Themes, Trends, and Influence: A Preliminary Study of the International Academic Discourse Power of Big Data Research in the Field of Chinese Social Sciences

Xue Meng and Tang Jingqian[(✉)]

Library of Jinan University, Jinan University, Guangzhou, Guangdong, China
187191942@qq.com

Abstract. Taking the big data-related literature included in SSCI as the data source, this paper applies the WoS analysis function and knowledge graph tool Citespace III to conduct bibliometric and visual analysis in social sciences. Quantitative and qualitative analyses are performed to uncover the international situation and existence of big data research in China, through evaluating key variables such as annual distribution, country/regional distribution, research institution distribution, hot spots, subject scope, paper influence, etc. The study aims to investigate China's internationalization in the subject of social sicence, and help build big data of academic discourse system in China.

Keywords: Social Science · Big Data · SSCI · Academic Discourse Power

1 Introduction

The emergence of big data promotes the digitization of the natural world and human thoughts and behaviors, and maps the natural, social and human thinking into a data world composed of data [1]. Nowadays, social science is not only a simple analysis and description of the object of study, but a complex, systematic and scientific study of things. In 2015, the State Council of China also issued the "Program of Action for Promoting the Development of Big Data", pointing out that adhering to innovation-driven development, accelerating the deployment of big data and deepening the application of big data have become the internal needs and inevitable choice for stabilizing growth, promoting reform, adjusting structure, benefiting people's livelihood and promoting the modernization of government governance capacity [2]. The unprecedented rich data resources and advanced data processing technology brought about by the big-data era have opened new horizons for empirical research in the social sciences field. Computational social sciences have paved the way for the acquisition and analysis of big data, multi-agent social simulation in internet society. The advent of the three method systems such as social surveys, statistical analysis, and social experiments have advanced the formulation of novel research methods [3].

At present, some scholars have analyzed the hot topics, evolution process and influencing factors of big data research in the field of social science in the United States, which provides a good reference for Chinese scholars to understand the research content of developed countries [4]. However, what is the status of big data research in social sciences in China? What are the trends and influences? What is the international academic discourse power in the world? To address these critical questions, in this article, we ought to consider the journal papers on big data research in the social sciences published by Chinese scholars included in the SSCI as the research object with the help of bibliometric methods and analysis tools such as Citespace.

2   Overview of SSCI Papers on Big Data Research in Social Sciences in China

2.1   Basic Characteristics

Since the first paper was published in 2010, the total number of papers in big data research journals in the field of social sciences was found to be 2,215. The annual distribution and changing trends were shown in Fig. 1, which implied the number of papers on big data research in the social sciences has multiplied. These papers have been published from around 85 countries/regions, showing a concentration trend in a few countries/regions.

The percentage of top 10 countries/regions distribution of SSCI papers on big data research in social sciences are presented in Table 1. From that we can see the United States, China, and the United Kingdom have found to be the top countries, of which the United States stands first with 797 papers, accounting for 35.98%, followed by China, accounting for 23.48%, and 286 in the UK, accounting for 12.91% (Table 1).

It is worth noting that China, a late publishing country in this field, has the fastest growth rate in publishing papers. From 2014 to 2016, there is a big gap with the United States, but it is almost neck and neck with the United Kingdom. After 2017, it has overtaken the United Kingdom, and the gap with the United States is getting smaller and smaller. China has become the largest contributor in this field in 2020.
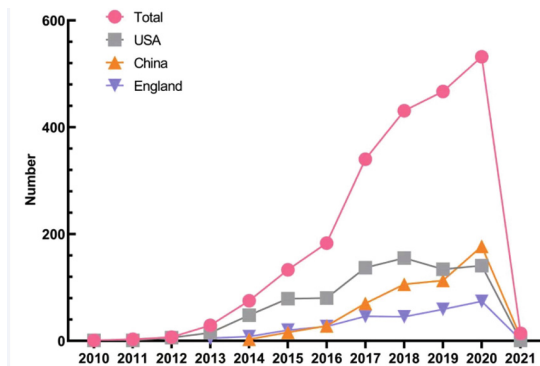

Fig. 1. Annual distribution of SSCI papers on big data research in the field of Social Sciences

Table 1. Country/Region Distribution of SSCI Papers on Big Data Research in Social Sciences (Top 10)

Country/Region	Article Number	Percentage
USA	797	35.98%
PEOPLES R CHINA	520	23.48%
ENGLAND	286	12.91%
AUSTRALIA	120	5.42%
ITALY	103	4.65%
SOUTH KOREA	101	4.56%
GERMANY	94	4.24%
CANADA	93	4.2%
NETHERLANDS	89	4.02%
FRANCE	87	3.93%

2.2 Themes and Subject Distribution

A series of related issues and concepts that the academic community pays attention to can usually be defined as the research field. Frequently, number and distribution of key words are often used to reveal the subject area of its research. CiteSpace's timeline view can show the specific research content involved in different knowledge subgroups, and reflect the development and connection between them from the time dimension [5]. Therefore, this part uses citespace to visually analyze keywords, reveal the internationally published topics of big data research in China's social sciences research, and bibliometrics to measure the subject areas of these papers.

According to statistic analysis, we have collected 3,019 keywords (average 5.8 keywords per paper) from 520 international papers in this field. They were used by the software Citespace for visualanalysis via classification, induction and collation after getting rid of the meaningless words in big data research such as numbers, symbols, etc., and as shown in Fig. 2.

As shown in Fig. 2, high-frequency keywords for big data research in social sciences in China are grouped into eight categories, they are network analysis, data envelopment analysis, risk, coordination, data mining, big data analytics, city, supply chain management, covid-19. Network analysis mainly includes keywords such as science, smart city and machine learning. Data envelopment analysis mainly includes keywords such as China, innovation, performance, dynamic capability. Risk mainly includes keywords such as social media, information model, risk, policy perspective. Coordination mainly includes keywords such as firm performance, policy, information analytics, capability, service quality. Data mining mainly includes keywords such as computational social science, technology. Big data analytics mainly includes keywords such as big data analytics, algorithm, business value, Internet, security, challenge. City mainly includes keywords such as carbon emission, land use, evolution, Supply chain management mainly includes keywords such as predictive analytics, bibliometric analysis, sustainability.

Fig. 2. Time line view of co-occurrence map of high-frequency keywords in big data research

Consequently, we can infer that big data research in social sciences in China mainly focuses on the following aspects: (1) Network analysis of big data in science, smart city, machine learning, etc.; (2) Research on China's innovation, performance and dynamic capability based on data envelopment analysis in a big data environment; (3) Big data research progress on the construction of social media and market information model and its risk evaluation and policy perspective under the environment; (4) Firm performance management strategy under the background of big data (policy and coordination capability, information analytics capability, service capability and quality research); (5) Computational social science (data mining) in communication science and the application of data mining technology; (6) The application of big data analytical algorithms in business value evaluation, and the challenges in internet security (security) analysis; (7) Urban research based on big data, including carbon emission, land use and its evolution; (8) Demand forecast for supply chain based on big data (Predictive analytics), bibliometric analysis, and sustainable development (sustainability) issues research, etc.; and (9) the impact of the new coronavirus (covid-19) based on big data on health or physical activity, and the relationship between the built environment and other factors on the infection of the novel coronavirus (covid-19).

3   Research Trends of SSCI Papers on Big Data Research in Social Sciences in China

Studying the frontiers of a subject can help us grasp the latest research trends of the subject in time and predict the development direction of the subject and the hot issues of future research [6]. The time zone view of CiteSpace presents the main research content in the coordinate system with time as the horizontal axis. The corresponding nodes are set in different time zones according to the time when the keywords first appear, and the node positions move upward along the time axis [7]. In this paper, the CiteSpace Time Zone view is used to analyze the time distribution of word frequency, and according to the changing trend of word frequency, the evolutionary trend and research frontier of big data research in the field of social sciences in China are explored (Fig. 3).
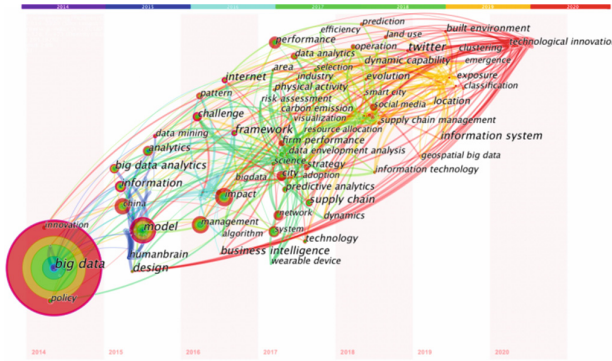
Fig. 3. Time zone view of evolution map in big data research in social sciences in China

As shown in Fig. 3, the number of nodes from 2014 to 2016 is relatively small, indicating that big data research in the field of social science in China is in the initial stage with few research achievements. This stage is mainly about the research of data subjects. "Big Data", "Big Data Analytics", "Innovation" and "Policy" are the representative keywords. Observed the connection situation, it can be found that such research contents appeared earlier in the time zone view and ran through the whole time line of big data research. It can be seen that scholars at this stage paid more attention to promoting policy formulation and innovation based on big data research for the purpose of national governance. After 2016, the nodes and connections in the time zone view gradually increased and intensives, indicating that the research on big data in the field of social sciences in China has developed rapidly and the research content has been further deepened. The emergence of topics such as "Smart City", "Location", "Social Media" and "Emergency" indicates that big data studies on topics such as Smart City, spatial information application, new Media integration and Emergency decision-making have received more attention from Chinese scholars. At this stage, Chinese scholars in the field of humanities and social sciences turned to focus on the impact of specific big data technologies and methods on humanities and social sciences research as well as the interpretation and knowledge discovery of relevant people's social sciences phenomena [8]. In addition, the data-driven service has also become a research hotspot in this stage.

At present, the most representative social science big data types mainly include electronic trace big data, social media big data, text big data, spatial location information big data [9]. Combined with Fig. 4, the author believes that the future research trend of big data in the field of social science in China may focus on the following aspects: (1) Research on the integration of social science big data with cloud computing, deep learning, artificial intelligence and other technologies. In recent years, some deep learning algorithms have been introduced into the field of public culture, and model algorithms such as Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) and Word2vec have been utilized [10]. On the basis of these algorithms, with the help of deep learning platforms such as Tensor Flow and PyTorch, in-depth analysis and mining of data in this field can be realized, and some intelligent applications can be realized through continuous iteration and optimization [11]. (2) Create the new service models
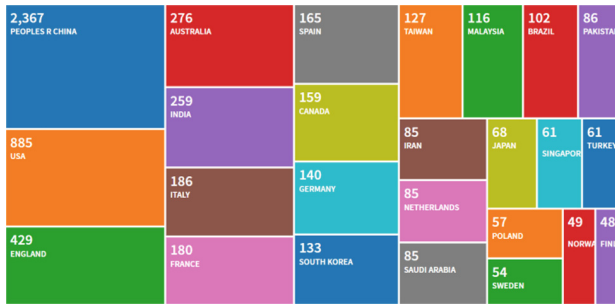
Fig. 4. Citations of big data research papers in social sciences in China

of social sciences based on "full data" knowledge mining. Actively pay attention to user portrait, science and technology portrait, urban portrait and other new field direction [8]. (3) Research on open sharing, multiple interaction and collaborative governance required for the construction of Society 5.0. Society 5.0 is the product of the further development of the smart society, smart city and information revolution [12]. Society 5.0 presents people with digital changes in life, industry, management, infrastructure, region and talent, We might call it a "data-driven society" [13].

4   The Influence of SSCI Papers on Big Data Research in Social Sciences in China

In this section, We analyzed the citation situation in this field in our country by using the Create Citation Report function of Web of Science, which showed that both the publication trend and the citation frequency of 520 papers increase year by year (Fig. 4). The figure shows that the total citation frequency is 6,355 times, and excluding the number of self-citation is 6,045 times, and the number of citation literature per article is 12.22. The h index is 39, that means 39 papers in these 520 documents have been cited more than 39 times. Besides, 520 papers were mainly cited by Chinese scholars, with 2367 citations by Chinese scholars, accounting for 37.25% of the total citations, followed by 885 citations by American scholars, 429 citations by British scholars and 276 citations by Australian scholars (Fig. 5). It is related to the total number of articles published by scholars worldwide in related fields, but these data clearly implied that China's research influence in this field needs to be improved.

Fig. 5. Distribution of source countries cited in big data research papers in social sciences in China

5   Conclusions

Through the bibliometric analysis and descriptive statistical analysis of the research papers on big data in the field of social sciences in China, this paper reveals the development of academic discourse in social sciences in China in recent years, there are: (1) Although the publication of big data in the field of social science in China started relatively late, from the perspective of the growth trend of papers, China has occupied the absolute advantage in the field of big data in social science represented by the United States and the United Kingdom. According to statistics, the output of international papers in this field in China has surpassed that of the United States and the United Kingdom in 2020. (2) From the perspective of publishing institutions, universities and research institutions, especially Wuhan University and Chinese academy of sciences, have made great contributions to improving the level of big data research in China's social sciences. (3) From research perspective, China's big data research reflects the interdisciplinary characteristics of modern social science research in China, mainly on the basis of computer science, the big data are applied to engineering, business, economic, environmental, research, management, and study of the problem of sustainable development (sustainability), and other fields, these areas are considered to be the important fields in public health, occupational health and sociology, On the other hand, research fields such as psychology have also received great attention.

Although China has been significant progress in the number of papers, high-frequency publishing institutions have emerged. However, China's international visibility and internationalization level still need to be improved, mainly reflected in the following points: (1) The influence of the research papers is low as reflected by the low average number of citations (12), and h-index (39) despite the increasing number of publications by China. (2) In terms of the source countries of citation literature, big data research in the field of social science in China is mainly cited by domestic scholars, and the international recognition rate needs to be improved. (3) There were only 24 highly cited papers signed by Chinese scholars, with a small number of highly cited papers. In addition, the first or corresponding authors of the four papers with the highest citation frequency were scholars from European and American countries. The highly cited papers with Chinese scholars as the responsible authors received relatively low attention.

References

1. Huang, X. R. & Shen, Q. L., "Big data promotes the integration of science and humanities", http://ex.cssn.cn/xspj/xspj/201809/t20180918_4562786.shtml, 2018.
2. The State Council Of the People's Republic of China., "Notice of the State Council on Issuing the Action Plan for Promoting the Development of Big Data", http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm, 2015.
3. Lou, J. J. & Zhang, D. C., "Computational social sciences and reconstruction of academic discourse system in the era of big data", Jishou University Journal (Social Science Edition), 2018, pp. 9–16.
4. Fang, H. J., Wang, X. Y., Liu Y. & Mu H. L., "A visualization analysis of the hot topics of big data research in the United States and their evolution", Library Development, 2020, pp. 293–297.
5. Shao, B. L. & Yang, Y., "Analysis of Personalized Recommendation Research Trends in Chinese Libraries from the Perspective of Knowledge Map",Library Work and Research, 2021, pp. 88–98.
6. Zhao R. Y., Li X. M., Zhang Z. Y., Chang R. R.& Li Y. S., "Progress of Graphic Files in the 13th Five-Year Plan Period: Analysis of Research Hotspot and Trend of Chinese and Foreign Periodical Papers", Library and Information Service, 2021, pp. 40–48 .
7. Chen Y., Chen C. M., Hu Z.G, et al., "Principles and Applications of Analyzing a Citation Space", Beijing:The Science Press, 2014.
8. Li Y. & Sun J. J., "Value Pursuit of Humanities and Social Sciences Big Data Research" ,[J]. Library & Information, 2019, pp. 1–7.
9. Hao L. & Li F. X., "Big Data Computing of Social Sciences: The key topic of Computational Social Science in Big Data Era", Research on Library Science, 2017, pp. 20–29+35.
10. Ruan G. C., Xie F., Tu S. W., "Application Research Based on Word2vec diversity in library recommender system", Library Journal, 2020, pp. 124–132.
11. Hua B. L., "Smart public cultural services from the perspective of Data, technology and application", Journal of Library Science in China, 2021, pp. 40–52.
12. Li M. Z. & Zhong H. H., "Society 5.0, super intelligent society and future prospects", Journal of social science research, 2019, pp. 1–9.
13. Prasetyo, Y. A., & Arman, A. A., "Group management system design for supporting society 5.0 in smart society platform", In 2017 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE, 2017, pp. 398–404.