. The overall connectivity of the co-authored network is weak and there are few core nodes that play a key role. In terms of cohesive subgroups, a total of 8 groups were obtained through the N-clique of UCINET, and all of the 8 groups were scientific research groups with a total number of 3-4 people. In terms of structure hole, from the calculation results of effective scale, efficiency, constraint and hierarchy, Jieping Liu and Liang Chen et al are located at the core of the network, have a relatively strong control over the network and have greater authority in the field of supply chain finance.

This study has significant theoretical and practical implications. From the perspective of theoretical implications, this study is the first academic paper to analyze and study the co-author networks of authors in the field of supply chain finance in China by using social network analysis method. This study, supported by abundant data, comprehensively discusses the major coauthor networks in the supply chain finance field in the past 15 years, which provides a foundation for future research. From the perspective of practical significance, researchers in the field of supply chain finance in China have less cooperation and exchanges, and the relationship is relatively loose. Therefore, academic exchanges and cooperation between various research institutions should be strengthened to promote the development and innovation of theoretical research on supply chain finance in China.

ACKNOWLEDGEMENT

This study was supported by the National Key R&D Program of China under Grant 2019YFC010167, and “User Information Behavior Research Team under Big Data and Intelligent Environment” by the academic team of young scholars of Humanities and Social Sciences of Wuhan University.

REFERENCES

- [1] Zhu Xingxiong, He Qingsu, Guo Shanqi. (2018). Application of blockchain technology in Supply Chain Finance. *China Business And Market*. 32(03):111-119.(in Chinese)
- [2] Song Hua, Chen Sijie. (2016). Development of Supply Chain Finance and Internet Supply Chain Finance:A Theoretical Framework. *Journal of Renmin University of China*. 30(05):95-104.(in Chinese)
- [3] Xie Shiqing, He Bin. (2013). Analysis of three typical modes of international supply chain finance. (04):80-86.(in Chinese)
- [4] Zhang T, Zhang C Y, Pei Q. (2019). Misconception of providing supply chain finance: Its stabilising role. *International Journal of Production Economics*, 213, 175-184.

- [5] Lam H K, Zhan Y, Zhang M, Wang Y, Lyons A. (2019). The effect of supply chain finance initiatives on the market value of service providers. *International Journal of Production Economics*, 216, 227-238.
- [6] Zhou Han. (2015). Study on supply chain finance and its risk of domestic commercial bank. Ms D Thesis. Yunnan University of Finance and Economics.(in Chinese)
- [7] Xia Yu, Fang Lei, Wei Mingxia. (2019). Supply Chain Finance: The Evolution and the Inner Logic of the Theory. *Management review*. 31(12):26-39.
- [8] Zhu Qinghua, Li Liang. (2008). Social network analysis and its application in Information Science. *Information Studies:Theory & Application*. (02):179-183+174.(in Chinese)
- [9] Lei Hui, Nie Shanshan, Huang Xiaobao, Ma Wei. (2015). The Behavior Characteristics of Network Disseminators:An Empirical Study Based on Social Network Analysis. *Journal of Intelligence*.34(01):161-168.(in Chinese)
- [10] Peng Xixian, Zhu Qinghua, Shen Chao. (2013). The Collaborative Research of Social Computing Based on Social Network Analysis. *Journal of Intelligence*. 32(03):93-100.(in Chinese)
- [11] Shao Yunfei, Ouyang Qingyan, Sun Lei.(2009). Social Network Analysis and Its Application to Innovative Research. *Chinese Journal of Management*. 6(09):1188-1193+1203.(in Chinese)
- [12] Li X, Kawachi I, Buxton O M, Haneuse S, Onnela, J P. (2019). Social network analysis of group position, popularity, and sleep behaviors among US adolescents. *Social Science & Medicine*, 232, 417-426.
- [13] Sinkovits R S, Moody J, Oztan B T, White D R. (2016). Fast determination of structurally cohesive subgroups in large networks. *Journal of computational science*, 17, 62-72.
- [14] Burchard J, Cornwell B. (2018). Structural holes and bridging in two-mode networks. *Social Networks*, 55, 11-20.
- [15] Dong Xuemei, Li Feiyang. (2011) Research on Rural Private Lending on Structural Hole Theory. *South China Finance*. (08):40-43.(in Chinese)
- [16] Liang Lujin.(2011) The Research and Application of Structural Holes Theory in Management. *China Management Magazine*. 04):52-62.(in Chinese)

Full Research Paper

Understanding Users' Response to Open Government Data Portal: an Exploratory Study

Ting Shao^{1}, Xiaofei Cui², Hongbo Jiang¹, Xiangling Huang³*

¹ School of Economics and Management, Xiamen University of Technology, Xiamen, 361024, China

² School of Natural Sciences, University of Manchester, Manchester, M13 9PL, UK

³ Teacher's College, Jimei University, Xiamen, 361021, China

Abstract: Open government data portal (OGDP) makes the government transparent and enables the public to obtain massive data. However, the public know little about OGDP and the value of government data is still waiting for exploration. This paper takes IS success model, TAM and the paradigm of SOR to study the factors affecting users' response to OGDP. It selects system quality, data quality, interactivity, policy environment, perceived usefulness, perceived ease of use and use intention as variables to build research model. After collecting 93 valid respondents, we utilized SmartPLS to conduct statistical analysis for the model. The exploratory study finds that data quality affects system quality, and system quality positively affects perceived usefulness and perceived ease of use. Policy environment affects users' interactivity with OGDP, then influences their perception of the system, ultimately responds as use intention. Finally, this paper confirms that the mediating effect only exists in interactivity and perceived ease of use on the relationship between policy environment and use intention. Then it proposes suggestions for improving data and system quality, enhancing interactivity to encourage the public to use OGDP more frequently, and create value with government data.

Keywords: Open government data portal(OGDP), Perceived usefulness, Perceived ease of use, Use intention

1. INTRODUCTION

Open government data (OGD) is termed as making public sector information freely available in open formats and ways that enable public access and facilitate exploitation^[1], and it constitutes an important resource due to its potential to empower citizens, businesses, and transform how government delivers public services. In the past decades, governments around the world have built data portals, which is of great significance for enabling greater accountability, delivery of higher quality and new services, reduction in operation costs, and stimulation of innovations in both government organizations and businesses^[2]. The US government launched its open data portal data.gov in 2009, and took OGD as national strategy. Afterwards, some other countries such as UK, Denmark and Australia established OGDP as well. In China, the first local government data portal "Shanghai government data service platform" was launched in 2011. From then on, OGD has become necessity for e-government development. However, most of the previous studies concentrate on mechanism of OGD, or the impact on government, but the public still know little about OGD in China, let alone use of OGDP.

Open government data is an important part of the implementation of e-government, there are some studies about continuous use intention of e-government application. Based on TAM and ECM-ISC(Expectation Confirmation Model of IS Continuance), perceived usefulness, perceived ease of use, perceived trust and perceived entertainment affect users' satisfaction and continuous use of e-government^[3]. In addition, others think

*Corresponding author. Email: syting112@163.com (Ting Shao), c932994936@163.com (Xiaofei Cui), hbjiang@xmut.edu.cn (Hongbo Jiang), 724117823@qq.com(Xiangling Huang)

that network externalities, service quality, expectation confirmation and satisfaction significantly affect the citizens' continuous use intention of government portal^[4]. A case study of Harbin showed that perceived ease of use, compatibility and trust of network are the main factors influencing use intention of government portal^[5]. Because e-government and its extended application are kind of novelty, most of the research on the use intention relied on TAM or IS success model.

By far, the research about users' response to OGDG is still scarce. Zhu et al. introduced "flow experience" as users' perception or emotion, attributes of OGDG as stimuli, and response as continuous use intention^[6], but they did not test the reliability and validity of the model. It is ubiquitous that users are with low cognition and utilization of open government data, Gao and Mo integrated TAM and UTAUT to find that perceived trust, performance expectation and convenience have significant impacts on the initial adoption intention of OGDG^[7], but they did not consider of data quality and interactivity, which may influence users' perception.

The goal of this research is to extend previous studies to a better understanding of user's response to open government data portal. We develop a research model that integrates IS success model, technology acceptance model (TAM) and stimulus-organism-response (SOR) paradigm, and then try to find the factors that would influence users' most important response "use intention", and then increase the usage of OGDG, ultimately release the value of government data.

2. RESEARCH BACKGROUND AND HYPOTHESES

In 1992, DeLone and McLean presented Information Systems (IS) Success Model as a framework and model for measuring the complex dependent variable in IS research. The model focuses attention on the information and system quality of specific IT systems. After ten years, they added "service quality" and used "net benefits" to the updated model for measuring IS success. In addition, information quality measures that have been used in e-commerce studies include accuracy, relevance, understandability, completeness, currency, dynamism, personalization, and variety. While data quality is a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and whether it's up to date. In big data era, since OGDG is kind of IS, this paper takes data quality as construct instead of information quality to the research model.

Technology acceptance model (TAM) was proposed by Davis^[8], he defined perceived usefulness (PU) as "the degree of which a person believes that using a particular system would enhance his or her job performance", and perceived ease of use (PEOU) as "the degree of which a person believes that using a particular system would be free of effort." Lots of research about e-commerce or e-government use TAM to do empirical study. For example, with regard to the government website of Shanghai, perceived usefulness and perceived ease of use have significant impact on the citizens' use behavior^[9]. Lan took TAM and ISC to study the impact of context-aware services on the continuous use of mobile government^[10]. It seems that perceived usefulness and perceived ease of use are suitable to study use intention of OGDG as well.

According to the Stimulus-Organism-Response (SOR) framework, environmental or situational stimuli(S) affect internal organism (e.g., cognition and emotion)^[11]. Stimuli may manifest in different representations, that affect the internal states (O) of people, then results in their behavioral responses (R). Since SOR is usually applied to investigate the relationship between environmental stimuli and response of users, perceived usefulness and perceived ease of use in TAM are two important factors affecting users' intention or behavior. In the context of open government data, the public are stimulated to visit OGDG in order to access government data for daily use or application development. Scholars confirmed that system quality directly affects user satisfaction and determines users' intention to continuous use^[12]. System quality means the clarity of page layout, ease of data retrieval, diverse function and stability of OGDG, and it will be stimuli to users. While data quality refers to data integrity, reliability, timeliness and clear classification, it also affects users' experience of data access from OGDG.

Interactivity means that users can receive feedback from OGDG in time, make comments about the data,

share data with others, and use the government data for application development. Trentin et al. found that customers would continue to use the website when their opinions are concerned^[13]. Therefore, this paper takes interactivity as kind of stimuli.

OGDP provides public services for citizens. Government should initiate policies or sponsor some activities to encourage citizens to use the portal. Therefore, policy environment also stands for a stimulus variable.

Internal state (O) is the process of cognition and emotion after stimulation. Xu based on TAM to study the factors affecting users' use intention of government applications, and found that perceived ease of use and perceived trust of government applications affect perceived usefulness^[14]. According to TAM and IS success model, Zhu et al. found that perceived usefulness, perceived ease of use and information quality positively affected continuous use intention of government website^[15]. Perceived usefulness means the government data may satisfy the users' demand, and perceived ease of use means the operation of OGDP is simple, users can search and download data in a suitable format freely. Therefore, this paper takes perceived usefulness and perceived ease of use as the "internal state" variables.

Response is use intention or behavior after users experiencing the stimulus and their internal state take some changes. So we takes use intention as the variable of "response" in the context of OGDP.

Scholars have done some research in information technology field. Zheng et al. based on the IS success model to find that information and system quality directly affect perceived individual benefits and user satisfaction, which ultimately determine continuous intention to consume and to provide information^[12]. Since data is kind of information, perceived usefulness and ease of use may stand for perceived benefits as well, this paper proposes the following hypotheses:

H1: System quality of OGDP is positively related to perceived usefulness.

H2: System quality of OGDP is positively related to perceived ease of use.

H3: Data quality of OGDP is positively related to perceived usefulness.

H4: Data quality of OGDP is positively related to perceived ease of use.

As for interactivity and policy environment, Yang proved that interactivity has positive impact on perceived usefulness and perceived ease of use of the virtual community^[16]. Zhang used TAM to study the influence of interactivity on the student's continuous learning intention. The result showed that interactivity positively affects perceived usefulness and perceived fun^[17]. In addition, Wang and Lo argued that the government and the public may influence government agencies throughout OGD adoption and implementation^[18]. Similarly, government may put forward some policies to encourage the public to use OGDP. Then the following hypotheses are proposed.

H5: Interactivity of OGDP is positively related to perceived usefulness.

H6: Interactivity of OGDP is positively related to perceived ease of use.

H7: Policy environment is positively related to perceived usefulness of OGDP.

H8: Policy environment is positively related to perceived ease of use of OGDP.

Based on TAM, the research on use intention of the government application (APP) shows that the public's perceived ease of use and trust of the government APP will affect perceived usefulness, while perceived usefulness and perceived ease of use will positively affect the public's use attitude of government applications^[14]. Eroglu et al. explored how environmental factors influence consumer's attitude with SOR, thus affects shopping outcome of online store^[19]. In the context of OGDP, the following hypotheses are proposed.

H9: Perceived ease of use is positively related to perceived usefulness of OGDP.

H10: Perceived usefulness of OGDP is positively related to use intention.

H11: Perceived ease of use of OGDP is positively related to use intention.

The research model is shown as Figure 1.

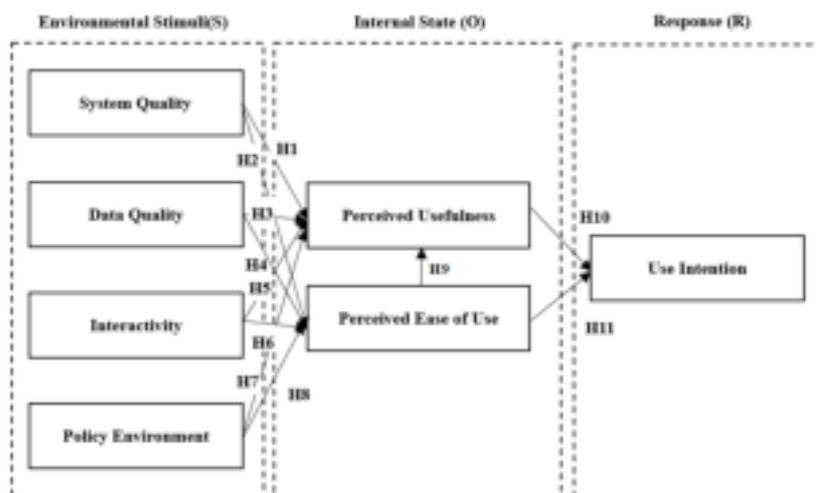


Figure 1. Research model

3. METHODOLOGY

3.1 Scale development

The questionnaire consists of two parts. The first is designed to investigate the demographic features of the respondents. The second part consists of 27 items to measure variables. The items were measured on Likert seven-point scale that ranges from 1 (strongly disagree) to 7 (strongly agree).

Table 1. Constructs and measurement items

Constructs	Item	Measurement	References
System Quality (SQ)	SQ1	Data engine of OGDP offers perfect function and is easy to use.	[6][10][15]
	SQ2	OGDP has clear page layout.	
	SQ3	OGDP is available in multiple formats for downloading data.	
	SQ4	Service provided by OGDP is completed.	
	SQ5	OGDP has smooth system access.	
Data Quality (DQ)	DQ1	OGDP has clear data classification.	[6][15]
	DQ2	OGDP updates data in time.	
	DQ3	The data of OGDP is reliable.	
	DQ4	The data of OGDP is completed.	
Interactivity (IA)	IA1	OGDP provides various forms of interaction.	[6][7]
	IA2	OGDP responds quickly to users' suggestion and comment.	
	IA3	OGDP provides instant data sharing through social media.	
	IA4	It is easy to use OGDP for application development.	
Policy Environment (PE)	PE1	The government ensures the standard and institution of OGDP by policy.	[20]
	PE2	The government encourages the public to participate in the innovative application of OGDP.	
	PE3	Government makes the public pay attention to OGDP.	
Perceived Usefulness (PU)	PU1	OGDP can effectively match my needs.	[6][10]
	PU2	OGDP enables me to learn the latest data released by the government.	
	PU3	OGDP can help me to solve problems and save my time and efforts.	

	PU4	OGDP can help to improve my efficiency.	
Perceived Ease of Use (PEOU)	PEOU1	It is easy to find OGDP.	[14] [21]
	PEOU2	It is easy to learn to use OGDP.	
	PEOU3	It is easy to follow the instructions provided by OGDP.	
	PEOU4	It is easy to find the data that I need by OGDP.	
Use Intention (UI)	UI1	I will try to use OGDP.	[15] [21]
	UI2	I will continue to use OGDP.	
	UI3	I will recommend OGDP to my friends.	

3.2 Descriptive statistical analysis

We distributed 115 questionnaires and received 93 valid responses from January to March 2019. Respondents were asked to visit some OGDP in China (e.g. <http://data.beijing.gov.cn/>, <http://data.sh.gov.cn/>) before they filled in the questionnaire. As shown in Table 2, more than half of the respondents used OGDP for the first time, 30.11% of them used occasionally, and only 16.13% of them used frequently. It is obvious that the public has low awareness of OGDP. Among the respondents, 51.61% were male and 48.39% were female, 55.91% of them were between 18 and 29 years old. As to job occupation, 33.33% were college students, 23.66% were company employees, and 18.28% were government staff. It seems that the respondents are relatively young and the job occupation distributed evenly, the sample is representative for OGDP users.

Table 2. The Demographics of Respondents (N=93)

Profiles	Options	Frequency	Percentage
Gender	Male	48	51.61%
	Female	45	48.39%
Age	18 or Younger	4	4.30%
	18-29	52	55.91%
	30-39	9	9.68%
	40-49	14	15.05%
	50 or Older	14	15.05%
Job Occupation	Professional(teachers/doctors/lawyers)	11	11.83%
	Company employees	22	23.66%
	Government staff	17	18.28%
	College students	31	33.33%
	Others	12	12.90%
Experience of OGDP	First time to Use	50	53.76%
	Frequently Use	15	16.13%
	occasionally Use	28	30.11%

4. DATA ANALYSIS

We employed SmartPLS, a type of partial least squares (PLS) technique for component-based structural equation modeling (SEM) to test the research model. Because PLS is suited for exploratory theory building and just requires the sample size to be at least 10 times the largest number of paths aiming at an endogenous construct^[22]. In our research model, the maximum number of paths that directing at an endogenous variable is 5. Therefore, the sample size of 93 was sufficient for the use of SmartPLS.

4.1 Measurement model

Reliability refers to the consistency or stability of the items, as shown in Table 3. This study employed Cronbach's α and composite reliability (CR) to reflect reliability. Most of Cronbach's α is greater than 0.7 except use intention, but the value 0.698 is very close to 0.7. In addition, CR of each construct is greater than 0.8, the reliability is acceptable. The AVE value of each variable is greater than 0.5, hence convergent validity is established. In addition, the VIFs of all items range from 1.213 to 2.529, far below the suggested threshold value 5. Therefore, there is no multi-collinearity in our study.

Table 3. Construct reliability and convergent validity

Constructs	Items	Loadings	Cronbach's α	CR	AVE	VIF
System Quality	SQ1	0.770	0.778	0.849	0.530	1.699
	SQ2	0.724				1.412
	SQ3	0.713				1.419
	SQ4	0.675				1.385
	SQ5	0.753				1.639
Data Quality	DQ1	0.851	0.820	0.881	0.650	2.019
	DQ2	0.788				1.611
	DQ3	0.737				1.552
	DQ4	0.844				2.147
Interactivity	IA1	0.794	0.773	0.854	0.595	1.600
	IA2	0.737				1.389
	IA3	0.802				1.567
	IA4	0.752				1.489
Policy Environment	PE1	0.883	0.843	0.905	0.761	2.245
	PE2	0.897				2.529
	PE3	0.836				1.715
Perceived Usefulness	PU1	0.799	0.777	0.857	0.600	1.763
	PU2	0.745				1.465
	PU3	0.734				1.355
	PU4	0.818				1.872
Perceived Ease of Use	PEOU 1	0.779	0.831	0.888	0.664	1.587
	PEOU 2	0.855				2.060
	PEOU 3	0.836				1.894
	PEOU 4	0.787				1.776
Use Intention	UI1	0.697	0.698	0.833	0.627	1.213
	UI2	0.824				1.529
	UI3	0.846				1.646

According to the method proposed by Fornell and Larcker^[23], if the Pearson correlation coefficient between each variable is smaller than the square root of the AVE on the diagonal, the discriminant validity is acceptable. As shown in Table 4, the discriminant validity meets the requirement.

Table 4. Results of discriminant validity

Constructs	DQ	IA	PE	PEOU	PU	SQ	UI
DQ	0.806						
IA	0.728	0.771					
PE	0.742	0.718	0.873				
PEOU	0.668	0.777	0.564	0.815			
PU	0.640	0.704	0.612	0.710	0.775		
SQ	0.788	0.736	0.630	0.685	0.709	0.728	
UI	0.660	0.646	0.552	0.692	0.689	0.698	0.792

4.2 Structural model

The relationships in the research model are examined by SmartPLS. After bootstrapping 5000 times, as shown in Table 5, 7 of 11 hypotheses, H2, H3, H4, H5, H7, H8 and H9 are not supported by significant at 0.05 level. It seems that the model needs some modification.

Table 5. Hypotheses testing results of research model

Hypothesis	Relationships	Path coefficient	P value	Results
H1	SQ→PU	0.333	0.013	Supported
H2	SQ→PEOU	0.170	0.192	Not Supported
H3	DQ→PU	-0.061	0.656	Not Supported
H4	DQ→PEOU	0.195	0.189	Not Supported
H5	IA→PU	0.127	0.510	Not Supported
H6	IA→PEOU	0.615	0.000	Supported
H7	PE→PU	0.175	0.176	Not Supported
H8	PE→PEOU	-0.040	0.712	Not Supported
H9	PEOU→PU	0.324	0.051	Not Supported
H10	PU→UI	0.397	0.002	Supported
H11	PEOU→UI	0.411	0.001	Supported

4.3 Model modification

After reviewing literatures again, we find some scholars believe that system quality includes data quality, and others think that data quality affects system quality, so we try to set data quality as an antecedent variable of system quality. Interactivity reflects users' communication with government and feedback of use experience of OGDP. Policy may increase the public's awareness of open government data, and then promote their interaction with government and OGDP. Therefore, policy environment may be an antecedent variable of interactivity. So we delete H3, H4, H7 and H8, afterwards develop two new hypotheses.

H3': Data quality of OGDP is positively related to system quality.

H7': Policy environment is positively related to interactivity of OGDP.

The modified research model is shown as Figure 2.

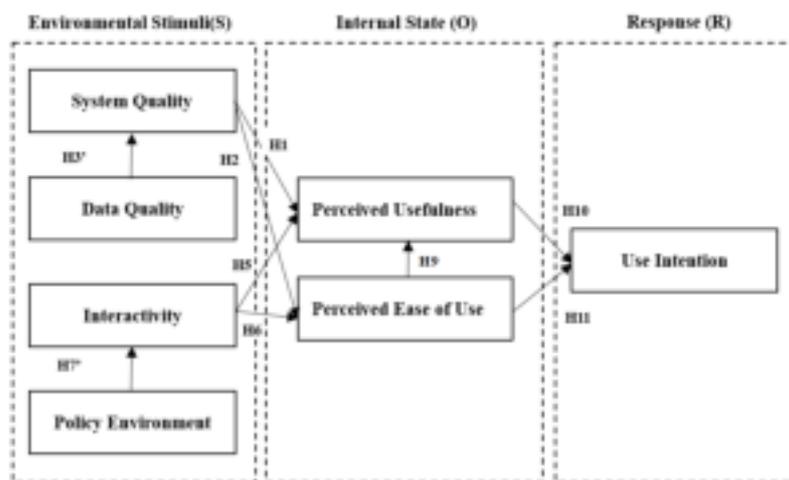


Figure 2. Modified research mode

4.4 Hypotheses testing results of the modified model

The hypotheses testing results of the modified model are shown in Table 6. Only H5 and H9 are not supported by significant at 0.05 level. It seems that the modified model is more suitable than the original one.

Table 6. Hypotheses testing results of the modified model

Hypothesis	Relationships	Path coefficient	P value	Results
H1	SQ → PU	0.339	0.016	Supported
H2	SQ → PEOU	0.247	0.036	Supported
H3'	DQ → SQ	0.788	0.000	Supported
H5	IA → PU	0.210	0.271	Not Supported
H6	IA → PEOU	0.595	0.000	Supported
H7'	PE → IA	0.718	0.000	Supported
H9	PEOU → PU	0.314	0.053	Not Supported
H10	PU → UI	0.397	0.002	Supported
H11	PEOU → UI	0.411	0.001	Supported

4.5 Mediating effect test

In the modified model, there may be some mediating effect between the variables. By comparing the direct effects of data quality and policy environment on use intention (Table 7) and the integrated indirect effects (Table 8) from bootstrapping results of SmartPLS, we find that most of the indirect effects of independent variables on use intention are not significant, indicating only the mediating role of interactivity and perceived ease of use on the relationship between policy environment and use intention exists.

Table 7. Direct effect test of variables on use intention

	Original	Sample(O)	Sample Mean(M)	T value	P value	Test result
DQ → UI		0.211	0.224	2.620	0.009	Significant
PE → UI		0.288	0.284	3.729	0.000	Significant

Table 8. Mediating effect test of variables

	Original Sample(O)	Sample Mean(M)	T value	P value	Test result
DQ→SQ→PU →UI	0.106	0.115	1.528	0.127	Not Significant
DQ→SQ→PEOU →UI	0.080	0.088	1.608	0.108	Not Significant
DQ→SQ→PEOU →PU→UI	0.024	0.022	1.517	0.129	Not Significant
PE→IA→PU →UI	0.060	0.058	0.987	0.324	Not Significant
PE→IA→PEOU →UI	0.175	0.177	2.691	0.007	Significant
PE→IA→PEOU →PU →UI	0.053	0.049	1.644	0.101	Not Significant

5.CONCLUSIONS AND DISCUSSIONS

This study investigated the issue of user's response to open government data portal and revealed that both perceived usefulness and perceived ease of use have positive effects on use intention, which means users are willing to use OGDG when they believe that it is convenient and will bring benefits. System quality has positive impact on perceived usefulness and perceived ease of use, which confirms that OGDG should provide easy and stable access to attract users. So government should increase investment in technology to enhance system reliability and make it convenient to operate. Data quality has significant impact on system quality, and then affects perceived usefulness and perceived ease of use. It seems that ensuring the reliability, timeliness and completeness of data is helpful to improve users' positive perception.

On the other hand, policy environment has positive effect on interactivity, and interactivity influences perceived ease of use, but perceived ease of use does not positively affect perceived usefulness. That is to say, users may prefer to use OGDG if the government offer more channels to interact with them and pay more attention to their feedback.

Finally, we test the mediating effect of the modified research model. Data quality(DQ) and policy environment(PE) have significant direct effect on users' use intention(UI). While there are respectively three path from DQ to UI and from PE to UI, the indirect effect only exists in the path of PE→IA→PEOU →UI, indicating the mediating role of interactivity and perceived ease of use on the relationship between policy environment and use intention.

5.1 Theoretical and Practical Implications

This study has both theoretical and practical implications for understanding users' response to OGDG. From a theoretical perspective, this study enhances current understanding of IS success model and TAM by focusing on OGDG. Considering the situation of open government data, it applies IS success model by taking data quality instead of information quality. More specifically, this study shows data quality of OGDG has significant impact on system quality, and then affects users' perceived usefulness and perceived ease of use. Our findings also explain the mediating role of interactivity and perceived ease of use on the relationship between policy environment and use intention. From a practical perspective, the study results show that both system quality and interactivity are important for improving perceived usefulness and perceived ease of use, and helps OGDG retain users. That means the government should focus on increasing data quality and system quality through continuous system upgrades. Moreover, it is a good strategy for the government to encourage users to visit and interact with the system frequently, and create value with the government data.

5.2 Limitations and Future Directions

This study still has some limitations due to the sampling methods and measurements. First, we distributed the questionnaires online. Respondents were asked to visit some OGDG in China before they filled in the questionnaire. There is no evidence that the sample is representative of the whole population of OGDG users. Future studies should investigate and compare different samples to increase external validity. Second, as

individual survey research, the sample size of 93 is limited. In order to make more sense of our findings, further research may validate the model by using a larger sample of more than 300 respondents. The last but not least, users of different age and job occupation may have different demand of data. In the future, we will discuss how to push accurate data to users by analysis of demographic information and access preferences, ultimately improve users' use intention of OGDP.

ACKNOWLEDGMENT

This research was supported by the Youth Project of the Humanities and Social Sciences of the Ministry of Education under Grant 18YJC630140, China; Social Science Fund of Fujian Province, China (FJ2020B023); Research climbing plan of Xiamen University of Technology, 2019(XPDKQ19005); Major Project of Industry-university-research Innovation Fund of Chinese Universities at New Generation Information Technology Innovation of China (2019ITA01011); Major Project of Social Science Research Base of Fujian Province, China (FJ2019JDZ051).

REFERENCES

- [1] Kalampokis E, Tambouris E, Tarabanis K.(2011). A Classification Scheme for Open Government Data: towards Linking Decentralised Data. *International Journal of Web Engineering and Technology*, 6(3): 266-285.
- [2] Janssen M, Charalabidis Y, Zuiderwijk A. Benefits.(2012).Adoption Barriers and Myths of Open Data and Open Government. *Information System Management*, 29: 258-268.
- [3] Dai L, Xu B Y.(2011). Research on public Continuous Use Behavior of Mobile E-government. *Journal of Information*, 30:186-195 (in Chinese).
- [4] Tang Z W, Han X, Wu S D.(2016). Analytical Framework of Public Use Intention on Government Websites: Based on the Perspective of Continuous Use. *China Administration*,4:27-33(in Chinese).
- [5] Wang H.(2006). Factors Affecting the Public's Willingness to Use Government Portals.Ms D Thesis. Harbin: Harbin Institute of Technology(in Chinese).
- [6] Zhu H C, Hu X, Wang X B.(2018). Research on the Continuous Use Intention of Open Government Data Portal Based on SOR Framework. *Modern Information*, 5:100-116(in Chinese).
- [7] Gao T P, Mo T L.(2018).An Empirical Study on the Initial Adoption Model of Government Data Open Platform. *E-Government*, 191(11):77-90(in Chinese).
- [8] Davis F D.(1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 13(3):319-340.
- [9] Cao P P, Zhao Y X, Xu Y X.(2008). Empirical Research on the Use Behavior of Government Websites Based on TAM. *Modern Library and Information Technology*, 2:76-81(in Chinese).
- [10] Lan Z H.(2011) Research on the Impact of Context-awareness Service on the Continuous Use of Mobile Government. Ms D Thesis. Shanghai: Shanghai Jiaotong University(in Chinese).
- [11] Mehrabian, A. and J.A. Russell(1974). *An Approach to Environmental Psychology*. MA: The MIT Press.
- [12] Zheng Y M, Zhao K, Stylianou A.(2013). The Impacts of Information Quality and System Quality on Users' Continuance Intention in Information-Exchange Virtual Communities: an Empirical Investigation. *Decision Support Systems*,56:513-524.
- [13]Trentin A, Perin E, Forza C.(2014).Increasing the Consumer-Perceived Benefits of a Mass-Customization Experience through Sales-Configurator Capabilities. *Computers in Industry*,65(4): 693-705.
- [14] Xu H Y.(2016). Research on the Influencing Factors of Public Interest in Government APP Based on TAM. Ms D Thesis.Wuhan: Central China Normal University(in Chinese).

- [15] Zhu D G.(2012).Research on the Continuous Use Behavior on Government Websites. *E-Government*, 12:114-121(in Chinese).
- [16] Yang Y.(2014). Research on the Relationship Between the Trust of Community Users and the Intention to Consume Online.Ms D Thesis. Wuhan: Central China Normal University,(in Chinese).
- [17]Zhang W.(2016). The Influence of Fun and Interactivity Perception on MOOCs Learners' Intention to Continuous Learning. *Journal of Education*, Renmin University of China, 2:122-138(in Chinese).
- [18] Wang H J, Lo J.(2016). Adoption of Open Government Data among Government Agencies. *Government Information Quarterly*,33:80-88.
- [19] Eroglu S A, Machleit K A, Davis L M.(2003).Empirical Testing of a Model of Online Store Atmospherics And Shopper Responses. *Psychology & Marketing*,20(2):139-150.
- [20] Zhao R Z.(2016). Research on China's Open Government Data Policy Environment from Multiple Perspectives. *E-Government*,6:97-104(in Chinese).
- [21] Zhou X Y.(2012). Empirical Research on Use Intention of Mobile E-Commerce Based on Technology Acceptance Model.Ms D Thesis. Guangzhou: South China University of Technology(in Chinese).
- [22] Chin W W.(1998). The Partial Least Squares Approach to Structural Equation Modeling. *Modern Methods for Business Research*, 295(2),295-336.
- [23] Fornell C,Larcker D F.(1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1):39-50.

Full Research Paper

Dual-channel Service Operations with the BOPS Option Considering Infection Risk Aversion Behavior

Xinyi Li¹, Yue Xu¹, Hui Li^{1*}, Xi Wang¹

¹School of Information, Central University of Finance and Economics, Beijing, 100000, China

Abstract: Nowadays many restaurants adopt a new retailing strategy of buying online and picking up in store (BOPS) to offer diversified service efficiently. During the epidemic period of coronavirus, customers behave with infection risk aversion, leading to the alteration in consuming demands for the three ways (pure online, pure offline and BOPS). Therefore, this paper establishes two models to investigate the problem on how restaurants allocate service contributions between online and offline channels. Firstly, we build utility equations to classify consumers with heterogeneous perceptions of infection risk. Considering the trait of infection risk aversion, we can obtain consuming demands of different channels and the aggregate revenues. Then, the optimal service levels of both online and offline can be calculated, as well as the optimal profits. We find that BOPS can help restaurants increase profits by adjusting the service level of both online and offline channels even in the case of consumer risk aversion. Finally, we draw out numerical experiments to verify our findings.

Keywords: dual-channel, BOPS, restaurant service levels, customer's infection risk aversion behavior

1. INTRODUCTION

Since its outbreak in December 2019, coronavirus has caused a severe blow to the catering industry. For fear of being infected with coronavirus, consumers eat out less frequently and turn to order online or takeout instead, generating a plummet in the revenue of the catering industry. Besides, restaurants are obligated to meet governments' anti-epidemic protests by adopting the measure of sitting at intervals to reduce the number of people gathering, exposing themselves to a worsening situation. Hence, it's not uncommon to witness more and more restaurants making efforts online to maximally compensate the losses brought by offline stores. They commonly take the action of strengthening take-out service, such as more delicate packaging and better appearance, boasting of having the same dining experience as in the store. This approach enhances consumers' stickiness but brings in more expense as well. In fact, a few restaurants raise the price instead because sometimes customers are intent to eat non-homemade food regardless of higher price, which obviously isn't feasible in the long run.

On top of the above measures, BOPS seems to be a better choice. BOPS, which literally means ordering goods online and then picking them up at a physical store, is a combination of online and offline service. For some customers, BOPS can reduce the risk of infection caused by in-store dining effectively, and is more affordable than take-out. In addition, restaurants can also benefit from this option. Studies show that people often have made a new purchase when they step into the store. For restaurants, they aren't supposed to throw away offline sales completely even in the critical epidemic period if they can ensure the consumers' physical safety or abide by epidemic prevention demands governments proposed strictly. However, once BOPS is implemented, among three purchasing means, which one do

* Corresponding author. Email: lihui@cufe.edu.cn (Hui Li), sherrylee1108@126.com (Xinyi Li), yuki.xu.0621@gmail.com (Yue Xu), xiwang@cufe.edu.cn (Xi Wang).

consumers averting infection risk choose and how should restaurants decide on online and offline service levels to maximize revenue when we are in epidemic period?

To answer these questions, we consider two models that contain main features needed, which are two restaurants who operate dual-channel and omnichannel respectively. Firstly, we analyze utility equations of two channels to classify heterogeneous customers with different choices. Based on that, we can calculate their optimal online and offline service levels and optimal profits further. For the channel with BOPS, two different cases are involved. Finally, some managerial propositions are drawn out for restaurants.

The rest part of this paper is organized as follows. Related literature reviews are given in Section 2. Section 3 describes the theoretical models. Section 4 contains model analysis in details about consumer classification. Then several numerical experiments are included in Section 5. Finally, Section 6 concludes the paper.

2. LITERATURE REVIEW

The paper is highly related with three streams of research: consumer behavioral modeling, omnichannel strategy management and service decisions.

2.1 Consumer behavioral modeling

Numerous literatures used consumer behavioral modeling on marketing studies. One of appealing and research-passionate branches is about the consumer choices on purchasing channels in retailing market. Usually, people choose purchasing online or offline first. Based on them, plenty of new cross-selling channels have appeared. Gu and Tayi^[1] examined one cross-channel strategy: pseudo-showrooming, where consumers inspect one product at a seller's physical store before buying a related but different product at the same sellers online store. They investigated optimal product placement strategies to achieve better coordination between retailers' online and offline channels. He, Xu and Zhao^[2] innovatively have added consumer's environmental awareness in behavioral model to explore the environmental impact reduction after BOPS is implemented.

2.2 Omnichannel strategy management

The second stream of studies are omnichannel strategy management. Currently, there exists two main genres in this field. One genre focuses on multichannel management in supply chain. These models involve the role of manufacturers, whose marketing demands are not simply affected by their traditional wholesales to retailers. In empirical researches, Chen, Kaya and Özer^[3] solved a manufacturer's problem of managing his direct online sales channel together with an independently owned bricks-and-mortar retail channel. By building a consumer channel choice model and conducting a sequence of controlled experiments, they identified optimal dual channel strategies. Tamer^[4] explored one manufacturer's channel inefficiencies induced by the presence of simultaneous vertical competition with double-marginalization and horizontal competition of substitutability similarly between an independent retailer and his wholly-owned channel. Additionally, Andy and Narendra^[5] added one more retailer into the model that one manufacturer supplies a common product to two independent retailers.

Other scholars focus on omnichannel choices on retailers. Among many cross-selling strategies, BOPS has grabbed most attentions, which describes the action of buying online and picking up in store for consumers. Santiago and Antonio^[6] used a data set and a stylized model, analyzing the impact of the implementation of BOPS. Like Gao and Su^[7], Liu and Xu^[8], they also investigated the impact of the BOPS initiative on store operations. Besides that, Gao and Su contributed greatly to other studies on omnichannel retailing. They did not only research the impact of online and offline self-order technologies on customer demand, employment levels and restaurant profits^[9], but also analyzed three information mechanisms: physical showrooms, virtual showrooms and availability information^[10].

Moreover, Bell, Gallino and Moreno^[11] took quasi-experimental methods to explore how offline showrooms benefit demand generation and operational efficiency for online-first and traditional retailers.

In contrast, we constructively introduce probability of available seats in restaurants into the models because the rule of sitting at intervals in epidemic period make vacancy much less probable. Sitting restrictions can avoid the risk of infection to a great extent. Whereas if no sitting availability, BOPS may be a much better choice.

2.3 Service decisions

There exists a substantial body of work concerning service decisions, which can be broadly classified into two streams. The first concentrates on marketing competitions about service levels as well as other influencing factors. In empirical investigation, scholars assumed two independent retailers competing for customers with service strategy, like Zhao and Atkins^[12], who focused on inventory service. They noticed that competing retailers have the service option of either agreeing in advance to transship excess inventory to each other or seeing unsatisfied customers switch to the competitor for a substitute. Some researchers conducted relevant studies on supply chain. Wu^[13] considered service competition in a closed-loop supply chain with remanufacturing. By a model with two manufacturers producing the new products and remanufactured products from used cores and a retailer, four competitive interactions are considered in terms of price and service. Raymond and James^[14] discussed the problem when consumer demand is stochastic. Other scholars explored service allocation purely. Tao, Gou and Zhang^[15] constructed first decision model which incorporates a seller's decision about its delivery service coverage within the framework of hoteling model. In the work of Jin, Li and Chen^[16], the optimal decisions on the product price and recommended service radius were derived for the retailer adopting BOPS.

In this paper, we grasp two influencing elements about consuming utilities, one being service level and the other, infection risk level. Our model is the first one that delineate consuming utilities by incorporating epidemic effects.

3. MODELS

We first consider such a scenario that a restaurant owns both online and offline channels. To research consumer behaviors, we assume that the choice of channel is up to themselves strategically. Struck by the rapid spread of coronavirus, they encounter varying degrees of infection perceived risk and the implementation of BOPS is likely to make a difference to consuming demands. Meanwhile we assume the order price remains constant at p while discrepancy exists between service levels in different channels. We assume all consumers have the same valuation v for an order and one consumer's utility in some channel is only related to dining service level s and risk level of infection r .

Following notations are used throughout the remainder of the paper:

Table 1. Notations

Notation	Definition
u_s, u_o and u_b	Consumer utilities of consuming online, offline and in BOPS channel in models respectively
v	Consumer valuation of the order
p	Price per order
s_s, s_o and s_b	Service levels of consuming online, offline and in BOPS channel in models respectively
r_s, r_o and r_b	Risk levels of consuming online, offline and in BOPS channel respectively
D_s, D_o and D_b	Consuming demands of consuming online, offline and BOPS channel in models respectively
φ	Probability of finding available seats in the restaurant

Notation	Definition
δ_1 and δ_2	Fraction of service costs of consuming online and offline in models respectively
α_1 and α_2	The fractions of BOPS service incorporate online and offline service respectively
k	Additional profit per order

3.1 Base model

Without BOPS put into effect, this restaurant sells an order by two channels, purely online or offline. Under the scenario of offline, consumers find an available seat for having deals in the probability φ and enjoy offline service level s_s while confronting face infection risk level r_s since exposed to other potentially infected individuals. Apropos of those who fail to find any vacancy, we assume they will leave this restaurant with utility being 0. Then we can obtain the equation:

$$u_s = \varphi(v - p + s_s - r_s) \quad (1)$$

If directly buying online, consumers will face infected risk level r_o since receiving the extraneous packaging bags possibly contaminated by virus and enjoy online service level s_o . Then we can obtain the equation:

$$u_o = v - p + s_o - r_o \quad (2)$$

To enable consumers to consider consuming in two channels, we assume both utilities are positive.

3.2 Model with BOPS

Adopting the strategy of BOPS, restaurants can operate through three channels: online, offline and BOPS. Through BOPS channel, people still need to get to the store and face offline risk. The avoidance of dine-in enables a risk level r_b smaller than that offline and the fraction is β . The BOPS service s_b consumers enjoy includes part of online service in fraction α_1 and part of offline service in fraction α_2 simultaneously. Then we can obtain the equation:

$$\begin{aligned} u_b &= v - p + s_b - r_b \\ r_b &= \beta r_s, \quad s_b = \alpha_1 s_o + \alpha_2 s_s \end{aligned} \quad (3)$$

If consumers can find a vacant seat in the restaurant, the utility is the same with that in the base model. But for those who fail to seat, they can turn to a new option, BOPS. We can obtain the equation:

$$u_s = \varphi(v - p + s_s - r_s) + (1 - \varphi)(v - p + s_b - r_b) \quad (4)$$

Concerning buying online, the utility equation is totally same as Equation (2).

4. ANALYSIS

We can use these above equations to depict some figures about utilities in different channels whose horizontal axis is the offline risk r_s and vertical axis is the online risk r_o . We assume they are both distributed uniformly with a range of from 0 to H , i.e., $r_s \sim U[0, H]$, $r_o \sim U[0, H]$. According to these utilities of relationship in size, we can classify several types of consumers in different cases. Next, we still illustrate the base model and then model with BOPS in order.

4.1 Base model

In dual-channel scenario, consumers have three choices, buying online or offline and leaving out directly. When $u_o > u_s$ and $u_o > 0$, customers will purchase online. Similarly, customers will purchase in offline channel if $u_s > u_o$ and $u_s > 0$. When $u_s < 0$ and $u_o < 0$, they will decide to leave out.

$r_{u_s=0}$ is the offline risk level derived from the Equation (1) and $r_{u_o=0}$ is the online risk level derived from the Equation (2). Given above class conditions, these lines $u_s = u_o$, $u_s = 0$ and $u_o = 0$ divide all consumers into three categories. We can demonstrate them as three parts in the Figure 1.

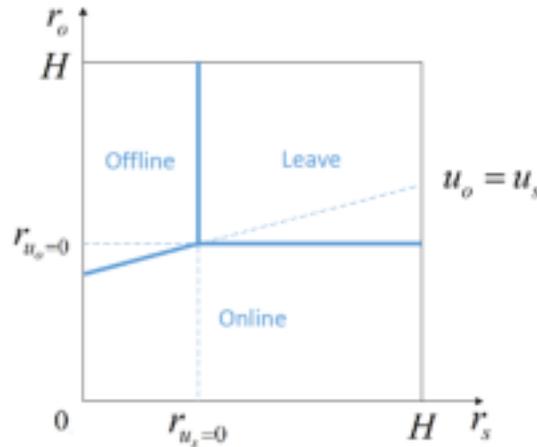


Figure 1. Consumer behavior in the dual-channel scenario

From this figure, the values of demands buying both online and offline, D_s and D_o , can be obtained.

$$D_s = \int_{r_s=0}^{r_s=r_{u_s=0}} \int_{u_o=u_s}^{r_o=H} dr_o dr_s$$

$$D_o = \int_{r_s=0}^{r_s=r_{u_s=0}} \int_{r_o=0}^{u_o=u_s} dr_o dr_s + \int_{r_s=r_{u_s=0}}^{r_s=H} \int_{r_o=0}^{r_o=r_{u_o=0}} dr_o dr_s$$

Then we assume consuming demands are identical with volumes of sales in different channels, so we can obtain the equations of restaurant’s aggregate profit and total volume of sales.

$$D_B = D_s + D_o \tag{5}$$

$$\Pi_B = p(D_o + D_s) - \frac{\delta_1}{2} s_o^2 - \frac{\delta_2}{2} s_s^2 + kD_s \tag{6}$$

Next, we can derive partial derivatives of Π_B about s_o and s_s from Equation (6).

Finally, we can calculate the optimal service level s_{oB}^*, s_{sB}^* . Then insert them into Equation (5) and (6) to obtain optimal demand D_B^* , optimal profit Π_B^* , which are shown in Proposition 1.

Proposition 1. *In the dual-channel scenario, the restaurant’s optimal service level s_{oB}^*, s_{sB}^* can be calculated from following equations.*

$$\left\{ \begin{aligned} s_{sB}^* &= \frac{\left(1 + \frac{k}{\delta_1} - \frac{p}{\delta_1}\right)(H - v + p)p + \left(1 - \frac{k}{\delta_1} + \frac{p}{\delta_1}\right)(v - p)k - kH}{\delta_2 - k\varphi - \frac{p^2}{\delta_1} + \frac{k^2}{\delta_1}} \\ s_{oB}^* &= \frac{p(H - v + p) - k(v - p)}{\delta_1} - \frac{p + k}{\delta_1} s_{sB}^* \\ s_{sB}^* &> 0, s_{oB}^* > 0 \end{aligned} \right.$$

The optimal demand D_B^* , optimal profit Π_B^* are as follows.

$$\begin{cases} D_B^* = D_{oB}^* + D_{sB}^* \\ \Pi_B^* = p(D_{oB}^* + D_{sB}^*) - \frac{\delta_1}{2}(s_{oB}^*)^2 - \frac{\delta_2}{2}(s_{sB}^*)^2 + kD_{sB}^* \end{cases}$$

4.2 Model with BOPS

In omnichannel scenario, consumers can be classified into four groups: buying online, offline or BOPS and leaving out directly. When $u_o > u_s$, $u_o > u_b$ and $u_o > 0$, customers will purchase online. Similarly, customers will purchase in offline channel if $u_s > u_o$, $u_s > u_b$ and $u_s > 0$ and consume by BOPS if $u_b > u_o$, $u_b > u_s$ and $u_b > 0$. When $u_s < 0$, $u_o < 0$ and $u_b < 0$, they will decide to leave out.

$r_{u_s=u_b}$ is the offline risk level derived from the equation $u_s = u_b$, while other two offline risks, $r_{u_b=0}$ and $r_{u_s=0}$, are from equation (3) and (4) respectively. We can't confirm the size of three values so there actually exist six cases: $r_{u_b=0} < r_{u_s=u_b} < r_{u_s=0}$, $r_{u_s=0} < r_{u_s=u_b} < r_{u_b=0}$, $r_{u_s=u_b} < r_{u_s=0} < r_{u_b=0}$, $r_{u_s=u_b} < r_{u_b=0} < r_{u_s=0}$, $r_{u_s=0} < r_{u_b=0} < r_{u_s=u_b}$, $r_{u_b=0} < r_{u_s=0} < r_{u_s=u_b}$. However, after analyzing, we can derive only two meaningful cases in practice. Case 1 is $r_{u_s=u_b} < r_{u_s=0} < r_{u_b=0}$ and Case 2 is $r_{u_b=0} < r_{u_s=0} < r_{u_s=u_b}$.

4.2.1 Case 1

In Case 1, if $\beta < \alpha_2$ and $\frac{s_o}{s_s} < \frac{1-\alpha_2}{\alpha_1}$, we can obtain $r_{u_s=u_b} < r_{u_s=0} < r_{u_b=0}$. Then according to above class conditions, these lines $u_s = u_o$, $u_b = u_o$, $u_s = u_b$, $u_s = 0$, $u_b = 0$ and $u_o = 0$ divide all consumers into four types. These four parts are demonstrated in the Figure 2.

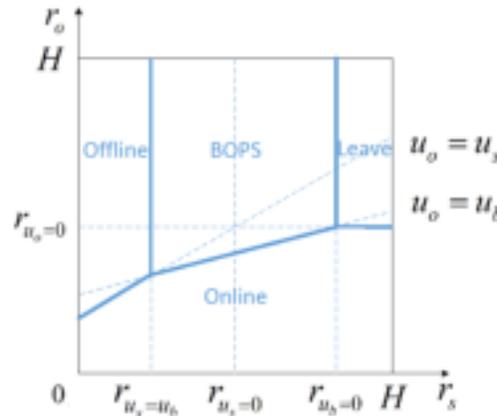


Figure 2. Consumer behavior in the omnichannel scenario in Case 1

From this figure, we can obtain the values of consuming demands in three channels, buying online or offline and by BOPS, which are D_s , D_o and D_b respectively.

$$\begin{aligned} D_s &= \int_{r_s=0}^{r_s=r_{u_s=u_b}} \int_{r_o=u_o=u_s}^{r_o=H} dr_o dr_s \\ D_o &= \int_{r_s=0}^{r_s=r_{u_s=u_b}} \int_{r_o=0}^{u_o=u_s} dr_o dr_s + \int_{r_s=r_{u_s=u_b}}^{r_s=r_{u_b=0}} \int_{r_o=0}^{u_o=u_b} dr_o dr_s + \int_{r_s=r_{u_b=0}}^{r_s=H} \int_{r_o=0}^{r_o=r_{u_o=0}} dr_o dr_s \\ D_b &= \int_{r_s=r_{u_s=u_b}}^{r_s=r_{u_b=0}} \int_{u_o=u_b}^{r_o=H} dr_o dr_s \end{aligned}$$

Like the base model, we still assume consuming demands are identical with volumes of sales in different channels.

So we can obtain the equations of restaurant’s aggregate profit and total volume of sales.

$$D_1 = D_s + D_o + D_b \tag{7}$$

$$\Pi_1 = p(D_o + D_s + D_b) - \frac{\delta_1}{2} s_o^2 - \frac{\delta_2}{2} s_s^2 + kD_s \tag{8}$$

Next, we can derive partial derivatives of Π_1 about s_o and s_s from Equation (8).

Finally, we can calculate the optimal service level s_{o1}^*, s_{s1}^* . Then insert them into Equation (7) and (8) to obtain optimal demand D_1^* , optimal profit Π_1^* , which are shown in Proposition 2.

Proposition 2. *In the dual-channel scenario, the restaurant’s optimal service level s_{o1}^*, s_{s1}^* can be calculated from following equations.*

$$\begin{cases} s_{s1}^* = \frac{\Lambda_5 \Lambda_3 - \Lambda_6}{\Lambda_4 - \frac{\Lambda_5 \Lambda_1}{\Lambda_2}} \\ s_{o1}^* = -\frac{\Lambda_1}{\Lambda_2} s_{s1}^* - \frac{\Lambda_3}{\Lambda_2} \\ s_{s1}^* > 0, s_{o1}^* > 0 \end{cases}$$

where

$$\begin{aligned} \Lambda_1 & p\left(-\frac{\alpha_2}{\beta}\right) + k \left[\frac{\beta(1 - \alpha_1 - \alpha_2) - \alpha_1(2\alpha_2 + 2\alpha_1 - 3) + \alpha_2 - 1}{2(1 - \beta)^2} \right. \\ & \quad \left. - \frac{1 - \alpha_1 - \alpha_2 + 2\varphi\alpha_1 + 2\alpha_1\alpha_2 - 2\varphi\alpha_1\alpha_2}{2(1 - \beta)} \right] \\ \Lambda_2 & p\left(-\frac{2\alpha_1}{\beta}\right) + k \left[\frac{\alpha_1(1 - \alpha_1 + \varphi\alpha_1)}{1 - \beta} - \frac{\alpha_1(1 - \alpha_1 - \beta)}{(1 - \beta)^2} \right] - \delta_1 \\ \Lambda_3 & p \left[H \left(\frac{\alpha_1}{\beta} + 1 \right) - \frac{(v - p)(1 + \alpha_1)}{\beta} \right] - k \frac{H\alpha_1}{1 - \beta} \\ \Lambda_4 & k \left[\frac{(1 - \alpha_2)(\varphi + \alpha_2 - \varphi\alpha_2)}{1 - \beta} - \frac{(\beta - \alpha_2)(1 - \alpha_2)}{(1 - \beta)^2} \right] - \delta_2 \\ \Lambda_5 & p\left(-\frac{\alpha_2}{\beta}\right) + k \left[\frac{(\beta - \alpha_2)\alpha_1 - (1 - \alpha_1 - \beta)(1 - \alpha_2)}{2(1 - \beta)^2} \right. \\ & \quad \left. - \frac{(1 - \alpha_2)(1 - \alpha_1 + \varphi\alpha_1) + \alpha_1(\varphi + \alpha_2 - \varphi\alpha_2)}{2(1 - \beta)} \right] \\ \Lambda_6 & p \frac{\alpha_2(H - v + p)}{\beta} + k \frac{H(1 - \alpha_2)}{1 - \beta} \end{aligned}$$

The optimal demand D_1^* , optimal profit Π_1^* are as follows.

$$\begin{cases} D_1^* = D_{o1}^* + D_{s1}^* + D_{b1}^* \\ \Pi_1^* = p(D_{o1}^* + D_{s1}^* + D_{b1}^*) - \frac{\delta_1}{2} (s_{o1}^*)^2 - \frac{\delta_2}{2} (s_{s1}^*)^2 + kD_{s1}^* \end{cases}$$

4.2.2 Case2

In Case2, if $\beta > \alpha_2$ and $\frac{s_o}{s_s} < \frac{1 - \alpha_2}{\alpha_1}$, we can obtain $r_{ub=0} < r_{us=0} < r_{us=ub}$. Next all analytical procedures are totally same with the case1. However, the only difference is we obtain three types of consumers: buying online or

offline and leaving out.

Then Figure 3 about consumer classification is as follows:

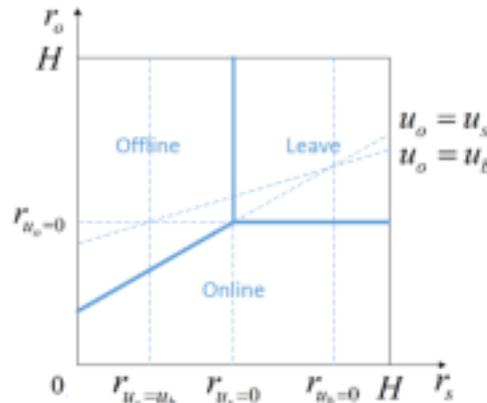


Figure 3. Consumer behavior in the omnichannel scenario in Case2

The following equations are the values of customer demands both online and offline, D_o and D_s .

$$D_s = \int_{r_s=0}^{r_s=r_{u_s=0}} \int_{u_o=u_s}^{r_o=H} dr_o dr_s$$

$$D_o = \int_{r_s=0}^{r_s=r_{u_s=0}} \int_{r_o=0}^{u_o=u_s} dr_o dr_s + \int_{r_s=r_{u_s=0}}^{r_s=H} \int_{r_o=0}^{r_o=r_{u_o=0}} dr_o dr_s$$

the equations of restaurant's aggregate profit and total volume of sales are presented next.

$$D_2 = D_s + D_o \quad (9)$$

$$\Pi_2 = p(D_o + D_s) - \frac{\delta_1}{2} s_o^2 - \frac{\delta_2}{2} s_s^2 + kD_s \quad (10)$$

Next, we can derive partial derivatives of Π_2 about s_o and s_s from Equation (10).

Finally, we can calculate the optimal service level s_o^* , s_s^* . Then insert them into Equation (9) and (10) to obtain optimal demand D_2^* , optimal profit Π_2^* , which are shown in Proposition 3.

Proposition 3. In the dual-channel scenario, the restaurant's optimal service level s_o^* , s_s^* can be calculated from following equations.

$$\begin{cases} s_{s_2}^* = \frac{\frac{\Lambda_{11}\Lambda_9}{\Lambda_8} - \Lambda_{12}}{\Lambda_{10} - \frac{\Lambda_{11}\Lambda_7}{\Lambda_8}} \\ s_{o_2}^* = -\frac{\Lambda_7}{\Lambda_8} s_{s_2}^* - \frac{\Lambda_9}{\Lambda_8} \\ s_{s_2}^* > 0, s_{o_2}^* > 0 \end{cases}$$

Where

$$\Lambda_7 = \frac{(\alpha_2 + \varphi - \varphi\alpha_2)(\alpha_1 k - \varphi\alpha_1 k - p)}{\beta - \beta\varphi + \varphi}$$

$$\Lambda_8 = \frac{\alpha_1^2(1-\varphi)^2 k - 2\alpha_1(1-\varphi)p}{\beta - \beta\varphi + \varphi} - \delta_1$$

$$\Lambda_9 = p \left[\frac{(1-\varphi)\alpha_1 H - (v-p)(\alpha_1 - \alpha_1\varphi + 1)}{\beta - \beta\varphi + \varphi} + H \right] + k \frac{\alpha_1(1-\varphi)(H + 2v - 2p)}{2(\beta - \beta\varphi + \varphi)}$$

$$\begin{aligned}
 \Lambda_{10} &= k \frac{(\varphi + \alpha_2 - \varphi\alpha_2)^2}{\beta - \beta\varphi + \varphi} - \delta_2 \\
 \Lambda_{11} &= \frac{(\varphi + \alpha_2 - \varphi\alpha_2)(\alpha_1 k - \varphi\alpha_1 k - p)}{\beta - \beta\varphi + \varphi} \\
 \Lambda_{12} &= p \frac{(\varphi + \alpha_2 - \varphi\alpha_2)(H - v + p)}{\beta - \beta\varphi + \varphi} + k \frac{(\varphi + \alpha_2 - \varphi\alpha_2)(H + 2v - 2p)}{2(\beta - \beta\varphi + \varphi)}
 \end{aligned}$$

The optimal demand D_2^* , optimal profit Π_2^* are as follows.

$$\begin{cases} D_2^* = D_{o_2}^* + D_{s_2}^* \\ \Pi_2^* = p(D_{o_2}^* + D_{s_2}^*) - \frac{\delta_1}{2}(s_{o_2}^*)^2 - \frac{\delta_2}{2}(s_{s_2}^*)^2 + kD_{s_2}^* \end{cases}$$

4.3 Discussion

From above analysis, we can find out in Case1, customers have four purchasing behaviors: ordering online, eating in restaurant, buying in BOPS and leaving, while in Base Model and Case2, customers can only choose from consuming online and offline or just leaving without BOPS option.

By comparing Base Model with Case1, absolutely after implementing BOPS, the consuming demand in online and offline channels have decreased for consuming in BOPS instead and some people choosing leaving in Base Model also transferred to BOPS channel. Thanks to the BOPS strategy, the holistic consuming proportion of the restaurant has risen obviously, which is mostly closed to the managerial facts as well.

In particular, the channel-choosing situations of Base Model and Case2 are the same even if in Case2 the restaurant has put BOPS into effect. In other words, though the restaurant has provided BOPS option, no guests choose it. But it still influences customers' options potentially in that more people leave the restaurant. The shifts of demand of consuming online and offline can be identified depending on some variables' values by further numeric experiments.

5. NUMERIC EXPERIMENTS

Given the two models are too complicated, we use some specific numbers to imitate them to prove the above all propositions. We divide numeric experiments into two parts. The first part is for the contrary between dual channel and case 1. Then in the second part, case1 and case2 get compared. all figures' horizontal axis is φ and vertical axes are optimal online service level s_o^* , offline service level s_s^* and profit Π^* respectively.

5.1 With and without BOPS under case1

when $v = 1$, $p = 0.5$, $H = 1$, $\delta_1 = 1$ and $\delta_2 = 1$, $k = 0.05$, $\alpha_1 = 0.5$, $\alpha_2 = 0.5$ and $\beta = 0.4$ we can draw following sub-figures (a), (b) and (c) in Figure 4.

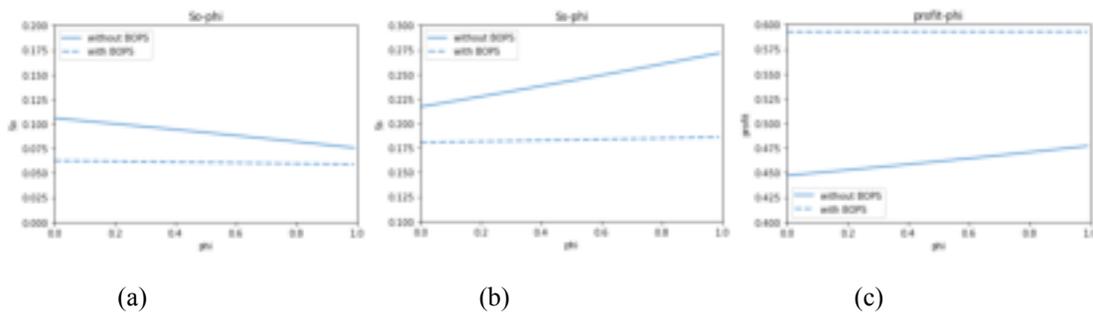


Figure 4. the values of s_o , s_s and Π in dual-channel and omnichannel

From them, we can find that under these conditions, s_o^* and s_s^* in dual-channel are both bigger than that in omnichannel at every value of φ ranging from 0 to 1, while Π^* is smaller. And as the value of φ increases gradually, s_o^* will reduce but s_s^* and Π^* will rise in dual-channel. In terms of omnichannel, s_o^* and Π^* fall down very slightly and s_s^* increase subtly. Hence, in channel management with BOPS, the values of s_o^* , s_s^* get lower but Π^* turns higher.

We can also find out in dual-channel, as the probability of available seats is increasing, the optimal offline service level will go up simultaneously so that the restaurants can reach the optimal revenue and the optimal revenue rises as well. On the contrary, the online service level changes conversely. In omnichannel, service levels and revenue experience subtle changes. The revenue always keeps relatively stable even with more vacancy in restaurants.

In closing, BOPS strategy boosts restaurant's revenue dramatically. Under certain conditions, BOPS strategy is an excellent implementation for restaurants in relatively both lower optimal online and offline service levels as a result of much higher optimal profits.

5.2 With and without BOPS under case2

when $v = 1$, $p = 0.5$, $H = 1$, $\delta_1 = 1$ and $\delta_2 = 1$, $k = 0.08$, $\alpha_1 = 0.2$, $\alpha_2 = 0.5$ and $\beta = 0.6$, we can draw following sub-figures (a), (b) and (c) in Figure 5.

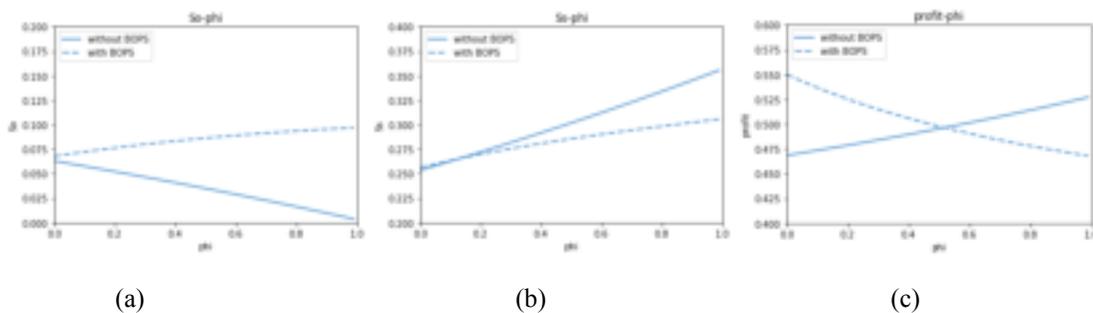


Figure 5. the values of s_o , s_s and Π in dual-channel and omnichannel

Accordingly, we can find that under these conditions, s_o^* in dual-channel is smaller than that in omnichannel at every value of φ ranging from 0 to 1. In terms of s_s^* , when φ is smaller than around 0.15, it in omnichannel is little more than that in dual-channel, while it falls behind the dual-channel strategy much more after φ is bigger than 0.15. The values of Π^* are similar with s_s^* , but the intersection of curve with BOPS and without BOPS is at approximately 0.5. Interestingly, the bigger the value of φ is, the bigger the value of s_s^* and Π^* is and the smaller s_o^* is in dual-channel. In addition, as the value of φ increases gradually, s_o^* and s_s^* will rise slowly but Π^* drop dramatically in BOPS scenario. Hence, after BOPS is implemented, the values of s_o^* , s_s^* and Π^* are all higher.

In Case 2 with BOPS, although the restaurant has provided BOPS option, no consumers choose it. Despite of the inappearance of BOPS option, it is still a brilliant inspiration for consuming ratio and optimal profits in restaurants especially when there are less available seats at higher optimal offline and online service levels.

6. CONCLUSIONS

In this paper, our whole research is based on the epidemic social background. Given this exogeneous condition,

BOPS as a type of omnichannel strategy plays an important role on consumers and retailers, decreasing consumers' risk of infection and sparking substantially higher revenue for restaurants. Therefore, we establish two models on respective scenarios to illustrate consumer purchasing behavior. After deriving the optimal values of service levels both online and offline, consuming demands and profits, we make comparisons between different scenarios in two cases by a series of numerical experiments. Eventually, we can draw out some enlightened conclusions in respect of channel management and omnichannel strategy. We cannot solely derive the optimal service levels and optimal profits in different scenarios by three cases, but also analyze significance of BOPS strategy implemented in epidemic background. In some situations, BOPS plays an important role in boosting restaurants' revenue but are not chosen apparently under certain conditions.

Despite of above research, some extensions about the models can be included in the future work. At the beginning, there is an assumption about offline utility in dual channel that if there are no available seats in the restaurant, the customers will leave it and their utility is 0. In fact, it's very likely to cause a negative utility. Upon the decision to leave the restaurant, the customer will probably turn to a new one or simply return home. Whatever the destination is, the utility is sure to be negative because of hassle costs consumed on transportation without enjoying deals in present restaurant and the amount of utility is hard to estimate. Given complexity, we just simply assume them as 0 in our models. Second, in realistic lives, judgements about risk levels are closely inseparable from governments' attitude toward current epidemic circumstances. In this paper, these influencing factors are ignored. Finally, in numerical experiments, we fixed all other variables' valuations except probability of finding available seats in the restaurant. For further research, changing different values about different variables can explore more managerial propositions. In Case 2 of model with BOPS, it puzzles that nobody chooses BOPS but profits are still much higher. All of these problems can keep digging deeply.

REFERENCES

- [1] Gu Z, Tayi K. (2016). Consumer Pseudo-Showrooming and Omni-Channel Product Placement Strategies.
- [2] He Y, Xu Q, Zhao D. (2020). Impacts of the BOPS Option on Sustainable Retailing. *Sustainability*, 12(20):8600.
- [3] Chen K, Kaya M, Özer Ö. (2016). Dual Sales Channel Management with Service Competition. *Manufacturing & Service Operations Management*, 10(4): 654-675.
- [4] Boyaci T. (2005). Competitive stocking and coordination in a multiple-channel distribution system, *IIE Transactions*, 37(5): 407-427.
- [5] Tsay A, Agrawal N. (2016). Channel Dynamics Under Price and Service Competition. *Manufacturing & Service Operations Management*, 2(4): 372-391.
- [6] Gallino S, Moreno A. (2014). Integration of Online and Offline Channels in Retail: The Impact of Sharing Reliable Inventory Availability Information. *Management Science*, 60(6):1434-1451.
- [7] Gao F, Su X. (2016). Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science*, 63(8): 2478-2492.
- [8] Liu J, Xu Q. (2020). Joint Decision on Pricing and Ordering for Omnichannel BOPS Retailers: Considering Online Returns. *Sustainability*, 12(4): 1539.
- [9] Gao F, Su X. (2017). Omnichannel service operations with online and offline Self-order Technologies. *Management Science*, 47, 1668-1680.
- [10] Gao F, Su X. (2016). Online and Offline Information for Omnichannel Retailing. *Manufacturing & Service Operations Management*, 19(1): 84–98.

- [11] Bell R, Gallino S, Moreno A. (2017). Offline Showrooms in Omnichannel Retail: Demand and Operational Benefits. *Management Science*, 64(4): 1629-1651.
- [12] Zhao X, Atkins D. (2009) Transshipment between competing retailers. *IIE Transactions*, 41(8): 665-676.
- [13] Wu G. (2012). Price and service competition between new and remanufactured products in a two-echelon supply chain. *international journal of production economics*, 140: 496-507.
- [14] Deneckere R, Peck J. (2016). Competition Over Price and Service Rate When Demand is Stochastic: A Strategic Analysis. *The RAND Journal of Economics*, 26(1): 148-162
- [15] Tao Z, Gou Q, Zhang J. (2018). A local seller's app channel strategy concerning delivery. *International Journal of Production Research*, 58(1):220-255
- [16] Jin M, Li G, Cheng T. C. E. (2018). Buy online and pick up in-store: Design of the service area. *European Journal of Operational Research*, 268(2): 613-623

Institutional Entrepreneurship and Acquiring Legitimacy of Social Commerce Platform

JingYang^{1}, Ruoyu Sun², Zeguo Yang³, Qianxi Zhang⁴*

¹School of Business Administration, Nanjing University of Finance and Economics, NanJing, 210023, China

² School of Business Administration, Nanjing University of Finance and Economics, NanJing, 210023, China

³ School of Management, Hefei University of Technology, Hefei, 230009, China

⁴ School of Business Administration, Nanjing University of Finance and Economics, NanJing, 210023, China

Abstract: Research on social commerce has ignored the process and mechanism of institutional entrepreneurship. Based on the institutional theory, we use qualitative research methods to study the process of institutional entrepreneurship of social commerce. This paper also analyzes the legitimacy acquisition process of social commerce from the perspective of institutional contradiction and institutional logic. We found that, firstly, institutional contradictions existing in traditional e-commerce organization field are the fundamental motivation for institutional entrepreneurship of social commerce platform. Secondly, social commerce entrepreneur proposed new institutional logics which are the solutions according to the institutional contradictions existing in traditional e-commerce organization field. Thirdly, because of the new institutional logics proposed by institutional entrepreneur, social commerce platform acquired cognitive legitimacy and normative legitimacy. Finally, the factors of organizational field influence the whole process of institutional entrepreneurship of social commerce.

Keywords: social commerce, institutional entrepreneurship, institutional contradiction, institutional logic

1. INTRODUCTION

Research shows that people's online behaviors are becoming more willing to actively acquire and share information with each other. Within this context, social commerce is developed and formed based on social media, websites, platforms. Compared with traditional commerce, users of social commerce are both buyers and presenters. As a new business model, social commerce is favored by entrepreneurs.

According to the new institutional theory, formation of the new ventures or organizational field is mainly reflected in achieving legitimacy^[1]. As a new organizational field developed from traditional commerce, the formation process of social commerce is also a process of achieving field acquire legitimacy through institutional entrepreneurship. The present research studies social commerce mainly from the evolution of social commerce and the development of driving factors. Wang & Zhang (2012) analyzed the development of social commerce from four dimensions of people, management, technology and information^[2]. However, these studies have not explored the internal mechanism of the formation of social commerce as a new organizational field and the process of its legalization from the perspective of institutional theory.

Accordingly, this study aims to answer two related questions:

Q1. How to propose new institutional logic in the process of institutional entrepreneurship of social commerce?

Q2. How can social commerce platforms gain legitimacy through new institutional logic?

This paper takes the institutional entrepreneurial process of social commerce as the research object. We also build the structural model of the institutional entrepreneurial process of social commerce.

* Corresponding author. Email: yangler713@hotmail.com(Jing Yang)

2. LITERATURE REVIEW

Social commerce is a form of business based on the social media, which integrates the online and offline environments. It provides various applications including forums and communities, ratings and reviews and referrals and recommendations which are all known as social commerce components^[3]. In such an environment, users can obtain social knowledge and experience and better understand their online purchasing intention, so as to make a more intelligent and accurate purchasing decision^[4]. Social commerce generally refers to e-commerce activities and transactions completed through social media environment, most of which are completed by using Web2.0 software on social networks^[5]. Therefore, social commerce can be regarded as a subset of e-commerce, but it includes the use of social media to assist in completing e-commerce transactions and behaviors, supporting by social interaction and user-generated content (UGC)^{[6][7]}. In essence, it is a combination of business and social behaviors.

In recent years, the domestic e-commerce industry has developed rapidly, more and more offline goods brands are being sold online, and a large number of agent-operated enterprises have achieved rapid development along with the transformation of offline enterprises^[8]. It also creates a number of star enterprises, such as Pinduoduo, Mogujie etc. Social commerce has become a market segment that cannot be ignored by e-commerce. But the existing research has mainly investigated the evolution, business model, platform design, user purchase intention and behavior of social commerce. However, the internal mechanism of the formation of social commerce platform, that is, how it conducts institutional entrepreneurship, propose new institutional logic and acquiring legitimacy, remains unexplored.

Dimaggio (1988) proposed the concept of institutional entrepreneurship which refers to the process in which actors under the institutional framework mobilize resources to create new institutions or transform existing institutions in pursuit of their own interests^[9]. The Seo-Creed model attributes the causes of institutional entrepreneurship to the category of institutional contradiction, specifically including four types of inefficiency, nonadaptability, interinstitutional incompatibility and misaligned interests^[10]. The Seo-Creed model was proposed by Seo and Creed in 2002. They identified four contradictions: gaps between the levels of performance arising from conformity to existing institutional prescriptions and from alternative opportunities in the marketplace (the “efficiency contradiction”); inability of a field to adapt to exogenous jolts because of “locked-in” patterns of behavior and thought (the “nonadaptability contradiction”); inconsistencies between values deeply held yet mutually inconsistent (the “interinstitutional incompatibility” contradiction); and divergence of the interests of those privileged and those disadvantaged by existing logics (the “misaligned interests contradiction”)^[11]. In traditional commerce, there are also several institutional contradictions mentioned above, such as the low efficiency contradiction. We proposed these institutional contradictions are the fundamental motivation for social commerce to initiate institutional entrepreneurship and establish new institutional logic.

Institutional logic is the stable existence of institutionalized norms and corresponding action mechanism in the organizational field^[12]. It is a system composed of a variety of institutional orders such as culture, values, beliefs and rules at the social level, which shapes the behavior of organizations and their members^[13]. The proposal of new institutional logic can help the social commerce gain legitimacy, but the new institutional logic should be compatible with the old institutional logic to some extent, so as to reduce the resistance to change and promote the realization of institutional change^[14]. Therefore, from the perspective of institutional logic, this paper can explain how social commerce acquires legitimacy.

Legitimacy is also a core concept of the new institutional theory. Suchman’s definition about legitimacy is widely accepted, legitimacy is a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions. Scott

(1995) classified legitimacy into three dimensions: cognitive, regulative, and normative [15]. The normative legitimacy comes from social norms and values, professional standards, customs and traditions etc., which are some codes of conduct and norms that organizations have to follow. Cognitive legitimacy, that is, the public's understanding and recognition of something depends on the cultural consensus and symbols that constitute the social reality and understanding framework [16]. We proposed that the key to the success of social commerce institutional entrepreneurship is to acquiring legitimacy, which are cognitive legitimacy and normative legitimacy.

3. RESEARCH METHODS

3.1 Source of data

Our purpose is to analyze the institutional entrepreneurial process of the social commerce. We used a single, exploratory case study, following a common research method for building theory. We began by collecting data in the form of secondary accounts to ensure convergence on events. From these, key primary sources were identified and a series of texts assembled for analysis (34 texts were selected). In the process of sample selection, this study adopts the method of theoretical sampling, that is, the selected sample data is for the purpose of constructing theories, rather than selecting samples from the perspective of statistical sampling. Table 1 shows the sources of text used in the coding process of this study.

Table 1. Source of information

Source of information	The number of documents
Social commerce platform website (Pinduoduo.com, Xiaohongshu.com, Mogu.com and so on)	17
News website (Ifeng.com, Sina.com, Yicai.com, 21st Century Business and so on. "social commerce" was used as the main keyword search, browsing in chronological order from 2011 to 2019)	16

3.2 Coding process

3.2.1. Open coding

Open coding refers to the information recorded on conceptualization and category, which is a large number of data records would be reduced step by step according to certain principles. It is the process of integrating the concept that is extracted from the data and aims to identify phenomenon and define the concepts. See Table 2 and Table 3 for some examples of this way.

Table 2. Labeling Process for Sample Data (Example)

Sample data	Definitions of phenomena
In recent years, with the advent of IP economy, web celebrity economy, and live streaming outlets, social e-commerce encourages consumers to purchase by utilizing social relations and personal influence. In 2017, Pinduoduo, a social e-commerce provider, quietly slipped into the top five, ranking the fifth with a 2.5% market share. Social e-commerce has become a rookie in the retail market.	a1 Knowledge economy a2 Web economy a3 Live outlets a4 Social facilitation of purchase

Table 3. Examples of Open Coding

Category	Concept	Definitions of phenomena
Credit logic	Trust endorsement	a23 Trust endorsement builds trust a20 We-Media build trust
	Deciding according to UCG	a5 Customers buy goods based on reviews

3.2.2 Axial coding

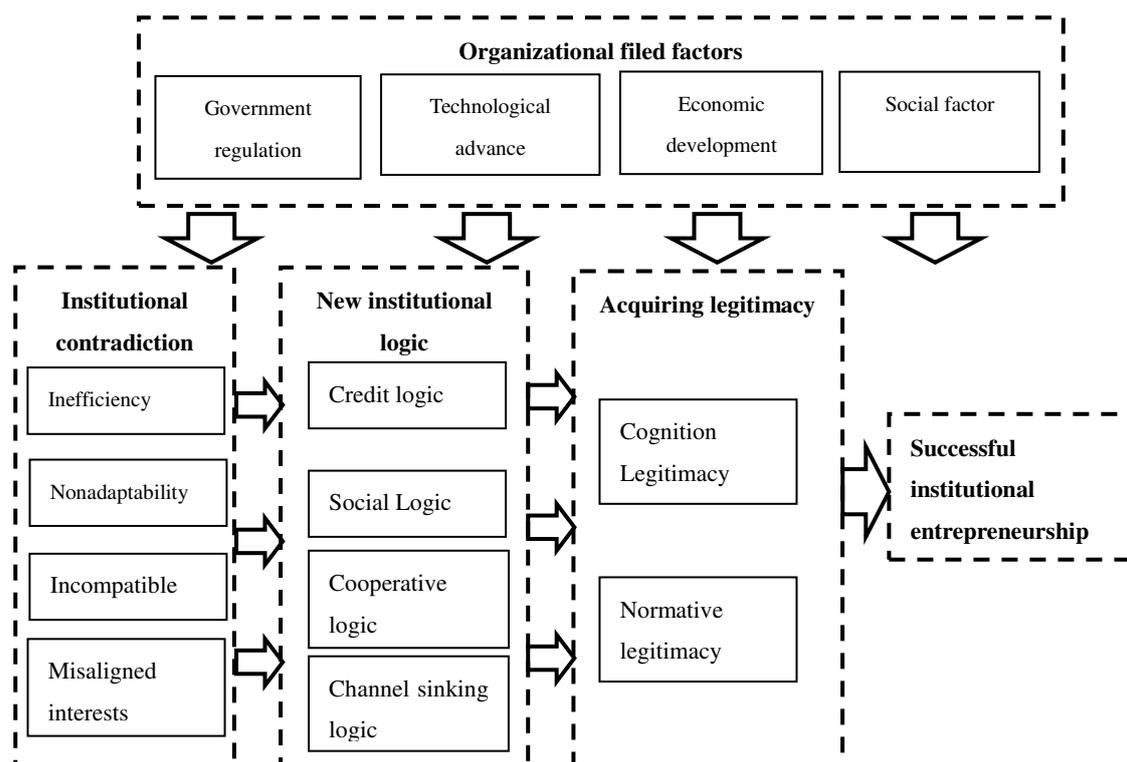


Figure1. The model of institutional entrepreneurship of social commerce platform

Based on open coding shows, we firstly get four institutional contradictions of traditional e-commerce, which are inefficiency, nonadaptability, interinstitutional incompatibility and misaligned interests. Secondly, we get four new institutional logic of social commerce, which are credit logic, cooperative logic, social logic and channel sinking logic. Thirdly, institutional entrepreneurs use these four new institutional logics to get cognitive legitimacy and normative legitimacy. Finally, institutional entrepreneurship successes. Government regulation, technical progress, economic development and social factors belong to organizational filed factors. Based on the "contradiction--praxis--institutionalization" of the Seo-Creed model, the axial coding is to establish the relationship between different categories to build a model for the process of institutional entrepreneurship of social commerce (Figure 1).

4 FINDINGS

4.1 Institutional contradictions

Based on qualitative data, we found four categories institutional contradictions of traditional e-commerce are consistent with the Seo-Creed model: inefficiency, nonadaptability, interinstitutional incompatibility and misaligned interests^[10].

Inefficiency. Because of the development of e-commerce in China, beside of JD and Alibaba, many small and medium-sized platforms also rush to fight for the limited market of e-commerce market. In order to obtain consumers, platforms bombard consumers with information, such as news, blogs, emails, text messages, phone calls, etc. These phenomena appear false marketing strategies. Consumers are deeply troubled by these marketing strategies, resulting in distrust of platform. These eventually lead to the collapse of a large number of small and medium-sized enterprises and inefficiency of e-commerce market.

Nonadaptability. Traditional e-commerce platform ignored the adaptation to the changing environment in

order to obtain short-term benefits and short-term adaptability. Firstly, in order to acquire customers and resources, e-commerce platforms preferred to rapidly expand and occupy the market, but ignoring the finance. These bring inventory pressure, capital chain rupture and other problems. Secondly, some e-commerce platforms make use of illegal behaviors to make profits to adapt to the short-term environment. As a resolute, government's punishment makes them suffer serious losses in capital and public trust, leading to the weakening of their long-term adaptability.

Interinstitutional incompatibility. Although the e-commerce industry institution can make it harmonized internally, problems such as difficult consumer rights protection and poor consumer shopping experience lead to the interinstitutional incompatibility between the e-commerce industry institution and consumer perception. The government's punishment of e-commerce industry also shows the interinstitutional incompatibility between the e-commerce industry institution and government regulations.

Misaligned interests. In order to acquire customers, resources and market, e-commerce platforms engaged in a price war. In this condition, it is difficult to meet the requirements of smaller platform, thus leading to profit mismatch. At the same time, the emergence of e-commerce oligopoly after the price war has raised the barriers of e-commerce industry, and it is difficult for new entrants to enter this field with the original business model.

Based on the above contradictions, e-commerce platforms start institutional entrepreneurship in order to find a way out. Social commerce as a new derivative model of e-commerce arises at the historic moment. Through the communication channels of social networking sites, SNS, Weibo, social media and network media, purchase and sales of goods can be by the way of social interaction, user-generated content and other means.

Proposition 1: Institutional contradictions existing in traditional e-commerce organization field are the fundamental motivation for institutional entrepreneurship of social commerce platform.

4.2 Institutional logic

Through the new institutional logic, social commerce entrepreneur solves the conflicts in the original contradictions, reconstructs the organizational field and acquires legitimacy. Therefore, in this study, the process of the new institutional logic is also the process of institutional entrepreneurship of social commerce. The open coding results show that the new institutional logic is divided into credit logic, social logic, Cooperative logic and channel sinking logic.

Credit logic. The credit logic comes from both the endorsement of trust established by key opinion leaders and the judgment of consumers on the generation of UCG. Trust deviation, false marketing and other factors make traditional e-commerce appear low efficiency contradiction. Social commerce entrepreneur re-establishes the trust between consumers and e-commerce platforms with the help of trust endorsement and UCG. Social commerce platforms can use more resources to improve functional efficiency. Therefore, credit logic solves the low efficiency contradiction of traditional e-commerce platform.

Social logic. Social logic is generated against the background of Internet decentralization based on the development of internet software and hardware and the change of consumer behavior. Through social activities based on the social commerce platform, users can read the advertorial and pictures of key opinion leaders, obtain the information shared by friends and UCG to make purchase decisions. Thus, social process also improves the online shopping experience of users. The emergence of new marketing methods based on social logic makes e-commerce platforms avoid low efficiency contradictions caused by false marketing and trust deviation.

Cooperative logic. Cooperative logic is mainly based on the cooperation between social platforms and traditional e-commerce platforms. According to grounded theory analysis, the cooperation of "social+ e-commerce" is divided into four kinds of circumstances. ①The cooperation between traditional e-commerce and pure social platforms, such as the cooperation between Weibo and Alibaba. ②The start of e-commerce on

pure social platform, such as WeChat e-commerce.③The self-built social platforms of pure e-commerce, such as Suning's development of "Yunxin" social platform.③The new entrants of social, such as Pinduoduo and Mogujie. The social platforms enable e-commerce platforms to obtain low cost and high stickiness of user, which greatly reduces the cost of e-commerce, and to some extent solves the interinstitutional incompatibility contradiction.

Channel sinking logic. The channel sinking logic aims to develop the market of small-town. The self-communication of users under social commerce can promote this process. The most typical example is that the industry unicorn Pinduoduo carried out viral marketing with the help of WeChat platform quickly entered into small-town. The development of small-town market makes some e-commerce platforms focus on lower-tier market to obtain long-term profit points. Therefore, the logic of channel sinking solves the conflict of misaligned interests caused by e-commerce oligopoly.

Proposition 2: Social commerce entrepreneur proposed new institutional logics which are the solutions according to the institutional contradictions existing in traditional e-commerce organization field.

4.3 Acquisition of legitimacy

According to the Seo-Creed model and the qualitative data, the result of institutional entrepreneurship practice is institutional change, or the acquisition of legitimacy.

Cognitive legitimacy comes from the public's understanding and awareness. In the institution of traditional e-commerce, factors such as false marketing and trust deviation undermine the public's perceived legitimacy of e-commerce. The social commerce model reconstructs the trust system of merchants and consumers through UCG and trust endorsement, making the social commerce platform gradually recognized by the public. In addition, based on the sharing of key opinion leaders and the self-dissemination of high-quality user experience, more and more consumers make reasonable purchase decisions. More and more consumers are more willing to use the social logic under the model of social commerce. From this perspective, the new institutional logic proposed by institutional entrepreneurship enables social commerce to gain cognitive legitimacy.

In the field of traditional e-commerce, the way that enterprises obtain customers led to a series of intra-industry conflicts within the industry. These indicates that the legitimacy of norms has been destroyed. The cooperative logic between social commerce platforms and other platforms reduces the conflicts in the e-commerce industry. Therefore, the logic of cooperation contributes to the legitimacy of the norms of social commerce. The existence of misaligned interests makes the e-commerce oligopoly formulate the industry norms and aggravate the conflict. Channel sinking logic uses the way of exploiting the small-town market. These ease the conflicts between industries, and solve the conflict of misaligned interests of traditional e-commerce. Therefore, the sinking of social commerce channels enables it to obtain normative legitimacy.

Proposition 3: Because of the new institutional logics proposed by institutional entrepreneur, social commerce platform acquired cognitive legitimacy and normative legitimacy.

4.4 Organizational field

From the open coding results, the organizational field factors in the process of institutional entrepreneurial of social commerce can be summarized into four dimensions: government regulation, technical progress, economic development and social factors. In the institutional entrepreneurship of social commerce, organizational field factors are constantly changing, and also always affect the whole process of institutional entrepreneurship.

Government regulation. From the perspective of the government, organizational field factors are mainly reflected in the regulation of the e-commerce industry by the government and the constraints of laws in the e-commerce industry. For example, the "Price war" in 2013 ended when the government punished several large e-commerce companies. Government regulation affects the whole process of institutional entrepreneurship of

social commerce.

Technical progress. Technical progress is reflected in internet finance, the development of software platform and the optimization of supply chain optimization. The emergence of internet finance represented by Alipay and Tenpay has facilitated online payment and ensured the development of social commerce. The social platform represented by WeChat and Weibo promotes user communication. When combined with e-commerce, it promotes the diversified marketing methods, and makes it more convenient for users to acquire and spread information.

Economic development. Besides the development of national economy, factors of economic development also include the disappearance of e-commerce dividend and supply-side structural reform. From the perspective of B2C market share, more than 80% of the market share is occupied by Tmall and JD. Suning have carved up most of the remaining market, and the e-commerce dividend has disappeared. The qualitative data shows that service-oriented consumption has a strong growth momentum. Service-oriented and high-quality consumption encourages consumers to obtain purchase information through multiple channels, which also stimulates the development of social commerce.

Social factors. Social factors include generational change, consumer psychology and consumer online shopping habits. The generations after 80s and 90s have become the major consumers and have the ability to improve their consumption level. They are also the largest internet participation group. Group psychology, curiosity psychology and increasingly consolidated online shopping habits urge consumers to obtain commodity information through social networking, thus becoming a key factor for the institutional entrepreneurship of social commerce.

Proposition 4: The factors of organizational field influence the whole process of institutional entrepreneurship of social commerce.

5. CONCLUSIONS

Based on grounded theory, this study constructed the institutional entrepreneurship process model of the social commerce platform. We found the following conclusions. Firstly, Institutional contradictions existing in traditional e-commerce organization field are the fundamental motivation for institutional entrepreneurship of social commerce platform. Institutional contradictions existing in traditional e-commerce organization field including: inefficiency, nonadaptability, interinstitutional incompatibility and misaligned interests. Secondly, Social commerce entrepreneur proposed new institutional logics which are the solutions according to the institutional contradictions existing in traditional e-commerce organization field. New institutional logics proposed by social commerce entrepreneur including: credit logic, social logic, Cooperative logic and channel sinking logic. Thirdly, because of the new institutional logics proposed by institutional entrepreneur, social commerce platform acquired cognitive legitimacy and normative legitimacy. Finally, the factors of organizational field influence the whole process of institutional entrepreneurship of social commerce. The factors of organizational field including government regulation, technical progress, economic development and social factors.

Traditional e-commerce platforms deviate from consumers' trust due to their wrong marketing methods, and adopt vicious price wars in order to obtain traffic, which leads to the deterioration of the industry environment and the overall inefficiency. In the process of institutional entrepreneurship of social commerce, social activities, UCG and trust endorsement of key opinion leaders make consumers re-establish solid trust in the e-commerce platform, which is also the key to the success of social commerce. As well as the function of social media, social commerce can gain market at a low cost and quickly and successfully enter small town. Therefore, e-commerce platforms should focus on the combination with social platforms in order to obtain

resources more quickly. At the same time, social commerce platforms should pay attention to the social atmosphere and UCG. When the social atmosphere and UCG are unfavorable to the development of the platform, measures should be taken as soon as possible to prevent the negative factors of the communication function of social platforms from losing more consumers' trust.

REFERENCES

- [1] Überbacher, F. (2014). Legitimation of New Ventures: A Review and Research Programme. *Journal of Management Studies*, 51(4): 667-698.
- [2] Wang C, Zhang P. (2012). The Evolution of Social Commerce: The People, Management, Technology, and Information Dimensions. *Communications of the Association for Information Systems*, 31: 105-127.
- [3] Hajli, Nick. (2015). Social Commerce Constructs and Consumer's Intention to Buy. *International Journal of Information Management*, 35(2):183-191.
- [4] Huang Z, Benyoucef M. (2015).references of Social Features on Social Commerce Websites: An Empirical Study. *Technological Forecasting and Social Change*, 95:57-72.
- [5] Dennison G, Bourdage-Braun S, Chetuparambil M. (2009).Social Commerce Defined. Research Triange Park, NC: IBM Systems Technology Group.
- [6] Caverlee J, Liu L, Webb S. (2010).The Social Trust Framework for Trusted Social Information Management: Architecture and Algorithms. *Information Sciences*, 180(1):95—112.
- [7] Fraser M, Dutta S. (2008).Throwing Sheep in the Boardroom: How Online Social Networking Will Transform Your Life. *Work and World*.
- [8] Li Y. (2019). Market Scale Prediction of e-commerce Agent Operation Based on Buffer Operator and GM (1,1) Model. *Journal of Lanzhou University of Arts and ence(Natural ence Edition)*.
- [9] Dimaggio P. (1988). Interest and Agency in Institutional Theory. *Institutional Patterns and Organizations: Culture and Environment*.
- [10] Seo M G, Creed W E D. (2002). "Institutional Contradictions, Praxis, and Institutional Change: A Dialectical Perspective." *Academy of Management Review* 27(2): 222-247.
- [11] GREENWOOD R, Roy S. (2006). Institutional Entrepreneurship in Mature Fields: The Big Five Accounting Firms. *Academy of Management Journal*, 49(1), 27-48.
- [12] Zhou Xueguang, Ai Yun. (2010).Multiple Logics in Institutional Change: Toward an Analytical Framework. *Social Sciences in China*,(4): 132 —150.
- [13] Thornton P, Ocasio W, Lounsbury M. (2012). The Institutional Logics Perspective: A New Approach to Culture, Structure and Process. *Leadership and Organization Development Journal*, 15(6):583.
- [14] Misangyi V F, Weaver G R, Elms H. (2008). Ending Corruption: The Interplay among Institutional Logics, Resources, and Institutional Entrepreneurs. *Academy of Management Review*, 33(3): 750-770.
- [15] Suchman M C. (1995).Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, 20(3):571-610.
- [16] Scott W R. (1995).Institutions and Organizations. Thousand Oaks: Sage Publication, 178.

Interactive Effects of Order Assignment Strategy and Information Disclosure on Courier's Continuance Participation Intention Arising in the On-Demand Logistics

Quanwu Zhao¹, Jiamin Yuan¹, Yuqing Liu^{2*}

1. College of Economics and Business Administration, Chongqing University, Chongqing, 400000, China

2. College of Mechanical Engineering, Chongqing University, Chongqing, 400000, China

Abstract: The application of crowdsourcing in logistics is a revolutionary change of urban distribution. The on-demand logistics is one of the typical crowdsourcing logistics modes which is popular in China since 2008, the demand volume of Chinese on-demand logistics was about 18.5 billion in 2019 with annual rate of growth about 37.6%, and the average annual growth rate in recent five years was 73.7%, the market scale reached 131.3 billion yuan with annual rate of growth about 33.8%, and due to the impact of Covid-19, the demand of on-demand logistics is increasing significantly, and the requirement for convenience and personalized services is changing sharply, the accelerating iteration of on-demand logistics patterns causes a downward trend of delivery income that it's difficult for the platform to balance the operation cost and labor motivation from the perspective of revenue. The prior research in crowdsourcing logistics focuses on the influence of continuance participation intention from the perspectives of external and internal incentive. However, incentive researches are unsuitable for current background of the downwards trend of delivery income completely. And the characteristics of instability, unsafety and unfairness of couriers occupation which were reported by social news not only the reasons of high turnover rate and mismatching of supply-demand, but also point to the problem of management mechanisms such as incentive mechanism, order mechanism and information mechanism.

As our knowledge, platform mechanism influence couriers' daily tasks immediately but the related researches are absent in crowdsourcing logistics, especially considering the order receiving and completing which is the first concern of courier, order mechanism and information mechanism will influence the order receiving and completing directly. Order mechanism means platform assigns orders to courier, it's the guarantee of matching supply and demand (couriers and orders), as well as the full coverage of the orders. There are different order assignment strategies, from the perspective of the human resources, evaluating performance and setting grades is common strategy, grades is the principles for orders assignment priority to encourage employees with good service quality to stay. In order to guarantee the timeless, some platforms will assign orders from perspective of space. This research summarized the two kinds of order assignment strategies as performance-priority strategy and distance-priority strategy. In performance-priority order assignment strategy, the courier will get an evaluation grade and differentiated orders assignment priority according to their total number of completed orders, total mileage, punctuality rate, praise rate and other indicators. The distance-priority order assignment strategy sends orders to the neighboring courier according to the distance from the orders and the stores, and give priority to the closer courier.

Information mechanism provides platform labor with assistance, which focuses on the information content of user interface that disclosure by platform. The quantity and accuracy of the information have significant influence on users' intention and behavior, and the richness of information content as an important dimension of information characteristics affects the information receivers' behavior immediately. Therefore, according to the operation information in Chinese on-demand logistics platform, this research summarized the information content and conclude two levels of information disclosure in on-demand logistics: one is overall layout information, the number of orders and courier, location layout and real-time traffic information; the other one is individual specific information, order details (distribution revenue goods category, merchant location and time requirements of each order) and transportation capacity details (the level and ranking of

* Corresponding author. Email: yolanda_617@outlook.com (Yuqing Liu)

each courier on the platform). Each information disclosure includes three aspects: order, capacity and transportation. Detailed information disclosure includes two levels of information: overall layout and individual specific information. Brief information disclosure only includes overall layout information.

To abstract differences in orders and information mechanisms specifically, and control other mechanisms (such as incentive mechanism) keep consistent better to highlights the effects of the research object, this research designed experiments based on justice theory for Chinese couriers, considered different orders assignment strategy (Performance-priority vs. distance-priority) and information disclosure (detailed information vs. brief information), introduced information justice and distribution justice perception as mediation to investigate the interaction between order assignment strategy and information disclosure, analyzed the moderating path of work experience and proactive personality. With a sample of 340 part-time and full-time couriers, experiment 1 results showed that the interaction between order assignment strategy and information disclosure in on-demand logistics significantly affects the continuance participation intention of couriers, especially under performance-priority order assignment strategy disclosure detailed information(vs. brief information), but under distance-priority order assignment strategy there were no difference in continuance participation intention between detailed and brief information disclosure which showed the ceiling effect of information disclosure, brief information already contains the basis information of distance-priority order assignment strategy, more detailed information would not improve continuous participation intention significantly. Information justice and distributive justice played a significant mediating role, and work experience and proactive personality moderate the above interaction relationships. However, the results of experiment 2 with a sample of 180 part-time couriers indicated that the mediating effect of distributive justice and the moderating effect of proactive personality were both insignificant, which showed the purposefulness and proactiveness characteristics of part-time couriers, .

The research reveals the mechanism of justice perception on continuous participation intention and the process of information understanding and utilization in crowdsourcing logistics platform, provided specific management suggestions for the platform to attract delivery capacity by setting order assignment strategy and information disclosure when the distribution income is gradually stabilized. Future, more mechanism can be introduced with appropriate justice perception, and more group characteristics of crowdsourcing logistics participants can be considered.

Keywords: On-demand logistics, Order assignment strategy, Information disclosure, Justice perception, Continuous participation intention

Full Research Paper**The Influence of Customer Co-creation Experience Value on Brand****Equity in the context of Social Media***Yunjun Jiang¹, Hong Zhang^{1*}, Guiming Lian², Siqian Wang¹*¹College of Management, Wuhan University of Science and Technology, Wuhan, 430065, China²College of Automation, Guangdong University of Technology, Guangzhou, 510006, China

Abstract: In recent years, with the widespread application of mobile internet and the rapid development of social media, the interaction between customers and enterprises has become another important way of brand operation. The concept of brand equity based on customer-brand relationship has been boldly put forward and concerned by scholars and industry. Although there are many empirical studies on its pre-influencing factors in the field of brand equity research, the relevant research based on social media situation combined with customer co-creation experience is rare. This study takes Club of huawei, an online community of fan communication set up by Huawei, as the research object. Through empirical testing, this study explores the impact of customer co-creation experience value (spiritual value, enabling value, practical value, hedonistic value) on brand equity based on social media environment. Through the empirical analysis of 245 sample data, it is found that the spiritual value, enabling value and hedonistic value of customer co-creation experience value have a significant positive impact on brand equity based on customer-brand relationship. However, the practical value obtained by customers in the process of interaction in the social media environment has no significant impact on brand equity based on customer-brand relationship. The research provides practical inspiration for brand community builders on social media how to improve brand equity by catering to the needs of customers to create experience value.

Keywords: Customer co-creation experience value, Brand equity, Social Media

1. INTRODUCTION

With the emergence of service-oriented business model, the traditional model dominated by product value has been greatly challenged. In the service-oriented business model, consumers are no longer only regarded as users of products, but also participate in the whole process of providing products and services, and become important value creation partners of producers. With the rapid development of social media platform, value co-creating has attracted more and more attention in industries other than service-oriented enterprises. More and more enterprises also begin to establish brand communities on social media, build direct interaction with customers and work with customers to create the market value of the brand^[1]. For example, Haier Community, Xiaomi Community and Club of Huawei. Customers can keep abreast of the latest developments of the enterprise, exchange product experience and participate in the design and evaluation of new products in these brand social media sites. At the same time, enterprises can also get customer feedback in real time, make reasonable adjustments, and make more targeted marketing plans and management decisions. On the other hand, according to The Special Research report on China's New consumption in 2018, people's consumption concept is changing with the rapid economic development and experiencing a phased evolution from "quantity" to "quality" to "experience"^[2]. Customers expect to gain diversified experiences in the process of consumption, and the popularity of social media has led to new changes in customer experience. The high degree of interconnection of social media not only enables isolated customers to carry out real-time, multi-channel and personalized social interaction, but also makes social interaction become the core realization path of value

* Corresponding author. Email: savagegardenzh@163.com (Hong Zhang) , 1499896794@qq.com (Yunjun Jiang),

co-creation between customers and enterprises in the social media environment. Based on this situation, the experience value of customers in the process of customer-enterprise interaction, that is, customer co-creation experience value, has been endowed with a new connotation. By building interactive channels between customers and enterprises on social media, customer rights and dominance are enhanced. Therefore, customers begin to actively participate in the interaction of brand social media sites, so as to achieve co-creation value.

As one of the sources of economic benefits and the most valuable assets of enterprises, brand equity has been the focus of academia and other circles since the concept was put forward. Among them, the CPC Central Committee with Comrade Xi Jinping as the core shows that it attaches great importance to brand building, such as the establishment of "China Brand Day", the launch of the "Brand Power Plan for the New era", the establishment of brand power strategic alliances and other major actions. China's brand building work presents a new normal. Academia holds that the essence of brand operation lies not in products but in customers. As the relationship between customers and enterprises changes as a result of value co-creation activities, brand equity based on the perspective of customer-brand relationship has been put forward and concerned by scholars. Stepping into the era of experience economy, enterprises pay more and more attention to meeting the needs of customer experience. Meanwhile, people also pay more attention to obtaining experience value from the process of product and service co-creation and interaction. This makes the promotion of customer co-creation experience value to become the inevitable choice to build the competitive advantage of enterprise brand^[3]. Therefore, in the process of interactive co-creation, how to cater to the needs of customer co-creation experience to promote customer-brand relationship and enhance brand equity is an urgent problem for social media platforms and brand builders.

Although there are many empirical studies on the pre-influencing factors in the field of brand equity research, there are few related studies in the field of social media context and value co-creation. In other words, the "black box" of the influence mechanism of customer co-creation experience value on brand equity from the perspective of customer-brand relationship in the context of social media has not been opened. Therefore, this study focuses on the current situation of social media, attempts to refine the measurement dimension of customer co-creation experience value, and discusses its impact on brand equity from the perspective of customer-brand relationship. Data were collected through questionnaires and verified by structural equation model. The research enriches the empirical study of brand equity from the perspective of customer-brand relationship in the field of brand equity, and analyzes the new connotation of customer co-creation experience value combined with social media situation. In the meantime, it also provides practical inspiration for brand community builders on social media how to improve brand equity by catering to the needs of customers to create experience value.

2. THEORETICAL BASIS AND RESEARCH MODEL

2.1 Customer co-creation of experience value

Value co-creation theories based on customer experience, service-leading logic and service ecosystem all emphasize that customer experience value is the intrinsic driving force of value co-creation. Research on the value of customer co-creation experience originates from the value co-creation theory based on customer experience proposed by Prahalad and Ramaswamy^[4]. These two scholars believe that co-creation of consumption experience is the core of customer and enterprise co-creation value. On the other hand, the service-oriented logic proposed by Vargo and Lusch^[5] also emphasizes that the value co-creation between customers and enterprises is based on the interactive experience of the two. From the perspective of service ecosystem, Lusch and Nambisan^[6] emphasized again that the core of value co-creation is to create experience value together between service providers and beneficiaries. Therefore, the author believes that customers

co-creation experience value is derived from the customer's evaluation of co-creation interaction experience, and is a new kind of experience value brought by the co-creation interaction behavior of both parties. It embodies the customer's satisfaction with the process of interaction, and is formed by changing situation as people consumption of a new type of consumption values.

For customers co-creation experience value dimensions, Holbrook^[7], a representative scholar, constructed the customer value classification framework according to three dimensions of experiential value. At the same time, he divides the customer value into efficiency value, superior value, status value, respect value, game value, aesthetic value, ethical value and soul value, which provides reference and thinking for the division of customer co-creation experience value. In the context of modern intelligent interconnection, two scholars, Novak and Hoffman^[8], regard intelligent products as independent ontology innovatively. They analyze customer co-creation experience from the perspective of ability empowerment, and divide customer co-creation experience into self-extension, self-expansion, self-limitation and self-reduction according to the two roles of customers and intelligent products. There are various ways to classify the value dimension of customer co-creation experience. Wang Xinxin and Wan Wenhai^[9] divided co-creation value into two dimensions: co-creation value in the field of production (rational cognitive value) and co-creation value in the field of consumption (emotional experience value) according to the different fields in which co-creation value occurs and the different roles played by customers. On the basis of Zhang Mingli and Tu Jianbo et al.^[10], Bu Qingjuan et al.^[11] proposed a three-dimensional division method to divide the customer experience value in the process of value co-creation into practical value, social value and entertainment value. Later, some scholars divided the experience value in co-creation into four dimensions. Wu Wenzhen and Chen Qijie^[12] divided the customer experience value into economic value, relationship value, enjoyment value and learning value based on the "use and satisfaction theory".

To sum up, the academic community has not formed a unified standard for the dimension division of customer co-creation experience value. Based on the analysis and conciseness of the existing literature and the Grounded theory research results of the author's team on the social media platform, according to the customer value classification criteria put forward by Holbrook (2006): external value and intrinsic value and Maslow's demand hierarchy theory, 11 main category theories are abstracted into four kinds of value themes: spiritual value, enabling value, practical value and hedonic value. Spiritual value comes from the intrinsic motivation of customers. It is a pure psychological and spiritual demand, which belongs to the intrinsic and transcendent value demand, including psychic value, emotional value and altruistic value; Empowering value stems from customers with enhanced control and self-awareness, high personalization needs, empowerment, and strong value together, belongs to external, high demand level value, including knowledge empowerment value, ability empowerment value; Practical value stems from the external motivation of the customer's participation in the value of economic benefits, social interests and function value, belongs to the value of external low demand, including social value, economic value; Hedonic value originates from customers' pursuit of pleasure, aesthetic enjoyment and novelty, which belongs to the high-level value of heart and id, including hedonistic, aesthetic value and novelty value.

2.2 Brand Equity

Brand Equity has been recognized as an important factor in building competitive advantage and differentiated marketing strategies. The definition of brand equity in existing studies mainly focuses on four perspectives: financial market output, product market output, customer perception and customer-brand relationship^[13]. Among them, brand equity research based on the perspective of customer-brand relationship holds that brand equity in essence does not come from customers' one-way perception and attitude towards the brand, but from the two-way interactive relationship between customers and the brand^[14]. In the first part of

brand equity based on the perspective of customer-brand relationship, the meaning of brand equity is defined: brand equity is created by the interaction between customers' attitude towards brand and brand's attitude towards customers^[15].

As brand equity is a multi-dimensional concept, the measurement methods of brand equity based on the perspective of customer-brand relationship also show a diversified trend: Fournier and Susan^[16] point out that brand equity comes from the quality of brand relationship and reflects the strength and development ability of the continuous connection between customers and brands. Xu Zhengliang and Gu Anwei^[13] believe that brand equity from the perspective of customer-brand relationship refers to customers' differentiated responses to brand marketing activities due to their relationship quality in the process of contact with brands, which includes six constructs: brand experience, brand function, brand symbol, brand attachment, brand trust and brand loyalty. And on this basis, Zhong Shuai and Zhang Qiyu^[14] further interpreted brand equity from the perspective of relationship interaction. Based on the theory of social psychology of interpersonal relations, it is defined as the differentiation effect of customers' response to marketing activities caused by the interaction perception of brand relationship. Hence, a five-dimension brand equity measurement scale based on relationship interaction is developed (brand personality, brand reciprocity, brand courtesy, relationship status, relationship energy). Because of its wide applicability, this scale has been widely adopted by scholars.

Based on the strength of these research, this study quotes Zhong Shuai et al. 's definition and scale of brand equity from the perspective of customer-brand relationship. Among them, brand personality is the customer's perception of the individual characteristics of the brand in the brand relationship, brand reciprocity refers to customers' perception of benefit exchange in the brand relationship, brand courtesy refers to customers' perception of emotional communication in the brand relationship, relationship status reflects customers' previous closeness to the brand, and relationship energy reflects customers' willingness to maintain relationship with the brand in the future^[14].

2.3 Research Models and Hypotheses

2.3.1 Relationship between spiritual value and brand equity

Spiritual value comes from the experience of satisfying spiritual needs such as belief identification, altruistic sense of achievement and emotional support obtained by customers' participation in social media interaction, including psychic value, altruistic value and emotional value. Based on the social media environment, customers gain spiritual value by interacting and exchanging ideas within the brand community, helping others solve problems and seeking emotional comfort. Through the co-creation and interaction within the brand community, customers' positive spiritual beliefs will be recognized and supported by others, and their spiritual needs will be satisfied, so as to deepen their brand cognition and further affect their perception of the brand. In addition, the spiritual value gained by customers encourages customers to continue to interact in the brand community, and the continuous positive interaction will make customers appreciate, love and rely on the brand, thus promoting the relationship between customers and the brand^[17]. Based on this, this study puts forward the following hypotheses:

Hypothesis 1(H1): The spiritual value acquired by customers in the process of co-creating interaction has a positive impact on brand equity based on the customer-brand relationship.

2.3.2 Relationship between enabling value and brand equity

Enabling value refers to the self-ability enhancement value gained by customers in the process of co-creating interaction on social media platforms, including knowledge enabling value and ability enabling value. Customers participate in the co-creation and interaction within the brand community through social media, and in the process of interaction, they improve their own abilities by acquiring information and knowledge, exchanging innovative ideas and expanding their own skills, so as to obtain enabling value. Research has found

that when customers receive guidance or educational knowledge from a certain brand, their cognition of brand products will have a positive effect^[18]. As the most critical element of brand equity, brand knowledge acquired by customers is affected by all aspects of the environment in the purchase process of customers, and has an impact on brand equity from the individual level of customers. Therefore, based on the social media platform, the knowledge gained by customers in the process of consulting product issues to brand enterprises in the brand community or participating in the interaction of brand marketing enriches their cognition of the brand, thus affecting their perception of the brand. On the other hand, when customers achieve self-improvement through co-creation and interaction, their sense of self-efficacy will be improved accordingly, thus creating emotional dependence on the brand and improving the quality of the relationship between customers and the brand. Based on this, this study proposes the following hypotheses:

Hypothesis 2(H2): The enabling value acquired by customers in the process of co-creating interaction has a positive impact on brand equity based on the customer-brand relationship.

2.3.3 Relationship between practical value and brand equity

Practical value is derived from the experience of customers achieving important goals through co-creating interaction in social media, including social value and economic value. By participating in the co-creating interaction in the brand community on the social media platform, customers are more closely connected with the members of the community, their social network can be rapidly extended, and the time cost of acquiring information and knowledge and the cost of perceived risk can be effectively reduced, thus obtaining practical value. Knox and Walker^[19] found that social value can influence the change of customers' cognitive attitude, thus influencing brand loyalty. Jin Yufang et al.^[20] verified the positive influence of customers' economic value on brand trust through empirical research. Brakus et al.^[21] put forward that brand owners can make customers rely on and feel affection for the brand by meeting customers' practical needs, and ultimately enhance brand equity. Based on the co-creating interaction within the brand community, the expansion of customers' social network promotes the integration of a variety of cognitive viewpoints. Therefore, customers' cognition of the brand is richer and more comprehensive, thus affecting their perception of the brand. In addition, in the process of interaction, the close connection with community members enhances the sense of belonging of customers, and the reduction of information acquisition cost also meets the needs of customers' actual interests, thus making customers have positive feelings towards the brand, and finally promoting the relationship between customers and the brand. Based on this, this study puts forward the following hypotheses:

Hypothesis 3(H3): The practical value acquired by customers in the process of co-creating interaction has a positive impact on brand equity based on the customer-brand relationship.

2.3.4 Relationship between hedonistic value and brand equity

The hedonistic value derives from the spiritual enjoyment, entertainment and pleasure experience gained by customers in the process of co-creating interaction with social media, including hedonistic value, aesthetic value and novelty value. In the process of co-creating interaction, the aesthetic visual design of the interface and the creative ideas burst out when communicating with others can bring enjoyment and pleasure to customers, thus enabling them to obtain the hedonistic value. Wang Xinxin and Fang Wenhai^[9] defined customer co-creation experience value as the customer aesthetic value composed of sensory pleasure, meaning of life and reconstruction experience. Through empirical research, they found that the customer aesthetic value created by consumers through consumption interaction in brand communities can improve brand loyalty. Zhang Mingli and Tu Jianbo^[10] found in their empirical study that satisfaction of hedonistic value has a significant positive impact on customers' cognitive behavior and attitude. Customers get fun and pleasure in the process of co-creating and interacting within the brand community, which satisfies their entertainment needs, promotes their attachment and affection to the brand, and enhances the relationship between customers and the brand. On the other hand,

the satisfaction of hedonistic value motivates the interaction enthusiasm of customers, and frequent interaction promotes the further cognition of customers to the brand, and then makes them have a good impression on the brand. Based on this, this study proposes the following hypotheses:

Hypothesis 4(H4): The hedonistic value acquired by customers in the process of co-creating interaction has a positive impact on brand equity based on the customer-brand relationship.

3. RESEARCH DESIGN

3.1 Questionnaire design

The questionnaire designed in this paper mainly covers three sections. The first part contains the basic personal information and relevant experience of the respondents. The second part mainly asks targeted questions about the real co-creation experience of the respondents. The third part evaluates the subjective perception of respondents by focusing on brand equity based on customer-brand relationship.

In the second part of the questionnaire, most of the variables used to measure the value of customer co-creation experience are from existing literature, and then reasonable modifications are made according to the research situation of social media in this study. The measurement items in this part are mainly from the scales used in 8 studies by scholars such as Kumar.M^[22], Nambisan et al^[23]. Subsequently, combined with reference to the measurement items in the existing scale and the questions reasonably supplemented by the author according to the research situation and practice, the scale of customer co-creation experience value based on the social media environment is finally organized and formed. In the third part of the questionnaire, the brand equity scale developed by Zhong Shuai et al was reasonably used for the measurement of brand equity based on the customer-brand relationship. The specific reasons for the selection have been explained in the previous section and will not be repeated here.

In order to measure whether the measurement items in the scale can reflect the problems that need to be investigated, two professional scholars in the field of value co-creation were invited to review the scale, and according to the feedback opinions, the scale was revised, finally forming the preliminary draft of the questionnaire of this research. Subsequently, 15 Club of huawei users were invited to participate in the pre-survey of the questionnaire in this study, and the initial questionnaire was supplemented and modified according to the questions and shortcomings of the respondents. Finally, the formal questionnaire of this study was formed after many improvements, as shown in Table 1.

Table 1. Questionnaire

Second-order Construct	First-order Subconstruct	Item
Spiritual Value (SPI)	Spirituality value (SPV)	SPV1. Participate in the interaction, seeing the incredible creativity and design of others, I have a positive belief in dealing with the problems of innovation.
		SPV2. Participate in the interaction, I feel very open-minded.
		SPV3. Participate in the interaction, I feel the optimistic atmosphere.
		SPV4. Participating in the interaction, I was inspired by a positive attitude towards life.
		SPV5. Participating in the interaction, I firmly adhere to the fairness, transparent and fair belief.
	Emotional value (EMV)	EMV1. When I pour out my difficulties in the Club of huawei, someone comfort and encourage me.
		EMV2. When I describe my predicament in the Club of huawei, someone accompany me to face it.
		EMV3. When I talk about my troubles in the Club of huawei, someone care me.
		EMV4. When I post my question in the Club of huawei, someone respond to me quickly.
	Altruistic value	ALV1. Participating in the interaction makes me feel that I am fulfilling my social responsibility.
		ALV2. Participating in the interaction makes me feel like I'm helping others.

	(ALV)	ALV3. Participating in the interaction makes me feel that I am contributing to the society.
Enabling Value (ENP)	Knowledge empowering value (KEV)	KEV1. Participating in the interaction, I have increased my knowledge about Huawei's products / services. KEV2. Participate in interaction, help me get solutions to product / service use related issues. KEV3. Participate in the interaction, so that I can have a deeper understanding of Huawei, Huawei products/services and technical improvements.
	Capability empowers value (AEV)	AEV1. Participating in interaction has improved my reputation in the Club of huawei. AEV2. Participate in interaction has improved my product use skills. AEV3. Participating in interaction has improved my ability to discern product performance. AEV4. Participate in the interaction can help me understand new ideas in Huawei products/services. AEV5. Participating in interaction can inspire me to come up with new ideas. AEV6. Participate in the interaction, I can better meet my preferences. AEV7. Participating in the interaction, I have professional skills related to product/service development. AEV8. Participating in the interaction, I gain a sense of accomplishment. AEV9. Participating in the interaction, I feel that I can influence the design and development of Huawei's new products/services (for example, the optimization of EMUI system functions, the appearance design of Huawei's new mobile phones).
Pragmatic Value (FUNC)	Social value (SOV)	SOV1. Participating in the interaction has enabled me to expand my social circle. SOV2. Participating in the interaction enhanced my sense of belonging to the Club of huawei. SOV3. Participating in interaction has strengthened my relationship with Huawei Club customer base. SOV4. Participating in the interaction gave me a sense of identity with the Club of huawei. SOV5. In the process of participating in the interaction, I feel like I'm a member of the Club of huawei.
	Economic value (ECV)	ECV1. Participate in the interaction, I can buy products/services at a more favorable price. ECV2. Participate in the interaction, I am likely to win prizes. ECV3. In return for participating in the interaction, I got a gift. ECV4. Because of participating in the interaction, I received virtual rewards (such as points and coins). ECV5. For participating in the interaction, I received a monetary reward.
Enjoyment Value (ENJ)	Hedonic value (HEV)	HEV1. Participate in the interaction, I get happiness. HEV2. Participate in interaction to solve problems and generate creativity, which makes me have fun. HEV3. Participating in interaction to solve problems and generate ideas makes me feel excited. HEV4. In the process of participating in the interaction, I felt joy and happiness.
	Aesthetic value (AV)	AV1. Various customer designs in the Club of huawei are visually attractive. AV2. I admire the aesthetic art of customer design in the Club of huawei. AV3. The customer design in the Club of huawei gave me visual enjoyment.
	Novelty value (NOV)	NOV1. Participate in the interaction and satisfy my curiosity. NOV2. I appreciate the creativity and design proposed by customers in the Club of huawei. NOV3. Participate in interaction and help me understand new trends. NOV4. Participation in interaction is an exploratory experience for me.
Brand Personality		BRA1. Huawei is an outstanding brand. BRA2. Huawei is a reliable brand.
Brand Discount		BRA3. Relative to the cost paid, Huawei is worth the money. BRA4. Buying Huawei is more cost-effective than other brands.
Brand Courtesy	/	BRA5. Huawei is very consumer friendly. BRA6. Huawei respects consumers.
Relationship Status		BRA7. The Huawei brand makes me feel safe and assured. BRA8. I think Huawei is trustworthy.
Relationship Energy		BRA9. I would rather spend some setbacks to buy Huawei. BRA10. I will always use Huawei in the future.

3.2 Data sources

In this study, the Club of huawei was selected as the research situation, and the users of the Club of huawei were the respondents of the questionnaire. Club of huawei has over 100 million users and is one of the most active platforms for brand fans to interact with each other in China. Users in Club of huawei can freely communicate and interact with others, publish ideas, share product related knowledge and experience, provide suggestions or opinions to customers and enterprises in the same industry, and answer customers' inquiries about products in the same industry, etc. Through the frequent interaction of Club of huawei, users will have a change in their cognition and emotion towards Huawei, thus promoting the relationship between users and Huawei. Therefore, the Club of huawei is consistent with the research content of this study.

In this study, the 4-week questionnaire data collection was started on April 15, 2019. The online questionnaire was mainly released through various channels such as the app of Club of huawei, related QQ groups and the comment area under the official Weibo. Members of Club of huawei were encouraged to fill in the questionnaire carefully by giving red envelopes to ensure the validity of the questionnaire data. In this study, after excluding incomplete and invalid questionnaires, there are 245 valid questionnaires to continue the next empirical test.

4. DATA ANALYSIS AND RESULTS

4.1 Reliability and validity test

In this paper, exploratory factor analysis(EFA) and confirmatory factor analysis(CFA) were used to evaluate the reliability and validity of the scale items in this study using SPSS17.0 and AMOS22.0 software respectively. The analysis results are shown in Table 2. Firstly, the reliability level of the scale was evaluated by Cronbach's α and Composite Reliability(CR). The results show that the Cronbach's α value and the CR value of each variable are in the range of 0.933-0.973, which are greater than the minimum acceptable standard 0.7, indicating that the internal consistency of the measurement items is good.

Secondly, the validity level of the scale was determined by content validity and construction validity, of which the latter was divided into Convergent Validity and Discriminant Validity. As the scale items in this study are extracted and summarized from a large number of existing literatures, and revised after expert review and feedback, it can be considered that this scale has certain content validity. Since the variables in this scale were constructed by combining the research context and literature analysis, the overall construction validity of the scale was first verified in this study. The indexes of the first-order model: $\chi^2=1820.66$, $df=886$, $RMSEA=0.066$, $TLI=0.924$, $CFI=0.932$, all of which meet the standard suggested by Bentley^[25], showing that the model has good goodness of fit. Since the value of customer co-experience proposed in this study (spiritual value, enabling value, practical value, hedonistic value) is a second-order variable, the goodness of fit of the second-order model is further tested in this study. The results showed that the estimated values of various fitting indexes in the second-order model were not significantly different from those in the first-order model ($\chi^2=1846.96$, $df=915$, $RMSEA=0.065$, $TLI=0.926$, $CFI=0.932$), which indicated that the setting of higher-order variables in this study did not affect the relationship between the first-order variables and the measured variables. Moreover, the second-order model was more simple and more convenient for theoretical research.

Subsequently, this study tested the aggregation validity of the scale by combining the factor load of each variable and Average variance Extracted(AVE). The results showed that the standard load of all the measurement items ranged from 0.757 to 0.991 (all greater than 0.7), and the AVE of each variable was greater than the minimum standard 0.5, indicating that the scale had good aggregation validity.

Finally, this study tested the discriminative validity of the scale by comparing the square root of AVE of each variable and the size of correlation coefficient between variables. The square root of AVE of each latent

variable in Table 2 is greater than the correlation coefficient with other latent variables, so the scale has good discriminative validity.

Table 2. Factor analysis results

	α	CR	AVE	1	2	3	4	5	6	7	8	9	10
SPV	0.965	0.955	0.812	0.901									
EMV	0.973	0.961	0.892	0.376	0.944								
ALV	0.953	0.954	0.873	0.373	0.491	0.934							
KEV	0.952	0.935	0.827	0.189	0.249	0.247	0.909						
AEV	0.933	0.947	0.668	0.051	0.067	0.067	0.290	0.817					
SOV	0.957	0.954	0.807	0.226	0.298	0.295	0.205	0.055	0.898				
ECV	0.955	0.955	0.811	0.235	0.310	0.307	0.213	0.057	0.259	0.900			
HEV	0.944	0.944	0.809	0.093	0.123	0.122	0.129	0.035	0.151	0.157	0.899		
AV	0.943	0.945	0.852	0.110	0.144	0.143	0.152	0.041	0.178	0.185	0.149	0.923	
NOV	0.940	0.936	0.785	0.102	0.135	0.134	0.142	0.038	0.166	0.173	0.139	0.163	0.886

4.2 Common method deviation

The sameness of data source or measurement environment and other anthropogenic covariation can easily lead to common method deviation and affect the final results of the research. In order to test whether there is serious common method variation in the data, this paper uses Harman's single factor test method to test the sample data. Among them, 10 factors with characteristic roots greater than 1 were extracted from the unrotated exploratory factor analysis results, and the maximum factor variance interpretation rate was 24.733%, lower than the critical value of 40%, so there was no serious common method bias in the sample data of this study.

4.3 Results of hypothesis testing

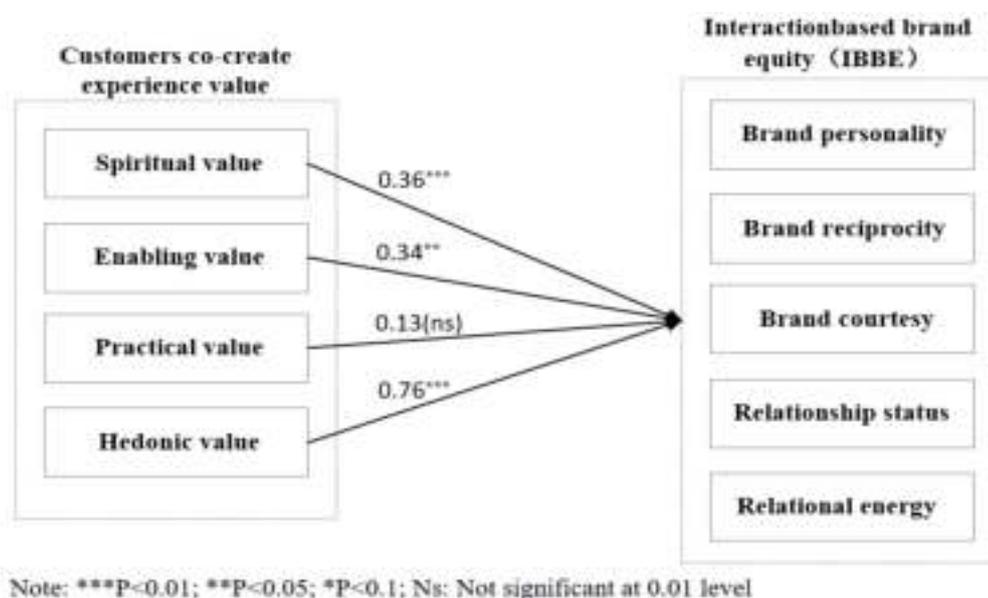


Figure.1 Results of hypothesis testing

In this study, AMOS 22.0 structural equation model analysis tool was used to test the model and research hypothesis proposed in the previous paper. As shown in figure 1, the spiritual value, enabling value and hedonistic value of customer co-creation value obtained by customers in co-creation interaction process have significant positive influence on brand equity from the perspective of customer-brand relationship. H1, H2 and H4 are supported. However, there is no significant relationship between the practical value gained by customers

in the process of co-creation and brand equity from the perspective of customer-brand relationship. H3 is not supported.

5. CONCLUSIONS

5.1 Research conclusion

This study takes Club of huawei as the research object and explores the influence of customer co-creation experience value acquired by customers in the co-creation interaction process on brand equity based on the customer-brand relationship from the perspective of social media context.

Research results show that the spirit value, can assign a value and hedonic value is based on the customer - brand relationship between effective predictors of brand equity (H1, H2, H4 hypothesis supported), suggests that customers through social media participation in the process of brand community to create interactive spirit value, can assign a value and hedonic value can significantly positive effect on customer perception of the brand, so as to promote the relationship between the customer and brand, enhance brand equity based on the customer - brand relationship.

Secondly, it is proved that utility value has no significant positive effect on brand equity based on customer-brand relationship (H3 hypothesis does not support this). The reasons for this result may have two kinds: first, with improvement of material life level, people demand for social contact and price sensitivity decreases, most respondents may result in the aspect of practical value, social value and economic value to create the experience of relatively thin, is not conducive to the research on the practical value of the measurement; secondly, it is understood that the members of Club of huawei set up by Glory pay more attention to product evaluation and other contents, the number of posts with group consciousness is less, and the community managers have less interaction with customers on the level of belonging and fans' welfare, which may have an impact on the evaluation of practical value in this study.

5.2 Research contribution

Based on the context perspective of social media, this study refines the new connotation of customer co-created experience value, explores the relationship between customer co-created experience value and brand equity in brand communities, and constructs the corresponding influence model for empirical test. The main theoretical contributions of this paper are as follows:

First, this paper builds a second-order factor model of customer co-created experience value based on the context of social media and existing literature. To a large extent, different from existing researches, scholars mostly focus on the emotional value/altruistic value/practical value/hedonistic value itself. This study further refines the customer co-created experience value as a second-order factor, and further improves the measurement of customer co-created experience value by existing scales.

Secondly, from the perspective of social media context, this paper explores the antecedent variables of brand equity by introducing the characteristics of interaction and co-creation of social media. Although previous studies have discussed the pre-influencing factors of brand equity, scholars mainly focus on indicators such as customer loyalty, while this paper considers the influence of customer co-creation experience value on brand equity based on the interactive co-creation characteristics of existing situations. Among them, the study shows that obtained in the process of creating the practical value to customers based on the customer - brand relationship there is no significant impact on brand equity, with general cognitive differences, this shows that the formation of new research situation for customers to create the experience value provides a new conclusion and view, enrich the existing research experience for customers to create value and brand equity related to explore.

This study has certain enlightening significance for the practitioners of brand building based on social media platforms.

First of all, brand community managers should provide customers with satisfactory experience value based on the interactive environment of social media. The results of this study show that brand equity based on the customer-brand relationship is not only affected by the spiritual value acquired by customers in the process of co-creating interaction, but also affected by the positive effects of empowering value and hedonistic value acquired in the interaction. In other words, brand community managers should also make full use of the technical enablement of social media platforms to meet customers' potential demands for enabling value and hedonistic value. For example, the managers of Club of huawei can set up a selection and recommendation mechanism on the home page, increase the exposure of more posts of higher quality, and optimize the visual design of the interactive interface.

Secondly, brand community managers should actively encourage customers to interact with other members and managers in a positive way, so as to increase their sense of brand identity and trust, thus driving the improvement of performance. And in the process of the customer to carry on the positive interaction with others, managers have access to the customer for the product novel ideas and Suggestions, which is advantageous to the enterprise new product development and improvement, such as millet can through the way of cultivating key opinion leaders to encourage the rice noodles are actively involved in the interaction in the millet community, express themselves for millet various ideas of new products, and millet can managers to collect back the value of information to the headquarters for the new product research and development.

5.3 Research limitations and prospects

The deficiencies of this study are as follows: First, due to the limitations of the survey environment and time and the large number of questionnaire questions, only 245 sample data were effectively collected in the end. Although the sample size reached the minimum requirement of the questionnaire survey, the number of respondents still needs to be increased in order to improve the reliability of the sample data. Second, this study draws on domestic and foreign scholars developed and verified the measurement scale, in order to combine in this paper, the research situation, the author in the customer to create the experience value and brand equity measurement item made reasonable modification, and because the domestic foreign language habits is different, in order to make respondents are more likely to understand the translated item, the author of the original item with the localization of the words change, so the questionnaire survey results of this study may with the original item is the error of measurement.

Based on the above shortcomings, the future research can be improved in a more scientific and reasonable way, and this study only consider based on social media perspective to create the customers experience value based on the customer - brand relationship between the influence of brand equity, future research can also from more diverse perspective to explore the customers to create the experience value to the brand influence, in theory reference and guidance for the industry to provide more.

ACKNOWLEDGEMENT

Philosophy and Social Science Research Youth Project of Hubei Provincial Department of Education: "Research on the influence Mechanism of value Co-creation on Enterprise Innovation performance in Social Media Environment" (19Q029, 2019.01-2020.12)

REFERENCES

- [1] Hong Zhang, Yaobin Lu, Yanling Yan. (2017). Research on the influence of Technical characteristics of Social Shopping Community on purchase intention. *Scientific Research Management*, 38 (2): 84-92(in Chinese)
- [2] iiMedia.(2018).2018 Special Research report on China's New consumption, Consult in Media Research(in Chinese)

- [3] Vila-López N, Rodríguez-Molina M. (2013). Event-brand transfer in an entertainment service: experiential marketing. *Industrial Management&Data Systems*, 113(5): 712-31.
- [4] Prahalad C K, Ramaswamy V. (2000). Co-opting customer competence. *Harvard Business Review*, 78(1): 79-87.
- [5] Vargo S L, Lusch R F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68(1): 1-17.
- [6] Lusch R F, Nambisan S. (2015). Service innovation: a service-dominant logic perspective. *MIS Quarterly*, 39(1): 155-76.
- [7] Holbrook, B. M. (2006). Consumption experience, customer value, and subjective personal introspection: An illustrative photographic essay. *Journal of Business Research*, 59(6): 714-25.
- [8] Hoffman D L, Novak T P. (2017). Consumer and object experience in the Internet of Things: an assemblage theory approach. *Journal of Consumer Research*, 44(3): 1178-204.
- [9] Wang X X, Wan W H. (2012). Research on the mechanism of co-creating value in the field of consumption and its effect on brand loyalty. *Management Science*, 25 (05): 52-65(in Chinese)
- [10] Zhang M L, Tu J B. (2014). The influence of virtual community co-creation user experience on user co-creation value. *Journal of Tongji University (Natural Science Edition)*, 42 (7): 1140-6(in Chinese)
- [11] Bu Q J, Jin Y S, Li Z H. (2016). Does interaction necessarily create value? -- the influence of customer value co-creation interaction on customer value. *Foreign economy and Management*, 38 (09): 21-37(in Chinese)
- [12] Wu W Z, Chen Q J. (2017). The influence of customer participation behavior on their satisfaction and behavior intention from the perspective of co-creation value. *Management Review*, 29 (09): 167-180(in Chinese)
- [13] Xu Z L, Gu A W. (2011). Research on brand equity-driven model from the perspective of relationship. *China Industrial Economics*, 11(10): 109~118(in Chinese)
- [14] Zhong S, Zhang Q Y. (2015). Brand equity concept, dimension and scale development based on relationship interaction. *Management Science*, 28(2): 69~79(in Chinese)
- [15] Blackston.M. (1995). The qualitative dimension of brand equity. *Journal of Advertising Research*, 35(4): 2~7.
- [16] Fournier, Susan. (1998). Consumers and Their Brands: Developing Relationship Theory in Consumer Research. *Journal of Consumer Research*, 24(4): 343~373.
- [17] Wu L L, Wang B. (2015). Research on the Influence Mechanism of Virtual Brand Community on Brand Relationship -- A Case Study of Xiaomi Community. *Management Case Studies and Review*, 8(1):71-83(in Chinese)
- [18] Li Q G. (2012). The formation of brand experience and its influence on brand equity. Ph D Thesis. Shanghai: Shanghai Jiao Tong University(in Chinese)
- [19] Knox S, Walker D. (2003). Empirical developments in the measurement of involvement, brand loyalty and their relationship in grocery markets. *Journal of Strategic Marketing*, 11(4): 271-286.
- [20] Jin Y F, Dong D H, Liu R M. (2006). An empirical study on the establishment of consumer brand trust mechanism and its influencing factors. *Nankai Management Review*, 11(5):28-35(in Chinese)
- [21] Brakus J J, Schmitt B H, Zarantonello L. (2009). Brand Experience: What Is It? How Is It Measured? Does It Affect Loyalty?. *Journal of Marketing*, 73(3): 52-68.
- [22] Kumar M, Noble C H. (2016). Beyond form and function: Why do consumers value product design?. *Journal of Business Research*, 69(2): 613-20.
- [23] Nambisan S, Baron R A.(2007). Interactions in virtual customer environments: Implications for product support and customer relationship management. *Journal of Interactive Marketing*, 21(2): 42-62.
- [24] Bateson J E G. (1985). Self-Service Consumer: An Exploratory Study. *Journal of Retailing*, 61(3): 49.

Full Research Paper

Exploring the Influence of Live Streaming in Social Commerce on Impulse Buying from a Affordance Perspective

Yan Mo¹, Qi Wang^{2*}

^{1,2}School of Economics and Management, Zhejiang sci-tech University, Hangzhou, 310018, China

Abstract: As a new form of social commerce, live streaming still has the phenomenon of consumers' impulse buying. Based on S-O-R model, this study builds a theoretical model from the perspective of affordance to investigate the impact of live streaming in social commerce on consumers' impulse buying. We obtained data from consumers who watched live streaming in the form of questionnaires, and conducted empirical tests on the model. The results show that visibility, expressiveness, guidance shopping and trading affect consumers' social presence and immersion, and thus let consumers make impulse purchases.

Keywords: live streaming; Impulse buying; Affordance; Social presence; Immersion

1. INTRODUCTION

As a new type of social commerce, live commerce has become the main form of shopping for consumers. AI media consulting data shows that more than 50% of users purchase through live broadcast every month, and the scale of China's live commerce market will reach 961 billion yuan in 2020, with a year-on-year growth of 121.5%. China Consumer Association conducted a survey on the current situation of live e-commerce, and 44.1% of the respondents said that there was a serious phenomenon of impulse buying in live commerce. However, most of the existing researches focus on the business model, marketing strategy and customers' purchase intention, and there are still few researches focusing on impulse buying. In addition, previous studies on online impulse buying mostly explored online store display, product type, price, time pressure, etc. Although the contribution of these studies is obvious, the technical characteristics of live commerce are different from traditional social commerce. First of all, live commerce is no longer confined to the traditional "graphic details" to introduce product attributes, uses, appearance and other information. Instead, live streaming technology is used by the anchor to complete a comprehensive display and introduction of the product, so that customers can get more comprehensive and rich product information. Secondly, in traditional social commerce, businesses communicate with customers through text, but in live e-commerce, customers communicate with anchors and businesses in real time in the form of bullet screen, so as to improve the purchase efficiency^[1]. Besides, in traditional social commerce, the lack of face-to-face interaction makes users doubt the authenticity of the goods, and increases the perceived risk of online shopping. In live-streaming room, the anchor can provide users with highly personalized services and guidance, thus having a meaningful impact on users' buying behavior.^[2,3]. Therefore, previous studies can't be completely used in the live commerce situation. We need to consider the formation mechanism of impulse buying behavior from the live streaming itself.

Live commerce has a high level of human-computer interaction, so it is necessary to consider the characteristics of real-time interaction, high visualization and consumer perception. The concept of affordance can help us to consider the technical characteristics of live commerce and consumers' perception when interacting with these characteristics^[4,5]. In live commerce, affordance comes from the interaction between users and technical features. Previous studies have demonstrated the effectiveness of affordance in social commerce. Therefore, we

* Corresponding author. Email: 463943168@qq.com (Qi Wang) , moyan@zstu.edu.cn (Yan Mo)

think that this concept is suitable for studying how live commerce, as a new form of social commerce, affects consumers' impulse buying behavior. The "stimulus-organism-response" (S-O-R) framework is widely used to explain impulse buying, which reflects people's internal cognitive and emotional processes in the face of external stimuli, resulting in positive or negative behaviors^[6]. As the concept of the possibility of influencing user behavior in a certain way, there is a "black box" between affordance and impulse buying behavior, that is, the subject's cognition and emotion. When looking up the literature on live broadcasting, we found that social presence and immersion have received great attention^[7,11]. Live shopping displays products in a direct and face-to-face way, creating a real-time experience for consumers. On this basis, the high interactivity provides consumers with an immersive experience, which makes it easier for consumers to perceive the brand and produce purchasing behavior. Although social presence and immersion have been identified as important antecedents of customers' purchase in live shopping, few studies have considered their role in consumers' impulse buying behavior at the same time.

The purpose of this paper is to study how live broadcast affects consumers' impulse buying behavior. In order to achieve this goal, based on the "S-O-R" model, this paper studies the influence of the technical characteristics of live shopping on consumers' impulse buying behavior from the perspective of affordance, and the mechanism of social presence and immersion in it. This study has practical significance for improving the product design of live broadcast platform and enhancing the effect of live broadcast marketing, and also helps to enrich the application of affordance theory in marketing and the theoretical research of consumer purchase behavior.

2. LITERATURE REVIEW

2.1 Impulse buying

In this article, impulsive buying was defined as the act of quickly making a decision to buy out of the plan due to the stimulation of the outside environment, and the customer's demand for the goods and inner emotions were stimulated. The main elements of this definition came from the existing online and offline impulsive purchase of documents.

Many researches shows that the basic feature of impulse buying is buying goods that are not originally planned and it can't be predicted^[8], which means that it is adorable on the spot, that is, when an individual is exposed to shopping scenes, he is stimulated by external stimuli to acquire a certain commodity desire. Many scholars believe that impulse buying is caused by emotions, and the whole purchase process is relatively rapid. That is, under the stimulation of shopping environment, consumers who are in a positive state are easily influenced by the warm shopping atmosphere, so it is difficult to examine the attributes of commodities carefully and make the purchase decisions quickly.

Impulsive shopping is highly stimulating. The type of products, the price of the products, and the environment of the online store play a key role in luring customers to buy on impulse. In live commerce, anchors display the goods to the customer in the form of real-time video. Customers interact and communicate with the anchor in real time through bullet screen. In addition, the live location is not restricted. Anchors can create different ambience to stimulate consumers to impulse buying^[9]. After sorting out, it was found that the existing studies had tested the impulsive buying behavior in the livestream by taking the clues of livestream scene atmosphere, online interaction, lively and interaction of the product display as the main factors. However, these studies doesn't directly explain the different features of the live streaming technique and the relationship between impulsive buying, nor could they provide direct and powerful advice to the businessmen and platforms. Therefore, under the stimulation of the technical features of the live commerce platform, this research would explore the inner experience of the consumers and the impulse to buy.

2.2 The affordance theory

The affordance theory originates from ecological psychology and it is a possibility of activity between the

subject and the environment. It is an object attribute independent of people and closely related to each person's perception ability. In terms of affordance, scholars have different definitions due to different research fields. According to Dong and Wang, the affordance of social commerce refers to the possibility of implementing actions provided by technology for both sellers and buyers who have specific abilities and aim to purchase^[10].

The concept of affordance is widely used to understand the relationship between IT technology and social practice. For example, IT affordance of social commerce platforms establishes strong and weak connections between customers and sellers^[10], the formation of rapid *guanxi* between customers and sellers from the perspective of social commerce affordance^[11]. Besides, the impact of live streaming on the purchasing intention of Chinese social commerce consumers, and the affordance of live streaming shopping platforms affects customers' purchasing intention^[7]. In social commerce, affordance arises from the user's perception of the platform's technical characteristics. When consumers watch live streaming, they can perceive the characteristics of it based on the use of platform functions. Therefore, we believe that affordance, as a concept that can comprehensively consider technical characteristics and consumer perception, can be used to explain how live streaming affects consumers' impulse buying.

2.3 Social presence theory

As one of the most important theories in communication, social presence refers to the salience of others in interaction and the consequent salience of interpersonal relationship^[12]. Social presence refers to the immediate sense of co-existence between an individual and other participants, as well as the degree of interpersonal relationship^[13]. This definition not only fully describes the connotation of social presence, but also focuses on immersive media and human-computer interaction scenes, which have a certain correlation with live streaming scenes.

With the development of computer technology and Internet, many scholars have found that social presence is not only an attribute of media. Therefore, many scholars have begun to try to divide social presence into multiple dimensions, so as to accurately describe the experience of users in various network scenarios. Previous studies showed that social presence can be expanded into three dimensions of awareness, affective social presence and cognitive social presence^[14]. In the field of live streaming marketing, Xie Ying explored the influence of social presence on online conformity consumption from the perspective of behavior and neurophysiology, and divided it into coexistence presence, communication presence and emotional presence^[15]. In the scene of live commerce, images are transmitted from one place to another using live streaming technology, which is of real-time reduction. In addition, consumers can obtain a large amount of real-time information from others on the spot through other people's behaviors such as thumb up, giving virtual gifts, and brushing bullet screens, so as to experience the feeling of physical presence of others, namely awareness. In the live streaming chat room, consumers can not only communicate with anchors and other consumers in real time, but also experience the emotional state of others through sending gifts and giving a like. So that customers can feel the salience of interpersonal relationship, that is, affective social presence. In addition, the individual in the live streaming chat room can have a certain understanding of others' ideas through communication, thus producing cognitive connection, that is, cognitive social presence. Therefore, this study takes awareness, affective social presence and cognitive social presence as three dimensions of social presence. Besides, when examining the multi-dimensional characteristics of social presence, we referred to the method of Shen and regard it as a second-order formative variable^[14].

2.4 Immersion

Immersion was first proposed by Csikszentmihalyi, which describes the overall feeling of an individual when he or she is engaged in something, including psychological perception characterized by deep sense of concentration and pleasure, time distortion and control^[16]. In the study of user behavior, immersion is often considered as an intrinsic motivation, which depends on the subjective feelings of the user and affects usage

behavior. Scholars at home and abroad have widely applied the conception of immersion in online games, online education [17], online shopping and other fields, and have confirmed its important role in user attitude and behavioral intention. As a highly interactive social commerce, live commerce is a combination of mobile live streaming and online shopping. When watching live e-commerce, due to high-frequency interaction, consumers often immerse themselves and forget the passage of time, thus creating a sense of immersion. Therefore, this study applies immersion as a mediating variable to the study of consumers' impulse buying behavior in live commerce.

2.5 S-O-R model

The "stimulus-organism-response" model (S-O-R) indicates that external environmental stimuli can affect the internal state of individuals and further affect individual behaviors [18]. In this model, stimulus refers to the effect of external environment on individuals, while organism refers to the internal state of each individual stimulated, and response refers to the behavior displayed by the individual. This model has been widely used in the study of consumer behavior in social commerce environment [3],[19],[20]. Considering the applicability of S-O-R model, this study regards it as the research model framework of consumers' impulse buying in live commerce.

To sum up, based on the S-O-R framework, it is proposed that live commerce affordance (S) can affect the social presence and immersion of consumers when watching live streaming(O), and then affect their urge to buy impulsively (R).

3. RESEARCH MODEL AND HYPOTHESES

3.1 Affordance and consumer internal experience

Existing studies have shown that in different scenarios, affordance has different attributes due to the different ways in which technical characteristics affect individual behavior to achieve goals. Dong and Wang identified its six dimensions: visibility, meta voicing, triggered attending, guidance shopping, social connecting and trading^[10]. The impact of live streaming on the purchase intention of Chinese social commerce customers, and indicates that the visibility, meta-voicing and guided shopping in the affordance of live streaming shopping platforms affected the purchase intention of customers^[7]. The visibility of the live streaming shopping platform shows the appearance and usage of the products to consumers in the form of real-time images, so that the products and brand information can be truly presented to customers in an all-round way. In the live streaming chat rooms, consumers can express their feelings and feedback through bullet screens, and the possibility that consumers can respond to the product or the content introduced by the anchor is expressiveness^[7]. In the live streaming chat rooms, anchors can directly respond to the questions raised by consumers and help consumers establish product demands. This technical feature, which enables consumers to make purchase decisions through providing personalized services, is guidance shopping^[7,10]. In addition, compared with traditional online shopping, the transaction process of live streaming is relatively complex, so the trading nature of the platform is the technical feature that makes the whole transaction process safe, smooth and efficient. Therefore, we believe that the affordance of live streaming shopping platforms can be divided into visibility, expressiveness, guidance shopping and trading.

3.1.1 Visibility affordance and consumer experience

Based on the visibility affordance, merchants can deliver detailed product information to consumers through live streaming. In this process, consumers can see all kinds of attributes and details of the product, and they will feel that they are observing the product on the spot. Through live streaming, consumers can also see the facial expressions and actions of anchors, further perceive their emotions. Thus, it can be argued that visual affordance helps customers perceive social presence. With the visualization technology characteristics of online video, live commerce has a high vividness, which can attract customers and make them immerse themselves in the constructed scene. At the same time, previous studies have pointed out that the vivid performance of product

display can bring customers a pleasant experience, so that customers are fully engaged to achieve the state of immersion^[21]. Therefore, the following hypotheses are made in this study:

H1a: In live commerce, visibility affordance has a positive influence on social presence of consumers.

H1b: In live commerce, visibility affordance has a positive impact on consumers' immersion.

3.1.2 Expressiveness affordance and consumer experience

Expressiveness affordance of living streaming provides the possibility of communication between consumers and anchors, which greatly improve the level of the interaction between them and let consumers perceive the anchor is a "real man" rather than the artificial intelligent robots. At the same time, consumers in the live streaming chat rooms have the opportunity to speak freely. They can communicate with other participants about the product or anchor's introduction in real time by using the bullet screen. Compared with the passive interactive environment, consumers can feel a higher level of social presence in this kind of active communication. In addition, high-frequency communication and interaction in the live streaming chat rooms can enhance consumers' perception of others, effectively shorten the distance between individuals, make the connection between each other more closely, and thus improve the sense of social presence^[14].

Live broadcasting providers provide real-time communication platform for consumers and anchors by means of expressiveness. Through empirical research confirms the positive effect of interaction in cloud logistics information platform on user immersion experience^[22]. Scholars like Hoffman and Nova believe that timeliness plays a crucial role in the audience's immersion experience in text based interaction^[23]. Bullet screen is the only way to interact with consumers and hosts in live commerce. Consumers use bullet screen to express their ideas and communicate with others^[11]. Pace found that the premise of feeling immerse was clear and timely feedback^[24]. The comments in live commerce are expressed in real time, which has a strong time effect. This kind of instant presentation shorten the time difference of message communication, making the customers more excited and more immersed in the live streaming. Therefore, this research made the following assumptions:

H2a: In live commerce, expressiveness affordance has a positive influence on social presence of consumers.

H2b: In live commerce, expressiveness affordance has a positive impact on consumers' immersion.

3.1.3 Guidance shopping affordance and consumer experience

In live commerce, guidance shopping affordance can provide consumers with products that match their interests to meet their needs^[7]. As the suggestions provided by anchors have a high degree of relevance to customer needs, consumers will focus their attention on watching live streaming. Guidance shopping affordance can effectively help consumers solve the problems they encounter in the purchase process, so that consumers will perceive a high value of use, and such satisfaction will make consumers feel happy^[25]. At the same time, anchors can provide purchase suggestions for customers according to their preferences and needs during the live streaming, which will make consumers feel that the services provided by live commerce are similar to those provided in offline shopping, so they will have an immersive feeling^[25]. Anchors often recommend products or answer questions raised by consumers, such interaction is personalized and tends to interpersonal communication, and the sense of social presence will be further enhanced. Therefore, the following hypotheses are made in this study:

H3a: In live commerce, guidance shopping affordance has a positive influence on social presence of consumers.

H3b: In live commerce, guidance shopping affordance has a positive impact on consumers' immersion.

3.1.4 Trading affordance and consumer experience

Trading affordance provides a secure way of payment to consumers, enabling transactions to be launched smoothly on the live electronic business platform, which greatly improves the convenience of consumers in the purchase process. When the live commerce platform provides transaction protection for the public, users are more likely to have positive psychological hints. The presentation of guaranteed information can deepen customers'

cognition of goods and bring consumers a sense of shopping security^[1]. When customers ask the anchor or business about the purchase process, their timely feedback is helpful for customers to form an overall understanding of the business and goods. When users click on the product link in the live room to pay the page, the live interface will show the word "user is going to buy". This prompts the customer to produce a co-existence society with others in their presence, and experience the positive feeling of accompanying shopping. In addition, the safety of the shopping process is closely related to the customer's immersion^[26]. The security and reliability of the live commerce, as well as the authenticity of the anchors and businesses, can reduce the perceived risk of consumers, make them integrate into the shopping environment faster and produce immersion. Therefore, this study puts forward the following hypotheses:

H4a: In live commerce, trading affordance has a positive impact on social presence of consumers.

H4b: In live commerce, trading affordance has a positive impact on consumers' immersion.

3.2 Consumer experience and urge to buy impulsively

3.2.1 Social presence and urge to buy impulsively

Studies have proved that social presence makes consumers feel that online shopping environment is more transparent and customers will feel more secure when making purchase decisions, so it will enhance consumers' purchase intention^[9]. The vitality brought by the technical features of live streaming enables consumers to learn more complete product information and more direct product experience, as well as the enthusiasm and friendliness of anchors, which can shorten the subjective distance between each other and let consumers have a sense of being on the scene. At the same time, the hypothesis of "social person" indicates that people have social attributes, and each individual does not want to exist in isolation. Through the communication with others in the live streaming chat rooms room, consumers will have positive interpersonal satisfaction and emotions. Under the guidance of positive emotions, they will reduce their perception of product risks and naturally tend to have a positive perception of the products recommended by the anchors so that to buy the products without a plan in advance. Therefore, the following hypotheses are proposed in this study:

H5: In live commerce, social presence has a positive impact on consumers' impulse buying.

3.2.2 Immersion and urge to buy impulsively

A number of studies have confirmed that consumers have a sense of immersion in the process of website browsing, which enhances unplanned purchase intention and the willingness to visit the website again. And immersion positively influences consumers' impulse buying intention^[9]. When watching live streaming, high-frequency interaction with others can bring consumers a pleasant and satisfying immersive experience. Dripped by such positive experience, consumers are more inclined to make positive behaviors so that buy products that are not originally planned. Therefore, this study puts forward the following hypotheses:

H6: In live commerce, immersion has a positive impact on consumers' urge to buy impulsively.

The research model is shown in Figure 1.

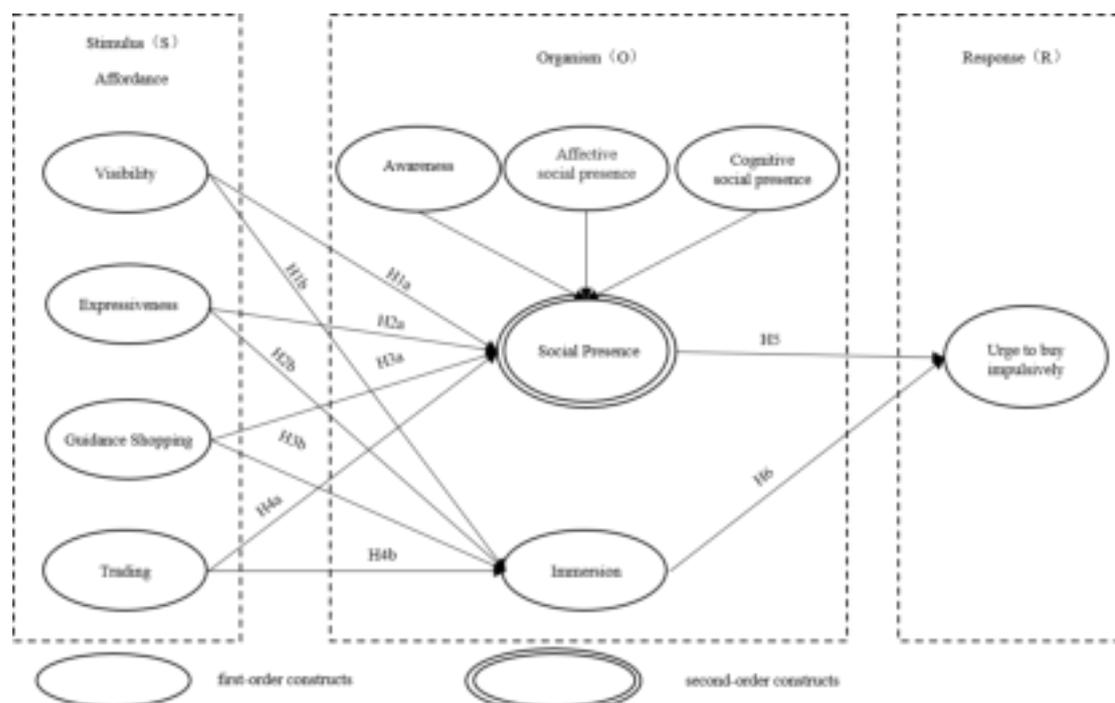


Figure 1. Conceptual model

4. DATA AND METHODOLOGY

This study mainly focuses on e-commerce platforms and social media platforms with live streaming functions, such as Taobao, Jingdong, Pinduoduo, Tik Tok etc. We adopted the method of questionnaire survey. The first part of the questionnaire is the information of the respondents, a total of 5 questions; In the second part, seven variables including affordance, social presence, immersion and impulse buying were measured, with a total of 32 items, all of which were measured by Likert 7-scale (1= strongly disagree, 7= strongly agree). Most of the measurement items in this study are based on foreign literature. In order to ensure situational adaptability, the variable items in this study were slightly adjusted for the scenarios of live streaming shopping. Considering the popularity of online shopping at present, we first released questionnaires through WeChat and Moments, and then collected large sample data through the questionnaire star platform. This survey lasted 50 days from early November 2020 to late December 2020.

A total of 454 questionnaires were collected, of which 387 were valid, with a recovery rate of 85.24%. Among all respondents, 60.4% are female and 39.6% are male. In terms of age, 7.9% were under 18 years old, 42.5% were between 19 and 25 years old, 21.8% were between 26 and 30 years old, 10.1% were between 31 and 35 years old, and 17.6% were over 36 years old. In terms of education level, 18.5% of them have high school degree or below, 24.4% have junior college degree, 41.6% have bachelor degree and 15.4 have master degree or above. In terms of user experience, 77.2% of users have placed orders in live streaming, and 73.9% of those who watch live streaming for less than 2 hours a day.

5. DATA ANALYSIS AND RESULTS

Since social presence in this study belongs to second-order formative variables, and partial least squares method based on variance (PLS-SEM) can deal with small sample data, non-normal data and high-order formative variables, we use this method to calculate the model. And SMART-PLS3.3 was used for reliability and validity analysis and hypothesis test of the model. We followed the Ou's study using a two-stage approach with a mix of repeated indicator methods and latent variable scores. First, we use first-order variables to model and obtain the

reliability and validity of the latent variables. Secondly, the first-order variables were directed to the second-order variables, and the latent variable scores of the first-order variables were obtained by using the repeated index method. Then, the latent variable score of the first order variable is put into the second order model for calculation as the measurement item of the original variable.

5.1 Measurement model

In Smart-PLS, we use Cronbach's alpha and combined reliability (CR) to measure the reliability of data. The reliability test results of this study were obtained by executing PLS algorithm on all variables, including first-order variables. Table 1 shows that Cronbach's alpha values ranged from 0.88 to 0.94, and CR values ranged from 0.93 to 0.95. Both metrics thus exceed the required value of 0.7, indicating that the construct has good reliability. In this study, average variance extraction(AVE) were used to test the convergence validity. As shown in Table 2, AVE values of all constructs range from 0.63 to 0.85, thus exceeding the acceptable level of 0.5, indicating that convergence validity is also satisfied. According to the Fornell-Larcker criterion, the AVE square roots exceed the correlation coefficients of each latent variable, which confirms discriminant validity.

Finally, we examine the multicollinearity of indicators. The VIF values of all indicators range from 2.033 to 3.759, lower than the standard of 5[27]. In addition, the VIF values of the first-order variables were 2.257, 3.171 and 3.027, lower than the 3.3[27], which indicates that there are no major multicollinearity problems.

Table 1. Cronbach's Alpha, Composite Reliability, AVE, and Correlations

	Cronbach's Alpha	CR	AVE	Furnell criterion										
				AW	CO	AF	IM	GS	UBI	EX	SP	TR	VI	
AW	0.914	0.939	0.795	0.892										
CO	0.902	0.939	0.836	0.614	0.914									
AF	0.909	0.943	0.846	0.608	0.795	0.92								
IM	0.891	0.925	0.754	0.596	0.64	0.602	0.868							
GS	0.912	0.938	0.792	0.584	0.585	0.57	0.648	0.89						
UBI	0.883	0.927	0.81	0.517	0.526	0.558	0.578	0.547	0.9					
EX	0.931	0.951	0.828	0.611	0.595	0.56	0.59	0.586	0.511	0.91				
SP	0.937	0.946	0.639	0.859	0.896	0.894	0.694	0.658	0.604	0.67	0.799			
TR	0.909	0.943	0.846	0.549	0.541	0.556	0.591	0.529	0.471	0.556	0.622	0.92		
VI	0.905	0.933	0.778	0.682	0.645	0.618	0.682	0.65	0.586	0.697	0.738	0.647	0.882	

Note: VI: visibility; EX: expressiveness; GS: guidance shopping; TR: trading; AW: awareness; AF: affective social presence; IM:

immersion; UBI: urge to buy impulsively

5.2 Common method bias

As the questionnaire survey method is adopted in this study, the common method deviation may be caused by the topic characteristics, content deviation and other factors. Therefore, in order to whether there is a problem of common method biases, we included in the PLS model a common method factor whose indicators included all the principal constructs' indicators and calculated each indicator's variances substantively explained by the principal construct and by the method[28]. The results demonstrate that the average substantively explained variance of the indicators is 0.808, while the average method based variance is 0.002. The ratio of substantive variance to method variance is large. In addition, most method factor loadings are not significant. Given the small magnitude and insignificance of method variance, we contend that the method is unlikely to be a serious concern for this study.

5.3 Structural model

In order to obtain empirical results, we bootstrap 5000 times when using Smart-PLS to calculate path

coefficient and T-values. We also ran PLS Algorithm to get R². Table 2 and Figure 2 show the structural equation model test results. First, visibility (0.334, p<0.001), expressiveness (0.208, p<0.001), guidance shopping (0.229, p<0.001) and trading (0.169, p<0.001) have positive influence on social presence, so H1a, H2a, H3a and H4a are supported. Second, visibility (0.297, p<0.001), expressiveness (0.109, p<0.05), guidance shopping (0.294, p<0.001) and trading (0.183, p<0.01) have positive impact on immersion, so H1b, H2b, H3b and H4b are supported. In addition, social presence (0.392, p<0.001), immersion (0.307, p<0.001) have positive impact on the urge for consumers to buy impulsively, so the H5, and H6 are supported.

R² represents the proportion of the dependent variable in the model that is explained^[27]. As shown in figure 2, the urge to buy impulsively R² value of 0.411, it shows that about 41.1% of the urge to buy impulsively is explained by social presence and immersion. Similarly, the R² of social presence is 0.638, and the R² of immersion is 0.561, indicating that the visibility, expressiveness, guidance shopping and trading of live streaming platforms have a good explanatory power for social presence and immersion. The overall model has above medium explanatory power.

Table 2. Hypotheses Testing Results

	Path	Path coefficients	Standard Deviation	T-values	P-values	Supported?
H1a	VI -> SP	0.334	0.054	6.158	0.000	yes
H1b	VI -> IM	0.297	0.061	4.896	0.000	yes
H2a	EX -> SP	0.208	0.045	4.659	0.000	yes
H2b	EX -> IM	0.109	0.050	2.153	0.031	yes
H3a	GS -> SP	0.229	0.043	5.387	0.000	yes
H3b	GS -> IM	0.294	0.050	5.880	0.000	yes
H4a	TR -> SP	0.169	0.048	3.560	0.000	yes
H4b	TR -> IM	0.183	0.049	3.766	0.000	yes
H5	SP -> UBI	0.392	0.046	8.429	0.000	yes
H6	IM -> UBI	0.307	0.052	5.865	0.000	yes

5.4 Post hoc assessment of mediating effects

In order to verify whether affordance influences impulse buying through social presence and immersion, the we use bootstrapping method in Smart-PLS to test the mediating effect of social presence and immersion. Table 3 shows the results of the mediating effect assessment test. It shows that social presence mediates the impacts of visibility (0.131, p<0.001), expressiveness (0.081, p<0.001), guidance shopping (0.09, p<0.001) and trading (0.066, p<0.01) on the urge to by impulsively and that immersion mediates the impacts of visibility (0.091, p<0.001), guidance shopping (0.08, p<0.001) and trading (0.056, p<0.001) on the urge to by impulsively. Meanwhile, immersion has no mediating effect of expressiveness (0.033, p<0.05) on the urge to by impulsively.

Table 3. Results from testing the mediating effects of immersion and social presence

Constructs			Indirect Effect(IV-M-DV)			Mediating Effect
IV	M	DV	Path coefficients	T-values	P-values	
VI	SP	UBI	0.131	4.853	0.000	Significant
EX	SP	UBI	0.081	3.827	0.000	Significant
GS	SP	UBI	0.090	4.485	0.000	Significant
TR	SP	UBI	0.066	3.422	0.001	Significant
VI	IM	UBI	0.091	3.652	0.000	Significant
EX	IM	UBI	0.033	1.956	0.050	Significant
GS	IM	UBI	0.080	4.000	0.000	Significant
TR	IM	UBI	0.056	3.269	0.001	Significant

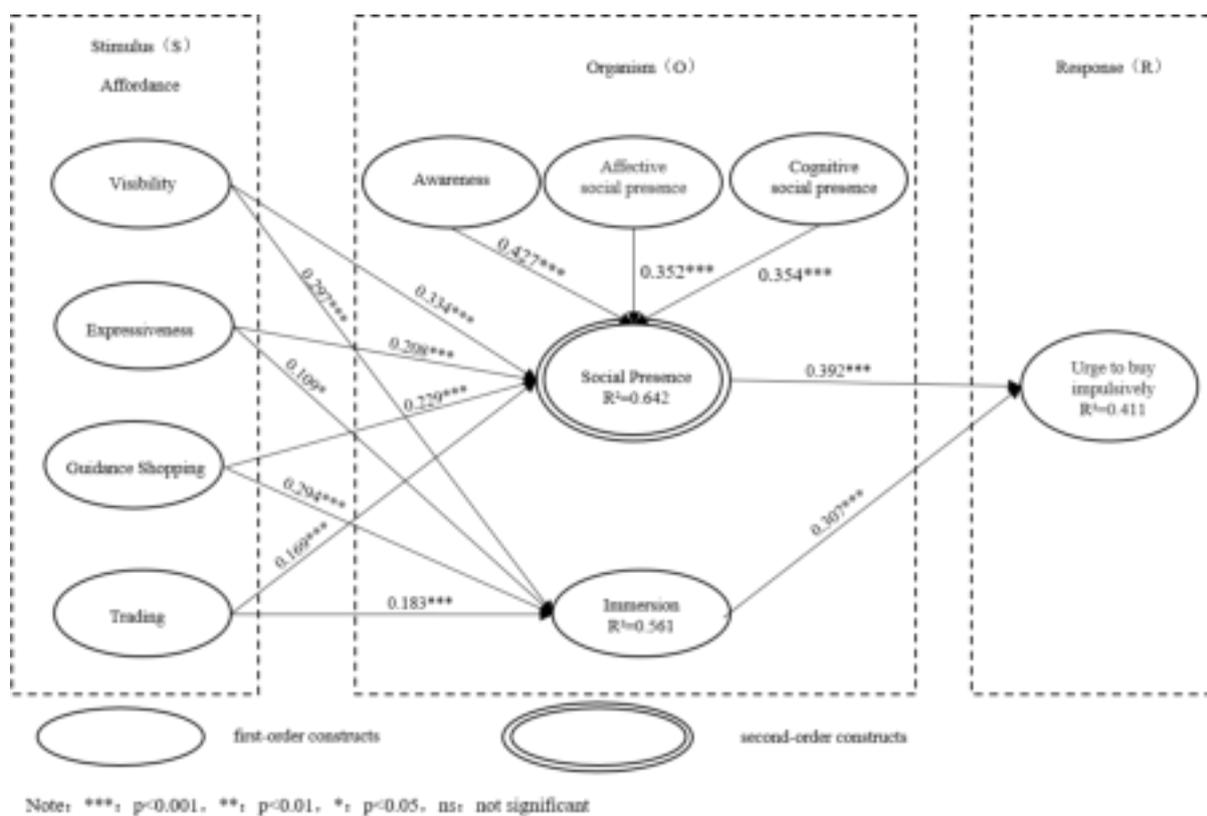


Figure 2. Model testing results

6. Conclusions and implications

This study mainly explores the impact of consumers' impulse purchase in live commerce, and proposes a theoretical framework for the impact of the affordance of live streaming platforms on consumers' impulse buying from the perspective of affordance. On this basis, combining with the existing theoretical research, We take the affordance of live streaming platform (visibility, expressiveness, guidance shopping, transactional) as independent variables, social presence (awareness, affective social presence, cognitive social presence) and immersion as mediating variables, and impulse buying behavior as dependent variables to establish a theoretical model. Then we conduct empirical analysis on the theoretical framework and the results show that all the hypotheses are valid.

6.1 Theoretical implications

Based on the perspective of affordance, this study explains the mechanism of consumers' impulse buying behavior in live commerce and provides a theoretical framework for future research. For the first time, this study attempts to define the multi-dimensional concept of social presence in the context of live streaming and divides it as a formative variable into awareness, affective social presence and cognitive social presence. This construction and analysis method is supported by past studies and empirical conclusions, and also provides an analytical approach for future multidimensional studies of social presence. Besides, our research makes up for the existing research vacancy on the impact of live streaming in social commerce on impulse buying of customers.

6.2 Practical implications

For live streaming shopping platforms, efforts should be made to improve interface visibility, communication with consumers and convenience of transaction process. In terms of visibility, live streaming shopping platforms should improve the clarity of live streaming images and adopt VR to enhance viewing experience. In view of expressiveness, live streaming shopping platforms should strive to provide consumers with a relaxed and pleasant communication environment; In terms of communication mode, the response time should be shortened so that

consumers can get a timely reply after expressing their questions or opinions. The platform should enhance the design expression of consumers' non-verbal behaviors, such as paying attention to the design of facial expressions, so that consumers can communicate their intentions more conveniently and vividly. Aiming at the guidance shopping, the platform can quickly classify a large number of bullet screens by keyword identification, so as to provide help for the personalized service of main live streaming. As for the transaction process in live streaming, the platform should enhance the security of the transaction process and simplify the transaction process, so that consumers will not weaken their positive emotions due to the complex and cumbersome transaction process.

6.3 Limitations and future research

There are still some shortcomings in this study. First, it is difficult to collect the actual purchase data of consumers and it is impossible to judge whether it is an impulse purchase only from the data, this research only adopts the form of questionnaire to investigate the impulse purchase of consumers. Therefore, future research could use experiments or the collection of actual purchasing data to study the impulse buying behavior in live streaming shopping. Secondly, this study takes social presence and immersion as mediating variables, but previous studies have shown that there may be other mediating effects between consumers' receiving of external stimuli and their impulse buying behavior. Therefore, future studies can consider whether there are other mediating variables and use longitudinal data to test the mediating effect. Thirdly, due to the complexity of the model, the study did not add regulating variables. However, in reality, consumers' impulse purchase in live e-commerce is influenced by many factors, so future research can consider personal impulse traits and commodity types as moderating variables for further study.

REFERENCES

- [1] Xin Yu, Zhengliang Xu. (2017). Research on Information Participation Behavior of Bullet Screen Users in Network Live streaming Platform: Based on the Perspective of Immersion Theory. *Information science*, 35(10): 147-151.(in Chinese)
- [2] Kim J B. (2015). The Mediating Role of Presence on Consumer Intention to Participate in a Social Commerce Site. *Journal of Internet Commerce*, 14(4): 425-454
- [3] Yu E, Jung C. (2018). Kim H, et al. Impact of viewer engagement on gift-giving in live video streaming[J]. *Telematics and Informatics*, 35(5): 1450-1460
- [4] Leonardi P.M., Huysman M.H., Steinfield C. (2013). Enterprise social media: definition, history, and prospects for the study of social technologies in organizations. *Journal of Computer-Mediated Communication*, 19 (1): 1–19.
- [5] Treem J, Leonardi M. (2013). Social media use in organizations: exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, 36 (1):143–189
- [6] Xiaoxiao Gong, Zuoliang Ye, Yuping Wu. (2019). A Study on the Influence Mechanism of Atmosphere Cues in Live Streaming Scene on Consumers' Impulsive Consumption Intention. *Chinese Journal of Management*, 16(06): 875-882.(in Chinese)
- [7] Sun Y, Shao X, Li X. (2019). How live streaming influences purchase intentions in social commerce: An IT affordance perspective. *Electronic Commerce Research and Applications*, 37.
- [8] Weinberg P, Gottwald W. (1982). Impulsive consumer buying as a result of emotions. *Journal of Business Research*, 10(1): 43-57.
- [9] Jun Fen, Mei Lu.(2020). Empirical research on impulsive purchase intention of live streaming marketing in the era of mobile Internet. *Soft Science*, 1-12.(in Chinese)
- [10] Dong X, Wang T. (2018). Social tie formation in Chinese online social commerce: The role of IT affordances. *International Journal of Information Management*, 42: 49-64.
- [11] Lin J, Luo Z, Cheng X, et al. (2019). Understanding the interplay of social commerce affordances and swift guanxi: An empirical study. *Information & Management*, 56(2): 213-224.

- [12] Lin J, Luo Z, Cheng X, et al. (2019). Understanding the interplay of social commerce affordances and swift guanxi: An empirical study. *Information & Management*, 56(2): 213-224.
- [13] Biocca F, Harms C, Gregg J. (2001). The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. *Interface & Network Design Lab*.
- [14] Shen K, Khalifa M. (2008). Exploring Multidimensional Conceptualization of Social Presence in the Context of Online Communities. *International Journal of Human-computer Interaction*, 24: 722-748.
- [15] Ying Xie, Chunqing Li, Gao Peng. (2019). Research on the Influence and Mechanism of Social Presence on Online Conformity Consumption in Live Marketing: Behavioural and Neurophysiological Perspectives. *Advances in Psychological Science*, 27(6): 990-1004. (in Chinese)
- [16] Csikszentmihalyi M. (1975). *Beyond Boredom and Anxiety*. San Francisco: Jossey-Bass, 1975: 79
- [17] Rodríguez-Ardura I, Meseguer-Artola A. (2016). E-learning continuance: The impact of interactivity and the mediating role of imagery, presence and flow. *Information and Management*, 53(4): 504-516.
- [18] Mehrabian, A. and Russell, J.A. (1974) *An Approach to Environmental Psychology*. The MIT Press, Cambridge.
- [19] Xiang L, Zheng X, Lee M K O. (2016). Exploring consumers' impulse buying behavior on social commerce platform: The role of parasocial interaction. *International Journal of Information Management*, 36(3): 333-347.
- [20] Hong Zhang, Yaobin Lu, Chunjie Xiang. (2017). Consumer Intention to Participate in Social Business: An Experience-Based Perspective. *Journal of Management Engineering*, 31(02): 40-46. (in Chinese)
- [21] Vonkeman C, Verhagen T, Van Dolen W. (2017). Role of local presence in online impulse buying. *Information & Management*, 54(8): 1038-1048.
- [22] Shaoping Zhu. (2014). Study on the influencing factors of community user acceptance behavior of cloud logistics information platform based on immersion theory. Ms D Thesis. Anhui. University of Anhui. (in Chinese)
- [23] Novak T, Hoffman D, Yung Y. (2000). Measuring the Customer Experience in Online Environments: A Structural Modeling Approach[J]. *Marketing Science*, 19: 22-42.
- [24] Pace S. (2007). A grounded theory of the flow experiences of Web users. *International Journal of Human-Computer Studies*, 60(1): 327-363
- [25] Ou C, Pavlou P, Davison R. (2014). Swift Guanxi in Online Marketplaces: The Role of Computer-Mediated Communication Technologies. *Mis Quarterly*, 38: 209-23
- [26] Lanqi Wang. (2014). Evaluation technology and application of online consumption behavior based on flow experience. *Psychology Technology and Application*, 5(1): 26-28. (in Chinese)
- [27] Hair Joseph F, Risher Jeffrey J, Sarstedt Marko J, et al. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31: 2-24
- [28] Huigang Liang, Saraf Nilesh, Qing Hu, Yajiong Xue. (2007). Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. *Mis Quarterly*, 31: 59-87

Short Research Paper**Visualizing Research Evolution in E-Commerce Field: Year 2001 to 2020***Yang Bai^{1*}, Hongxiu Li², Jianyao Chen¹*¹School of Information Management, Central China Normal University, China²Department of Information and Knowledge Management, Tampere University, Finland

Abstract: The paper aims to provide a comprehensive understanding of the evolution of the research themes and trends in e-commerce research based on the co-word analysis. 17,417 keywords were collected from 3,289 academic articles published in the selected seven leading journals in the field of e-commerce from 2001 to 2020 to examine the core research themes and trends over the two-time spans of years 2001-2010 and 2011-2022. Some research themes have kept their continuity in 20 years, indicating that the research themes in the e-commerce research have maintained temporal continuity. At the same time, the research themes in the e-commerce field have evolved with technology development and diffusion. This research provides useful suggestions for the future research direction for scholars in the e-commerce field.

Keywords: Co-word analysis, Research trends, Research theme, Research evolution, E-commerce

1. INTRODUCTION

Electronic commerce (e-commerce) refers to the transaction of goods and services by means of electronic communication^[1]. The emergence of e-commerce allows customers to ignore the obstacles of physical distance to shopping, which greatly promotes global commodity transactions. With the fast development of information technologies various types of e-commerce, such as non-business e-commerce, intra-business e-commerce, business-to-consumer (B2C), consumer-to-business (C2B), business-to-business (B2B), consumer-to-consumer (C2C), mobile-commerce, social-commerce, have been developed^[2], and attracted great attention from both researchers and practitioners.

Prior research has researched on e-commerce literature to explore the research development status and trend in e-commerce field and to provide a reference for researchers on their research topic decisions in this field. Such as some studies have studied the research trend of e-commerce in a specific theme, including identifying the dimensions of e-commerce to develop a framework for e-commerce^[3], reviewing e-commerce research from the perspective of economic analysis^[4], the journals and people central to e-commerce research^[5], and the literature characteristics and author productivity distribution in e-commerce research^[6]. Although prior research has provided an understanding of some specific research areas in e-commerce research, there is a lack of research to provide a comprehensive understanding of the development of research themes in the field based on recent literature, such as what are the dominant research themes in the field and how the research themes in the field have evolved with technology development. Thus, there is a call for research to provide a comprehensive understanding of e-commerce research based on recent literature to provide the knowledge on the core research themes to scholars in the e-commerce field and guide their future research.

To address the above research gap, 3,289 articles published in seven leading journals in the e-commerce field between 2001 and 2020 are analyzed to identify the evolution of research themes in e-commerce research.

The rest of the paper is organized as follows: Section two summarizes the past research in the area of e-commerce to provide the research background for this study. Section three provides a description of the research methods applied in this study. After that, the results of the study are presented. Finally, we discuss the

* Corresponding author. Email: baiy@mails.ccnuc.edu.cn(Yang Bai)

main research findings in this study and summarize the contributions of the current research.

2. RESEARCH BACKGROUND

In the late 1990s, the Internet had promoted the formation of e-commerce. Later web2.0 has driven the development of platform-based e-commerce such as B2C, B2B, and C2C [7]. In recent years, emerging technologies have scaled e-commerce development, such as social media, cloud computing, big data, and AI [8]. Obviously, e-commerce has been driven by the rapid advancement of technologies, which might have also steered the evolution of research themes in this field.

Some prior research has provided the knowledge structure and research evolution in the field of e-commerce based on literature. Prior research has noticed the interdisciplinary nature of e-commerce based on the review of the research trends in the e-commerce field. Ngai and Wat [9] found that inter-organizational systems, corporate strategy, and security were popular research topics in e-commerce research. Urbaczewski et al. [10] argued that e-commerce research needs cross-disciplinary theoretical support, such as theories in the fields of organization science, economics, technology, social science, behavioral study, and legal studies. DeLone and McLean [11] adapted the well-established DeLone & McLean IS Success Model to investigate the challenges in e-commerce and identified six dimensions as e-commerce success metrics based on e-commerce literature. Wareham et al. [12] studied on 582 academic articles in e-commerce field published in academic journals between 1997 and 2003, and found that B2B, strategy, B2C, trust, and technology adoption were the popular research topics in this field and survey has been a dominant research method in this field. They also posited that the research topics in this field should have been developed accordingly in e-commerce research when new technologies have been continuously applied in the field. Based on articles published at the six leading journals in the e-commerce field between 2006 and 2010, Shiau and Dwivedi [13] conducted a statistical analysis and identified five core research themes in the e-commerce field, including e-markets, acceptance and application of technology, task-related application of e-commerce, identity and evaluation, and trust.

Some scholars have applied bibliometric analysis in research to examine the research themes as well as the research trends in the e-commerce field. Lin et al. [14] reviewed the papers published in the top 10 journals in the e-commerce field and provided a general understanding of e-commerce research via identifying the most influential articles, the intellectual connection between authors and the published articles as well as the main research themes in e-commerce research in information systems (IS) domain. Lee et al [15] conducted a diversity measurement analysis of 1,103 articles in e-commerce, information systems (IS), and marketing journals between 1996 and 2005 and found that IS journals tend to rely more on economic theories whereas e-commerce journals focus more on redefining e-commerce research area. Based on a study on the 4,948 articles published in the Science Citation Index Expanded (SCIE) and 2,875 articles in the Social Science Citation Index (SSCI) from 1999 to 2008, Wang and Chen [16] found that e-commerce research in SCIE focuses more on the computer and information science, whereas that in SSCI database is more business management focused. Based on 5,429 articles in the e-commerce field published from 1996 to 2015, Tsai [6] found that the distribution of author productivity in e-commerce research does not fit Lotka's Law and e-commerce is closely related to disciplines such as economics, engineering, computer science, and information science. By analyzing over 1,000 e-commerce articles published between 1987 and 2017, Yoo and Jang [17] suggested promising research themes for future e-commerce research via a bibliographic survey. Obviously, though previous studies have provided an understanding of the research theme in the e-commerce field, there is a lack of more comprehensive research on recent literature to reflect the overall evolution of research topics and research trends in e-commerce field.

3. RESEARCH METHODS

3.1 Co-word analysis

Co-word analysis is a bibliometric method based on co-occurrence analysis to identify the subject areas that have characterized fundamental research at different time periods^[18]. Co-word analysis has advantages in discovering connections between subjects in a specific research field and tracing its scientific development. With the ability to discover knowledge in databases, co-word analysis has been widely applied to identify research topics in a specific research field, including the relationships between research topics, the extent to which these topics are central to the research field as well as are internally structured^[19].

Co-word analysis has been widely employed in the literature to detect the knowledge structure in various research fields, such as tourism research^[20], e-learning^[21], information retrieval^[22], library and information science^[23], and IS^[24]. In this study, co-word analysis is applied in line with these studies to evaluate the research themes and knowledge structure of e-commerce research.

3.2 Data

In this study, we collected the research data from 7 leading e-commerce journals. The journals in the “Business, Management, and Accounting” category in SCImago Journal Rank (SJR) 2019 were identified in order to get a complete pool of literature in the e-commerce field. We have excluded the journals which are not associated with e-commerce and finally, 7 journals were included in this study. The articles published between 2001 and 2020 were extracted. In total, 17,417 keywords were obtained from 3,289 articles collected from the 7 journals. More details are presented in Table 1.

Table 1. E-commerce journals included in this study

Journals	Starting year	Keyword availability (year)	Articles collected	Keywords collected
International Journal of Electronic Commerce	1996	2000	408	2,360
Electronic Markets	1991	2000	557	3,382
Electronic Commerce Research and Applications	2002	2002	856	4,892
Electronic Commerce Research	2001	2001	444	2,177
Journal of Electronic Commerce Research	2000	2000	363	981
Journal of Theoretical and Applied Electronic Commerce Research	2006	2006	303	1,698
Journal of Electronic Commerce in Organizations	2003	2003	358	1,927
Total	—	—	3,289	17,417

To obtain more accurate outcomes and to guarantee the diversity of topics, we standardized the keywords by merging synonyms (e.g., online market and electronic market; customer satisfaction and user satisfaction), singular and plural forms of nouns and gerunds (e.g., website and websites), and abbreviations (e.g., word of mouth and WOM; user-generated content and UGC) and filtering out some general keywords which do not represent research topics (e.g., general, online, etc.). In line with previous studies^[25], the keyword of “e-commerce” has also been filtered out since “e-commerce” is a common keyword applied in e-commerce articles. The process of keyword merging was conducted by two researchers together.

We split the sample data into two sub-datasets for the time periods of 2001–2010 and 2011–2020, aiming to compare the differences of the research themes in the two different time periods and detect the paradigm changes in the e-commerce field in these 20 years. There are 1,353 articles published between 2001–2010 and 1,936 articles in the period of 2011–2020. More articles have been published in the latter period, indicating that e-commerce has been a growing research field in the last 10 years.

Only the keywords that appeared at least five times in the studied 20 years have been retained to represent the research topics. Finally, 344 keywords (total frequency: 2,167) covering 1,054 (77.90%) of 1,353 articles

With regard to the changes of keywords in the studied two periods, 75 of the 411 keywords (18.24%) have been found to be new in the latter period of 2011-2020, suggesting the development of new research themes in addition to traditional major research themes during the past 20 years as the business, technological and social environments have evolved in the last 20 years. Such as the research topics like “social media”, “cloud computing”, “sentiment analysis” and “big data” are emerging new topics in e-commerce research in the period of 2011-2020, which reflects these topics have gained their importance in e-commerce research [27]-[28].

4.2 The evolution of e-commerce research themes

In this study, we employed CorText to generate evolution maps for research topics in the e-commerce field [29], which helps visualize the details and trajectories of themes in the past two decades (shown in Figures 3 and 4) and clearly present how the research themes have evolved in e-commerce research the studied 20 years.

As shown in Figure 3, online business development and technology application are parts of the identified major research themes which keep their continuous development in the first period of 2001-2010. Online business development embodied in the migration of e-commerce business model development and the digitalization and management in e-commerce as its core elements. For example, “B2C” business keeps as a research topic from 2001 to 2003 and merged with “network effects” and “perceived ease of use” and “perceived usefulness” of technologies in e-commerce research and merged with “online retailing” in 2006 and “accessibility” in 2009. In addition, the “B2B” business model was developed from the research topics of “website” and “internet” in 2005 as a core research topic and further developed as a couple of research themes in the following 5 years, such as “online banking services” and “mobile payment”, “online service” and “usability”, and “economic analysis” and “information privacy”, which are essential in B2B business development. Specifically, “online banking” and its “perceived risk”, and “mobile payment” have gradually been emphasized in e-commerce research after 2007 [30]. “C2C” appeared as another business model in 2006 and further developed as the “online auction” business model in 2010. Though the research topics such as “B2C”, “B2B”, “C2C” are scattered research topics, which reflects the business model development as a core research theme in the e-commerce field.

From the year 2001 to 2010, as shown in Figure 3, various technologies have been studied in e-commerce research. In 2001, some topics in regard to technologies appeared, such as “intelligent agent”, “Internet”, “online payment” and “machine learning”, but these topics have not been developed as main research themes in the e-commerce field yet. In 2002, the topic “technology adoption” gradually became prominent, and new emerging information technologies have constantly been studied in e-commerce research. For example, technology research was divided into “XML” and “search engine” in 2004. With the update of technologies, short messaging service (“SMS”), “automated negotiation system”, “blind signature”, “recommendation system”, “Web 2.0”, “multi-agent system” and “intelligent agent” gradually appeared and developed as a dominant research topic in the studied ten years.

In the period of 2011-2020, the e-commerce research has seen a diversification of research themes, but still highlights the convergence of new technology applications. With the development of web2.0 technology, “social media”, “online community” and “social network sites” have gradually emerged, and “WOM”, “UGC”, and “sentiment analysis” have also developed as research themes. As a new business model, “social commerce” emerged in 2011 and gained great attention in research. But these topics have been scattered in e-commerce research in the early period of years 2011-2020 and have not developed as a core research theme yet. Similarly, the application of “big data” also promoted the innovation of e-commerce models (2014-2016). Only in recent years, a couple of core research themes have emerged in e-commerce research. Such as “e-satisfaction” and “sharing economy” has been developed as a core research theme from 2016 to 2020 due to the popularity of “big data”, “UGC”, and “recommendation systems”.

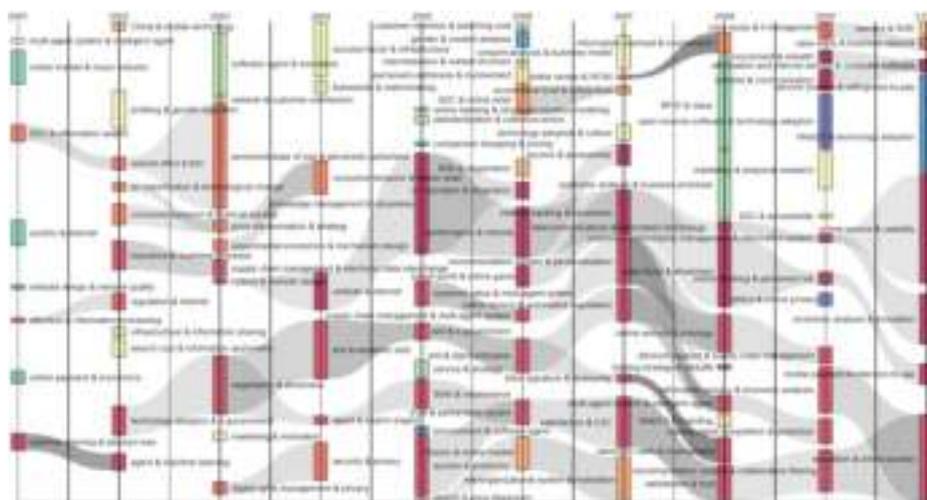


Figure 3. Research evolution in e-commerce (2001-2010)

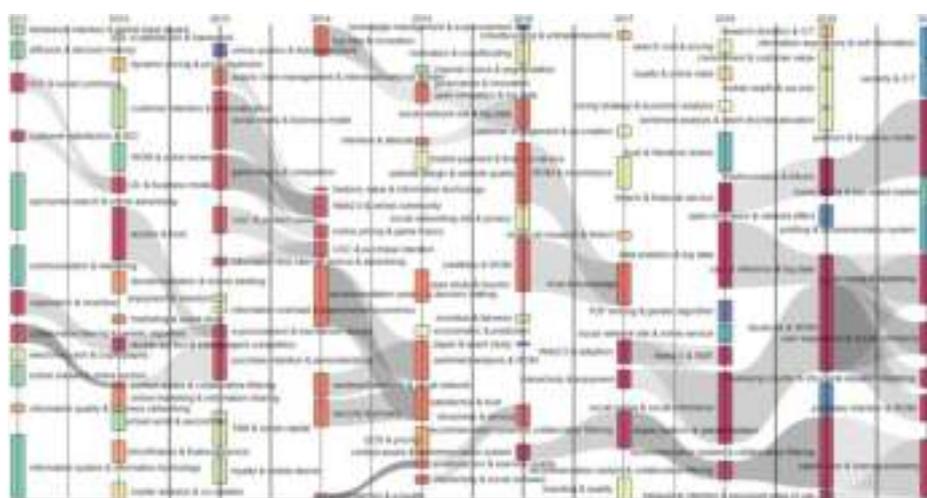


Figure 4. Research evolution in e-commerce (2011-2020)

Web 2.0 and interaction have meerged as a new research topic of the digital platform and further absorbed the topics of “big data” and “WOM”, and developed to a couple of new topics, such as “social media” and “marketing”, “user experience” and “social commerce”, “WOM” and “purchase intention”. From 2018 to 2020, the platform and business model also emerged as a new research topic. As shown in Figure 4, “consumers” has become a research focus on e-commerce, such as different consumer-related research themes are dispersed in the studied period, including “satisfaction”, “loyalty”, “trust”, “purchasing intention”, “product review”, “WOM”, “UGC”, “user experience” as well as “security” and “privacy” of users.

5. DISCUSSION AND CONCLUSIONS

This study identified the major research topics as well as their evolution in e-commerce research from 2001 to 2020 by employing co-word analysis. The research findings in this study provide some insights into the e-commerce field. First, by comparing the research topic evolution during the two periods, we found that some research themes have become a cumulative tradition in the field of e-commerce, such as the research topics of “trust”, “recommendation system”, “online retail/shopping”, “business model”, “consumer behavior”, “privacy”, “security”, and “supply chain management”. Throughout the 20 years, these research topics have retained their continuity in e-commerce research. The findings indicate that though research themes in e-commerce are also emerging and fading in the past 20 years, there is a temporary continuity of research themes in the research

field.

Simultaneously, the visualization of e-commerce research themes in the two different studied periods shows the characteristics of the research themes at that time. In the first phase, 2001-2010, “e-commerce business model” seemed to receive more attention in e-commerce research, and “B2B”, “B2C”, and “C2C” research has been in combination with other research themes, while this phenomenon dimmed in 2011-2020. This might be due to the fact that the e-commerce business model has already become mature in 2001-2010 and no need for much research on these topics. This phenomenon also occurs in “XML” technology and “Internet”, which have lost their original core position in the last 10 years due to their common application. Instead, the application of new technologies in e-commerce has become a core research theme in the e-commerce field. In 2011-2020, “social media” has become a hot research topic since understanding consumers, such as their behavior, should rely more on their “UGC”, “WOM” in social media ^[31]. In addition, e-commerce is also developing and innovating itself through leading-edge technologies such as “big data”, “mobile platforms” and “cloud computing”.

In this paper, we applied the co-word analysis method to reveal the distribution of research topics and the evolution process of research themes based on the articles published in the leading e-commerce journals in the past 20 years. The findings in this study imply that new technologies are being integrated into e-commerce and have absorbed into the existing knowledge landscape in e-commerce research to advance the knowledge structure of the field. The findings of the study complement and refine the work on the status of e-commerce research based on recent literature and should help scholars obtain an in-depth understanding of the research in the past 20 years and guide their future research in the e-commerce field.

ACKNOWLEDGEMENT

This research was supported by the National Social Science Foundation of China under Grant 19ZDA345.

REFERENCES

- [1] Tian, Y., & Stewart, C. (2008). History of e-commerce. In *Electronic commerce: concepts, methodologies, tools, and applications*. Hershey: IGI Global, 1-8
- [2] Esmaeili, L., & Hashemi G, S. A. (2019). A systematic review on social commerce. *Journal of Strategic Marketing*, 27(4), 317-355
- [3] Michael J. Shaw, David M. Gardner, Howard Thomas (1997). Research opportunities in electronic commerce, 21(3), 149-156
- [4] Kauffman, R. J. , & Walden, E. A. . (2001). Economics and electronic commerce: survey and directions for research. *International Journal of Electronic Commerce*, 5(4), 5-116
- [5] Hsu, Chien-Lung, Westland, J. Christopher, Chiang, Chun-Hao (2015). Editorial: Electronic Commerce Research in seven maps. *Electronic Commerce Research*, 15(2), 147-158
- [6] H. Tsai (2015). The research trends forecasted by bibliometric methodology: a case study in e-commerce from 1996 to July 2015. *Scientometrics*, 105(2), 1079-1089
- [7] Yu, L. (2019). E-Commerce Models, Players, and Its Future. In *Advanced Methodologies and Technologies in Digital Marketing and Entrepreneurship*. Hershey: IGI Global, 193-204
- [8] Ilmudeen, A. (2021). Big Data, Artificial Intelligence, and the Internet of Things in Cross-Border E-Commerce. *Cross-Border E-Commerce Marketing and Management*, 257-272
- [9] Ngai, E. W. T., & Wat, F. K. T. (2002). A literature review and classification of electronic commerce research. *Information and Management*, 39(5), 415-429

- [10] Urbaczewski, A., Jessup, L. M., & Wheeler, B. (2002). Electronic commerce research: A taxonomy and synthesis. *Journal of Organizational Computing and Electronic Commerce*, 12(4), 263-305
- [11] DeLone, W. H., & McLean, E. R. (2004). Measuring e-commerce success: Applying the DeLone and McLean Information Systems Success Model. *International Journal of Electronic Commerce*, 9(1), 31-47
- [12] Wareham, J., Zheng, J. G., & Straub, D. (2005). Critical themes in electronic commerce research: A meta-analysis. *Journal of Information Technology*, 20(1), 1-19
- [13] Shiau, W. L., & Dwivedi, Y. K. (2013). Citation and co-citation analysis to identify core and emerging knowledge in electronic commerce research. *Scientometrics*, 94(3), 1317-1337
- [14] Lin, A. J., Hsu, C. L., & Chiang, C. H. (2016). Bibliometric study of Electronic Commerce Research in Information Systems & MIS Journals. *Scientometrics*, 109(3), 1455-1476
- [15] Lee, S. M., Hwang, T., & Kim, J. (2007). An analysis of diversity in electronic commerce research. *International Journal of Electronic Commerce*, 12(1), 31-67
- [16] Wang, C., & Chen, C. (2010). Electronic Commerce Research in Latest Decade: A Literature Review. *International Journal of Electronic Commerce Studies*, 1(1), 1-14
- [17] Yoo, B., & Jang, M. (2019). A bibliographic survey of business models, service relationships, and technology in electronic commerce. *Electronic Commerce Research and Applications*, 33, 100818
- [18] Callon, M., Courtial, J.-P., Turner, W. A., Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191-235
- [19] He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, 48(1), 133-159
- [20] De la Hoz-Correa, A., Muñoz-Leiva, F., & Bakucz, M. (2018). Past themes and future trends in medical tourism research: A co-word analysis. *Tourism Management*, 65, 200-211
- [21] Bai, Y., Li, H., & Liu, Y. (2021). Visualizing research trends and research theme evolution in E-learning field: 1999–2018. *Scientometrics*, 126(2), 1389–1414
- [22] Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management*, 37(6), 817-842
- [23] Hu, C. P., Hu, J. M., Deng, S. L., & Liu, Y. (2013). A co-word analysis of library and information science in China. *Scientometrics*, 97(2), 369-382
- [24] Liu, Y., Li, H., Goncalves, J., Kostakos, V., & Xiao, B. (2016). Fragmentation or cohesion? Visualizing the process and consequences of information system diversity, 1993-2012. *European Journal of Information Systems*, 25(6), 509-533
- [25] He, Q., Wang, G., Luo, L., Shi, Q., Xie, J., & Meng, X. (2017). Mapping the managerial areas of Building Information Modeling (BIM) using scientometric analysis. *International journal of project management*, 35(4), 670-685
- [26] Landauer, T. K., Peter W. Foltz, & Darrell Laham. (2009). An Introduction to Latent Semantic Analysis. *Behavior Research Methods*, 41(3), 944-950
- [27] Javid, E., Nazari, M., & Ghaeli, M. (2019). Social media and e-commerce: A scientometrics analysis. *International Journal of Data and Network Science*, 3(3), 269-290
- [28] Norian, J. H., Jama, A. M., Eltaieb, M. H., & Adam, A. A. (2020). Data-Driven E-commerce Techniques and Challenges in the Era of the Fourth Industrial Revolution. *Scientific Journal of Informatics*, 7(2), 291-302
- [29] Leydesdorff, L., & Goldstone, R. L. (2014). Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal *Cognitive Science*. *Journal of the Association for Information Science and Technology*, 65(1), 164-177
- [30] Ngai, E. W., & Gunasekaran, A. (2007). A review for mobile commerce research and applications. *Decision support systems*, 43(1), 3-15
- [31] Zhao, Y., Wang, L., Tang, H., & Zhang, Y. (2020). Electronic word-of-mouth and consumer purchase intentions in social e-commerce. *Electronic Commerce Research and Applications*, 41, 100980

Short Research Paper**Understanding User Contributions in Smoking cessation****Online Health Communities***Chenglong Li^{1*}, Hongxiu Li², Yuting Jiang³, Reima Suomi¹*¹Turku School of Economics, University of Turku, Turku, Finland²Department of Information and Knowledge Management, Tampere University, Tampere, Finland³School of Information Management, Wuhan University, Wuhan, China

Abstract: Users make contributions to online communities in different ways. Prior literature has rarely investigated how different user groups make contributions to smoking cessation OHCs. To illuminate the contribution of different user groups in smoking cessation OHCs, this study aims to evaluate user contribution from two dimensions (Content-contribution and popularity) associated with users' questioning and answering behaviors. Based on the user log data collected from a smoking cessation OHC in Finland (Stumppi.fi), we plan to assess user contribution level for four different user groups (lurker, conversation-starter, conversation-replier, and Conversation-starter & replier) based on user activity data via applying entropy method. The research might make theoretical contributions to the literature on user contribution and offer practical implications to smoking cessation OHC service providers.

Keywords: user contribution, smoking cessation, online health community

1. INTRODUCTION

Smoking is still an important health concern around the world, with the World Health Organization estimating that over 8 million deaths are the result of direct tobacco use ^[1]. Recently, smoking cessation online health communities (OHCs) have become a popular platform to assist smokers who want to quit their smoking habits to achieve abstinence ^[2, 3]. In such OHCs, users can start a conversation to seek help, to share smoking cessation experience and tips, and to respond to others' postings. This encourages the creation and exchange of user-generated content (UGC). However, in OHCs, most UGCs are contributed by a small number of key users ^[4]. The majority of users never or only occasionally contribute UGC to the OHCs ^[5]. Understandably, most of prior studies on user contributions focus on the content contribution of users in OHCs as UGC is important for the sustainable development of OHCs. The other user engagement activities in OHCs have been largely ignored in user contribution research, such as reading, following, sharing, and voting ^[6]. From an ecosystem perspective, in OHCs like smoking cessation online communities, different users play different roles and make contributions in different ways, thereby complementing each other, and maintaining the ecosystem. Such as lurkers in smoking cessation online communities make contributions to the online communities mainly via reading, following, voting, or sharing UGC with other users though they make no content (UGC) contribution to the online communities at all. Therefore, ignoring the other forms of user contributions other than content contribution might hinder our comprehensive understanding of user contribution in smoking cessation OHCs.

Although recent research has investigated distinct types of users with their own patterns of engagement in OHCs, such as posters and lurkers ^[5, 7], the understanding of contributions from different user groups is still unclear. The multiple functions of OHCs allow users to contribute to OHCs not only by lurking and posting, but also by following a topic or voting for a topic like thumbs-up. We still lack a comprehensive understanding of how users' different activities in smoking cessation OHCs contribute to the communities. For instance, how do

* Corresponding author. Email: chenglong.li@utu.fi (Chenglong Li)

users who do not generate UGC contribute to the OHCs? Thus, an in-depth investigation on user contributions to OHCs via analyzing users' different behavior patterns in OHCs is needed to get a full picture of the contribution of all the users in smoking cessation OHCs, not only the UGC contributors as posters, but also lurkers who never make or reply a request in smoking cessation OHCs.

To address the above research gap, this study will investigate user contributions based on users' different activity behaviors in smoking cessation OHCs. Based on the features of the smoking cessation OHCs and associated user activity behaviors, we aim to investigate user contributions among four different user groups (lurker, conversation-starter, conversation-replier, and Conversation-starter & replier) from two dimensions, namely content-contribution and popularity. The research model will be empirically tested through log data collected from a smoking cessation OHC in Finland (Stumppi.fi).

Specifically, based on the four types of user behaviors in Stumppi.fi, we suggest two dimensions of user contributions in smoking cessation OHCs: content-contribution and popularity. And we categorize the Stumppi.fi users into four groups in accordance with their behaviors of initiating a conversation or responding to others: lurker, conversation-starter, conversation-replier, and Conversation-starter & replier. Log data from Stumppi.fi are used in this study to empirically test our proposed model in explaining how different users contribute to the OHC with different activity behaviors.

The rest of this paper is organized as follows. Section 2 provides a literature review on smoking cessation OHCs and user contribution. Then, Section 3 presents the research method that will be employed in this research, including a description of the variables included in the user contribution assessment model, the methods for data analysis, and the data collection plan. Finally, we discuss the potential theoretical and practical implications and highlight the limitations of the current study and future research directions.

2. LITERATURE REVIEW

2.1. Research on smoking cessation OHCs

Extant studies have been conducted to investigate smoking cessation OHCs. These studies have revealed two main research streams. The primary research stream focuses on the effectiveness of such OHCs for smoking cessation. This research stream often conducts experiments to test whether the usage of smoking cessation OHCs leads to positive smoking cessation outcomes. For instance, Graham et al. [2] found that people who use smoking cessation OHCs are more likely to stop use tobacco products in three months than non-users. In a Facebook OHC for young adult smokers, 12-month usage was found to be associated with increased attempts to quit and decreased cigarette consumption [8].

The other research stream focuses on social support in smoking cessation OHCs. This research stream typically analyzes UGC in smoking cessation OHCs to identify social support provided to users via content analysis. Informational support, emotional support, and companionship activities have been found to be common in smoking cessation OHCs [9, 10]. This is consistent with research findings in OHCs regarding various health concerns, such as breast cancer [11], Autism Spectrum Disorders [12], and HIV/AIDS [13]. In addition, some studies have applied UGC to detect users' smoking status. For instance, Wang et al. [3] designed an approach to identify smoking status by applying machine learning techniques to classify UGC in smoking cessation OHCs.

Summarizing these research streams, we find that few studies have investigated how different users contribute to smoking cessation OHCs. Previous studies on online communities mostly examined posters' behavior, with less attention paid to other types of users, such as lurkers. In order to ascertain whether all users contribute to smoking cessation OHCs in different ways, it is necessary to assess how different users make contributions to smoking cessation OHCs based on their activities in smoking cessation OHCs.

2.2. User contribution

Many studies have been conducted to investigate different roles of users in OHCs, mainly focusing on posters and lurkers^[4, 5, 7]. For instance, the research by van Mierlo^[4] has confirmed that the 1% rule can explain the participatory patterns in OHCs. In other words, 90% of users lurk and do not contribute, 9% of users contribute sparingly, and 1% of users contribute the vast majority of UGC in OHCs. In the work of Yang et al.^[5], perceived social support from OHCs have been found to exert different influences on commitment between posters and lurkers. Specifically, both perceived recognition for contribution and perceived freedom of expression have a stronger positive influence on the commitment of lurkers than that of posters.

However, this dichotomy unable to capture the complementary nature of different user contributions. In a smoking cessation OHC, postings only reveal the content contribution of users. The more important in online communities is that the posts will be read, shared, followed or voted by other users. In other words, how the posts will influence other users is more important than generating posts. In addition, users can not only post messages, but also offer feedback to posters' postings via different functions provided by the OHC. For instance, a smoking cessation OHC might have an "Add to Favorites" button. Some users who never post messages might use this function to mark and collect their favorite topics purposively. This kind of behavior has often been viewed as lurking in prior studies^[6, 14]. Lurking has been found to be a significant form of user engagement^[15]. Lurkers could also make contributions to the OHCs as they are "consumers" of the UGC generated by active users^[5, 15]. They can contribute to the OHCs in their own ways, such as voting on content, following topics, or adding a poster to favorites^[6].

In order to understand how people participate in and contribute to online communities, some researchers have moved further beyond the dichotomy of posting and lurking. For instance, Deng et al.^[6] have applied the concept of immersion from gaming field to study user contributions in SQA, and defined immersion as the degree to which a user involves in an SQA in which content is a focus. They categorized users of an SQA into four different user groups: questioners who raise questions, answers who offer answers, lurkers who only read postings, and questioner-answers who offer both questions and answers^[6]. They proposed three dimensions to measure user contributions^[6]: (1) content-contribution, refers to a state of deep involvement in an SQA via generating contents, such as asking or answering a question. (2) Activeness, refers to a state of influencing users' opinions by interacting with other users, such as following a topic or a user, and adding a favorite to a topic. (3) Influence, refers to a state of influencing other users in an indirect way, such as voting and thanking. Through analyzing user activity data of an SQA, they found that lurkers have higher community-immersion scores than questioners who only ask questions, indicating that lurkers contribute more than questioners^[6]. In the context of OHCs, Carron-Arthur, Ali, Cunningham and Griffiths^[16] conducted a systematic review and identified 41 participation patterns, from topic-focused responders who respond to others' requests rather than initiate posts to influential users who are leaders in OHCs. They summarized three dimensions to identify user participation patterns in OHCs, including content-based dimension, activity-based dimension, and network-based dimension. In the content-based dimension, prior research mainly applied content generation patterns as metrics to evaluate users' content-contribution, such as information seekers and providers, emotional support seekers and providers^[11]. In the dimension of activity, users' various online activities (e.g., time logged in, read, post, and thread initiation) are considered as metrics to identify engagement types, such as caretaker and discussant^[17]. In terms of network-based dimension, the social network analysis (SNA) metrics (e.g., out-degree, in-degree) have been used to categorize user types, such as key players, moderate users, and takers^[18, 19].

This study aims to quantify the contribution of all users of smoking cessation OHCs. Such OHCs provide a variety of interaction functions, such as posting messages, response, comments, and favorites, to exchange

information and social support ^[9, 20]. Taking Stumppi.fi as an instance, users can “create a conversation” to disclose current situation and ask for assistance, other users can “reply” by offering informational and emotional support. Some users can purposely select their favorite conversation by clicking the “add to favorite” button. Recognizing the features of the smoking cessation OHC environment as a socialized helping and supporting community and the importance of user involvement for OHCs, we applied two dimensions of user contributions: Content-contribution and popularity, which are all closely related to user involvement. Specifically, we take content-contribution dimension to represent the state of visible and active engagement of OHC users via starting conversations or replying to posts to exchange social support ^[16, 21]. Besides, popularity reflects a state of being influenced by UGC in a way of browsing, reading, and setting personal favorites, thereby affecting such elements as the number of audiences or UGC consumers ^[15, 22].

3. RESEARCH METHOD

3.1. Study case

The research case is a smoking cessation OHC in Finland: Stumppi.fi (<https://stumppi.fi/>). Stumppi.fi is an Internet-based portal for assisting smokers to achieve abstinence. It was initially funded by Finland’s government-supervised not-for-profit Slot Machine Association in 2004, but since January 2017 the funding has come from the Ministry of Social Affairs and Health and the Funding Centre for Social Welfare and Health Organizations. The language of Stumppi.fi is Finnish. It has operated an OHC for ex-smokers and current smokers in Finland since 2007. It had more than 9,000 registered members by the end of 2018.

Unlike general social network sites (SNS), such as Facebook and Instagram, Stumppi.fi does not offer too many functions (e.g., vote, like, and share) for users to interact with others. A simple and effective operation interface is helpful for users to focus on social support they need and avoid social overload ^[23]. The functions provided by Stumppi.fi include reading posts, creating a new conversation, replying to others’ posts, and adding conversations to personal favorites. Figure 1 shows the front page of the OHC in Stumppi.fi.

Before the data collection, we have obtained an ethical permit from the Ethics Committee of the home university of the authors. With the help of the staff of Stumppi.fi, we have obtained the log data of registered users of Stumppi.fi from 2007 to 2018. Stumppi.fi maintains a complete transactional history of all online events, such as asking questions and responding to others in the threads. This provides literal evidence of user contributions in the OHC, such as reading posts, asking questions, and offering answers to others’ questions. All identifiable information was deleted to protect users’ privacy. In total, we have obtained 5,131 users, 4,514 threads, and 193,089 posting messages.

3.2. The model for assessing user contribution

In Stumppi.fi, users can exchange social support through following ways: (1) create a new conversation to start a discussion, (2) reply to others’ conversations, (3) read others’ conversations and corresponding replies, (4) send private messages to others, and (5) set others’ conversations as favorite topics. Based on these functions provided by Stumppi.fi, we select 6 user-behavior attributes related to user behaviors for the measurements in our proposed contribution-assessment model. Due to privacy issues, we exclude the private message function in our study. Table 1 presents the details of these attributes.

We associate each of these 6 attributes with more appropriate of the two dimensions of user contribution: content-contribution and popularity.

First, on the content-contribution dimension, referring to generating and sharing content in Stumppi.fi ^[16, 21], two attributes, including conversation created and replies, are selected as the indicators for the model we propose here.

Second, on the popularity dimension, referring to being influenced in a way of reading and setting personal favorites [15, 22]. In this context, we intend to measure the degree to which audiences consume and appreciate the UGC provided by posters, we consider reads count, participant count, and replies count and favorites as proxies for the popularity in Stumppi.fi. Figure 2 presents the model for assessing user contribution.

Table 1. Items of behavior data and their explanations

Items	Explanation
Conversation created	The number of conversations that a user has created.
Replies	The number of replies that a user posted.
Reads count	The total number of reads of conversations that a user has created.
Participants count	The total number of participants of the conversations that a user has created.
Replies count	The total number of replies of the conversations that a user has created.
Favorites	The number of conversations that a user marked as “favorite”. The user can directly join in the marked conversation by clicking the favorite topics button.

3.3. Classification of Stumppi.fi users

We classify Stumppi.fi users into four categories on the basis of functions of seeking and sharing social support via conversations and replies – that is, in terms of the conversation-creating and conversation-replying activities of users. These are (1) conversation-starters, refers to those who start a conversation by creating a new discussion. (2) Conversation-repliers, refers to those who only offer replies to others’ conversations. (3) Lurkers, refers to those do not post messages. And (4) Conversation-starters & repliers, refers to those generate content by both starting and responding to conversations. We assume that creating conversations and responding to them are the core methods by which users share social support and help others in the smoking cessation OHC. To a certain extent, the number of conversation created by users reflects their demands for personal disclosure and social support acquisition, and the number of replies posted by users reflects their desire to respond to others’ requests and offer informational and emotional support. The major difference between the creators and respondents is the users’ preference regarding obtaining and offering social support in smoking cessation OHCs. Lurkers refer to users who do not generate content, but contribute the OHCs via other behaviors such as reading and following other’s postings [5, 24]. Table 2 presents the descriptions for our four user classes.

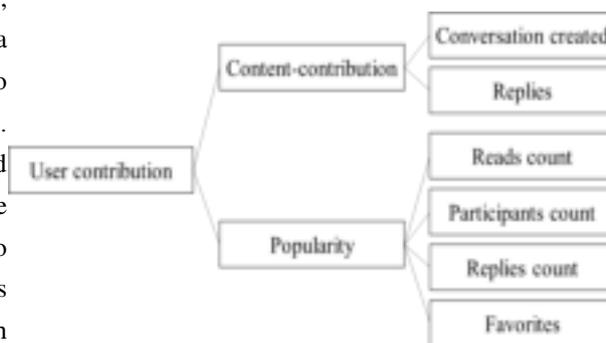


Figure 2. The model for assessing contribution of Stumppi.fi users

Table 2. The classification of users

Class of users	Description
Conversation-starters	Number of conversations a user started > 0; Number of a user’s replies = 0
Conversation-repliers	Number of conversations a user started = 0; Number of a user’s replies > 0
Lurkers	Number of conversations a user started = 0; Number of a user’s replies = 0
Conversation-Starters & repliers	Number of conversations a user started > 0; Number of a user’s replies > 0

3.4. Calculation of user contribution

Entropy method is an objective weighting method, which determines the weight of each index according to the information provided by the observed value of each index. We can use entropy value to judge the dispersion degree of an index. The greater the dispersion degree of an index, the greater the influence of this index on the

comprehensive evaluation. Specifically, for an index, the greater the gap between each observation value of the index, the greater the role of the index in the comprehensive evaluation. The smaller the difference between the observed values of the index, the smaller the role of the index in the comprehensive evaluation. If all the observed values of an index are equal, the index has no effect in the comprehensive evaluation. In order to objectively assign index weights and avoid errors caused by human factors, we plan to use entropy method to calculate index weights in this study.

The realization process of entropy method is as follows:

(1) Constructing data matrix:

$$S = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}_{n \times m}$$

Note: x_{ij} is the observed value of the j th index in the i th sample.

(2) Non-negative processing of data

If there is a negative number in the data, the data need to be non-negative processing. In addition, in order to avoid the meaninglessness of logarithms in entropy calculation, data translation is required.

For bigger, better indicators:

$$x'_{ij} = \frac{x_{ij} - \min(x_{1j}, x_{2j}, \dots, x_{nj})}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})} + 1$$

($i = 1, 2, \dots, n; j = 1, 2, \dots, m$)

For smaller, better indicators:

$$x'_{ij} = \frac{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - x_{ij}}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})} + 1$$

($i = 1, 2, \dots, n; j = 1, 2, \dots, m$)

(3) Calculate the entropy value of the j th index:

$$e_j = -k \sum_{i=1}^n P_{ij} \log(P_{ij})$$

Where $P_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}$ ($j = 1, 2, \dots, m$). So k is equal to $1/\ln m$.

(4) Calculate the coefficient of difference.

$$g_j = 1 - e_j$$

Note: the larger the g_j , the more important the index.

(5) Calculate index weight.

$$w_j = \frac{g_j}{\sum_{j=1}^m g_j} \quad (j = 1, 2, \dots, m)$$

(6) Calculate the combined score.

$$S_i = \sum_j^m w_j * P_{ij} \quad (i = 1, 2, \dots, n)$$

4. POSSIBLE IMPLICATIONS AND LIMITATIONS

Our study may provide theoretical contributions to the literature regarding user contribution in online communities. First, this study attempts to investigate the contribution of all the users of a smoking cessation OHC from two dimensions (content-contribution and popularity) based on their activities in smoking cessation OHCs, including conversation-starters, conversation-repliers, lurkers, and Conversation-starters & repliers. Our findings may add new insights into user contribution in OHCs via providing detailed explanations on how different users contribute to OHCs from their popularity and content-contribution in smoking cessation OHCs. As we know, this study might be the first study to use all user activity data in smoking cessation OHCs to explain user contribution in OHCs, which will provide research findings with high validity and reliability.

Second, this study aims to investigate the contribution of different users from a role complementarity perspective. Different from prior studies focusing on users who generate content while ignoring other contributors such as lurkers, this study considers all different user groups, including lurkers and posters. Our findings might offer a comprehensive picture of how different user groups complement each other in smoking cessation OHCs.

Third, this study purposes to explain the user contributions in smoking cessation OHCs by adopting a user contribution assessment model derived from the characteristic features of smoking cessation OHCs. Our research might offer an approach to evaluate users' engagement in smoking cessation OHCs.

This study may also offer practical implications to smoking cessation OHCs service providers on individual users' engagement and their contributions to the OHCs. First, the findings of different forms of contributions by different user groups might assist service providers to adopt customized strategies to promote the contributions. Second, the user contribution assessment model might offer a tool to monitor users' engagement in different dimensions, and then develop personalized management for users based on their user contribution scores. Third, smoking cessation OHC service providers should view smoking cessation OHCs as an ecosystem as different users play different roles and support each other in the OHCs. Service providers should pay attention to the user contribution of both positive users (content generators) and passive users (lurkers).

This research has following limitations. Firstly, the data are only collected from a smoking cessation OHC in Finland (Stumppi.fi). Future work could replicate this study in countries with different cultures to increase the generalizability of the research findings. Second, this study is only conducted in the context of smoking cessation OHCs. Future research could consider other OHC contexts, such as OHCs focusing on alcohol, cancer, and fit, to test the proposed contribution-assessment approach in different OHCs.

REFERENCES

- [1] World Health Organization. WHO report on the global tobacco epidemic 2019. 2019. Available from: <https://www.who.int/teams/health-promotion/tobacco-control/who-report-on-the-global-tobacco-epidemic-2019&publication=9789241516204>.
- [2] Graham, A.L., G.D. Papandonatos, B. Erar, and C.A. Stanton, Use of an online smoking cessation community promotes abstinence: Results of propensity score weighting. *Health Psychology*, 2015. 34(0): p. 1286-1295.
- [3] Wang, X., K. Zhao, S. Cha, M.S. Amato, A.M. Cohn, J.L. Pearson, G.D. Papandonatos, and A.L. Graham, Mining user-generated content in an online smoking cessation community to identify smoking status: A machine learning approach. *Decision Support Systems*, 2019. 116: p. 26-34.
- [4] van Mierlo, T., The 1% rule in four digital health social networks: an observational study. *Journal of Medical Internet Research*, 2014. 16(2): p. e33.
- [5] Yang, X., G.X. Li, and S.S. Huang, Perceived online community support, member relations, and commitment: Differences between posters and lurkers. *Information & Management*, 2017. 54(2): p. 154-165.
- [6] Deng, S., Y. Jiang, H. Li, and Y. Liu, Who contributes what? Scrutinizing the activity data of 4.2 million Zhihu users via

- immersion scores. *Information Processing & Management*, 2020. 57(5): p. 102274.
- [7] Han, J.Y., J.H. Kim, H.J. Yoon, M. Shim, F.M. McTavish, and D.H. Gustafson, Social and psychological determinants of levels of engagement with an online breast cancer support group: posters, lurkers, and nonusers. *J Health Commun*, 2012. 17(3): p. 356-71.
- [8] Ramo, D.E., J. Thrul, K. Chavez, K.L. Delucchi, and J.J. Prochaska, Feasibility and Quit Rates of the Tobacco Status Project: A Facebook Smoking Cessation Intervention for Young Adults. *Journal of Medical Internet Research*, 2015. 17(12): p. e291.
- [9] Zhang, M. and C.C. Yang, Using content and network analysis to understand the social support exchange patterns and user behaviors of an online smoking cessation intervention program. *Journal of the Association for Information Science and Technology*, 2015. 66(3): p. 564-575.
- [10] Rocheleau, M., R.S. Sadasivam, K. Baquis, H. Stahl, R.L. Kinney, S.L. Pagoto, and T.K. Houston, An observational study of social and emotional support in smoking cessation Twitter accounts: content analysis of tweets. *Journal of Medical Internet Research*, 2015. 17(1): p. e18.
- [11] Wang, X., K. Zhao, and N. Street, Analyzing and Predicting User Participations in Online Health Communities: A Social Support Perspective. *Journal of Medical Internet Research*, 2017. 19(4): p. e130.
- [12] Mohd Roffeei, S.H., N. Abdullah, and S.K. Basar, Seeking social support on Facebook for children with Autism Spectrum Disorders (ASDs). *International Journal of Medical Informatics*, 2015. 84(5): p. 375-85.
- [13] Flickinger, T.E., C. DeBolt, A.L. Waldman, G. Reynolds, W.F. Cohn, M.C. Beach, K. Ingersoll, and R. Dillingham, Social Support in a Virtual Community: Analysis of a Clinic-Affiliated Online Support Group for Persons Living with HIV/AIDS. *AIDS and Behavior*, 2017. 21(11): p. 3087-3099.
- [14] Cranefield, J., P. Yoong, and S.L. Huff, Rethinking lurking: Invisible leading and following in a knowledge transfer ecosystem. *Journal of the Association for Information Systems*, 2015. 16(4): p. 213-247.
- [15] Edelman, N., Reviewing the definitions of "lurkers" and some implications for online research. *Cyberpsychol Behav Soc Netw*, 2013. 16(9): p. 645-9.
- [16] Carron-Arthur, B., K. Ali, J.A. Cunningham, and K.M. Griffiths, From Help-Seekers to Influential Users: A Systematic Review of Participation Styles in Online Health Communities. *Journal of Medical Internet Research*, 2015. 17(12): p. e271.
- [17] Jones, R., S. Sharkey, J. Smithson, T. Ford, T. Emmens, E. Hewis, B. Sheaves, and C. Owens, Using metrics to describe the participative stances of members within discussion forums. *Journal of Medical Internet Research*, 2011. 13(1): p. e3.
- [18] Cobb, N.K., A.L. Graham, and D.B. Abrams, Social network structure of a large online community for smoking cessation. *American Journal of Public Health*, 2010. 100(7): p. 1282-9.
- [19] Bambina, A., *Online social support: the interplay of social networks and computer-mediated communication*. 2007: Cambria press.
- [20] Naslund, J.A., S.J. Kim, K.A. Aschbrenner, L.J. McCulloch, M.F. Brunette, J. Dallery, S.J. Bartels, and L.A. Marsch, Systematic review of social media interventions for smoking cessation. *Addictive Behaviors*, 2017. 73: p. 81-93.
- [21] Ballantine, P.W. and R.J. Stephenson, Help me, I'm fat! Social support in online weight loss networks. *Journal of Consumer Behaviour*, 2011. 10(6): p. 332-337.
- [22] Giermindl, L., F. Strich, and M. Fiedler, How do they differ? Analyzing the motivations of posters and lurkers for participation in enterprise social networks. *Lost in Digital Transformation? The role of Enterprise Social Networks in facilitating digital collaboration*, 2018. 19(2): p. 89-120.
- [23] Maier, C., S. Laumer, A. Eckhardt, and T. Weitzel, Giving too much social support: social overload on social networking sites. *European Journal of Information Systems*, 2015. 24(5): p. 447-464.
- [24] Lai, H.-M. and T.T. Chen, Knowledge sharing in interest online communities: A comparison of posters and lurkers. *Computers in Human Behavior*, 2014. 35: p. 295-306.

Short Research Paper

Are Live Streamers Similar or Different? Grid Typology Model of Live Streamers in E-commerce

Dingtian Zhang¹, Yi Wang^{2}, Qinyi Zhou³*

¹School of Business Administration, Southwestern University of Finance and Economic, Chengdu, 611130, China

²School of Business Administration, Southwestern University of Finance and Economic, Chengdu, 611130, China

³Andrew and Erna Viterbi School of Engineering, University of Southern California, Los Angeles, 90089, USA

Abstract: As a new type of consumption in the digital economy era, live streaming shopping has become popular among e-commerce consumers. The success of live streaming e-commerce is largely determined by the qualities of live streamers, and how to evaluate live streamers is a critical topic in e-commerce. Researchers have paid little attention to the complex nature of live streamers and mainly view them as salespersons. Furthermore, no efforts have been conducted to comprehensively identify their typology. To address these gaps, in this study examining live streamers simultaneously from the perspectives of salespersons and spokespersons is proposed. A Live Streamer Grid Typology Model is developed by assessing the qualities of live streamers in terms of their abilities as salespersons and spokespersons.

Keywords: live streaming shopping, live streamer, sales, spokesperson, Live Streamer Grid Typology Model

1. INTRODUCTION

According to the China Retail Traffic Index report conducted by Linkshop (2020)^[1], the overall retail traffic in January 2020 dropped by 22% year-on-year because of the impact of the COVID-19 epidemic. Live streaming shopping, on the contrary, was prosperous in 2020. According to the Ministry of Commerce of China (2020)^[3], live streaming e-commerce sales reached 10 billion in the first half of 2020. There are 400,000 active live streamers, 50 billion streaming viewers, and more than 20 million products in the live streaming market. Promising live streaming e-commerce, created by Li Jiaqi, the Lipstick Big Brother, and Wei Ya, the Taobao live streaming host, have attracted extensive attention among Internet celebrities, traditional celebrities, brand owners, and online stores^[3]. Many companies offering diverse products and services, such as traveling, financial services, and food delivery services, are reaching consumers through live streaming. The era in which “everything can be live streamed, everyone can live stream, everywhere can be live streamed” is coming. Live streaming e-commerce has provided a more effective environment for the interaction between streamers and consumers. Moreover, live streaming e-commerce based on user-generated content is more trustworthy for consumers compared with traditional e-commerce^[4].

In July 2020, the Ministry of Human Resources and Social Security of the People’s Republic of China^[5] added the “live streaming salesperson” post in the Internet marketing profession, which means that the live streaming e-commerce industry has entered a stage of standardized and orderly development. After the market dividend period of live streaming, the live streaming e-commerce industry has challenged the comprehensive strength of live streamers and the degree of refined operation of the team behind them. The professionalization of the live streamer also promotes the development of the live streaming e-commerce industry. For example, Yiwu has released the assessment of “Specialized Occupational Ability Assessment Certificate for E-commerce Live Streaming Practitioners”^[6]; more multi-channel networks (MCNs) are joining to train professional hosts^[7], and a

* Corresponding author. Email: wangyi63@swufe.edu.cn(Yi Wang)

considerable number of universities in China have established live streaming professional training programs and made live streaming an official course^[6]. These changes urgently require us to solve various problems, such as how to measure and train the comprehensive ability of live streamers. In this context, considering the endorsements and sales attributes of the live streamer simultaneously would provide directions for solving these problems.

Prior studies have been focused on this problem, trying to distinguish the types of live streamers and study the impact of different characteristics of live streamers on consumers^{[8][9]}. Some scholars divide the skills of live streamers into four categories, i.e., transaction, persuasion, content generating, and relationship building, using them to explain the various activities of live streamers^{[10][10]}. This research mainly focuses on the salesperson nature of live streamers. Other scholars have also studied live streamers by emphasizing their roles as spokespersons^{[11][12]}. They argue that live streaming shopping can create a more effective environment for celebrities to interact with their fans. Of course, live streamers should pay attention to matching the product with their own image; otherwise, fans may think that he or she is a commercially motivated marketer^[12]. In such studies, scholars devote more attention to the live streamer's spokesperson ability. Overall, research in this area is still in its infancy, and includes the following gaps.

First, the dimensions of evaluating live streamers are not comprehensive. Some scholars^{[12][13][14]} only focus on the characteristics of live streamers in some ways. Wongkitrungrueng and Assarut (2020)^[14] examine the main live streamers' sales methods in live streaming rooms without evaluating the influence of live streamers. Kang (2014)^[13] and Kim (2013)^[15] examined the factors of live streaming shopping, such as live streamer reputation, social presence, and service quality. The main aspect that has been neglected is the instant interaction between live streamers and consumers, which is one of the most salient characteristics of live streaming e-commerce. Therefore, how to comprehensively evaluate live streamers is a major gap in this research area.

Second, the scope of live streamers to be evaluated is not sufficiently comprehensive. Some scholars^[8] have limited their focus to top-tier live streaming hosts and ignored other types of live streamers which limits the applicability of their conclusions.

Third, the dual attributes of live streamers as spokespersons and salespersons are not simultaneously considered. In defining live streamers, some scholars consider live streamers as online sellers^[14]. Some scholars also consider the role of live streamers as representatives or spokespersons in live streaming e-commerce^[12]. One must consider these two abilities at the same time to better understand live streamers.

After analyzing the literature on live streaming e-commerce, the present research, through focusing on the sales and spokesperson abilities of live streamers, intends to address the following questions.

1. How to comprehensively measure live streamers' abilities?
2. How to develop a Live Streamer Grid Typology Model?

2. THEORY DEVELOPMENT

Live streaming is considered as a form of synchronous social media, and contains simultaneity^[16] and authenticity^{[17][17]}. The combination of live streaming and e-commerce has generated live streaming e-commerce, which contains not only social commerce properties^[18], but unique media properties as well^[15]. In this study, both the commerce and media attributes are considered.

With live streamers at the core of live e-commerce, the question "what kind of live streamers are more popular among live streaming shoppers?" has been widely discussed by scholars. Wongkitrungrueng et al. (2020)^[14] studied marketing strategies adopted by live streamers in live streaming rooms, and summarizes 4 marketing methods and 12 marketing strategies. Assarut and Wongkitrungrueng (2020)^[10] focus on live streamers as small individual sellers and argue that live streaming has become a direct selling tool for many individual sellers, such

as Taobao merchants. Huang et al. (2020)^[19] analyze the sales expression skills of successful live streamers from the perspective of discursive strategies. Park and Lin (2020)^[12] examine the fit between live streamers and products by applying theories related to celebrity endorsement. This line of research treats live streamers as “WangHong”^[20], those that are influential and have gained a certain number of fans on social media platforms. Along with the increasing popularity of social media, “WangHong,” with numerous fans, generate commercial value and serve as key opinion leaders (KOLs). Just as live streaming has both social business and social media attributes, live streamers should have both the abilities of salespersons and spokespersons. Therefore, in this study live streamers are examined as salespersons and spokespersons simultaneously. A Live Streamer Grid Typology Model is developed by assessing the qualities of live streamers in terms of their abilities as salespersons and spokespersons.

2.1 Live streamers as salespersons

Retailing literature has a long history of classifying salespersons. In the present study, the primary literature in this field is used as a theoretical basis to construct a classification system for live streamers based on sales abilities. In 1986, Moncrief^[21] empirically developed a classification system for sales jobs by conducting clustering analysis of 121 sales activities, and categorized salespersons into institutional sellers, trade service providers, missionary sellers, order takers, and residuals. However, due to dramatic changes in sales work in the 21st century^[22], it is slightly inappropriate to apply this typology to classify live streamers. Based on Moncrief (2020)^[22], unfit indicators, i.e., travel, office, and recruiting, which are specific to offline traditional sales channels, are excluded herein. Indicators are selected to evaluate the sales abilities of live streamers, e.g., relationship (whether the relationship with customers is good), entertaining (online and offline interactive activities with customers), product support (self-support or agent), and promotional activities and sales services (live broadcast room activities and pervasive skills). Other important indicators for measuring the sales abilities of live streamers are expression skills (the ability to introduce products) and cooperation (e.g., whether there is an assistant in the live broadcasting room). These six indicators are used comprehensively measure the qualities of live streamers in playing the salesperson role.

2.2 Live streamers as spokespersons

Live streaming e-commerce is classified as either traffic economy or fan economy. Live streamers intensively communicate with consumers and provide extensive content on social media aiming to establish credibility and influence consumers’ decisions and behaviors^[20]. In this way, live streamers act as spokespersons and ambassadors of products or brands. Some studies have focused on the spokesperson nature of live streamers. For example, Liu et al. (2021)^[23] evaluate live streamers based on live streamer-customer intimacy and focus on several attributes of live streamers, e.g., authenticity, similarity, and customer response capability. Xu et al. (2020)^[9] also apply the literature on spokespersons and introduce attractiveness as a key characteristic of live streamers. Wang and Yu (2020)^[8] further examine whether live streamers are celebrities and argue that this feature could impact live streamer performance. Consistent with the above mentioned studies, the celebrity endorser scale developed by Ohanian (1990)^[24] is explicitly used in this paper to measure the spokesperson abilities of a live streamer in terms of attractiveness (the live streamer’s facial and physical attractiveness), trustworthiness (the extent to which consumers trust the live streamer), expertise (the extent to which the live streamer is perceived by consumers as a source of effective claims), and influence (the live streamer’s appeal among fans).

2.3 Live Streamer Grid Typology

In the present study, live streamers are classified by assessing the qualities of live streamers in terms of their abilities as salespersons and spokespersons. In particular, a Live Streamer Grid Typology Model is proposed to identify five types of live streamers, i.e., grassroots, sales, expert, traffic, and omnipotent (shown in Figure 1).



Figure1. Live Streamer Grid Typology

Grassroots live streamers. Grassroots live streamers are those at the lower levels of the live streaming pyramid. These live streamers have mediocre language skills and sales techniques to attract consumers, which means they are hardly superior salespersons or influential spokespersons. Grassroots live streamers have a limited number of fans and thus their live streaming rooms are usually not crowded. They are live streamers more out of their own personal interests and live streaming provides them opportunities to express themselves^[25].

Sales live streamers. Sales live streamers are becoming increasingly more crucial in the live streaming market. Compared to grassroots live streamers, they have a clearer understanding of products. In China, major Taobao e-commerce and well-known brands (Bai Cao Wei, Li Ning, and others) are cultivating live streamers for their products. These live streamers usually have excellent sales skills. For example, they can accurately describe the advantages of products and solve consumers' problems quickly. In addition, most of the products sold in the live broadcast room are promotional products. Although sales live streamers are not as powerful as omnipotent live streamers and will not interact with consumers too much, they can perform well regarding the appeal of the brand itself and low-cost advantages of products.

Expert live streamers. Expert live streamers are those who have achieved some previous successes in a certain domain, and, as a result, have a large number of fans in their respective fields. These live streamers intend to leverage their reputation and influence in specific domains to sell related products (e.g., e-gaming bloggers). In terms of sales attributes, these live streamers can clearly describe product features and accurately understand consumers' needs by virtue of their knowledge ability. In terms of endorsement attributes, they have a stable fan base^[11]. These live streamers are usually KOLs and influence consumers by using their own V-blogs, specialties, and attractiveness.

Traffic live streamers. Traffic live streamers comprise the most unstable group in the entire body of live streamers. Li Xiang, a well-known hostess with a large fan base, is an example. They can not only easily initiate a hot and viral topic among fans, but also provide credible endorsements for products they sell^[2]. Traffic live streamers are only responsible for attracting consumers with service activities conducted by their assistants. They are best suited to sell products that match their personalities and features. When the products do not match their personalities or features (for example, frying pans sold by a female movie star often portraying royalty), it may disappoint consumers^[12]. In summary, traffic live streamers could be perfect spokespersons, but are inferior in sales because of their poor sales skills.

Omnipotent live streamers. Omnipotent live streamers are those standing at the top of the live streaming hierarchy. Wang and Yu (2020)^[8] believe that Wei Ya and Li Jiaqi are the representatives of this type in China. They are often listed as Taobao top-tier live streaming hosts. They have excellent sales skills, which help them to improve customers' purchase intentions, describe products, and maintain a close relationship with their fans. Many unknown brands often become popular overnight because of their recommendations. In the eyes of consumers,

they are the spokespersons for high-quality and inexpensive products.

By constructing the Live Streamer Grid Typology Model, most of the live streamers in the live streaming e-commerce market can be classified, their strengths and weaknesses clearly pointed out, and future research on the interaction between live streamers and consumers continued to be developed. In addition, the model can provide a useful frame of reference for studies in the field of live streaming shopping.

3. FUTURE RESEARCH

Based on the proposed Live Streamer Grid Typology Model, we will further explore how the features of different types of live streamers and the live streamer-consumer interaction influence consumers' decisions and behaviors. This study can clearly illustrate the interaction mechanism between live streamers and consumers, showing how various types of live streamers influence consumer purchasing in different ways. A conceptual model will be developed to examine these effects. It is believed that this study can offer useful implications for practitioners.

To evaluate the model (shown in Figure 2), data from various sources will be collected and analyzed. First, archival data (such as consumer comments, number of consumer likes, number of consumer orders, number of active consumers, and so on) of live streamers will be collected. These data can facilitate the analysis of patterns of interactions between live streamers and their consumers. For instance, the sales attributes of streamers can be measured by order volume and their spokesperson attributes can be measured by multiple indicators, such as consumer likes and the contents of streamer-consumer interaction. Next, live streaming clips of streamers will be captured and analyzed to assess sales skills of streamers and to identify their styles. Furthermore, a questionnaire will be used to collect data to measure consumers' attitude toward streamers and behaviors.

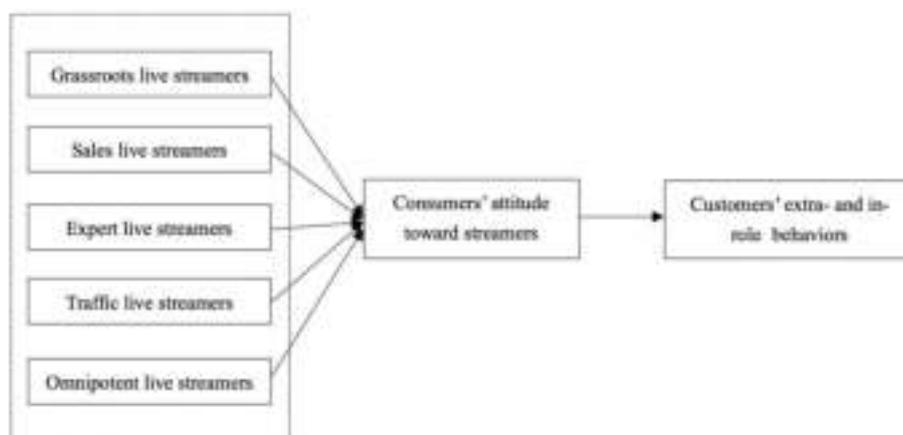


Figure 2. Future research model

REFERENCES

- [1] ShopperTrak (2020). China retail traffic index report. <http://www.linkshop.com.cn/web/archives/2020/455054.shtml>. (in Chinese)
- [2] Souhu(2020). Live Streaming with Goods Industry Insight Report 2020. https://www.sohu.com/a/403847596_712171. (in Chinese)
- [3] KPMG and Aliresearch (2020), Towards a trillion dollar market for live e-commerce. <https://pandaily.com>. (in Chinese)
- [4] Gong, Wanqi; Li, Xigen (2017): Engaging fans on microblog: the synthetic influence of parasocial interaction and source characteristics on celebrity endorsement. In *Psychol. Mark*, 34 (7): 720–732.
- [5] Ministry of Human Resources and Social Security of the People's Republic of China (2020). Notice on the release of occupational information such as blockchain engineers and technicians. <http://www.mohrss.gov.cn>. (in Chinese)
- [6] News Agency (2020) .The "live streamers" has a certificate of professional competence in e-commerce live.

- http://csj.xinhuanet.com/2020-05/13/c_139051851.htm. (in Chinese)
- [7] iiMedia Research (2020). Statistics of Livestreaming: Taobao Livestreaming Platform Boasts 20,000 Hosts in 2019. <https://www.iimedia.cn/c1061/69517.html>. (in Chinese)
- [8] Wang, Xinmin; Yu, Siqi (2020): What Elements Drive the Sale in E-Commerce Live Broadcasts Most in China? In *Frontiers in Economics and Management*, 1 (12): 281–288.
- [9] Xu, Xiaoyu; Wu, Jen-Her; Li, Qi (2020): What Drives Consumer Shopping Behavior in Live Streaming Commerce?. In *Journal of Electronic Commerce Research*, 21 (3): 144–167.
- [10] Wongkitrungrueng, Apiradee; Dehouche, Nassim; Assarut, Nuttapol (2020): Live streaming commerce from the sellers' perspective: implications for online relationship marketing. In *Journal of Marketing Management*, 36 (5-6): 488–518.
- [11] Zhao, Qun; Chen, Chun-Der; Cheng, Ho-Wen; Wang, Jin-Long (2018): Determinants of live streamers' continuance broadcasting intentions on Twitch: A self-determination theory perspective. In *Telematics and Informatics*, 35 (2): 406–420.
- [12] Park, Hyun Jung; Lin, Li Min (2020): The effects of match-ups on the consumer attitudes toward internet celebrities and their live streaming contents in the context of product endorsement. In *Journal of Retailing and Consumer Services*, 52(C): 1-6.
- [13] Kang, Ju-Young M.; Johnson, Kim K. P.; Wu, Juanjuan (2014): Consumer style inventory and intent to social shop online for apparel using social networking sites. In *Journal of fashion marketing and management*, 18 (3): 301-320.
- [14] Wongkitrungrueng, Apiradee; Assarut, Nuttapol (2020): The role of live streaming in building consumer trust and engagement with social commerce sellers. In *Journal of Business Research*, 117 (C): 543–556.
- [15] Kim, Sanghyun; Park, Hyunsun (2013): Effects of various characteristics of social commerce (s-commerce) on consumers' trust and trust performance. In *International Journal of Information Management*, 33 (2): 318–332.
- [16] Scheibe, Katrin; Fietkiewicz, Kaja J.; Stock, Wolfgang G. (2016): Information Behavior on Social Live Streaming Services. In *Journal of Information Science Theory and Practice*, 4 (2): 6–20.
- [17] Tang, John C.; Venolia, Gina; Inkpen, Kori M. (05072016): Meerkat and Periscope. In Jofish Kaye (Ed.): *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI'16: CHI Conference on Human Factors in Computing Systems. San Jose California USA, 07 05 2016 12 05 2016. New York NY: ACM, 4770–4780.
- [18] Shadkam, Mahdi; O'Hara, James (2013): Social commerce dimensions: The potential leverage for marketers. In *The Journal of Internet Banking and Commerce*, 18 (1): 1–14.
- [19] Huang H, Blommaert J, Van Praet E(2020): "OH MY GOD! BUY IT!" a Multimodal Discourse Analysis of the Discursive Strategies Used by Chinese Ecommerce Live-Streamer Austin Li, *International Conference on Human-Computer Interaction*: 305-327.
- [20] Li, Ruohan (2018): The Secret of Internet Celebrities: A Qualitative Study of Online Opinion Leaders on Weibo[C]//51st Hawaii International Conference on System Sciences: 533-542.
- [21] Moncrief III, William C. (1986): Selling activity and sales position taxonomies for industrial salesforces. In *Journal of Marketing Research*, 23 (3): 261–270.
- [22] Moncrief, William C.; Marshall, Greg W.; Lassk, Felicia G. (2006): A Contemporary Taxonomy of Sales Positions. In *Journal of Personal Selling & Sales Management*, 26 (1): 55–65.
- [23] Liu, G. H., Sun, M., & Lee, N. C. A. (2021). How can live streamers enhance viewer engagement in eCommerce streaming?. In *Proceedings of the 54th Hawaii International Conference on System Sciences*: 3079-3089
- [24] Ohanian, Roobina (1990): Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. In *Journal of advertising*, 19 (3): 39–52.
- [25] Yang,Guangying (2019): The Production of "Exquisite and Poor Quality Images" and the Formation of "Virtual Community Consciousness" - On the Generation Mechanism of Youth Subculture in ShakeYin Short Video Community. In *China Youth Study*, 19 (6) :79–86. (in Chinese)

Short Research Paper

Traffic Monetization on Community E-commerce Platforms: Trust in Virtual Communities and Its Effect on Actual Purchase

Yi Li^{1*}¹School of Information, Central University of Finance and Economics, Beijing, 102206, China

Abstract: In China, major community e-commerce websites are focusing on increasing customer bases by converting members in their virtual communities into online buyers, which is one of the important ways of monetizing social traffic. Based on the Stimulus-Organism-Response (S-O-R) theory, this paper proposes and empirically tests a model of trust in virtual communities. We will collect data through online questionnaire surveys with users of Little Red Book, followed by structural equation modeling to analyze the influence mechanism of community e-commerce platform characteristics on consumers' trust. The expected results will show the full mediation effects of interactivity and presence in the relationship between the quality of user-generated content (UGC) and the influence of key opinion leader (KOL) in virtual communities and trust. And consumers' identity trust, cognitive trust as well as emotional trust will all significantly promote their actual purchases. This paper fills the research gap of trust mechanism on community e-commerce platforms. It also guides the community e-commerce platforms to encourage users to generate more high-quality content and promote major brands to cooperate with KOLs to accelerate the monetization of social traffic in virtual communities.

Keywords: community e-commerce platforms, traffic monetization, trust, interactivity, presence

1. INTRODUCTION

Community e-commerce refers to the business model in which consumers with common preference and values recommend, share and discuss products in virtual communities to convert social traffic into e-commerce transactions. Nowadays, with the popularization of the mobile Internet, social communities have gathered a large number of users and become new portals for e-commerce to acquire customers. Community e-commerce has emerged in this context. Different from traditional e-commerce platforms, it has the following characteristics: (1) Stronger sense of participation. Traditional e-commerce platforms always implement precise marketing to customers unilaterally, but in community e-commerce platforms, every user can generate content to influence the decisions of other consumers; (2) Higher trust value. Trust is a key factor in the embedding of social networks into e-commerce. Users can reduce perceived risks in e-commerce transactions with the sense of trust in social networks ^[1]. With the continuous integration of social traffic and e-commerce transactions, China's community e-commerce scale has exceeded 1 trillion in 2019, an increase of 59% year-on-year.

Founded in 2013, Little Red Book is a typical community e-commerce platform. It adopts the "community + e-commerce" business model, which not only operates a virtual community for users to share their lives, but also operates the B2C self-operated e-commerce platform named Welfare Agency. Currently, Little Red Book has more than 150 million monthly active users and more than 8 billion notes are exposed daily. The reason why it succeeds is that, on one hand, users of Little Red Book can generate content (UGC) spontaneously. A group of young people share their consumption experiences in virtual community which makes Little Red Book the birthplace of consumer trends. On the other hand, major brands cooperate with key opinion leaders (KOL) such as Internet celebrities and well-known bloggers to reach consumers in the form of short videos and live broadcasts in the community of Little Red Book, so as to precipitate super users and form a closed loop of

* Corresponding author. Email: liyibtbu@163.com

consumer purchases. The e-commerce transactions on Little Red Book present the characteristics of high repurchase rate (48.7%), high conversion rate and low return rate. Community e-commerce has changed the role of users in traditional e-commerce. The online interaction and content generation of users in virtual communities will directly affect their trust ^[2]. Therefore, the **Research Question (1)** is: Compared with traditional e-commerce platforms, how do the characteristics of UGC and KOL in virtual community enhance users' trust and promote their purchase intention?

Although operating e-commerce platform is a feasible way to monetize social traffic, the road for community e-commerce is still tough. Compared with traditional e-commerce platforms, the self-operated Stock Keeping Unit (SKU) is limited in number and the quality of the introduced third-party channels is not guaranteed, so consumers have to choose other online stores and it is difficult for traffic to convert in their own stores. In addition, after over-commercialization, the sharing experience of users in virtual communities will be weakened. Based on this, the **Research Question (2)** is: After being recommended in virtual community, users will give priority to purchasing in online stores operated by community e-commerce platforms, or on other third-party platforms (such as Taobao)?

Community e-commerce platforms integrate the connection between users into the e-commerce platform, which can improve the trust environment of traditional e-commerce to a certain extent. However, the current research in the field of community e-commerce mainly focuses on users' satisfaction, sharing of knowledge, value creation, etc. ^[3] and there are few studies on platform trust mechanisms. In addition, how to convert social traffic into online shopping traffic and increase users' loyalty in community e-commerce platforms is rarely involved.

The purpose of this study is to analyze how the quality of user-generated content (UGC) and the influence of key opinion leaders (KOL) are associated with consumers' trust and actual purchases in community e-commerce platforms. We will take the users of Little Red Book as the research subject of the questionnaire, and divides the trust of users into three sub-dimensions: identity trust, cognitive trust and emotional trust. Based on the S-O-R theory, we will build a model of actual purchase of consumers on the community e-commerce platform. To the best of authors' knowledge, though several studies have been carried out to identify the trust mechanism in online shopping, how to increase trust in the scenario of the community e-commerce platforms with mediating effects of interactivity and presence has been studied by this article first. This article is also the first to study about how to monetize the social traffic gathered in the virtual community on its own e-commerce platform.

The paper is arranged as follows. The theoretical background of community e-commerce and trust mechanism will be discussed in the next section. Subsequently, theoretical framework and hypotheses development will be discussed. Then the method to analyze how the quality of UGC and the influence of KOL are associated with consumers' trust and actual purchases on community e-commerce platforms is proposed. This article will conclude with the implications and future work.

2. THEORETICAL BACKGROUND

2.1 Virtual community

Traditional community is a geographical concept. There are content and interaction in the community, which realizes the connection between people and things. With the development of network technology, virtual community (VC) is a kind of new community formed by people connected via the Internet ^[4]. In other words, virtual community is a group of people with similar interests and practices who communicate regularly and continuously on the Internet by sharing cyberspace ^[5].

Virtual community is not only for sharing the ideas, facts or information, but it is useful for developing

business or a better relationship. For that reason it attracts e-commerce people to expand their business [6]. Community e-commerce came into being under this background.

2.2 Community e-commerce

Community e-commerce is composed of two parts: a virtual community integrating content generation and interpersonal social functions and an online shopping platform (Figure 1) [7]. It is essentially a new type of e-commerce model that promotes user interaction and user-generated content (UGC) through virtual communities and drives business transactions. It can be traced back to the online auction community with 1.6 million registered online members established by eBay in 2000.



Figure 1. Community e-commerce.

The research on community e-commerce mainly focuses on the following aspects: (1) User behavior. This kind of research mainly studies the purchase intention of customers and analyzes the mechanism of factors such as trust, loyalty and incentives in the purchase decision [7][8]. (2) Business models. This type of research attempts to analyze the characteristics of business models in the community e-commerce scenario and differences between business models in the traditional business environment [9]. (3) Network analysis. This part of research focuses on the characteristics of social media and its impact on community e-commerce, users and organizations [10]. In addition, there is a small amount of research focusing on privacy policies.

Most scholars just regard community e-commerce as a new customer acquisition and marketing method, emphasizing its advantages in terms of efficiency, cost and experience. However, they lack of understanding of the internal mechanism of community e-commerce development. At present, there are still vacancies for the characteristics of the community e-commerce platform, such as the quality of UGC and the influence of KOL on the user's sense of trust, and how to monetize the traffic gathered in the virtual community into e-commerce transactions.

2.3 Trust

Because of the complexity and abstraction of trust, there is no uniform definition of it in academia. In the field of sociology, trust is regarded as the basis for human participation in social life. In the field of e-commerce, trust is defined as a set of beliefs regarding the seller's honesty, ability and benevolence towards the buyer [11].

A large number of literature have focused on consumer trust in online shopping. When transactions are converted from a traditional face-to-face situation to a virtual environment, trust is even more important. Users need trust to reduce the uncertainty and complexity in transactions [12]. Regarding the influencing factors of consumer trust in online shopping, the trust system in B2C e-commerce consists of institutional trust, direct trust and recommendation trust [13]. In addition, consumers' personal factors, system environment factors, website factors, company factors and relationship marketing methods can increase consumer trust in e-commerce platforms [14]. In online marketplaces, maintaining and improving users' loyalty requires the maintenance and strengthening of trust [15].

After the rise of community e-commerce, scholars compared it with traditional e-commerce and found that it integrates the functions of social media and features such as product recommendation and experience sharing will have an impact on users' trust and behavior [16]. Traditional e-commerce users take products or services as

their primary needs, and users' trust comes from factors such as the quality of the platform, the credit rating of the merchant, the prior feedback of products and after-sales guarantees. However, due to the embedded social network, community e-commerce has changed the structure of consumer demand. In this context, the users' trust source channels are broadened and the sharing of information in virtual community can reduce users' risk perception. Trust in community e-commerce platforms can be divided into identity trust, behavioral trust, cognitive trust as well as emotional trust [17]. The open, transparent and effective information in the interaction helps to build an orderly trading environment and trust can be improved. The interaction between consumers and the website as well as the interaction between consumers can increase the spatial presence of consumers which can increase consumers' trust in the capabilities of B2C online stores. And the interaction between consumers and online sellers as well as the interaction between consumers can increase the social presence of consumers which can increase consumers' trust in the integrity and goodwill of B2C online stores [18].

The trust on the community e-commerce platform serves as the link between users and between users and the platform, reducing customer acquisition costs for the platform and transaction costs for users so as to create greater commercial value. Hence, trust is an important part of the research field of community e-commerce platform. However, research on the formation mechanism of trust in community e-commerce platforms is very scarce and shallow. In this context, this article explores how UGC and KOL in virtual communities promote the generation of users' trust with mediating effects of interactivity and presence.

3. THEORETICAL FRAMEWORK AND HYPOTHESES DEVELOPMENT

The well-known "Stimulus-Organism-Response" (S-O-R) theory is widely used in the research of consumers' behavior in online shopping environment. According to the S-O-R theory proposed by psychologists Mehrabian and Russell, stimulus (S) is an external environmental factor of the body, and organism (O) is a psychological transformation mechanism of users internalizing stimulus into information, response (R) represents the users' relevant behavior to the external stimulus. We make full use of this research paradigm to establish the theoretical framework (Figure 2) of the influence mechanism of UGC quality and KOL influence on consumer trust and actual purchase.

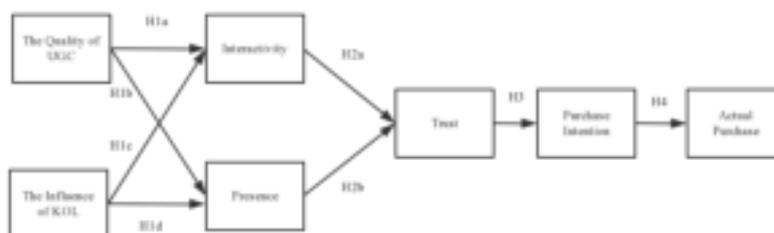


Figure 2. Theoretical framework.

High-quality reviews can describe product features and usage experience in detail, while low-quality reviews are simple and unclear content. The higher the quality of user-generated content, the stronger the promotion of consumers' purchasing intention [19]. Compared with reviews with low emotional intensity, reviews with high emotional intensity can stimulate consumers' desire to purchase. To conclude, the quantity, quality as well as information and interpersonal interaction of user-generated content (UGC) can affect consumers' trust. To operate the community e-commerce platforms, whether the UGC is attractive directly determines consumers' lock in target products.

Although major e-commerce platforms are trying to use big data technology for precise marketing, how to satisfy consumers' personalized information needs is still difficult. Especially in the era of information explosion,

the cost of searching for information is very high. Therefore, listening to the suggestions of key opinion leaders (KOL) has become the better choice of consumers. Existing research points out that the influence of key opinion leaders mainly depends on product-related and individual-related factors. Among the factors related to products, Nicola et al. (2009) found that opinion leaders showed stronger influence when they were more involved in a specific product category. Moreover, the continuous high degree of product involvement is conducive to opinion leaders collecting knowledge related to this type of product from different information channels (Venkatraman, 1990), and thus better affecting information recipients. Robertson (1971) found that consumers who prefer to purchase complex, high-risk and difficult-to-test products are more likely to be influenced by opinion leaders. Among the factors related to individual characteristics, Cosmos and Sheth (1980) subdivided 24 characteristics of consumer opinion leaders and classified them as practical, authoritative, expert, territorial, accessible, and self-centered and stubborn. Opinion leaders are not only at the center of the social community, but also possess the characteristics of accurate product knowledge and not easy to be restrained. The speed at which they deliver new information will significantly affect the number of new information adopters, so using key opinion leaders is one of the effective marketing strategies ^[20].

Compared with advertising media, communication between individuals has been considered more trustworthy and influential. Both the user-generated content and key opinion leaders can provide consumers with product information and shopping recommendations through frequent WOM communication, thereby affecting their interactivity and presence. Interactivity refers to the buyer's subjective perception of high-quality interaction with the seller and presence refers to the perception of intimacy or being close to another person ^[21]. In community e-commerce websites, details, such as product features, promotion information and seller's services, can be specified during the interaction process in virtual community. In addition, since users in social community may deal with other users' questions, a higher degree of presence is more likely to result in a buyer feeling that his/her expectations and service needs are met.

Hypothesis 1a: The quality of user-generated content (UGC) in virtual community contributes to users' interactivity with the e-commerce platform.

Hypothesis 1b: The quality of user-generated content (UGC) in virtual community contributes to users' presence with the e-commerce platform.

Hypothesis 1c: The influence of key opinion leaders (KOL) in virtual community contributes to users' interactivity with the e-commerce platforms.

Hypothesis 1d: The influence of key opinion leaders (KOL) in virtual community contributes to users' presence with the e-commerce platforms.

Hypothesis 2a: A user's interactivity contributes to his/her trust in the platform.

Hypothesis 2b: A user's presence contributes to his/her trust in the platform.

In the context of online marketplaces, Gefen and Straub (2004) suggest that trust can be nourished by interaction, justifying the contribution of social presence to trust building, which subsequently leads to higher purchase intentions. Online trust in general leads to trust-related outcomes, such as transaction intentions and actual transactions ^[22]. We thus account for the effect of trust on purchase intentions in our context.

When a user is recommended for a certain product in virtual community and has the willingness to buy, he/she can either choose to purchase in online stores operated by the community e-commerce company or a third-party e-commerce platform. Therefore, how to improve user stickiness and keep users on the platform to complete real purchases is particularly critical.

Hypothesis 3: A user's trust in the platform contributes to his/her purchase intentions.

Hypothesis 4: A user's purchase intentions lead to his/her actual purchases on community e-commerce

platform.

4. PROPOSED METHOD

In order to test the proposed hypotheses, we will conduct a survey study with users of Little Red Book, followed by psychometric analysis using the structural equation modeling (SEM). It is particularly suitable for testing causal relationships among psychological perceptions that are not directly observable to researchers.

To develop the questionnaire (Table 1), we reused some existing quality questions designed by others for the same construct and slightly modified them according to the research logic of this article. In addition, some other items were formed by qualitative interviews and expert suggestions. Items for the quality of UGC and influence of KOL were adapted from related literature ^[23], and identity trust, cognitive trust and emotional trust were adapted from trust mechanism literature ^[17]. Items were designed with a 7-point Likert scale. 1 means strongly disagree, 7 means strongly agree.

To ensure that items reflect the intended construct, content validity should be checked first. Most of the questions we used were the rephrasing of different aspects of a construct as defined in related literature, which provided the basis for content validity. In addition, two professors familiar with the research project were invited to discuss the phrasing of the questions to ensure that they had at least face validity. Minor changes have been made on the basis of the feedback.

Table 1. Questionnaire.

Constructs	Items	References
The Quality of the User-generated Content (UGC)	The content shared by users in virtual community provides views on different aspects of the product.	Zhao, (2020)
	The content shared by users in virtual community is helpful for me to increase my understanding of the product.	
	The content shared by users in virtual community is helpful for me to learn professional knowledge in related fields.	
The Influence of the Key Opinion Leaders (KOL)	The key opinion leaders of Little Red Book have a certain leading position in this product field.	Zhao, (2020)
	I think the key opinion leaders have experience in buying and using this product.	
	The key opinion leaders usually pay close attention to such products.	
Interactivity	The users in virtual community respond to my questions very quickly.	Liu, (2003) ^[24]
	The virtual community facilitates two-way communication between users.	
Presence	I felt I was more in the “real world” than the “computer world” when I was browsing the content in virtual community.	Khalifa & Shen, (2004) ^[25]
	There is human warmth in virtual community.	
Identity Trust	I trust the virtual community of Little Red Book.	Yan & Peng, (2019)
	I trust the members in virtual community of Little Red Book.	
Cognitive Trust	I believe that Little Red Book is valuable.	
	I believe that Little Red Book is safe.	
Emotional Trust	My loyalty to Little Red Book is very high.	
	I have the willingness to share in virtual community of Little Red Book.	
Purchase Intention	I would like to buy products shared by users of Little Red Book.	Davis &
Actual Purchase	I choose to buy the product from the Welfare Agency operated by Little Red Book.	Vankatesh, (2003)

5. IMPLICATIONS AND FUTURE WORK

The theoretical implication of this article is to improve the mechanism of UGC and KOL on user trust on the community e-commerce platforms. In addition, it fills the research gap on how to monetize the social traffic in virtual community.

The practical implication of this article is to guide the community e-commerce platforms to encourage users in virtual community to generate more high-quality content and give more support to the KOLs to promote the monetization of traffic on this e-commerce platform.

From now until the WHICEB conference, our major focus is collecting data via surveys. After that, we will use SPSS to analyze the reliability and validity of the questionnaire. In addition, we will use the structural equation modeling method to test the hypotheses in the theoretical framework and examine the mediating effects shown in Figure 2. At last, we intend to present the preliminary results of our study at WHICEB.

REFERENCES

- [1] Hou Lisong, Chen Zeying, Tong Junjie. (2017). Influencing factors and mechanism of customer trust in the environment of WeChat. *Financial Science*, 2: 75-87(in Chinese)
- [2] Qin Fen, Li Yang. (2018). Review and prospect of user-generated content incentive mechanism. *Foreign Economy and Management*, 8: 141-152(in Chinese)
- [3] Dai Debao, Liu Xiyang, Fan Tijun. (2015). Research on the factors influencing the adoption intention of Internet personalized recommendation in the era of Internet +. *Chinese Soft Science*, 8: 163-1672(in Chinese)
- [4] Balasubramanian S, Mahajan V. (2001). The economic leverage of the virtual community. *International Journal of Electronic Commerce*, 5(3): 103-138.
- [5] Ridings C M, Gefen D, Arinze B. (2002). Some antecedents and effects of trust in virtual communities. *Journal of Strategic Information System*, 11(3-4): 271-295.
- [6] Chirag S. (2012). Virtual community: the new hope for e-commerce. *Indian Journal of Computer Science and Engineering*, 3(1): 20-23.
- [7] Sun Tianhui, Ding Xiaodong, Liu Zheng. (2021). Research on the mechanism of enhancement of community e-commerce user loyalty based on behavior analysis. *Management Modernization*, (1): 85-87(in Chinese)
- [8] Anderson R H P. (2015). Trust development and transfer in social commerce: Prior experience as moderator. *Industrial Management & Data Systems*, 31: 1012-1030.
- [9] Abuduaini Abudurehman, Xie Ying. (2020). Research on the business model innovation of social e-commerce based on multi-case comparison. *Economy Forum*, (10): 123-132(in Chinese)
- [10] Sashi C M. (2012). Customer engagement, buyer-seller relationships, and social media. *Management Decision*, 50(2): 253-272.
- [11] David G, Elena K, Detmar W S. (2003). Trust and TAM in online shopping: an integrated model. *MISQ*, 27(1): 51-90.
- [12] Pan Yu, Zhang Xing, Gao Li. (2010). Research on factors affecting consumers' purchase intention in online retailing: an analysis based on trust and perceived risk. *Chinese Industrial Economy*, (7): 115-124(in Chinese)
- [13] Tian Bo, Qin Zheng. (2008). Research on online trust system model in B2C e-commerce. *East China Economic Management*, 22(10): 120-124(in Chinese)
- [14] Xiang Luquan. (2020). The theoretical model construction of consumer trust factors in B2C e-commerce. *Chinese Business*, (3): 120-124(in Chinese)
- [15] Yim H K, Swaminathan A S. (2004). Customer relationship management: strategy, process and technology customer relationship management: its dimensions and effect on customer outcomes. *Journal of Personal Selling and Sales Management*, 24(4): 263-278.
- [16] Hajli N, Sims J. (2015). Social commerce: the transfer of power from sellers to buyers. *Technological Forecasting and*

- Social Change, 1(12): 350-358.
- [17] Yan Huili, Peng Zhengyin. (2019).Research on trust mechanism of social e-commerce platforms from the embedded perspective——exploration based on grounded theory. *Scientific Decision*, (3): 47-72(in Chinese)
- [18] Zhao Hongxia, Wang Xinhai, Zhou Baogang. (2015).Research on online interaction, presence and consumer trust in B2C online shopping. *Management Review*, 27(2): 43-54(in Chinese)
- [19] Wu Pengfei, Yan Qiang. (2015).Research on the impact of online comments on the adoption of E-WOM in social networks. *Journal of Beijing University of Posts and Telecommunications (Social Science Edition)*, 17(1): 52-61(in Chinese)
- [20] Van-Eck P S, et al. (2011).Opinion leaders' role in innovation diffusion: a simulation study. *Product Development & Management Association*, 28(2): 187-203.
- [21] Carol Xiaojuan Ou, et al. (2014).Swift guanxi in online marketplaces: the role of computer-mediated communication technologies. *MIS Quarterly*, 38(1): 209-230.
- [22] Gefen D, Straub D W. (2004).Consumer trust in B2C e-commerce and the importance of social presence: experiments in e-products and e-services. *Omega*, (32): 407-424.
- [23] Zhao Xiuyun. (2020).Research on the impact of user generated content (UGC) on consumers' purchasing intention——take Little Red Book as an example. *Yunnan University of Finance and Economics*, 1-51(in Chinese)
- [24] Liu Y. (2003).Developing a scale to measure the interactivity of websites. *Journal of Advertising Research*, 43(2): 207-216.
- [25] Khalifa M, Shen N. (2004).System design effects on social presence and telepresence in virtual communities. *Proceedings of the 24th International Conference on Information Systems*. Washington DC: 547-558.

Short Research Paper**Influence of Social Network Integration on the Online Review Helpfulness***Jiaxuan Wu¹, Shengli Li^{1*}*¹ Department of Information Management, Peking University, China

Abstract: Online consumer reviews are important for consumers when they make purchasing decisions. However, the large volume of online reviews makes it difficult for consumers to identify those helpful reviews. The influencing factors on online review helpfulness have drawn great attention from different research fields. In recent years, online review websites start to exhibit more features of social media. For example, some websites allow users to integrate with other social media accounts. The influences of such social factors, however, are rarely studied in the literature. Drawing on a dataset from Qunar.com, this paper explores how social network integration and reviewer network centrality influence online review helpfulness through a negative binomial regression model. Our results show that both factors have a positive effect on review helpfulness, and that network centrality positively moderates the effect of social network integration. Our research results provide important implications for reviewers, industry practitioners, and online review websites.

Keywords: online consumer reviews; social network integration; network centrality; online review helpfulness

1. INTRODUCTION

Online consumer review is usually defined as product evaluations published by consumers on e-commerce or third-party websites, mainly in the form of star rating and open review text^[1]. According to the research survey provided by Saleh (2015), 90% of consumers read online reviews, and 88% of consumers believe that online reviews are as trustworthy as personal recommendations^[2]. This shows that most consumers will read online reviews before making a purchasing decision to reduce their uncertainties. Consequently, online reviews have important influences during consumers' product selection and purchase process.

The rapid increase number and uneven quality of online reviews, however, also pose a great challenge to consumers. They may find it difficult to identify reviews that provide the most truthful and valuable information. Therefore, to address such an issue of information overload, it is crucial to help consumers quickly identify the most "helpful" reviews from tons of reviews. Online review websites usually allow readers to give feedback on reviews by clicking the "useful" vote buttons. The helpfulness of a review thus can be measured by the number of votes it received.

Online review helpfulness has drawn great attention from researchers from different fields. The concept of online review helpfulness was first proposed by Chatterjee (2005)^[3], which refers to the influence degree of information use. Mudambi (2010) regards it as the subjective perceived value of whether consumers are helpful to online reviews in the decision-making process^[1]. In addition, many scholars have studied how different factors may affect the helpfulness of online reviews, mostly focusing on reviewer characteristics and review characteristics^[4-8].

Since online review websites exhibit more and more social media features, a few scholars also established that social factors may influence online review helpfulness as well [9,10]. In recent years, online review websites start to allow users to integrate with their social network accounts (such as Twitter, Facebook, and Weibo). Such a mechanism may have important influences on review helpfulness but has not been well studied in the existing literature. This paper aims to supplement relevant literature by considering social network integration. In particular, based on a dataset drawn from Qunar.com, we explore the influence of social network integration on online review

* Corresponding author. Email: lishengli@pku.edu.cn

helpfulness and how such influence is moderated by reviewer network centrality. Results show that social network integration positively influences online review helpfulness and this relationship is positively moderated by reviewer network centrality.

2. LITERATURE REVIEW

2.1 Online review helpfulness

As above mentioned, identifying high quality reviews from a large volume of reviews is important to consumers. Thus, online review helpfulness has become a hot topic in different diplomacies, including information systems and marketing. Many scholars studied the influencing factors of online reviews helpfulness. Scholars mainly considered factors from four types. The first type of factors come from reviewer characteristics. Reviewers' identity disclosure, reviewer expertise, and reviewer reputation are all considered to have a significant impact on online review helpfulness^[4-7,11]. The second is review characteristics. Review content (including review length, review readability, review pictures, etc.), review emotional tendency, review valence, review time and review response have also been shown to influence online review helpfulness^[8-12]. The third type relates to reader characteristics. For example, Laura (2011) points out that readers' prior knowledge has a positive effect on review helpfulness^[13]. The last type of factor comes from product characteristics. For example, prior scholars have studied the influence of different product types on review helpfulness^[14-15].

2.2 Social network integration

Social network integration refers to the phenomenon in which many online platforms seek to supplement their user communities through integration with well-known social networking sites such as Facebook, Twitter, and Weibo^[16]. In other words, social network integration facilitates the creation and login of user accounts and provides a more personalized user experience, but it also weakens the anonymity of users^[17].

Several scholars have examined the influence of social network integration during the process of online review generation. For example, Lee (2015) found that, in the presence of social network integration, consumer reviews on products will be influenced by existing reviews through observational learning and peer pressure mechanisms^[18]. Huang (2017) pointed out that social network integration would increase the review number and emotional languages, while decrease cognitive languages and negative languages^[16]. Pu (2020) examined the influence of social network integration on the generation of online reviews from the perspective of social presence theory^[17].

Despite the important role of social network integration in the process of review generation as pointed by these pioneering previous studies, rare research has been done to examine the influence of social network integration on online review helpfulness.

Our paper contributes to the literature by considering how social factors including social network integration and reviewer network centrality will influence online review helpfulness. To the best of our knowledge, we are the first to explore the influence of social network integration on the helpfulness of online reviews and how this effect is moderated by reviewer network centrality.

3. HYPOTHESES DEVELOPMENT

3.1 Social network integration

Sussman (2003) proposed that argument quality and source credibility are the direct factors that affect information usefulness^[19]. Social network integration can increase the readers' perceptiveness of review helpfulness through both argument quality and source credibility. First, reviewers can share their reviews with other social platforms through social network integration, and these reviews may be seen by friends in the reviewer's social circle. Reviewers thus are motivated to improve the review quality to establish a positive social

image and to build a reputation. Consequently, consumers will perceive these high-quality reviews as more helpful. Second, social network integration enables consumers to obtain the identity information of the reviewer. When consumers read a review, they can see the source of the information, which improves the credibility of the review. This may also lead to more useful votes for this review.

According to the anonymity effect and social norm theories, the psychological pressure on individuals from group norms or the influence of others will be reduced under the condition of anonymity. As a result, individuals will show greater autonomy and independence to express their opinions more fully^[20]. That is, identity anonymity makes reviewers express their evaluation of consumption experiences more truthfully. However, social network integration leads to disclosure of reviewers' identities and thus weakens their anonymities. As a result, in the presence of social network integration, reviewers are more likely to publish reviews that conform to social norms, and readers will consider these reviews as less useful information since they don't represent reviewers' truthful opinions about products or services.

In sum, the influence of social network integration on review helpfulness depends on whether the effect of argument quality and source credibility dominate the anonymity effect and social norm or otherwise. High quality reviews can help readers to make quick purchasing decisions while source credibility can help mitigate readers' uncertainty upon the identity of reviewers. On the contrary, whether the reviewers choose to conform to social norms when writing a review is not observable to readers and thus may play a less important role. Thus, it is reasonable to hypothesize that the former effect dominates the latter effect. That is, social network integration will improve online review helpfulness. Therefore, we propose the following hypothesis:

H1: When reviewers choose social network integration, their reviews will be perceived as more helpful.

3.2 Reviewer network centrality

As above mentioned, online review websites show more and more social media features in recent years. For example, one reviewer can follow another reviewer on a review website. Usually, some reviewers may have relatively more followers while some reviewers might follow a large number of other reviewers. This unevenly distributed follower-followee relationship can be measured by centrality, a key concept in social network analysis. Centrality refers to the degree to which a user in a social network is centered in the whole network and thus can measure the influence of users in social networks^[21]. In a social network, centrality can be divided into two types: external network centrality, which refers to the number of followees a user has in a social network, and internal network centrality, indicating the number of followers a user has in a social network^[10].

On one hand, reviewers with a high external network centrality thus have more information sources than those with a low external network centrality, since they followed more other reviewers. They may spend more time than others to read followees' reviews and this may make them being more experienced in writing reviews. Consequently, high external network centralities might lead to higher quality reviews. In addition, reviewers with a high external network centrality may be also motivated to write high quality reviews in order to receive returned attention from other reviewers. Therefore, it is reasonable to conjecture that high external network centrality is an indicator of "usefulness".

On the other hand, high internal network centrality can be viewed as an indicator of reputation in a social network^[22]. From a social psychology perspective, reputation is essentially a signal of enhanced credibility^[23]. To attract a large number of followers on a review website, the reviewer must have provided many trustworthy and high-quality reviews in the past. In addition, reviewers with a high internal network centrality are considered more influential since their activities are immediately observed by a large number of followers. But this also requires these reviewers to provide trustworthy and high-quality reviews to maintain a large number of followers. Therefore, reviews posted by reviewers with a high internal network centrality are more likely to be perceived as more trustworthy.

Base on the above discussions, we propose the following hypothesis:

H2a: Reviews posted by reviewers with high external network centrality will be perceived as more helpful.

H2b: Reviews posted by reviewers with high internal network centrality will be perceived as more helpful.

In addition, reviewer network centrality may also moderate the influence of social network integration on online review helpfulness. First, as abovementioned, social network integration will reveal reviewer identity information and enhance source credibility. A higher reviewer network centrality means a higher influence on review websites. When such a reviewer chooses social network integration, her information will be open to more users compared to a reviewer with low network centrality. Thus, the effect of enhanced source credibility would be more salient. Second, reviewers with a high network centrality usually are active users in social media. When they share their reviews, they might care more about their social images than reviewers with a low network centrality and thus are more motivated to improve review quality.

Therefore, we propose the following hypothesis:

H3a: Reviews' external network centrality positively moderates the impact of social network integration on review helpfulness.

H3b: Reviews' internal network centrality positively moderates the impact of social network integration on review helpfulness.

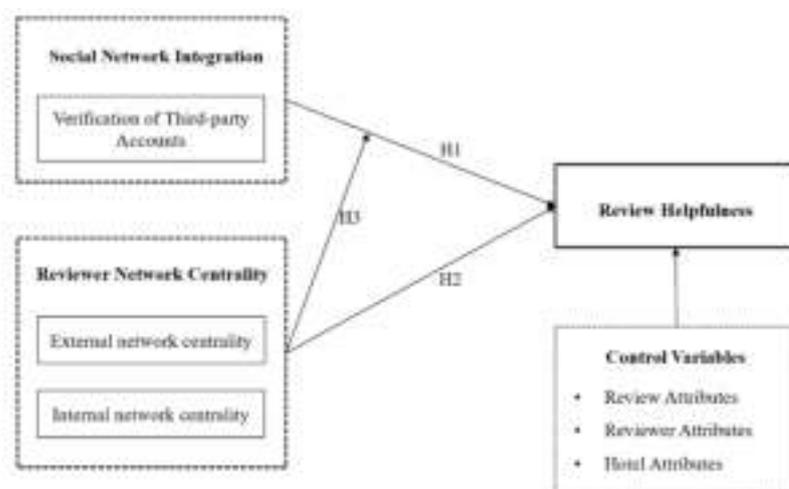


Figure 1. Research Model

4. RESEARCH METHODOLOGY

4.1 Data collection

Qunar.com is a leading Chinese online travel agency. It provides consumers with one-stop booking services for travel products and also allows consumers to post reviews for their purchasing. Qunar.com now enables consumers for social network integration, making it an ideal website for our research. We collected all consumer reviews for hotels in the city of Beijing. The dataset includes in total 2,489,626 reviews of 10,459 hotels. In addition, a large number of reviews on Qunar.com are posted by anonymous users. We deleted these reviews since reviewer characteristics cannot be obtained for them. This leaves us a number of 587,661 reviews. Variables used in this research are explained in detail as follows.

4.2 Variable design

To verify all hypotheses proposed in this paper, we designed the following regression model:

$$\begin{aligned}
 \text{Review Helpfulness} = & \alpha + \beta_1 \text{Social Network Integration} + \beta_2 \text{External Network Centrality} + \beta_3 \\
 & \text{Internal Network Centrality} + \beta_4 \text{External} * \text{Social Network Integration} + \beta_5 \text{Internal} * \text{Social} \\
 & \text{Network Integration} + \theta \text{Controls} + \varepsilon
 \end{aligned} \tag{1}$$

The dependent variable *Review Helpfulness* is the number of “useful” votes that one review receives. The independent variable *Social Network Integration* is a dummy variable. If the reviewer enables social network integration by verifying a third-party social media account, it equals 1. Otherwise, it takes a value of 0. Therefore, H1 is tested by β_1 , which captures the influence of social network integration on review helpfulness. *External Network Centrality* is measured by the number of reviewers’ followees. *Internal Network Centrality* is measured by the number of reviewers’ followers. The estimation of β_2 and β_3 thus are used to test H2a and H2b. β_4 and β_5 are the coefficients of interaction terms of *External Network Centrality* and *Internal Network Centrality* with *Social Network Integration*, which are used to test H3a and H3b, respectively.

The control variables in this paper included the number of reviews posted by the reviewer, hotel level, and rating difference between the current review with the average rating for the same hotel. As mentioned above, Qunar.com offers four hotel levels to choose from, including "two-star or less/ economical", "three-star/comfort", "four-star/premium" and "five-star/luxury". To avoid multicollinearity, three dummy variables – "HotelLevel_2", "HotelLevel_3" and "HotelLevel_4" are set to represent four levels. In addition, existing research points out that a higher number of reviews posted by reviewers can improve their expertise and reputation and their reviews will be perceived as more helpful [11]. We thus include the number of reviews posted by the reviewer to control this effect. Further, the gap between a review rating with the average rating of the hotel may also influence the review’s perceived helpfulness [10]. We thus include rating difference as a control variable too.

5. RESULTS AND DISCUSSIONS

5.1 Descriptive statistics

Table 1 reports the descriptive statistics. Review helpfulness is a count variable with a minimum value of 0 and a maximum value of 46. Through further frequency analysis, we also found that its skewness equals $17.55 > 0$, indicating a right skewed distribution. The regression models of the count variable should be adopted. As is known to all, Poisson regression fits well when the variance of count dependent variable is equal to its mean. However, the variance (0.1616) of the dependent variable in this study is greater than the mean (0.0663). Therefore, we choose the Negative Binomial regression model. We choose the software Stata 16.0 for regression analysis.

Table 1. Descriptive statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
Review helpfulness	587,661	0.0663	0.402	0	46
Social Network Integration	587,661	0.0735	0.261	0	1
External network centrality	587,661	0.0178	1.428	0	322
Internal network centrality	587,661	0.0323	2.834	0	761
Number of reviews posted	587,661	11.52	19.99	0	2,138
Rating difference	587,214	0.696	0.690	0	3.900
HotelLevel_2	426,927	0.396	0.489	0	1
HotelLevel_3	426,927	0.271	0.444	0	1
HotelLevel_4	426,927	0.222	0.416	0	1

5.2 Regression analysis results

Before the regression analysis, we conduct the multicollinearity test of each variable and the results were shown in Table 2. All the variance inflation factors (VIF) are less than 10, indicating that there is no multicollinearity issue among variables.

Table 2. Test for Multicollinearity

Variable	VIF	1/VIF
Social Network Integration	1.03	0.971564
External network centrality	1.51	0.661173
Internal network centrality	1.54	0.648464
Number of reviews posted	1.06	0.945675
Rating difference	1.03	0.972928
HotelLevel_2	2.79	0.358234
HotelLevel_3	2.52	0.397102
HotelLevel_4	2.35	0.426300
Mean VIF	1.73	

The regression results are shown in Table 3. Model 1 is a regression model involving only control variables. Model 2 adds the variable of social network integration to Model 1. Model 3 adds two variables of reviewer network centrality to Model 1. Model 4 is the regression model of all variables except interaction effects, and Model 5 is the regression model of all variables including interaction effects. As shown in Table 3, the Likelihood ratio test p value for each model is 0.000, indicating a strong fit of the Negative Binomial regression model with our dataset.

The results of Model 2 ($\beta_1=0.379$, $p=0.000$) and Model 4 ($\beta_1=0.374$, $p=0.000$) consistently demonstrate that social network integration has a significant positive effect on review helpfulness. This suggests that social network integration may increase source credibility as we have hypothesized, and reviewers are motivated to improve the review quality when their reviews will be shared through other social platforms. Thus, the review helpfulness is increased. H1 is strongly supported.

The results of Model 3 and Model 4 show that both external network centrality ($\beta_2=0.0434$, $p=0.000$; $\beta_2=0.0346$, $p=0.000$) and internal network centrality ($\beta_3=0.0277$, $p=0.000$; $\beta_3=0.0294$, $p=0.000$) have significant positive effects on the review helpfulness. Higher network centralities of reviewers usually represent their high reputation and high professional skills. Thus, H2a and H2b are also supported.

Model 5 shows that external network centrality ($\beta_4=0.0799$, $p=0.000$) and internal network centrality ($\beta_5=0.0231$, $p=0.000$) can positively moderate the influence of social network integration on review helpfulness. This result indicates that the higher the reviewer's network centrality is, the stronger influence of social network integration on review helpfulness. H3a and H3b are supported.

Table 3. Negative Binomial Regression Analysis Results

	Model 1	Model 2	Model 3	Model 4	Model 5
Social Network Integration		0.379***		0.374***	0.373***
External network centrality			0.0434***	0.0346***	-0.0545***
Internal network centrality			0.0277***	0.0294***	0.0236***
External \times Social Network Integration					0.0799***
Internal \times Social Network Integration					0.0231***
Number of reviews posted	5.90e-0.5	-0.000820**	-0.00425***	-0.00539***	-0.00574***
Rating difference	0.586***	0.586***	0.586***	0.586***	0.586***
HotelLevel_2	-1.130***	-1.120***	-1.122***	-1.112***	-1.111***
HotelLevel_3	-0.539***	-0.533***	-0.533***	-0.527***	-0.526***
HotelLevel_4	-0.152***	-0.146***	-0.146***	-0.141***	-0.140***

	Model 1	Model 2	Model 3	Model 4	Model 5
Inalpha	2.083***	2.074***	2.064***	2.064***	2.052***
Constant	-2.765***	-2.796***	-2.731***	-2.758***	-2.755***
chibar2(01)	1.9e+04	1.9e+04	1.9e+04	1.9e+04	1.9e+04
Prob >= chibar2	0.000	0.000	0.000	0.000	0.000
R-squared	0.0356	0.0366	0.0375	0.0384	0.0385

6. CONCLUSIONS

Online review websites are showing more and more social media features. One notable mechanism that emerged in recent years is social network integration. However, little research has been done to study its implications on online review helpfulness. Based on a dataset drawn from Qunar.com, we show that social network integration and reviewers' network centrality have a positive impact on review helpfulness and network centrality positively moderate the influence of social network integration. The research significance of this study includes the following two folds:

This research has certain theoretical significance. This paper is a supplement to the research in the field of online reviews. It provides new variables for the study of the influencing factors of online reviews helpfulness from the perspective of social networks. Examining the impact of social network integration on online review helpfulness, we also enrich the literature in the field of social media, by adding discussion on the potential value and impact of social media.

This research also has practical implications for consumers, reviewers, industry practitioners, and third-party review sites. The empirical results of our study can help consumers find high quality reviews. Reviewers also have clearer guidance on how to post more helpful reviews. Industry practitioners such as hotel managers can use incentive measures to encourage reviewers to post more useful reviews. Third-party platforms can use our results to update their algorithms to improve the review sorting mechanism.

As with other literature, our study has several limitations. First, the dataset in this paper is cross-sectional data, so it may not be able to observe the dynamic behaviors of individuals. Future studies may use panel datasets and conduct further studies. Second, this paper uses the hotel review data from Qunar.com. Future studies may extend this research to other industries such as online shopping, entertainment, and online education.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China (71972004, 71832011).

REFERENCES

- [1] Mudambi S.M, Schuff D. What makes a helpful online review? A study of customer reviews on Amazon.com[J]. MIS Quarterly, 2010, 34(1): 185–200.
- [2] Saleh K. The importance of online customer reviews [R/OL]. [2018-04-04]. <https://www.invespro.com/blog/the-importance-of-online-customer-reviews-infographic/>.
- [3] Chatterjee P. Online Reviews: Do Consumers Use Them? [J]. Advances in Consumer Research, 2001, 28.
- [4] Korfiatis N, Garcia B E, Sanchez A S. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content[J]. Electronic Commerce Research and Applications, 2012, 11(3):205-217.
- [5] Racherla P, Friske W. Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories [J]. Electronic Commerce Research & Applications, 2013, 11(6):548-559.
- [6] Schindler R M, Bickart B. Perceived helpfulness of online consumer reviews: The role of message content and style [J]. Journal of Consumer Behaviour, 2012, 11(3):234-342.

- [7] Filieri R, Hofacker C F, Alguezaui S. What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score[J]. *Computers in Human Behavior*, 2018,80:122-131.
- [8] Swagato C. Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach[J]. *International Journal of Hospitality Management*,2020,85:102356.
- [9] Yin Dezhi, Bond S D, Zhang Han. Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews[J]. *MIS Quarterly*,2014, 38(2):539-560.
- [10] Yin Guopeng. What online reviews do consumers find more useful? -- The influence of social factors[J]. *Management World*,2012(12):115-124 (in Chinese)
- [11] Zhiwei Liu, Sangwon Park. What makes a useful online review? Implication for travel product websites[J]. *Tourism Management*, 2015, 47(47):140-151.
- [12] Pan Yue, Zhang J Q. Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. [J]. *Journal of Retailing*, 2011, 87 (4):598-612.
- [13] Laura Connors, Susan Mudambi, David S. Is It the Review or the Reviewer? a Multi-Method Approach to Determine the Antecedents of Online Review Helpfulness[C]//44th Hawaii International Conference on Systems Science (HICSS-44 2011), Proceedings, 4-7 January 2011, Koloa, Kauai, HI, USA. IEEE Computer Society, 2011.
- [14] A. Ghose and P. G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1498-1512, Oct. 2011, doi: 10.1109/TKDE.2010.188.
- [15] Zhang Yanhui, Li Zongwei. Analysis of the Factors that Influence Online Reviews Helpfulness: Based on the regulating Effect of Product Type[J]. *Management Review*, 2016,28(10):123-132 (in Chinese)
- [16] Huang N, Hong Y, Burtch G. Social network integration and user content generation: Evidence from natural experiments[J]. *MIS Quarterly*, 2017, 41(4):1035-1058.
- [17] Pu J, Chen Y, Qiu L, et al. Does Identity Disclosure Help or Hurt User Content Generation? Social Presence, Inhibition, and Displacement Effects[J]. *Information Systems Research*, 2020, 31.
- [18] Lee, Young-Jin, Hosanagar, Kartik, Tan, Yong. Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings[J]. *Management Science*, 2015, 61.
- [19] Sussman S W, Siegal W S. Informational influence in organizations: An integrated approach to knowledge adoption [J] .*Information Systems Research*,2003,14(1):47-65.
- [20] Kang, R., Brown, S., and Kiesler, S. Why do people seek anonymity on the internet? informing policy and design[C] // Sigchi Conference on Human Factors in Computing Systems. ACM, 2013: 2657-2666.
- [21] Smith T, James, et al. Reconsidering Models of Influence: The Relationship between Consumer Social Networks and Word-of-Mouth Effectiveness[J]. *Journal of Advertising Research*, 2007, 47(4):387-397.
- [22] Susarla A, Oh J H, Tan Y. Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube[J]. *Information Systems Research*, 2012, 23(1):23-41.
- [23] Cialdini, Robert, B. *Harnessing the Science of Persuasion. (cover story)* [J]. *Harvard Business Review*, 2001.Bird R B, Stewart W E. (1960). *Lightfoot E N. Transport Phenomena*. New York: John Wiley & Sons Inc, 75

Short Research Paper

The Influence of Perceived Trust and Time Moderator on the Purchase Intention of Consumers in the Context of E-Commerce Livestreaming

Kun Jia*

School of Information, Central University of Finance and Economics, Beijing, 100000, China

Abstract: Due to the immediate surge of digital economy brought by the COVID-19 pandemic, e-commerce livestreaming industry in China has been quickly emerging as an important platform for online shopping. However, the factors that affect the trust of consumers in online shopping are different from those in offline economy. E-commerce livestreaming hosts are the key figures for the sales of products in the livestreaming room. Therefore, this paper will adopt Stimulus-Organism-Response (SOR) model to study the influence of e-commerce livestreaming hosts' professionalism, interactivity and popularity on the perceived trust of consumers, further affecting the purchase intention of consumers. It is worth noting that the change of day and night also plays a moderating role in the purchase intention of consumers. In this paper, empirical research method will be adopted and questionnaires will be distributed through Internet. The data will be analyzed in Structural Equation Model(SEM) through SPSS software and LISREL software. Suggestions are expected to be brought up to promote the development of e-commerce livestreaming.

Keywords: E-commerce livestreaming, perceived trust, e-commerce livestreaming hosts, purchase intention

1. INTRODUCTION

Due to the inconvenience of travel and the risks of infection, many offline enterprises have suffered heavy losses or even went bankrupt. Yet opportunities emerged while the challenges prevailed. During the coronavirus pandemic, online shopping reached a new “golden era” and e-commerce livestreaming boomed rapidly. AliResearch published that the transaction amount of livestreaming on “Double 11” in 2019 exceeded 20 billion yuan, with 177 e-commerce livestreaming hosts' annual transaction amount exceeding 100 million yuan^[1].

E-commerce livestreaming refers to the delivery of e-commerce activities and transactions via a live streaming platform^[2]. Up to June 2020, the number of netizens in China has reached 940 million, 309 million of which are e-commerce livestreaming users, accounting for 32.9% of the total netizens^[3]. With the development of technology, hundreds of thousands of consumers can watch a livestreaming room at the same time. When the livestreaming hosts introduces the products, the link of the products will be displayed on the livestreaming room page. If consumers want to buy the products directly, they just need to click the link and pay for the bill. In addition, consumers can communicate with livestreaming hosts through online comments cause the hosts will select some comments to answer in real time.

Some scholars are very interested in this phenomenon. The rise of livestreaming economy, however, has taken place in the recent few years. The research period of scholars is still short. The current aspects involved include the marketing mode of e-commerce livestreaming, industry analysis, purchase intention, etc. Most importantly, the significance of e-commerce livestreaming hosts is only noticed in a few previous studies, and the hosts are often confused with opinion leaders^[4].

In real life, all products are introduced and promoted by the livestreaming hosts. In this context, which feature of hosts will affect the perceived trust of consumers? Besides, Ministry of Human Resources and Social

* Kun Jia. Email: jk13718506828@163.com

Security in China also stipulated the job requirements of “live salesman” as a new profession in 2020^[5], which provided a clearer basis for the concept of e-commerce livestreaming hosts. Therefore, this paper use "Stimulus-Organism-Response" model (SOR model) to analyze the influence of e-commerce livestreaming hosts' professionalism, interactivity and popularity on the perceived trust of consumers, further affecting the purchase intention of consumers.

In addition, previous studies noticed the virtual situation in livestreaming room. For example, in the process of recommending products, livestreaming hosts build beautiful virtual situation through exquisite words and expressions, such as "This is the perfect lipstick for your lips in dates!"^[6]. However, there are few studies focus on the actual situation of consumers, which is the missing point of previous studies. If the time has changed, will consumers' purchase intention be different? So this paper will also analyze the moderating effect of the change of day and night on the purchase intention of consumers, draw conclusions through empirical analysis and put forward practical suggestions for the special situation. This paper will remind the practitioners to sale strategically instead of following the other people blindly.

As a short paper, the structure is arranged as following: first review the previous literature on this topic, then put forward the hypotheses and research model include the moderating effect. This paper will adopt a survey method to get the questionnaire data, which is going to be analyzed in Structural Equation Model (SEM). The final part is the further work of this paper.

2. LITERATURE REVIEW

2.1 SOR model

Mechrabian and Russell proposed the famous "Stimulus-Organism-Response" model (SOR model) based on environmental psychology in 1974^[7]. The SOR model points out that the external environment produces impacts on the individual's cognitive and emotional state, which has an impact on individual behavior. Eroglu used SOR model for the first time in online shopping, and constructed an impact model of online shopping environments on consumers' purchases^[8]. The study concluded that different online store atmospheres will produce different effects on the internal state of consumers, which will result in different shopping results.

In the context of e-commerce livestreaming, the only subjects that consumers can see are e-commerce livestreaming hosts. Therefore, e-commerce livestreaming hosts have become the only source of stimulus, which affects the perceived trust of consumers, and further affects the purchase intention of consumers.

2.2 Perceived trust

Regarding the definition of trust, Morgan and Hunt pointed out in 1994 that trust means that one party has enough confidence in the reliability and honesty of the other party in the cooperation^[9]. In the business context, Das and Ten defined trust as the buyer's trust in the reliability of the seller in the transaction, which is the approval of buyers toward the seller in a risky transaction^[10].

This nature of e-commerce livestreaming relative to the traditional commerce context gives rise to information asymmetry and transaction risks: identity uncertainty of partners, fear of their opportunism and product quality uncertainty^[11]. Especially in e-commerce industry, trust is one of the most important factors that affect the purchase intention^[12].

2.3 E-Commerce livestreaming hosts

Since e-commerce livestreaming is an emerging economic form. It is until June 2020 that the authority confirmed the “live salesman” as an official profession. That is the reason why most of the existing literature do not treat e-commerce livestreaming hosts as a real profession, and confuse them with opinion leaders.

Hosts with powerful influences can indeed be regarded as opinion leaders, but the detailed definitions of them are profoundly different. The earliest study on opinion leaders originated from Lazarsfeld et al, which

believes that information always flows from traditional media to opinion leaders^[13]. The opinion leaders got the information, stepped into re-dissemination, and then the information flowed to the general public. As Goldsmit stated, opinion leaders can directly influence other people's search and purchase through various methods of recommendations^[14]. However, e-commerce livestreaming hosts and opinion leaders are different. According to the requirements of the new profession as “live salesman” from the Ministry of Human Resources and Social Security in China, the task of this occupation includes not only the duties of opinion leaders, namely the promotion of products to the general public, but also includes the review of product information such as corporate qualifications and product quality. Besides, livestreaming hosts are supposed to participate in the post-sales service. If there is a quality problem with the goods purchased in the livestreaming room, consumers can request the hosts for return or compensation. Therefore, opinion leaders and e-commerce livestreaming hosts should not be confused as the same.

At present, there are very few papers that deliver substantial study on e-commerce livestreaming hosts. During the exploration on the attributes of e-commerce livestreaming hosts, professionalism is often related to appearance attractiveness, charm attributes, and recommended attributes. But these attributes are not the essence of livestreaming hosts. Therefore, this paper will study the impact of the livestreaming hosts' professionalism, interactivity and popularity on the perceived trust of consumers based on the government's definition of this new profession.

2.4 Purchase intention

Most of the previously cited documents involve the purchase intention of consumers. Regarding the definition of purchase intention, Schiffma believes that purchase intention can measure the possibility of consumers buying a product^[15]. The higher the purchase intention is, the more likely it is for consumers to eventually buy products.

Specifically in the e-commerce livestreaming context, the purchase intention may either be a deliberate decision or impulse buying. Both purchase intentions are worth noting. Due to the short duration of the livestreaming, the goods will be sold out in a few seconds, impulsive consumption accounts for a large proportion in online shopping. For example, perceived pleasure and perceived arousal can affect impulsive purchase intentions of consumers^[16].

3. RESEARCH MODEL AND HYPOTHESES DEVELOPMENT

Based on the discussion about the background and theoretical model, the research model of this paper is as follow:

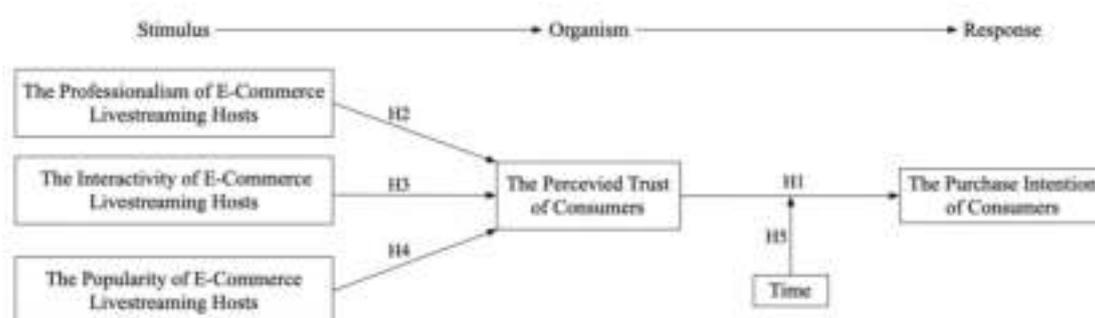


Figure 1. Research model

Yongsheng Zhou and Shihua Tang lately applied SOR model in the field of e-commerce livestreaming based on the perspective of social presence^[17], they pointed out that emotional social presence will affect the purchase intention of consumers through the mediating variable of perceived trust. According to SOR model, the

stimulus first affects the body and then produces a response. Therefore, emotion changes have an important impact on actions. In e-commerce, trust is one of the most important factors that affect the purchase intention. The stronger consumers' trust in livestreaming hosts is, the easier it is to purchase products through recommendations. The following hypothesis is proposed:

H1: The perceived trust will positively affect the purchase intention of consumers.

With the rapid development of e-commerce livestreaming, more and more e-commerce livestreaming hosts have emerged. Ying Xie and Chunqing Li defined e-commerce livestreaming hosts as a new type of group that provides consumers with products display through product trials and experience sharing on livestreaming platforms, thereby promoting consumers' purchase^[18]. Jiaqi Li, the leading e-commerce livestreaming hosts in China, builds up his reputation as a host with fair evaluation through real feedback and criticism of products, so as to gain the trust of consumers.

But the task of e-commerce livestreaming hosts is not only to introduce in front of the camera, but also requires numerous work before and after the livestreaming. Therefore, based on the regulations of "live salesman", this paper innovatively divides the measurement items of "the professionalism of e-commerce livestreaming hosts" into three aspects, namely, products selection planning before livestreaming, products recommend during livestreaming and related services after livestreaming. Because a truly professional livestreaming hosts will not only concentrate on the product introduction during the livestreaming, but also strictly control the product selection process before the livestreaming. Participating in after-sales service is also part of the duties for e-commerce livestreaming hosts. Such a comprehensive service will arouse the perceived trust of consumers.

H2: The professionalism of e-commerce livestreaming hosts will positively affect the perceived trust of consumers.

Compared with offline commerce where consumers can actually experience storefronts, products, and service personnel, the only subject that consumers can see in e-commerce livestreaming room is the livestreaming hosts, which means that all consumers' sources of trust come from the livestreaming hosts, so the livestreaming hosts needs to maintain real-time interactions to generate a sense of intimacy and closeness with the consumers^[19]. Yingxin Chen indicates that there is a significant positive correlation between the interaction of sellers and consumers and the establishment of trust^[20], which means that the more interaction produced between sellers and consumers, the higher the level of customer trust establishment will be.

For example, the well-known e-commerce livestreaming hosts Ya Wei builds an Internet celebrity brand by cultivating stickily loyal fans. The development of modern technology allows the live commenting to be displayed in real-time on the pages of the hosts and the consumers. Consumers can ask questions through internet, the livestreaming hosts can answer and chat in the livestreaming room. Even though the livestreaming hosts and consumers are in different spaces, they can still achieve full interaction. It greatly reduces consumers' uncertainty about the product and increases their trust.

H3: The interactivity of e-commerce livestreaming hosts will positively affect the perceived trust of consumers.

In modern society, "attention" is a scarce resource. When multiple livestreaming hosts are live at the same time, consumers often choose the most popular livestreaming hosts. Because of the herd effect, consumers will follow the choices of most people, believing that the more popular the hosts are, the better their reputation might be.

Some small hosts with limited fame have not established their reputation among the public. Consumers do not know the actual situation of these small hosts and cannot be sure of the quality of the products they introduce, so consumers will not trust the hosts without reputation. Consumers usually choose to trust the well-known livestreaming hosts in comparison.

H4: The popularity of e-commerce livestreaming hosts will positively affect the perceived trust of consumers.

It should be noted that even if consumers have perceived trust on livestreaming hosts, their purchase intentions do not remain unchanged. Consumers are often busy with work during the day and do not have enough time to watch e-commerce livestreaming. As time changes, at night, consumers are more likely to relax when they return to the comfortable environment at home, and it is easier for them to reward themselves. Guopeng Chen used a psychological perspective to explore that day and night preferences can predict impulsive shopping tendencies. Therefore, in the night dominated by perceptual thinking, the possibility of impulsive consumption is greatly increased, and the purchase intention is also higher than that during the day. Time changes play a great moderating role in this process.

Although there is no previous research about the moderating role of time in this topic, there are precedents in other aspects about time as moderating role. For example, Viswanath Venkatesh regard time as a moderating variable. As time changes, the impact of the health consultation kiosks on women' health in rural India also change^[21]. The introduction of time as a moderating variable into the topic is also one of the innovative thinking of this paper.

H5: Time plays a moderating role in the process of perceived trust affecting purchase intention of consumers.

4. METHODOLOGY

This paper will conduct empirical research through survey. The questions in the questionnaire will be developed around the research model. The questionnaire design is mainly based on the previous studies and focus group discussions. Each measurement item is denoted by Likert Scale, which contains five types of responses: "strongly agree", "agree", "not necessarily", "disagree", and "strongly disagree".

This paper will distribute the questionnaires through the Internet. Firstly, a small part of the questionnaires will be distributed for the exploratory factor analysis. If the initial result is reasonable, the scope of the distribution will be expanded. After collecting the questionnaire data, this paper will use SPSS to analyze the reliability and validity of the questionnaire. Then the measurement items and factors in the questionnaire will be matched in Structural Equation Model through LISREL. Factor loading and path coefficients are tested to analyze the degree of fitness between the questionnaire data and the model in this paper. The test of moderating effect will be carried out based on the suggestions of Zhonglin Wen and Zhen Zhang^[22].

5. IMPLICATIONS AND FURTHER WORK

It is easy to ignore the interesting phenomena happening around us, such as why "Double 11" starts in the middle of the night? Why is it easier to eat snacks at night? It is because this special time has a moderating effect on consumers' purchasing intentions in the evening. As mentioned before, most of the previous studies focus on the impact of the virtual context constructed by the livestreaming hosts on the consumer's purchase intention, such as drawing the beautiful application scenario of lipstick to attract consumers to buy, but there is little discussion about the physical environment of the consumer. The academic value of this article is to innovatively propose the moderating effect of time in e-commerce livestreaming. This paper discusses the

factors that affect perceived trust in the context of e-commerce livestreaming. It also gives a clearer view on the characteristics of e-commerce livestreaming hosts according to the Ministry of Human Resources and Social Security in China.

In real life, many practitioners try to market through livestreaming, but the number of their audience is so few that results in minimal sales. This article will use empirical analysis to draw conclusions and provide practical suggestions for e-commerce practitioners to improve the situation.

The further work is to design the questionnaire based on the review and thinking of the previous studies, and then determine the measurement items in the questionnaire through focus group discussion, then distribute the questionnaire through Internet. After collecting the questionnaire data, this paper will use SPSS to analyze the reliability and validity of the questionnaire. Then the data will be analyzed using Structural Equation Model through LISREL. At the final step, this paper will draw conclusions and put forward practical suggestions.

REFERENCES

- [1] AliResearch. (2020). The new economy report of Taobao e-commerce livestreaming in 2020. <http://www.aliresearch.com/ch/information/informationdetails?paperCode=56954219873308672&type=%E6%8A%A5%E5%91%8A> (in Chinese)
- [2] Xiaoyu Xu, Jen-Her Wu, & Qi Li. (2020). What Drives Consumer Shopping Behavior in Live Streaming Commerce? *Journal of Electronic Commerce Research*, 21(3): 144–167.
- [3] CNNIC. (2020). The publication of China Statistical Report on Internet Development. *China Broadcast*, 2020(11): 54 (in Chinese)
- [4] Lu Meng, Fengjun Liu, Siyun Chen, Shen Chen. (2020). The capability to pick you up ——the research on the influence mechanism of the different information source characteristics from various Internet celebrities. *Nankai Business Review*, 23(01): 131-143. (in Chinese)
- [5] The Ministry of Human Resources and Social Security of China. (2020). The notification on publishing the new careers. http://www.mohrss.gov.cn/gkml/zcfg/gfxwj/202007/t20200706_378490.html (in Chinese)
- [6] Xiaoyi Han, Zhengliang Xu. (2020). The implication from the attributes of e-commerce livestreaming hosts on the purchase intention of consumers ——based on Grounded Theory. *Foreign Economics and Management*, 42(10): 62-75. (in Chinese)
- [7] Jacoby J. (2002). Stimulus-Organism-Response Reconsidered: An Evolutionary Step in Modeling Consumer Behavior. *Journal of Consumer Psychology*, 12(1): 51-57.
- [8] Sevgin A. Eroglu, Karen A. Machleit, Lenita M. Davis. (2001). Atmospheric qualities of online retailing. *Journal of business research*, 54(2): 177-184.
- [9] Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(3): 20-38.
- [10] Das, T. K & Teng, B. S. (1998). Between trust and control; developing confidence in partner cooperation in alliances. *Academy of Management Review*, 23(3): 491-512.
- [11] Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26(3): 243–268.
- [12] Ganguly, B., Dash, S. B., & Cyr, D. (2009). Website characteristics, trust and purchase intention in online stores: An empirical study in the Indian context. *Journal of Information Science & Technology*, 6(2), 22-44.
- [13] Lazarsfeld P.F., Berelson B., Gaudet H. (1944). *The People's Choice: How the Voter makes up his Mind in a Presidential Campaign*. New York: Columbia University Press, 99-100.
- [14] Goldsmith R.E., Flynn L.R., Goldsmith, E.B. (2003). Innovative consumers and market mavens. *Journal of Marketing Theory and Practice*, 11(4): 54-64.

- [15] Schiffman L.G. Kanuk L.L. (2000). *Consumer Behavior*. 7th. Wisconsin: Prentice Hall, 135-137.
- [16] He Xu, Hongjian Qu, Jianzhong Cai. (2021). The influencing factors of impulsive purchase intention of clothing consumers in the context of livestreaming. <http://kns.cnki.net/kcms/detail/31.1865.N.20201110.1813.011.html>. (in Chinese)
- [17] Yongsheng Zhou, Shihua Tang, Jing Xiao. (2021). The study on the purchase intention of consumers in e-commerce livestreaming platform—from the perspective of social presence. *Contemporary Economy Management*,2021(01): 1-11. (in Chinese)
- [18] Ying Xie, Chunqing Li, Peng Gao. (2019). The study on the implication of livestreaming market on online group consumption—from the perspective of conducts and neuroscience. *Advances in Psychological Science*,27(06): 990-1004. (in Chinese)
- [19] Sjoblom, Max, Hamari, Juho, Hilvert-Bruce, Zorah, et al. (2018). Social motivations of live-streaming viewer engagement on Twitch. *Computers in human behavior*, 84(Jul.):58-67.
- [20] Yingxin Chen, Xutong Gao.(2021). Research on mutual trust between buyers and sellers in e-commerce livestreaming. *Chinese Management Science*, 29(02):228-236. (in Chinese)
- [21] Venkatesh V, Sykes TA. (2020). Ict for Development in Rural India: A Longitudinal Study of Women's Health Outcomes. *MIS Quarterly*, 44(2):605-629.
- [22] Zhonglin Wen, Lei Zhang. (2006). The moderated variables with mediating effect and the mediate variables with moderating effect. *Psychological News*, 2006(03): 448-454. (in Chinese)

Full Research Paper**Solution Exemplars and Sales Performance of Crowdsourcing****Solvers: the Moderating Role of Reputation and Competence***Quanwu Zhao¹, Zhiyuan Zhou², Xi Yang^{3*}*^{1 2 3}School of Economics and Business Administration, Chongqing University, Chongqing, 400044, China

Abstract: An increasing number of signals are using to entice solvers to make online purchases by seekers in the competitive online markets today. However, how solution exemplars are in terms of their reputation or competence to improve sales performance has not yet been investigated. Extending signal theory to the online service marketplaces, we analyses the effect of solution exemplars' structural characteristics on seekers' sales performance such as quantity, diversity and popularity, exploring the moderating impact of seeker's reputation and competence. We test the model using data from ZBJ.com, a popular crowdsourcing contest platform in China. Our analysis conducts a series of interesting findings, the impact of exemplar quantity and popularity on sales performance is positively significant, contrary to solution exemplar diversity. Regarding the moderation effects, reputation is proved to be negative, which is opposite to competence. We also elaborate on the theoretical contribution and practical significance.

Keywords: solution exemplars, online service marketplaces, reputation, competence

1. INTRODUCTION

Online service marketplaces are online transaction platforms hosted by a third party to facilitate exchanges for solutions between solvers and seekers (e.g. Innocentive (www.innocentive.com) and Zhubajie (www.zbj.com)), which is different from traditional online product marketplaces. While in online service marketplaces, the final solution often does not exist yet, which needs to be customized according to the needs of seekers. Therefore, and potential solvers still face the problem of information asymmetry, which deters them from purchasing ^[1]. For solvers, the key to online service marketplaces success is be better than competitors at addressing such information asymmetry, in order to entice seekers to buy.

One common strategy is to give seekers extrinsic cues, such as warranties, online reputation indexes and a visually appealing web site quality ^[2], which would be called "signals" in the signaling literature ^[3]. Some scholars have certified that effective signals can reduce information asymmetry and enhance a seeker's trust in transacting with a seller, thereby promoting purchase behavior ^[4]. In the online service marketplaces, the platforms offer the "solution exemplars" display interface as a kind of signals, which can be used to build professional-looking websites and highlight the solvers' experience visually. It is an important signals in online service marketplaces. However, little work has been done to research the signals of solution exemplars in online service marketplaces.

Solvers are commonly found to compete by using solution exemplars to initiate many diverse and distinctive actions, as they can display different types of solution exemplars and their cooperators.

Thus, the structural characteristic of solution exemplars and solver-specific factors could impact sales performance in a highly nascent and competitive online service marketplaces ^[5]. This study aims to examine these two research question on which the literature is resoundingly silent: what is the different effects of solution exemplars' structural characteristic on the actual sales performance of solver and how does solvers' reputation and competence moderate these relationships in online service marketplaces?

* Corresponding author. Email: 1055075128@qq.com(Xi Yang)

To approach these research questions, we depicts how three dimensions of solution exemplars' structural characteristic—exemplar quantity, exemplar diversity and exemplar popularity— impact solvers' performance, contingent up on reputation and competence. Then we test the model using data from 244 solvers in May 2020 on ZBJ.com, which is a popular online service marketplaces in China.

2. RESEARCH CONCLUSIONS

The results show that the impact of exemplar quantity and exemplar popularity on sales performance is positively significant. Contrary to our expectation, the relationship between exemplar diversity and sales performance is non-significant. Regarding the moderation effects, reputation is proved to moderate the impact of exemplar diversity and popularity on performance negatively. As predicted, at high levels of competence, sales performance is positively related to exemplar structural characteristics.

This study has important practical significance for both service platforms and solvers.

For platforms, a complete exemplar display function system should be built to encourage solvers to display solution exemplars, reducing information asymmetry, and attracts more seekers to trade on the platform finally.

Our results suggest that solvers should be more concerned with the quantity and quality (popularity) of the solution exemplars rather than the diversity, especially seekers with high reputation. Simply put, the seeker creates an active and friendly image, as well as a need to be professional in presenting valuable exemplars, which will attract more seekers.

Compared with solvers' reputation, they should pay more attention to the impact of competence. That is, the structural characteristics of solution exemplar is always positive impacting solvers' sales performance under the high competence. On the contrary, our research results also show that when the solver's reputation is relatively high, the solver should be very cautious and not spend too much time in providing a variety of exemplar displays.

3. RESEARCH CONTRIBUTIONS

Compared with previous studies, our research has three theoretical contributions. First, although the exemplar display of solvers is crucial to performance in online service marketplaces, it is rarely studied. We study the impact of the structural characteristics of solvers' exemplar on sales performance for the first time. Secondly, we divide exemplar structural characteristics into three categories through literature research: exemplar quantity, exemplar diversity and exemplar popularity. Finally, we also discuss the moderating effect of reputation and competence of solvers.

REFERENCES

- [1] Wells J D, Valacich J, Hess T J. (2011). What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions. *MIS quarterly*, 35(2): 373-396.
- [2] Li H, Fang Y, Wang Y, et al. (2015). Are all signals equal? Investigating the differential effects of online signals on the sales performance of e-marketplace sellers. *Information Technology & People*, 28(3): 699-723.
- [3] Michael S. (1973). Job Market Signaling. *Quarterly Journal of Economics*, 87(3):355-374.
- [4] Connelly B L, Certo S T, Reutzell C, et al. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1): 39-67.
- [5] Li H, Y Fang, Lim K H, et al. (2019). Platform-Based Function Repertoire, Reputation, and Sales Performance of E-Marketplace Sellers. *MIS Quarterly*, 2019, 43(1): 207-236.

Short Research Paper**Research on Influencing Factors of User Satisfaction of Knowledge****Paid Products -- from the Aspect of Heuristic-Systematic Model***Yufei Han¹, Chuyu Jiang², Yanping Zeng^{3*}*¹School of Information, Central University of Finance and Economics, China²School of Information, Central University of Finance and Economics, China³School of Economics and Finance, University of International Relations, China

Abstract: Zhihu Live, a knowledge paid product, is a special virtual product of real-time interactive voice experience. Information seekers cannot obtain complete content before making payment decisions. Therefore, they need to evaluate the perceived usefulness of Live information through systematic and heuristic clues to make payment decisions. After acquiring complete knowledge, users will score or evaluate user satisfaction based on the difference between perceived usefulness before the payment stage and actual quality of knowledge. Zhihu Live allows speakers to describe knowledge in text and voice. Based on this, this paper explores how heterogeneous information features transmit different clues, enhance users' perceived usefulness of information, and then influence the final user satisfaction of the product. Combining the Information Foraging Theory, Information Acceptance Model and Heuristic-Systematic Model, we set up knowledge characteristics of heterogeneous information on the impact of user satisfaction measurement model of knowledge paid product, and to test text and voice characteristics effect on user satisfaction, and the moderating effect that theme features have on user satisfaction.

Keywords: user satisfaction, Information Acceptance Model, Heuristic-Systematic Model, Zhihu Live

1. INTRODUCTION

With the development and promotion of the concept of knowledge sharing, combined with the promotion of Internet and other new technologies, knowledge sharing has broken the inherent characteristics of the original free, and started the road of monetization. From then on, the era of "paying for knowledge" in China has officially kicked off. After several years of precipitation, knowledge payment industry has begun to take shape, but it has also entered the bottleneck period of development. According to the Annual Report on the Development of China's Sharing Economy, the direct financing scale of knowledge and skills in China's sharing economy was about 46.4 billion yuan in 2018, up about 75.4% year on year. In 2019, the financing scale reached 31.4 billion yuan, a year-on-year increase of -32.3 percent. From the data performance, 2018 has been the peak year of knowledge sharing, and after 2019, the scale of direct financing in the market has shown a trend of decline, which shows that investors are not optimistic about the future of knowledge sharing, which may be related to the use experience of investors. Therefore, how to improve the user satisfaction of knowledge paid products, enhance the perceived usefulness of users' information, and then motivate them to make knowledge payment decision has become a difficult problem for knowledge payment service providers to solve.

Zhihu Live is a kind of real-time Live interactive knowledge paid products. Speakers share knowledge and skills through real-time voice, pictures, Live broadcast and other diversified information forms. The audience can interact with the speaker through questions and answers. Compared with other information, the right to use the product can only be acquired after the cost of knowledge acquisition is paid. In order to prevent the problem of information asymmetry, Zhihu Live added relevant information to the product page to help users understand the applicability of the product, including content introduction, outline information, rating stars, popularity,

* Corresponding author. Email: 13581956128@163.com (Yanping Zeng)

feedback and evaluation clues ^[1]. Under the limitation that users can't get all the information, the Live speakers convey the quality clues of knowledge paid products by carefully designing the content introduction information and audio audition information of the Live page, and enhance users' perceived usefulness of knowledge so as to arouse users' impulse of knowledge payment. Users will form a first impression of Live based on the text clues provided by Live speakers, and combine with the audio audition function to form a perceived usefulness of knowledge of the audio quality. When the user's usefulness reaches a certain threshold, they choose to make a payment decision to accept this information. After paying for the information, users can submit comments or score the Live event to express their satisfaction with the quality of the information or experience ^[2].

2. LITERATURE REVIEW

2.1 Heuristic-Systematic Model

There is no agreement on the specific definition of heuristic behavior and systematic behavior. Chaiken (1980) believes that Heuristic-Systematic Model divides individual information behavior process into heuristic behavior and systematic behavior ^[3]. Heuristic behavior is based on intuition, usually based on the external clues of information to make simple judgments; Systematic behavior, on the other hand, is based on rationality and requires the use of sufficient cognitive resources to evaluate information. Shi et al. (2020) use the Heuristic-Systematic Model to distinguish these signal clues ^[4]. This model holds that users will process and obtain relevant information through heuristic clues and systematic clues, in which heuristic clues are usually related to external information such as information release source and content creator, and knowledge seekers need to pay less cognitive effort. Systematic cues are relevant to the content itself and require sufficient cognitive effort.

The Heuristic-Systematic Model provides a theoretical explanation for users to search for information, process information, evaluate information, receive information and make payment decisions in the context of knowledge payment. In Zhihu Live page, knowledge seekers can learn about the scope of the Live argument and the story elaboration style of the Live presenter from the text information such as the content introduction and outline. They can also learn about the content expressive force of the Live presenter in the actual voice interaction process from the audio information with the help of the trial listening function of Live. Therefore, in this paper, we regard text information and voice information as systematic clues in this study. In addition, knowledge seekers can clearly know the reputation clues such as the number of likes and fans received by the speakers. They can also judge the popularity of the Live course by the number of participants and the number of comments. Here, the number of paying users and the speaker's reputation information are regarded as heuristic clues in this study.

2.2 Research status of user satisfaction of knowledge paid products

For the study on user satisfaction of knowledge paid products, scholars mostly use questionnaires or interviews and other research methods to analyze and empirically test the model through structural equation model. Jin Xiaopu et al. (2020) designed relevant questionnaires to analyze the current situation of user satisfaction of knowledge payment platforms from four aspects ^[5]. In addition to the indicators in questionnaires and interviews, Huang Jiahui et al. (2018) measured the service utility with two variables: the number of appointments and the number of post-consultation reports ^[6]. Wang Hui (2019) uses the high praise rate as the proxy variable of user satisfaction ^[7].

The influencing factors of previous researches on product user satisfaction include non-product variables such as review information, information communicator and information source. There are few researches on quantifying the influence of information characteristics on user satisfaction with actual secondary data, which makes it difficult to match product user satisfaction at the level of actual product quality. Just study also

indicated that, information readability ^[8], persuasive information ^[9] and emotional tendency ^[10] affect product performance text information characteristics, such as, characteristics of information about their goods is introduced, the influence of user satisfaction research is still less, Live speaker itself provides information of text and voice features can affect user satisfaction of knowledge paid product and influence mechanism path is lack of exploration. As an important information clue that affects users' subsequent purchasing decisions and feedback evaluation, text information bears the task of introducing information content context, text structure and key knowledge points to information seekers. Voice messages can highlight context and semantics more, and information seekers can better grasp the real feelings of Live speakers and the actual experience during the Live broadcast through voice messages.

Perspective, so this article is based on user satisfaction of knowledge paid product to praise rate as proxy variable of user satisfaction, focusing on the analysis of knowledge distributors to provide knowledge information quality characteristics, to explore the influence factors of user satisfaction and the characteristics of different information to the user satisfaction with the role of the path.

2.3 Research status of Heuristic-Systematic Model and user satisfaction

Heuristic-Systematic Model has been applied more and more to the research of network information behavior, but less to the field of knowledge sharing in virtual community. Chen Minghong et al (2015) empirically concluded that knowledge sharing satisfaction is influenced by both systematic factors and heuristic factors based on structural equation model and questionnaire survey ^[11]. Based on stepwise regression and structural equation model, Tao Xiaobo et al. (2020) empirically found that the major factors of classification and recognition of heuristic cues and systematic cues have a positive effect on information adoption behavior ^[12]. Therefore, based on the Information Acceptance Model, when the user's perceived usefulness of the product reaches the threshold value and carries out the knowledge payment operation, the subsequent knowledge payment operation caused by the perceived usefulness can be regarded as the information acceptance behavior. The user's perceived usefulness of the product is caused by systematic factors and heuristic factors. After the user adopts the operation of information adoption, they can get the opportunity to experience the complete product. However, there will be a certain gap between the feeling after the experience and the perceived usefulness at the beginning, which is defined as the user satisfaction.

2.4 Summary of this chapter

To sum up, the marginal contribution of this paper is expected to be as follows: First, HSM model is applied to the study on the user satisfaction of knowledge paid products, from the dual-processing theoretical model of individual behavior process, namely Heuristic-Systematic Model, to better understand the influence and mechanism of heterogeneous information characteristics on user satisfaction; Secondly, voice information, a quantifiable index, is introduced into the heterogeneous information feature analysis to quantitatively analyze the influence of voice information features on user satisfaction from multiple dimensions, such as voice duration and voice clarity. Thirdly, in order to further understand the influence mechanism of user satisfaction, it explores whether text or voice features are more suitable for Zhihu Live under different theme types.

Based on zhihu Live as the empirical object, combining the Information Foraging Theory, Information Acceptance Model and Heuristic-Systematic Model, from the heuristic characteristic clues and systematic features two perspective to study its effect on product user satisfaction, and on this basis to explore the perceived usefulness and theme type in this influence to play a role. Specifically, this paper aims to solve the following problems : (1) How do heterogeneous information characteristics affect product user satisfaction? Do systematic information and heuristic information have different action mechanisms or influence degrees on user satisfaction of knowledge paid products? (2) For Zhihu Live under different theme types, what text/voice features should speakers adopt to effectively improve user satisfaction of the product?

In this paper, the measurement index of heterogeneous information characteristics is constructed at the theoretical level to promote the evolution of machine learning methods such as text analysis in the context of knowledge payment. In practice, it is helpful for paid knowledge lecturers to further improve the form and style of information, determine a reasonable range of information price and cost and release information clues that can satisfy users. Service developers to better understand the user's information, and thus targeted to improve the quality of information service, help to consider what information the extension form knowledge payment platform, and platform should provide voice value-added services to provide decision-making reference for detection, and strengthen the social attribute of knowledge platform.

3. RESEARCH HYPOTHESIS AND MODEL BUILDING

3.1 Theoretical basis

3.1.1 Information Foraging Theory and Research

According to the Information Foraging Theory, users' evaluation of information content quality depends on the background of Information provided. They will actively search, select and identify Information. After selecting Information, they will evaluate the actual information quality through signals and finally decide whether to adopt the Information. The theory of information seeking explains the adaptive relationship between information seekers and information providers. Information signals refer to clues, such as text or visual clues, that help information seekers to determine the potential value of specific information in a specific background. For Zhihu Live, information seekers, that is, users, can evaluate the quality of information provided by information providers on the page, and then decide whether to accept relevant information or not.

3.1.2 Information adoption model and its research

Information adoption behavior is a process in which the subject purposefully selects, evaluates, accepts and uses information, and this process will ultimately affect the subject's subsequent behavior. In order to explain the Information Acceptance behavior of users, Sussman et al.(2003) proposed the Information Acceptance Model, which regarded the process of Information influencing people's decision as the process of Information Acceptance, and regarded the quality of arguments and the credibility of information sources as the central path and the edge path of influencing Information Acceptance respectively ^[13].

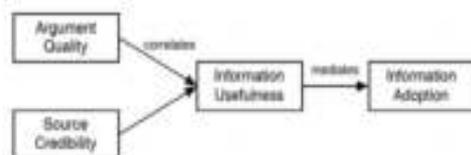


Figure 1. Information Acceptance Model

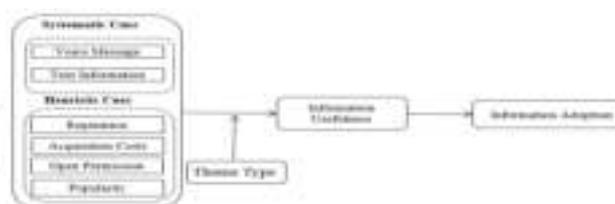


Figure 2. Model constructed12

3.2 Research hypothesis and model proposed

3.2.1 The influence path of systematic cues on user satisfaction

(1) The influence of text information characteristics on user satisfaction

Thematic correlation degree refers to the cosine similarity degree between Zhihu Live outline and title. When the outline description is more consistent with the theme, it indicates that the outline plays a good supplementary role for the theme, which helps users to grasp the theme content and avoid too big psychological gap for the theme content after paying for the operation, thus helping to improve user satisfaction.

In addition, the number of knowledge points displayed in an article, whether the sentences are concise and

whether the sentences are lengthy will affect the user's reading experience. Especially when the information provider is verbose and obscure, it tends to affect the user's understanding, resulting in low perceived usefulness. Based on the above analysis, the following hypotheses are proposed:

H1a: Thematic relevance is positively correlated with user satisfaction.

H1b: The amount of information is negatively correlated with user satisfaction.

(2) The influence of voice message characteristics on user satisfaction

Voice interactivity refers to the number of messages that are answered to a paying user's questions during each Live audio session. If users' questions in Live can be answered quickly and effectively, it can enhance users' sense of participation and contribute to the improvement of user satisfaction. For different real-time Live services, a short voice tone provided on a page before payment will also appeal to different groups of users. For Live, which is about entertainment, a voice that is too serious and rigid is less popular with users, while for topics related to stating facts, such as history, it is more expected. In this study, the voice timbre was measured from the main timbre category and timbre score of the Live speakers. And the voice clarity of the speaker is also an important dimension of the effect of speech information on user satisfaction. The higher the mandarin level and pronunciation score of Live speakers is, the better product experience will be brought to users, which will help to improve user satisfaction. Based on the above analysis, the following hypotheses are proposed:

H1c: Voice interactivity is positively correlated with user satisfaction.

H1d: Voice timbre is positively correlated with user satisfaction.

H1e: Voice clarity is positively correlated with user satisfaction.

3.2.2 The influence path of heuristic cues on user satisfaction

Presenter reputation is measured by the number of fans and the number of approvals of the presenters. The higher the reputation of the keynote speaker, the stronger the authority and presentation ability of the keynote speaker, which is conducive to the user's consumption experience. Open access refers to Live's open access to users, that is, whether to listen according to the chapter, seven days for no reason refund, unlimited listening back and other permissions. The more open access, it will help users to listen to Live lectures again after the event, which can well improve user's consumption experience and enhance user satisfaction. Popularity refers to the popularity of Live, or how popular Live is. The more popular Live is, the higher the quality of Live is, which is worth viewing for users with potential demand, and user satisfaction will not be very low. Based on the above analysis, the following hypotheses are proposed:

H2a: Speakers' reputation is positively correlated with user satisfaction.

H2b: Open permissions are positively correlated with user satisfaction.

H2c: Live popularity is positively correlated with user satisfaction.

3.2.3 The moderating effect of theme type

Live in a social Q&A community is organized by topic for easy retrieval. Li Wu et al. (2019) also found that theme type can moderate the influence of content fun on content expression^[14]. Based on the above analysis, heuristics and systematic information characteristics of different topic types may have great differences in their impact on user satisfaction. Therefore, the following hypotheses are proposed:

H3: Theme type can moderate the influence of heterogeneous information characteristics on user satisfaction.

4. RESEARCH DESIGN

4.1 Data capture and cleaning

In this paper, the crawler is used to collect the relevant information of 931 Live events and Live speakers from September to December 2020 at an interval of 10 days. The data were preprocessed, and the evaluation star

level was eliminated (when the number of Live participants was small, the evaluation star level was not displayed) and the Zhihu Live sessions that had not been started, and 893 Live sessions were retained as the research data. On the basis of crawling the text data of Live real-time page, the data is classified and then text analysis is carried out, including text semantic analysis, text word frequency analysis and text style recognition.

4.2 Variable measurement and description

The specific variable measure is shown in Table 1, but some individual variables need to be specified.

4.2.1 Explained variable measurement

In Zhihu Live, paying users can give star ratings to Live content after buying knowledge Live products. The star score can represent the psychological gap between the user's perceived usefulness before acquiring the product and the psychological gap after paying for the product. Here, the number of Live reviews is used to measure the user satisfaction of paid knowledge in the user group, and the Live praise rate is used as the proxy variable of user satisfaction in the robustness test.

4.2.2 Explanatory variable measurement

Presentation style measure is by reading about the Live text, based on the theory of Aristotle's rhetoric and grounded theory, and based on social computing to Harbin Institute of Technology and information retrieval research center of synonym word Lin (extension) on the semantic analysis of persuasive language can be divided into five kinds: resorting to credible, resort to return, resort to emotional appeal to logic and state the fact. With the help of sound recognition software, the voice characteristics of Live in the samples were analyzed in terms of voice duration, voice interactivity, voice timbre and speech clarity.

4.2.3 Adjusting variable measurement

Zhihu Live topics include 17 categories, which are subdivided into two categories for simplicity. Theme division is based on reference to the theme division method of Fu Shaoxiong et al (2019)^[15]. The final topic partition results are shown in Table 1. The financial and economic, law, business, professional subject delimit social field. Design, art, and reading and writing delimit the cultural field. Areas of life contains travel, food, music, film and television and games, lifestyle, medical and health and sports. Education field consist of education, science and technology, Internet and psychology.

Table 1. An introduction to variable measure methods

Variable Types	Measuring Dimensions	Variable Name	The Meaning
Explained Variable	User Satisfaction	Review_volume	The number of people who paid for and commented on the Live event.
		Review_valence	Ratio of the number of people giving the Live event 5 stars to the number of total reviews
Explanatory Variables (Systematic Clues)	The text Information characteristics	Thematic_relevance	Cosine similarity between Live outline and Live title
		Outline_count	Number of Live syllabus points displayed
		Presentation_style	Text presentation style for Live content profiles
	voice Information characteristics	Live_duration	The duration of Live real-time speech
		Live_interactivity	The number of messages that were answered by a paying user during Live Live
		Live_timbre	Live the dominant tone of the speaker
		Timbre_score	Live presenter's tone rating
Live_clarity	Live speakers' mandarin level		
Moderating Variable	The topic type	Live_genre	The subject category in which Live is located

Variable Types	Measuring Dimensions	Variable Name	The Meaning
Explanatory Variables (Heuristic Clues)	Presenter reputation	Reputation	The sum of the number of fans and the number of approvals of the keynote speaker is standardized
	Acquisition costs	Live_Price	Live's price tag
	Open access	Open permissions	[Listen according to the chapter] [refund for seven days without any reason] [free listening for members] [unlimited listening back] and other Live open permissions
	popularity	Live_popularity	Live heat

4.3 Empirical model construction

In order to study the influencing factors of user satisfaction based on Heuristic-Systematic Model, the model is constructed as follows:

$$Review_volume = \beta_1 \times Heuristic + \beta_2 \times Systematic + \varepsilon_1 \quad (1)$$

$$Review_volume = \beta_1 \times Heuristic + \beta_2 \times Systematic + \beta_3 \times Live_genre + \varepsilon_2 \quad (2)$$

$$Review_volume = \beta_1 \times Heuristic + \beta_2 \times Systematic + \beta_3 \times Live_genre + \beta_4 \times (Heuristic \times Live_genre) + \beta_5 \times (Systematic \times Live_genre) + \varepsilon_3 \quad (3)$$

Among them, to be sure, because the part has not been empirically explore, just explain that are relevant to all possible variables listed in table 1, but in the end is not necessarily all use, so not to formula (1) the introduction of the interpretation of the specific variables, but expressed as a heuristic clues to the heuristic feature variables, systematic on behalf of the entire system type variable characteristics of cues. Equation (2) introduces the moderator variable of topic type on the basis of Equation (1), and Equation (3) introduces the interaction terms of moderator variable and all heuristic characteristic cue variables and the interaction terms of moderator variable and all systematic characteristic cue variables on the basis of Equation (2). Each β represents the standardized regression coefficients of their respective variables, $\varepsilon_1 \sim \varepsilon_3$ is the residual term.

Firstly, the structural equation model can be selected to verify the research hypothesis proposed above, or ordinary OLS regression or negative binomial regression can be selected to test the hypothesis, and the different influence paths of heuristics and system variables on user satisfaction can be obtained. Secondly, the moderating effect of theme category is analyzed by general regression. On this basis, a simple slope test and an interaction graph are also conducted to compare the influence of heuristic information cue and systematic information cue on user satisfaction under the moderating effects of topic type. Finally, the robustness test of the model is carried out. The preliminary idea is to use another proxy variable of the explained variable to conduct regression analysis again, and analyze whether the conclusion of the model is robust according to the results.

5. SIGNIFICANCE AND DEFICIENCY OF THE RESEARCH

5.1 Research significance

According to the research content of this paper, it can be of great reference value to information providers. Specifically, the two points are as follows: First, according to the relationship between heuristic and systematic variables and user satisfaction, the core factors affecting user satisfaction can be clearly recognized, and the corresponding information requirements can be further improved to bring better user experience; Secondly, the influence mechanism of heuristic and systematic characteristic variables is not necessarily the same under different theme categories, so that information providers can better manage their Live topics according to different topic categories.

5.2 Shortage of research

As Zhihu Live platform is selected for research and analysis, the externality of the research conclusion could not be guaranteed. Later, we can try to study the mechanism of Heuristic Systematic Model on user satisfaction under different knowledge payment communities.

REFERENCES

- [1] Cao Q, Duan W, Gan Q, et al. Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach[C]. *decision support systems*, 2011, 50(2): 511-521.
- [2] Zhao Hua, Xia Xiulan. Customer Lifetime Value Improvement Model Based on User Satisfaction [J]. *China Commercial Economics*,2005(12):42-45. (in Chinese)
- [3] Chaiken, S.. Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, 1980, 39(5): 752-766.
- [4] Shi X,Zheng X & Yang F.Exploring payment behavior for Live courses in social Q&A communities: An information foraging perspective[J].*Information Processing & Management*,2020.
- [5] Jin Xiaopu, Xu Fang, Bi Xin.Knowledge payment platform user satisfaction survey and ascension strategy [J/OL]. *Theory and practice of intelligence*: 1-14 [2021-02-02]. (in Chinese)
- [6] Huang Jiahui. An empirical study on the influencing factors of service utility of online medical platform [D]. Southeast University,2018. (in Chinese)
- [7] Wang Hui. Research on User Needs Mining Based on Feature Extraction and Sentiment Analysis [D]. Zhejiang Sci-Tech University,2019. (in Chinese)
- [8] Ghose A, Ipeirotis P G Li B. Designing ranking systems for hotels on travel search engines by mining user-generated and crowd-sourced content[J]. *Marketing Science*,2012,31(3):493-520.
- [9] Wang Wei, Chen Wei, Zhu Xiaoguo, Wang Hongwei.The success rate of crowdfunding and persuasion of language style: An empirical study based on Kickstarter [J].*Management World*,2016(05):81-98. (in Chinese)
- [10] Yin D, Bond S D, Zhang H. Anxious or angry? effects of discrete emotions on the perceived helpfulness of online reviews[J]. *MIS Quarterly*, 2014, 38(2): 539-560.
- [11] Chen Minghong, Liu Ying, Qi Xianjun.Research on the willingness of continuous knowledge sharing in academic virtual communities: A heuristic-system model perspective [J].*Library Forum*,2015,35(11):83-91. (in Chinese)
- [12] Tao Xiaobo, Xu Pengyu, Fan Chao, Yang Yiweng.Research on the influence mechanism of new product developers' information adoption behavior in innovation community [J]. *Management Review*,2020,32(10):135-146. (in Chinese)
- [13] Sussman S W, Siegal W S. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption[J]. *Information Systems Research*, 2003, 14(1): 47-65.
- [14] Li Wu, Ai Pengya, Bin Feng.Fan power and knowledge payment: A study of online question-and-answer platform users' paying behavior [J].*Library Journal*,2019,38(04):74-81. (in Chinese)
- [15] Fu Shaoxiong, Chen Xiaoyu.Analysis on the influencing factors of the content expressiveness of knowledge web celebrity: Taking Zhihu as an example [J]. *Information and Data Work*,2019,40(06):81-89. (in Chinese)

INDEX OF AUTHORS

- Bangchen Pan 367
Bei Wu 15
Bin Yang 95
Bingchen Guo 345 377
Bingqing Ji 56
Bo Jia 46
Bohong Li 294
Cai Lihua 498
Chang Liu 317
Chang Wang 365
Changyu Wang 163
Chengcheng Cai 461
Chenglong Li 583
Chunxiao Yang 186 398
Chuyu Jiang 622
Cui Guo 81
Cuicui Liu 390
Cuiping Wang 252
Daqing Zheng 347
Delin Hou 206
Di Xu, Yan Li 79
Dingtian Zhang 591
Fei Wang 345 377
Gaoyong Li 56
Ge Liu 325
Ge Zhang 56
Guangqian Peng 93
Guiming Lian 551
Guoxin Wang 461 81
Guoyin Jiang 29 354
Haiying You 122 132
Hao Zhou 68
Hexin Zhang 186 398
Hong Peng 303
Hong Zhang 551
Hongbo Jiang 110 518
Hongquan Li 459
Hongxiu Li 575 583
Huang Yingpei 498
Huanlian Yang 81
Hui Li 529
Huiying Wei 461
Jiahua Jin 46
Jiali Li 367
Jiali Zhao 15
Jiamin Yuan 549
Jian Tian 271
Jianfeng Luo 416
Jiang Wu 38 290
Jiangping Wan 186 398 416 500
Jianyao Chen 575
Jiaxin Chen 244
Jiaxuan Wu 605
Jiayi Xu 390
Jin Lin 264
Jin Yan 163
Jinfeng Wu 206
Jing Fan 459
Jing Zhao 345 377 496
Jingxuan Cai 290
JingYang 541
Jingyi Zhao 232
Jun Liu 388
Junfang Cao 303
Junfeng Yin 93
Junjie Zhou 81 461
Junli Lei 29
Keyi Zhang 390
Kun Jia 613
Lewei Hu 496
Li Li 244 284
Li Wang 252
Liang Yang 174
Liangqiang Li 174
Lifeng He 1
Lijun Li 271
Lin Ni 406
Liyi Zhang 448
Mei Zhang 152
Meng Zhang 22 174
Mengru Zhou 317
Mengyan Liu 186
Min Zhang 232
Minghao Li 315
Minglei Li 29
Mingxing Shao 459

- Miyan Liao 174
 Ning Wang 95 508
 Ning Zhang 1
 Ningyue Cao 163
 Pengjie Wu 500
 Ping Gao 93
 Ping Wang 365
 Qi Wang 563
 Qianxi Zhang 541
 Qinyi Zhou 591
 Quanwu Zhao 549 620
 Reima Suomi 583
 Ronglin Hou 345 377
 Ruoyu Sun 541
 Shengli Li 605
 Shengyou Liu 365
 Shihui Shuai 406
 Shixuan Fu 325
 Shuang Chen 102
 Shuang Zhang 315 472
 Sijia Guo 390 424
 Siqian Wang 551
 Suya Hu 79
 Tao Xiong 38
 Tianmei Wang 140
 Tienan Wang 3
 Ting Shao 110 518
 Tingshun Xiong 38
 Wanqiang Yang 354
 Wei Li 406
 Wei Luo 365
 Wei Zhang 152
 Weiguo Fan 1
 Weiwei Liao 244
 Weiwei Yan 232
 Wenbin Qu 194
 Wendi Wang 194
 Wenfei Zhao 1
 Wenhao Wang 461
 Wen-Lung Shiau 317
 Wenping Liu 29
 Wenqian Xu 424
 Xi Wang 529 620
 Xiang Yang 436
 Xiangbin Yan 46
 Xiangbing Chen 122 132
 Xiangcheng Yang 212 424
 Xiangling Huang 518
 Xiaofei Cui 518
 Xiaofeng Wang 194
 Xiaojiang Zheng 325
 Xiaojie SHI 484
 Xiaoli Li 68
 Xiaolin Li 174
 Xiaomin Zhan 461
 Xiaoping Liu 102
 Xiaoying Li 224
 Xiaping Xin 140
 Xie Yu 152
 Xin Zhang 56
 Xinling Lao 461
 Xinru Zheng 290
 Xinyi Li 529
 Xinying Guo 81
 Xu Gechen 498
 Xu Liu 3 416
 Xuefeng Li 152
 Xuejian Li 398 500
 Xueyan Dong 3
 Yan Mo 563
 Yan Tu 140 224
 Yan Wan 459
 Yanchun Zhu 152
 Yang Bai 575
 Yang Zhang 93
 Yang Zhao 95 508
 Yanmei Zhang 294
 Yanping Zeng 622
 Yi jiang 212
 Yi Jiang 424 472 496
 Yi Li 597
 Yi Wang 22 591
 Yi Yang 102
 Yihang Zuo 416
 Ying Chen 284
 Ying Liu 472

Yiping Ma	38	Zhaotao Chen	500
Yu Wang	347	Zhen Zhu	315 345 377
Yuan Chen	38	Zhenkun Wei	472
Yue Xu	529	Zhibin Cui	294
Yueying Li	354	Zhichao Zhang	110
Yufei Han	622	Zhiliang Pang	1
Yulan Lin	284	Zhiping Hou	365
Yunjun Jiang	551	Zhiwei Du	500
Yuqing liu	549	Zhiyuan Ge	46
Yuting Jiang	583	Zhiyuan Zhou	206 620
Yuting yang	448	Zhongyi Hu	38
Yuzhen Liu	333	Zhou Xingjian	498
Zeguo Yang	541	Zidie Chen	3
Zengyuan Wu	15	Zihan Peng	68
Zhao Du	388	Zishan Huang	459
Zhao Li	436		

Alfred University

Our Mission:

In all of our actions, we seek to assure that our students and faculty attain distinction to their personal, public and professional lives

Our primary vehicle for this is the development and delivery of programs of instruction and scholarship in professional management. Recognizing that we live in world in which technical advances, political shifts and social changes are all occurring rapidly, our principal objectives is to be a dynamic learning organization that prepares our students for leadership roles in their professions. We will design our systems and provide for incentives that build ongoing commitment to improving the quality of what we do.

In support of this mission our educational objectives are to:

1. Provide undergraduate and graduate programs that are excellent in quality, innovative in delivery, and relevant to current business practices.
2. Develop leadership and lifelong leadership skills.
3. Provide an environment, which fosters understanding and appreciation of cultural diversity and ethical conduct.
4. Support, conduct and disseminate scholarship in business.
5. Serve the community through programs and partnerships that enhance the intellectual quality of the region, and that enhance our core intellectual activities.
6. Provide active learning opportunities, which develop distinction through the acquisition of professional business skills.

Phone 607-871-2124

Fax 607-871-2114

Email businesscollege@alfred.edu

Website <http://business.alfred.edu>

China University of Geosciences

China University of Geosciences has gone through many changes and significant difficulties since 1952. The academic culture of universities like Beijing University, Qinghua University, etc. has greatly influenced the creation and the direction of this university. There are successively over 60 academicians who have ever industriously carried out teaching in the university; even more middle-aged and young experts have started their careers here. The university became one of the key universities in 1960, and was authorized to set up Graduate School in 1986. In 1997, it entered the 211 Project as one of the first group of universities in our country. It is therefore a key university being mainly created in China.

The university has splendid glories brought by outstanding records. It has shaped its own traditions and demeanors of being self-supporting and realistic. Up to now, the university has cultivated over 80,000 graduates. They are engaged in their own fields and dedicated themselves to the development of the country. Many of them have become academic elites, political genius, business tycoons, and excellent athletes. Premier Wen Jiabao and 21 academicians of Chinese Academy of Sciences and Engineering are typical representatives among them.

Over more than fifty years, people of CUG have been forging ahead, keeping pace with the times, and carrying out innovations. They have caught hold of the opportunities brought about by the reforms, and achieved rapid progress with their intelligence and efforts. Today, CUG has changed from just a geological college with single discipline to a comprehensive university offering multi-disciplines with geology, resources, environment, and geological engineering technology as the main features along with the development of sciences, engineering, liberal arts, management, economics and law.

Presently, the university has raised the goal of constructing itself into a world-class the field of earth sciences. In order to achieve the goal, we will inherit and carry forward the excellent traditions of CUG. We will regard people as the resource, discipline construction as the guide, improvement of teaching strength as the core, and talent cultivation as the lifeline. At the same time, we will intensify the conception of competition, quality and services, carry out reforms while remaining innovative, and aim to achieve outstanding credits in order to meet the demands of both the history and the future.

Yanxin Wang

President of China University of Geosciences, Wuhan, China