

Overview of tourism industry research based on data software analysis

Chanjuan Wang¹

¹School of Economics and Commerce, South China University of Technology, Guangzhou China Guizhou Minzu University, Guizhou China

Abstract—The rapid development of tourism promotes the study of tourism industry. It accumulates and summarizes a lot of relevant knowledge and theories. This paper takes the research literature of tourism industry retrieved and collected on the Web of Science as the research object. The research tool is the Citespace.5.6 R3 Software. The author combined data analysis and literature reading to summarize the knowledge of tourism industry research in the past five years from 2016 to 2020. Through the research, we get some knowledge and conclusions about the tourism industry research. We hope that these results will provide some basic references for travel industry practitioners and scholars.

1 INTRODUCTION

Scholars have been vetting the epistemology, knowledge domain, and/or intellectual structure of disciplines to elucidate how they have evolved over time [1]. The first is to promote the theoretical maturity, and at the same time, to further improve the discipline knowledge system, which can strengthen its theoretical guiding significance. The second is to provide a simple and clear introduction reference for the practitioners of specific industries and scholars in the field of initial involvement in the face of massive data and literature resources. At present, the representative tourism research summaries include: tourism research progress [2], progress in tourism management [3], information technology and tourism management [4], event tourism [5], tourism demand modeling and forecasting [6], tourism innovation [7], sustainable tourism [8] etc. It reflects the professionalism and depth of the research. Compared with other fields of tourism research, the overall study of tourism industry is limited. It is mainly specific industry research, such as hotels [9-11], cruise [12], medical tourism [13].

With the emergence of information visualization tools, large-scale non-numerical information resources can be visually presented, providing help for people to understand and analyze data [14]. There are a wide variety of information visualization tools, among which Citespace, RefViz and HistCite have attracted wide attention.

This paper will be analyzed with the help of software Citespace. It is not only a summary of the tourism industry research, but also a new attempt to analyze the tourism industry research with the help of software.

2 SURVEY OF RESEARCH SAMPLES

In this paper, “tourism industry” was taken as the retrieval condition to retrieve the two parts of the Web of Science, subject and title. The retrieval scope is the core collection, and the time span is from 2016 to 2020. A preliminary search of 5,007 English literatures (data capture date: July 3, 2020) was conducted. Then deleted meetings, comments and reprints, a total of 3,064 literatures were finally marked. In terms of the number of literatures, it shows an increasing trend from 2016 to 2019. The number rose slowly, in 2016 (511 articles), 2017 (593 articles), and 2018 (644 articles), In 2019 (842 articles), there was a sharp increase, reaching an ascending peak. By 2020 (475 articles), although the number is declining, there is only half a year data of 2020 at present. By the end of this year, the number of documents is likely to exceed that of 2019, according to data estimates. Research focuses on hospitality leisure sport tourism, management, environmental studies, environmental science, green sustainable science technology, economics, business, sociology etc.

3 METHODS AND DATA PROCESSING

3.1 Methods

CiteSpace is an information visualization software which based on the Java language, developed by Chaomei Chen, a professor at Drexel University's School of Computing and Information Metrology [15]. CiteSpace is a program that focuses on the underlying knowledge of scientific analysis. It is a citation visualization analysis software developed in the background of Scientometric and Data

e-mail: 1638218019@qq.com

and information visualization. With the constant update of CiteSpace, it has not only provided the excavation of citation space, but also provided co-existing analysis functions among other knowledge units, such as the cooperation among authors, institutions, and countries/regions [16]. The software has been widely used and concerned. In view of the powerful data analysis ability of this software, this study attempts to apply it to the specific tourism industry research. So that we will have an overall overview of tourism industry research with pictures and texts.

3.2 Data processing

Because the number of samples taken exceeds the limit of a single Web download, we exported the literature several times. The data is exported in other file formats. We merged multiple exports and saved them in a file named download_1.txt. Then, the author ran citespaces.5.6R3, creating new data processing task, selecting data to be processed, and setting basic parameters. The period is from 2016 to 2020. The node types are author, institution, country and keyword. Other parameters default. In the process of specific analysis, the parameters were adjusted.

4.1 Author and Country

Software can be used to study the number of articles published by authors and the cooperative relationship between authors. We selected the author analysis link in the software, setting the corresponding parameters and generated the results. The adjusted mapping results are shown in Figure 1. The node represents the author of the article. The size of the node represents the number of articles. The color represents the time of the article. Lines are the embodiment of the cooperative relationship between authors. As shown in Figure 1, the number of nodes is 462, the number of connections is 251, and the network density is 0.0024. These indicates that there are many prolific authors, 31 of whom have published more than 5 articles (including 5 articles). However, the centrality of these prolific authors is relatively low, and there is no core author yet. Compared with prolific authors, there are more independent authors. There were 251 times of cooperation between the authors. The thickness and color of the connecting lines reflects the cooperation intensity and cooperation year. Due to the low network density, it can be seen that although there were a lot of cooperation, the overall cooperation relationship was not close.

4 THE ANALYSIS PROCESS

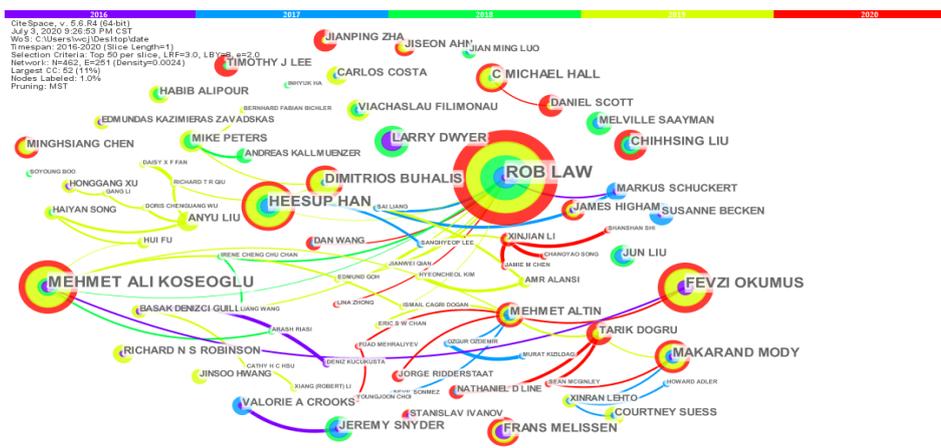


Fig 1. Author network

Figure 2 shows the countries with more literature: China, United States, Australia, United Kingdom, Spain, Taiwan, South Korea, Italy, Canada, Turkey, etc. It shows high network density and more cooperation between

countries. Among them, the intermediary centrality of United Kingdom, United States, South Korea, Australia and China are relatively large. It indicates that these countries play a strong role in connection.

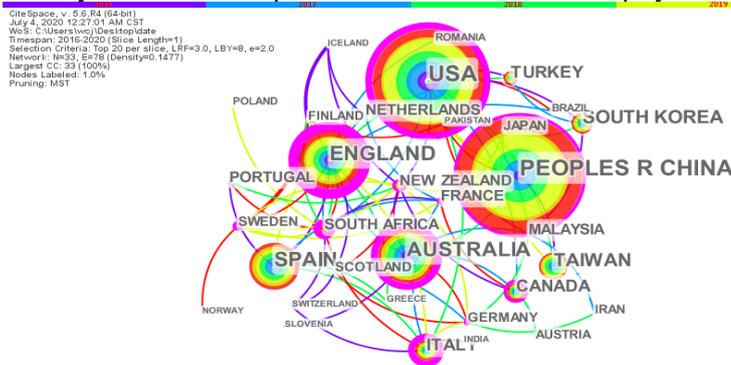


Fig 2. Analysis of the country

4.2 Institution

According to the data of publishing institutions in the software, the institutions with more papers are Hong Kong Polytech Univ (101), Griffith Univ (65), Univ Queensland (45), Univ Cent Florida (45), Chinese Acad Sci (44), etc. These institutions not only have a large number of papers, but also have strong intermediary centrality. This suggests that they have close contact with other institutions. Some institutions, such as Colorada State Univ, Univ Oulu, Purdu Univ and Washington State Univ, did not publish a large number of articles. Because of the cooperative relationship, they have a strong intermediary centrality. It is still very important in the whole network.

4.3 keyword

4.3.1 High frequency words and clustering:

Keywords are highly summarized and condensed in the research theme and content of the paper. Accurate grasp of keywords can understand the general content of the literature research, and the statistics of the word frequency of keywords can clearly understand the subject, institution and research hotspot in a specific period of time [17]. According to the software results, keywords with the same concept or similar meaning were combined and sorted out as shown in Table 1. Table 1 interceded the keywords with word frequency of more than 100 times and their centrality.

Table 1. KEYWORDS FREQUENCY AND CENTRALITY

Keyword	frequency	centrality	Keyword	frequency	centrality
tourism	920	0.18	innovation	150	0.19
impact	483	0.20	demand	144	0.21
industry	402	0.23	service	135	0.02
sustainability	365	0.10	china	133	0.11
management	342	0.19	climate change	132	0.12
hospitality	286	0.11	quality	129	0.05
performance	280	0.09	attitude	126	0.07
model	279	0.12	determinant	118	0.08
satisfaction	278	0.21	travel	109	0.08
behavior	229	0.03	framework	107	0.11
hotel	218	0.14	firm	104	0.11
perception	213	0.09	conservation	104	0.03
destination	192	0.08	strategy	102	0.00
experience	181	0.06	motivation	100	0.03
growth	169	0.08			

The high-frequency words map was adjusted in the software, as shown in Figure 3.

a) Large nodes indicate that there are many research achievements, such as tourism, impact, industry, sustainability, management, hospitality, which are the focus of the research. Many and thick connections indicate their strong mediating centrality, which are closely related to other studies. They have a strong importance in the whole study.

b) There are also some high-frequency words, although the frequency are relatively low and the nodes are small, but the lines are thick and more. Their high mediating centrality connects many research topics, such as mode, satisfaction, hotel, innovation, demand, china, climate, change, framework, firm. It indicates that they are not research hotspots at present, but play an important mediating role in the whole research.

c) Although some high-frequency words have high frequency, large nodes, but a few thin connections and low intermediate centrality. It indicates that there are concentrated research results on this issue, but they are not closely related to other studies.



Fig 3. The map of keywords

Figure 4 shows the keywords clustering. It is mainly divided into 7 clustering: hospitality (0.745), sustainability (0.694), tourism demand (0.786), destination Image (0.849), resilience (0.756), gender (0.889), smart tourism (0.906). The seven clustering only represent the highest keyword in its clustering, not the theme. For each clustering, except the sustainability, the clustering contour value S of all is greater than 0.7. It indicates that the clustering results are convincing.

a) The clustering results show that the value of clustering module Q=0.5017 is greater than 0.3. The result of clustering is significant.

b) The average clustering contour S=0.5114 is greater than 0.5. The result of clustering is reasonable.

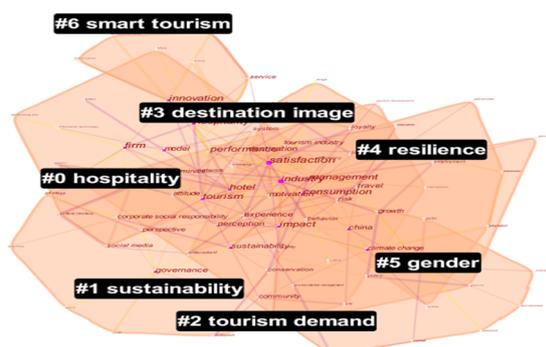


Fig 4. The figure of clustering

The keywords that are most closely related to hospitality (top 5) are environmental sustainability, technology, hotels, sustainability, sustainable tourism. For sustainability are sustainability, sustainable tourism, sustainable development, qualitative research, stakeholder theory. For tourism demand are connectedness to nature, frugality consciousness, environmental concern, tourism and hospitality industry, religiosity. For destination image are satisfaction, tour guide, tourism, trust, experience. For resilience are climate change, disaster, adaptation, nature-based tourism, business failure. For gender are medical tourism, youth, tourist accessibility, culture-led development, size. For smart tourism are Airbnb, sharing economy, entrepreneurship, grounded theory, newspaper. Even with the division of clustering, the plates overlap a lot. It indicates that there is a lot of common research among the categories.

Through the clustering situation and specific literature content reading, the summary is as follows.

Table2. THE RESEARCH TOPICS

topic	The specific content
tourism consumer	behavior, experience, perception, satisfaction, quality of service
tourism enterprises	hotels, cruise, tourism catering, tourist attraction
sustainability research	carbon emission, energy use, climate change, conservation of resource
management	corporate social responsibility, performance, competition, stakeholders
marketing	Price research, forecast market, new tourism products, channels
new topic	Tourism innovation, social media, management innovation, research of research

a) The research topics can be divided into 6 parts, as shown in Table 2.

b) The research objects include regions, countries, specific destinations, specific tourism projects. The hospitality industry is a prominent research object, which has strong practical operation and practical guiding significance.

c) A lot of research methods are quantitative research methods, which are represented by the traditional structural equation methods. There are a few qualitative research methods. Many mixed methods were used. Due to the update of questionnaire technology in data collection, information acquisition in social media, big data analysis and other factors, the research object's data are from a broader source.

d) Research theories mainly involve specific theories of management, psychology, sociology and economics etc. There are many interdisciplinary researches, but the use of theories is not in-dept.

4.3.2 The evolution of keywords:

Keywords can be represented by timezone diagrams in Citespace. Parameter Settings retains the settings of the previous keywords. The resulting graph is sorted out as shown in Figure 5. The font size of keywords reflects the frequency of occurrence. The place where the keyword appear was the year it first appeared. Everything that followed accumulates where it first appeared. The link shows the strength of the connection between the keyword and the keyword that followed. The evolution of keywords can reflect the change of knowledge points in the research field.

a) There are the most keywords in 2016. One of the main reasons is that the research on the tourism industry had attracted much attention and covered a lot of contents before 2016.

b) The large frequency of a certain type of study in a single year indicates that the research hotspot is representative in that year. Such as intention in 2017, governance and efficiency in 2018, sharing economy in 2019, firm performance and decision making in 2020.

c) The traditional research problems in the field have attracted the continuous attention of many scholars.

d) There are not many new problems and innovative researches in the field in 2017 to 2020. Most of them are small-scale researches extended from the original problems.

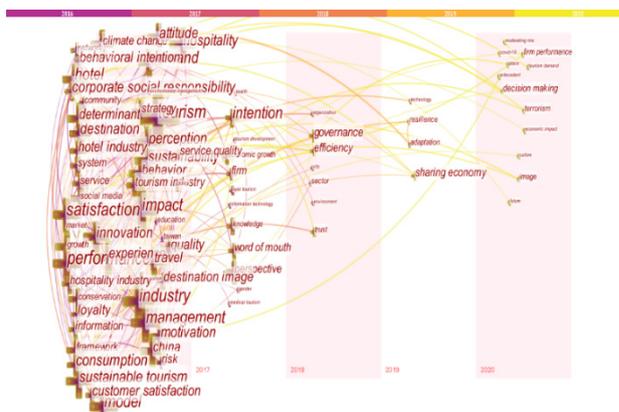


Fig 5. The timezone of keywords

4.3.3 The bursts of keywords:

As shown in Figure 6, there are only a few emergent keywords.

a) The emergence of risk, behavioral intention, rural tourism and perspective lasted for two years. Within two years, they became a research hotspot and then ended. It reflects the rapid change of the hotspot of tourism research, and their sustainability is not strong.

b) Relevant researches on perspective and knowledge were carried out as early as 2016, but they were not highlighted until 2017. In the course of discipline mature development, more attention is paid to the accumulation of knowledge and multiple perspectives.

c) Rural tourism is a kind of tourism advocated by many countries and regions. The explosive research on it has strong characteristics of the times.

d) The emergence of community and knowledge have been for three years. Although community research has been reduced, there is plenty of research space in the content. Research on knowledge will remain hot for some time. On the whole, the emergence of keywords has both the natural law of discipline development and the practical significance of the times. In the future, due to the reality of tourism development, there may be a lot of knowledge of disciplines and hot research on practical development.

Top 6 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2016 - 2020
risk	2016	3.252	2016	2017	
community	2016	5.7972	2016	2018	
behavioral intention	2016	1.8773	2016	2017	
rural tourism	2016	7.6251	2017	2018	
perspective	2016	2.3223	2017	2018	
knowledge	2016	2.8538	2018	2020	

Fig 6. The bursts of keywords

5 Conclusions

5.1 Author and Country

The study of tourism industry has attracted the continuous attention of scholars and also attracted many independent scholars to join in. However, the cooperative relationship between authors is not close, and core authors have not yet been formed. The countries with a large number of papers are mainly concentrated in European and American tourist source countries and Southeast Asian tourism host countries. Tourism practice promotes the study of tourism theory. Cooperation in theoretical research has strengthened academic links and communication between regions and countries. Since the web of Science is mainly composed of academic articles, the research institutions are mainly universities. The university has become the representative of the key research on this topic, playing an important role. However, there is a lack of collaborative

research with other tourism industry organizations or tourism administrations to promote academic research.

5.2 Research overview

The research topic not only focuses on the whole, but also expands its scope. Many keywords are connected. It indicates that they appear together in the same literature. It further indicates that there are few studies on independent problems, but more studies on multiple problems around a research object. The theme is embodied in six aspects: tourism consumer, tourism enterprises, sustainability research, management, marketing and new topic. The research mainly focuses on the traditional tourism research scope, with a few new areas explored. The research is rich in specific content and goes deep into specific problems of specific industries. The research methods have certain reference. However, due to the specific object of study, their generalization and universality are weaker. The theoretical research results are relatively less, and the construction of tourism knowledge is relatively lacking. With the help of other disciplines of theory research are not sufficient, not in-depth.

5.3 Future

The research of tourism industry has gone through a period of rapid development just like the development of tourism. Now, it's a slow growth period. The tourism industry suffered a setback this year under the influence of COVID-19. The industry's travails have aroused or stimulated much attention in the research field. A lot of tourism research on the impact of particular events is likely to be formed in the future. This has a strong period memory and may form a research peak of the topic.

The communicative nature of tourism will further promote practical and academic international exchanges and cooperation. All kinds of organizations are closely linked. Academic research cooperation between academic organizations and tourism industry organizations or tourism administrations is particularly encouraged.

In the future, the research in specific fields will be further expanded and enriched. Due to the introduction of new technologies and knowledge, the tourism industry will face new challenges and opportunities in the future. Promoting the whole industry research is the theoretical demand of objective and practical development. This will gradually improve the tourism industry macro research. For those regions that take tourism as a key industry for regional development, the macro-tourism industry research is more forward-looking. The development of tourism industry emphasizes economy, but the soul of tourism development is its cultural connotation. Integrating cultural elements into industrial research may be a new entry point for future research.

The direction of research development reflects the sustainability of tourism research development. While the content of the early research is increasing, many new research directions are also found. At present, the research content is still relatively complex. The future research may

develop towards the direction of theoretical development and systematic integration of knowledge content. Tourism innovation field needs to integrate the original basic knowledge to explore a new starting point.

On the whole, this paper analyzes the basic elements of the current tourism industry research with software. It embodies the advantages and characteristics of software analysis of massive academic literature. Its advantage is the huge amount of data to process and the results are straightforward. For the tourism industry practitioners, academic research beginners, it is a good way to quickly understand the tourism industry research situation. However, this is only a simply overview according to the data and part of the content. It is not sufficient to study a field only through these indicators. Further data analysis results should be combined to further read relevant literatures. The same data should be put into other software to check the results and compare the results. The results of the data analysis are only a preliminary outline. The learning of knowledge still needs a process.

Acknowledgements

This paper is the stage learning outcome of the author's doctoral research. It embodies the learning process. I would like to thank my tutor and classmates for their help in this process.

References

1. Koseoglu, Mehmet Ali, et al, "Bibliometric studies in tourism," *Annals of Tourism Research*, vol. 61, pp 180-198, November 2016.
2. Leong, Lai-Ying, et al, "Tourism research progress—a bibliometric analysis of tourism review publications," *Tourism Review*, June 2020.
3. Law, Rob, Shanshan Qi, and Dimitrios Buhalis, "Progress in tourism management: A review of website evaluation in tourism research," *Tourism management*, vol. 31.3, pp 297-313, 2010.
4. Buhalis, Dimitrios, and Rob Law, "Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research," *Tourism management*, vol. 29.4, pp 609-623, 2008.
5. Getz, Donald, "Event tourism: Definition, evolution, and research," *Tourism management*, vol. 29.3, pp 403-428, 2008.
6. Song, Haiyan, and Gang Li, "Tourism demand modelling and forecasting—A review of recent research," *Tourism management*, vol. 29.2, pp 203-220, 2008.
7. Hjalager, Anne-Mette, "A review of innovation research in tourism," *Tourism management*, vol. 31.1, pp 1-12, 2010.
8. Buckley, Ralf, "Sustainable tourism: Research and reality," *Annals of tourism research*, vol. 39.2, pp 528-546, 2012.
9. Kusluvan, Salih, et al, "The human dimension: A review of human resources management issues in the tourism and hospitality industry," *Cornell Hospitality Quarterly*, vol. 51.2, pp 171-214, 2010.
10. Tsang, Nelson KF, and Cathy HC Hsu, "Thirty years of research on tourism and hospitality management in China: A review and analysis of journal publications," *International Journal of Hospitality Management*, vol. 30.4, pp 886-896, 2011.
11. Madera, Juan M., et al, "Strategic human resources management research in hospitality and tourism," *International Journal of Contemporary Hospitality Management*, vol. 29, pp 48-67, 2017.
12. Sun, Xiaodong, Xuegang Feng, and Dinesh K. Gauri, "The cruise industry in China: Efforts, progress and challenges," *International Journal of Hospitality Management*, vol. 42, pp 71-84, 2014.
13. Connell, John, "Contemporary medical tourism: Conceptualisation, culture and commodification," *Tourism management*, vol. 34, pp 1-13, 2013.
14. Jun Tian, "Comparative analysis of information visualization analysis tools: take CiteSpace, HistCite, and RefViz for examples," *Library Science Research*, vol.14.54, pp 90-95, 2014.
15. Guangyang Liu, "The spread Track of CiteSpace Application in China: Based on the Statistical and Visual Analys of Cross-Database Date from 2006 to 2015," *Library information knowledge*, vol.02, pp 60-74, 2017.
16. Jie Li, Chaomei Chen, CITESPACE: Technology text Mining and visualization, Capital University of Economics and Business Press, Beijing, 2016, pp 3
17. Junfang Zhao, Fan Jiang, "Bibliometrics and Mapping of Scientific Knowledge in Modern University Education," *University Education Science*, vol.1, pp 115-123, 2014