

An Overview of Research on Big Data in Economics

Based on Bibliometric Software Citespace and CSSCI Database

Limei Guo

Economics School
Sichuan University
Chengdu, China

Abstract—Big data itself and big data technology are affecting the development of society, and thus affect the economic research. In order to fully understand the research hotspots and development trends of big data in the field of Chinese economics, this article shows the overview of the big data researches based on the Chinese Social Science Citation Index (CSSCI), with the software CitespaceV, which takes “big data” as the key word, and “economics” as the literature category. The progresses of big data researches in the field of economics are reviewed, and the existing literatures from three aspects, namely research hotspots, evolution of research hotspots and research basic knowledge are compared and analyzed. This research shows that: 1) The hotspots of the researches are “big data”, “internet finance”, “cloud computing”, “data mining”, “smart cities”, “precision poverty alleviation”, etc.; 2) Big data appears in the “Internet of Things” clustering at first and gradually evolves to “online shopping”, “internet finance”, and “classified networks” clustering; 3) The high-cited literatures on big data have gone through the stages from theory to application, and applications have focused on taxation and auditing.

Keywords—big data; research hotspots; the evolution of hotspots; basic knowledge

I. INTRODUCTION

Big data is affecting social development in all aspects, especially in the economic field. In 2007, American scientist Jim Gray predicted that data-intensive science was being separated from computer science and became the fourth paradigm of scientific research.[1] In March 2012, the Obama administration launched the "Big Data Research and Development Initiative". Six federal agencies announced that they invested 200 million dollars to improve technologies which can extract information from big data to make big data an important national asset. In August 2015, the State Council of the People's Republic of China issued the "Outline for the Promotion of Big Data Development" ("Outline" for short). The "Outline" shows that it is a trend for using big data to promote economic development, improve social governance, and improve government services and supervision capabilities worldwide. Big data is becoming important resources and technology for all to compete for.

This paper shows the researches of big data in the field of Chinese economics with CitespaceV and knowledge map

which can provide information for researchers in the future. This paper is organized as followed: the first part is the introduction. The second part is the previous studies, research method and data sources. The third part is the domestic research situation, which is analyzed from three perspectives: the literature research hotspots, the evolution of research hotspots, and the basic knowledge of the research. The fourth part is the conclusion.

II. PREVIOUS STUDIES, RESEARCH METHOD AND DATA SOURCES

A. Previous Studies

There are many related researches on big data, and different scholars have reviewed it from different perspectives. Li He explained the current researches from the perspectives of big data basic theory, storage and analysis processing technologies, and big data application research [2]; Wang Qian analyzed the historical evolution of the global big data from 1993 to 2016, and summarized the evolution path of “the theory of embryonic stage—the technical exploratory stage—the practical application stage”[3]; Tao Xuejiao reviewed the related concepts, characteristics, and big data technologies of big data[4]; Shahriar Akter and Samuel Fosso Wamba overviewed big data from the perspectives of definitions, characteristics, types, commercial values, and challenges[5]; Jie Sheng et al. systematically organized big data from the perspective of management[6]. The above studies have reviewed big data from different perspectives. This article attempts to review the research situation of big data from the perspective of economics.

B. Research Method

Citespace is citation visualization software developed by Prof. Chen Chaomei of the School of Computer and Information Science at Drexel University in the United States. The software focuses on scientific analysis and information visualization. Its functions keyword co-occurrence, literature co-citations, and keyword timeline view can analyze current research, and research trends. At present, Citespace is mainly used in 60 fields such as computer science, information science, and medicine in international scientific research.[7] CitespaceV is used in this

article to conduct a corresponding bibliometric analysis of big data researches in the field of economics.

C. Data Sources

The CSSCI database is chosen and the keyword is set as “big data”. The literature category is set as “economics”. The time zone division is set between 1998 and 2017 and the data is got on May 29, 2018. Literature data of the search structure is put into the CitespaceV, and the pathfinder is selected.

III. DATA ANALYSIS

A. Research Hotspots—Analysis of Keywords Co-occurrence

Citespace can determine the hot areas and trends by keywords clustering. The word frequency analysis method is to extract the frequency distribution of keywords or subject words that can express the core content of the literature in

the literature information, and thus to study the development trends and research hotspots in the field. The keywords with high degree of frequency indicate that they are research hotspots. The time slicing is set from 1998 to 2017 and 1 year per slice. The node type is set as keyword. Top 50 levels of most cited or occurred items are selected from each slice. The running result shows that the number of nodes N is equal to 38 and the number of links E is equal to 59 (Density=0.0839). In the keyword co-occurrence map generated by Citespace, 38 keywords and 59 links were selected in total. From the table 1, it can be seen that the keyword "big data" occurs most frequently, which is 180 times and the centrality of it is 1.52. The next one is "Internet Finance" with 14 occurrences and the centrality is 0.11. Other keywords include the “big data era”, “cloud computing”, “Internet+”, “Internet of things”, “data mining”, “smart city”, “big data analysis”, “transformation”, “accurate poverty alleviation” and “internet”. These keywords indicate that they are the research hotspots.

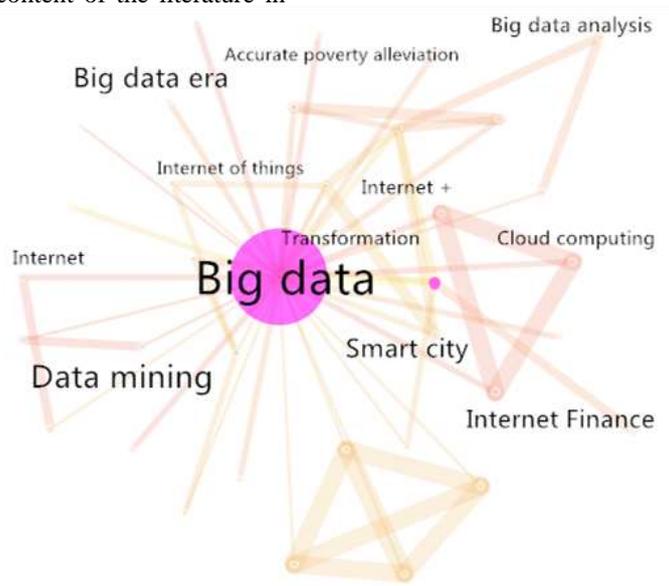


Fig. 1. The visualization map of keywords big data in the field of economics based on CSSCI(keywords co-occurrence).

TABLE I. THE TOP 12 KEYWORDS SITUATION IN THE FIELD OF ECONOMICS

NO.	Frequency	Centrality	Key word
1	180	1.52	Big data
2	14	0.11	Internet finance
3	14	0.01	Big data era
4	7	0.08	Cloud computing
5	5	0	internet+
6	4	0.03	Internet of things
7	4	0	Data mining
8	4	0	Smart city
9	3	0	Big data analysis
10	3	0.02	transformation
11	3	0	Accurate poverty alleviation
12	3	0	internet

^a. Data sources: the results of running Citespace based on CSSCI

The principle of co-word analysis is to count the number of occurrences of words in the same group of literatures. The intimacy between them is judged by the number of co-occurrences. The lines between nodes indicate co-occurrence relationships, and their thickness indicates the intensity of co-occurrence. The co-word method is used to analyze topics such as hot topics, topic distribution, and subject structure. Intermediary centrality is an indicator of the importance of a node in the network (not less than 0.1). Literatures with high centrality are usually the key pivots connecting two different fields, also known as turn points. The figure 1 shows the relationship of the co-occurrence of keywords. It can be seen that the keyword “big data” share lines with many other keywords such as “internet”, “data mining”, “the era of big data” and so on. The keyword “big data” also have high centrality based on the table 1, which is 1.52 and it is much

higher than 0.1. Thus it is the key pivot connecting many other fields.

B. The Evolution of Research Hotspots

As mentioned above, the timeline view mainly measures the relationship between clusters and the historical span in a

cluster. The evolution of clustering can be seen from the timeline view. After the co-occurrence analysis of the keywords, the layout is set as timeline view and then the timeline view of keyword can be got. The clustering is named by keywords.

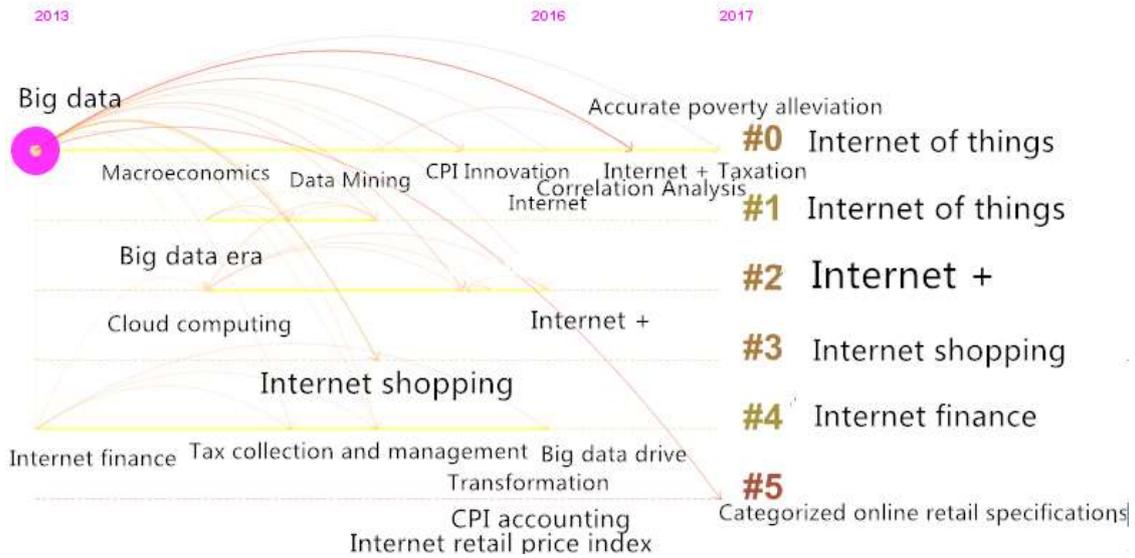


Fig. 2. The timeline view of the co-occurrence of the keywords.

It can be seen from the "Fig. 2" that the researches mainly focus on six directions: "Internet of Things", "Internet +", "Internet shopping", "Internet finance", and "categorized online retail specifications". The clustering zero and the first clustering share the same name. It can be seen from the figure 2 that the clustering zero "internet of things" and the fourth clustering "internet finance" are the earlier areas enjoying development. The keywords in the clustering zero ("Internet of Things") from 2013 to 2017 include "Macroeconomics", "Data Mining", "CPI Innovation", "Internet", "Correlation Analysis", "Internet + Taxation", and "Precise Poverty Alleviation". The fourth Clustering ("Internet finance") has discussed the keywords "Internet finance", "tax collection and management", "transformation" and "big data drive". From the keyword in 2017, it can be seen that "precise poverty alleviation", "categorized online retail specifications", "CPI accounting", and "Internet retail price index" are the frontiers of research.

It can be seen from the "Fig. 2" that the keyword "big data" co-occurs with all the keywords in the clustering and it co-occurs most frequently in the "internet of things" clustering.

C. Analysis on High Cited References

The high cited references can represent the basic knowledge. Selecting "Cited Reference" in Node Types and running Citespace, the co-cited references map of the big data research in the field of economics can be got. The result shows that the number of nodes N is equal to 60 and the number of links E is equal to 91 (Density=0.0514). In the co-cited references knowledge map, 60 articles and 91 links between the literatures were selected. Detailed information can be found in "Table II".

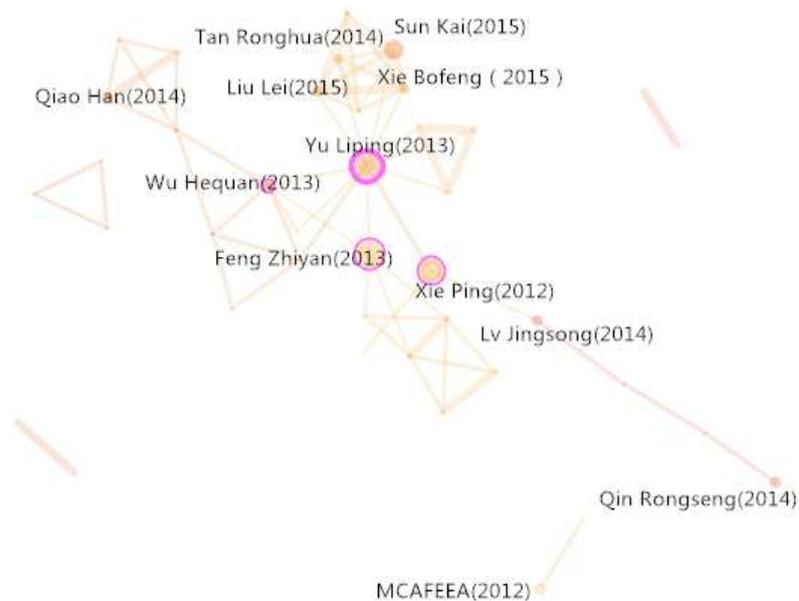


Fig. 3. The visualization map of the co-cited references.

TABLE II. THE TOP 12 REFERENCES WITH HIGH CITED FREQUENCY

NO.	Cited Frequency	Cetrality	Author	Year	Article title(literature information)
1	8	0.15	Feng Zhiyan	2013	On the research frontiers of business management in the context of Big data
2	8	0.15	Xie Ping	2012	The Research of Internet Based Finance Mode
3	7	0.36	Yu Liping	2013	Big Data and Big Data Economics
4	6	0.01	Sun Kai	2015	Big Data--Propeller for Constructing Modern Tax Administration System
5	4	0	Qiao Han	2014	Research on CPI Compilation in China With Scanner Data in the Context of Big Data
6	4	0	Qin Rongsheng	2014	Research about Influence of Big Data and Cloud Computing to Audit
7	4	0.09	Lv Jinsong	2014	Research on Data Analysis in Audit over Financial Institutions
8	4	0.14	Wu Hequan	2013	Opportunities and Challenges in the Era Of Big Data
9	3	0.01	Tan Ronghua	2014	Some Thoughts on Big Data Application In Tax Administration
10	3	0	Mcafee A	2012	STRATEGY&COMPETITION Big Data: The Management Revolution
11	3	0.01	Xie Bofeng	2015	Construction of Tax Microdata System in Big Data Era
12	3	0.01	Liu Lei	2015	An Analysis of Tax Management in the Era of Big Data

The references listed in the "Table II" belong to the classic literature. From a time point of view, the high-cited references mainly appear between 2012 and 2015. From the "Table II", it can be seen that two pieces of references appeared in 2012. The author of *The Research of Internet Based Finance Mode* put forward the concept of the Internet financial model, and studied its payment methods, information processing, and resource allocation. He believed that the Internet financial model can promote economic

growth by increasing the efficiency of resource allocation and reducing transaction costs, which the society benefits from.^[8] In *STRATEGY & COMPETITION Big Data: The Management Revolution*, the author thinks that with data, managers can quantify management, so they can better understand their business affairs and can transform knowledge to improve the decision-making.^[9] In 2013, there were 3 articles. The authors of *On the research frontiers of business management in the context of in the context of Big*

data pointed out four directions for future research of big data: the behavioral mechanism and social capital structure in the social network environment, the corporate network ecosystem and their collaboration Symbiosis mechanism, innovation of business model based on big data and the customer insight and marketing strategy in big data environment.^[10] The author of *Big Data and Big Data Economics* discussed the challenge of big data to traditional economics and put forward the concept of big data economics for the first time. The author believed that big data economics not only integrates theoretical sciences, experimental sciences, and complex phenomenon simulations together, but also integrates natural sciences and social sciences together, and integrates theoretical research and practical applications together.^[11] In the *Opportunities and Challenges in the Era of Big Data*, Wu Hequan discussed the big data from several aspects: the advent of the era of big data, the application of big data, and the challenges and implications of big data technology. He believed that big data is a new generation of information technology. This is a precious opportunity for China's development.^[12] The high-cited articles in 2014 and 2015 focused on taxation and audit research, indicating that the research on big data has gradually transitioned from theoretical research to applied research. The areas of applied research are mainly focused on taxation, auditing, and CPI compilation.

As the literatures with high centrality are the pivots connecting different fields. According to "Table II", the literature of NO.1, NO.2, NO.3 and NO.8 are the turning points as their centrality is higher than 0.1. It also can be seen from Figure 3 that the nodes of authors Feng Zhiyan, Yu Liping, Xie Ping and Wu Hequan are bigger than others and they have more lines linking other literatures. Based on the figure 3, the literature of author Yu Liping co-occurred most with other literatures, which shows that *Big Data and Big Data Economics* is the classical literature.

IV. CONCLUSION

The impacts of big data are reflected in all aspects. This paper analyzed the research status and development trends of big data in the economic field from three aspects: the research hotspots, the evolution of research hotspots, and the basic knowledge of research. The research findings are 1) the research hotspots are "big data", "internet finance", "cloud computing", "data mining", "smart cities", "precision poverty alleviation", etc. It shows that big data research gradually moves from theory to application; 2) big data appeared in the "Internet of Things" originally. Then the research trend gradually evolved towards "online shopping", "internet finance", and "classified networks". 3) The high-cited references on big data have gone through the stages from theory to application, and applications have focused on taxation and auditing. From the stage of the historical evolution of global big data research drawn by Wang Qian, comparing China's hotspots and the evolution of hotspots, it shows that China's related research of technology exploration and practical application stage needs further study.

REFERENCES

- [1] Wang B, Li X. 2017. Big data, platform economy and market competition a preliminary construction of plan-oriented market system in the information Era[J]. *World Review of Political Economy*, 8(2): 138-161.
- [2] He L, Cuimin Y, Yafeng L. A Review of Big Data Research Based on Bibliometrics[J]. *Information Science*. 2014, 32(06): 148-155.
- [3] Qian W, Tianzhu L, Xiaoqin L. Historic Evolution of Global Big Data Research: 1993-2016[J]. *Forum on Science and Technology in China*. 2017(07): 33-39.
- [4] Xuejiao T, Xiaofeng H, Yang L. Overview of Big Data Research[J]. *Journal of System simulation*. 2013, 25: 142-146.
- [5] Akter S, Wamba S F. Big data analytics in E-commerce: a systematic review and agenda for future research[J]. *Electronic Markets*. 2016, 26(2): 173-194.
- [6] Sheng J, Amankwah-Amoah J, Wang X. A multidisciplinary perspective of big data in management research[J]. *International Journal of Production Economics*. 2017, 191: 97-112.
- [7] Chaomei C, Jie L. *CiteSpace: Text Mining and Visualization in Scientific Literature*[M]. Beijing: Capital University of Economics and Business Press, 2016.
- [8] Ping X, Chuanwei Z. The Research of Internet Based Finance Mode[J]. *Journal of Financial Research*. 2012(12): 11-22.
- [9] McAfee A, Erik B. *Strategy & competition Big Data_ The Management Revolution*[J]. *Harvard Business Review*. 2012: 61-68.
- [10] Zhiyan F, Xunhua G, Dajun Z, et al. On the research frontiers of business On the research frontiers of business management in the context of Big data[J]. *Journal of Management Sciences in China*. 2013, 16(1): 1-9.
- [11] Liping Y. Big Data and Big Data Economics[J]. *China soft science magazine*. 2013: 177-183.
- [12] Hequan W. Opportunities and Challenges in the Era of Big Data[J]. *QIUSHI*. 2013(3): 47-49.