DSI DATA SCIENCE AND INFORMETRICS
VOL.1, NO.3, AUG. 2021

RESEARCH ARTICLES

China´s research contribution in big data

Shiji Chen[a,b], Junping Qiu[a,b], Bo Yu[a,b]

a. Chinese Academy of Science and Education Evaluation (CASEE) , Hangzhou Dianzi University, Hangzhou, China

b. Academy of Data Science and Informetrics (ADSI) , Hangzhou Dianzi University, Hangzhou, China

ABSTRACT

Big data is one of the current and future research frontiers. It has received international attention, and some countries have even upgraded big data research to a national strategy. Therefore, it is interesting to understand the status quo of big data research and identify the status and contribution of a country. Our study is divided into two parts. The first part of this study combines core lexical query and expanded lexical query to get relatively integral publications' data sets on big data. Citation relationships and a maximum connected subgraph algorithm are used to clean and filter unrelated publications. Then the Leiden algorithm is selected to cluster the citation network for big data and VOSviewer is used to map the big data knowledge structure. In the second part of this study, we analyze China's research contribution in terms of research output and highly-cited papers. In order to better show the distribution of big data research in China, we utilized science overlay mapping to visualize the status quo of China's research in big data. Our study shows that China is one of the most important countries in big data research and the research covers almost all areas of big data. However, the research performance is relatively low. In terms of knowledge structure with science overlay mapping, China's research mainly focuses on cloud computing, the Internet of Things (IoT), and social media. However, research topics with a greater rate of highly-cited papers are mainly found in cloud computing, big data medicine, and Industry 4.0. These topics are also the dominant areas of China's big data research.

KEYWORDS

Big data; Science overlay mapping; China; Research contribution

Introduction

With the development of information technology, data is growing and accumulating at an unprecedented rate. From scientific research to medical insurance and from banking to the internet, the amount of data generated in various fields is exploding. In the era of big data, data has become an important commercial capital, a valuable resource, and an important economic investment. More importantly, big data research has invoked a major societal transformation, bringing about change to life, work, and thinking. It also brings new development opportunities and challenges to scientific research. Victor et al. mentioned in their book "The

Era of Big Data" that the essence of big data lies in three changes (Mayer-Schönberger & Cukier, 2014). Firstly, we can analyze more data and all data related to a phenomenon, not just rely on random sampling. Secondly, there is so much research data that the obsession with the pursuit of accuracy no longer exists. The thinking mode of pursuing the accuracy of data is suitable for the situation of mastering a "small amount of data." The third change is that we are no longer trying to find causal relationships. Instead, correlation relationships are more important (Anderson, 2008). Big data research has reshaped our understanding of data and has initiated the shift from logic and experiment-driven research paradigms to data-driven ones. As Jim Gray pointed out, with the continuous integration of scientific computing and as multi-disciplinary, massive scientific data continues to be produced, scientific simulation and research based on data will gradually become one of the mainstream trends in scientific development. In other words, data-intensive scientific discoveries or the fourth paradigm of scientific research will become mainstream (Hey, 2012).

Big data is one of the current and future research frontiers. It has received attention from multiple disciplines, such as computer science, statistics, management science, medicine, social science, etc. At the same time, big data research is also valued by various countries, and some countries, such as the USA and China, have even upgraded big data research to a national strategy. Therefore, many studies try to explore the status quo of big data research. As big data related-documents expand rapidly, the bibliometric method is an effective way to explore the research and development trends of big data. Singh et al. (2015) analyzed the research output data on 'Big Data' from 2010 to 2014, which was indexed in the Web of Knowledge and Scopus. Liu et al. (2020) analyzed the articles obtained from the Scopus database published between 2013 and 2018. Raban and Gordon (2020) explored the evolution of data science and big data research in terms of bibliometric analysis. These studies mainly explore research output, collaboration, major contributors (countries, institutions, and individuals), publication sources, and thematic distributions of big data research. In addition, some other studies that analyze big data research are limited to specific areas due to the bibliometric method. These studies are mainly manifested in analyzing the research trend and applications of big data in computer science (Kalantari et al., 2017), business intelligence (Liang & Liu, 2018), circular economy (Nobre & Tavares, 2017), and supply chain management (Lamba & Singh, 2017; Sheng et al., 2017; Mishra et al., 2018).

For bibliometric analysis, a challenging initial step is to get a reliable data set for theme-related publications. There are several bibliometric search strategies used in the literature to harvest publications, including lexical queries, evolutionary lexical queries, citation analysis, and the use of core journal sets to locate relevant publications in research databases (Huang et al., 2011). The most common search method is the core lexical query which usually applies a core of related keywords. Most studies in big data bibliometric analysis use this search strategy. This strategy is simple and can quickly obtain relevant research literature, but a major drawback is its susceptibility to subjectivity when experts are used to defining the keyword set. So the type method possibly leads to omitting relevant research literature. Of course, these strategies are not isolated. They can be used in combination. Zitt and Bassecoulard (2006) proposed a hybrid lexical-citation method that combines lexical queries and citation analysis. Huang et al. (2015) also devised a systematic methodology to help identify research relating to big data. The method combines four search approaches, namely core

lexical query, expanded lexical query, specialized journal search, and cited reference analysis. However, as we know, publications obtained by these search approaches, regardless of one strategy or a combination of multiple strategies, may also include documents that are weak or even irrelevant to the analysis subject. Therefore, our study explores the cleaning of the publications data that is irrelevant or weakly relevant to the analysis subject.

Furthermore, the bibliometric analysis of big data research is all from the whole analysis based on publications data. In fact, big data research is typically an interdisciplinary field. This means that big data research contains multiple sub-fields and research topics. Different research entities, including different countries, institutions, journals, and authors may have different performances and distributions in sub-fields or research topics. Therefore, previous research does not identify these differences. Additionally, the first author and corresponding author are generally considered as the most important in the author list. To identify the contribution of a country more accurately, our study identifies the first author and corresponding author for each publication. In recent years, the scientific research of China has developed rapidly, so this study tries to explore China's research contribution and the distribution in big data research.

The rest of this paper is organized as follows. Section 2 proposes the research framework of our study, describes the construction method of publications data set, and how to measure and present research contributions. Section 3 presents the science mapping of big data and China's research output, distribution, and scientific impact on big data. We conclude our research and avenues for future study in Section 4.

Data and Methodology

Publications Data Set Construction

At present, the data set constructed based on the subject-based bibliometric analysis or science mapping is usually obtained through a lexical query. As mentioned above, the papers obtained through lexical query may contain some documents that are not related to the research topic. Obviously, these irrelevant or weakly relevant documents will cause interference or deviations in bibliometric analysis and science mapping construction. Documents related to a topic will eventually form a relatively close citation network through citation relationships. If there is no citation relationship between a paper and the topic-related citation network, then theoretically, the paper has little to do with the topic. The current bibliometric analysis or science mapping construction does not take it into consideration, and the publications data set cannot be cleaned and processed effectively, especially when there is a large amount of data. Our study attempts to use citation relationships to clean and filter the publications data set. Specifically, we establish a citation network by the connection between publications and their references in the publications data set, obtaining the largest connected component of the citation network through the maximum connected graph algorithm. The most connected component is used as the core citation network. Then we use expanded lexical query to retrieve more documents from Web of Science as candidate publications and include related publications by the expanded publication's citation to core citation network. The specific process is as follows (Figure 1):

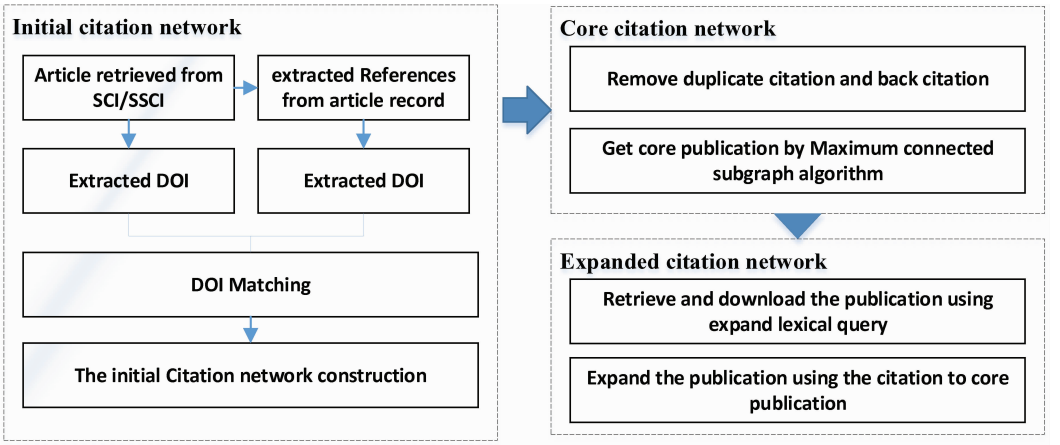

Figure 1 The framework of the final citation network formation

(1) Initial citation network

Firstly, we retrieve relevant SCI/SSCI papers of big data research from the Web of Science database, only including Science Citation Index (SCI) and Social Science Citation Index (SSCI), by core lexical query. The full record format, including references, is downloaded from Web of Science. In order to improve the efficiency of data analysis, we write a computer program in Java to analyze and process the publication records. The DOI of each paper and reference are extracted. The DOI of the reference and of the paper are matched to construct the citation network between publications.

For big data research, unsurprisingly, "Big Data" itself is central. We directly used "Big Data" as core lexical applied in Web of Science. Here we only choose two databases, SCI and SSCI. The search field is the subject, the time period is from 2003 to April 27, 2021, and document types are limited to Article and Proceeding Paper. In the end, 22,666 papers were obtained using this search method, which also involved 740,982 references.

(2) Core citation network

In the process of constructing the citation network, repeated citations and back citations were discovered. Repeat citations may be due to data errors, and back citations may be caused by citations when the paper was published online without an issue and page number assigned. Therefore, repeated citations and back citations in the citation network needed to be eliminated. As mentioned above, for these publications recorded by core lexical query, it is difficult to form a fully connected citation network because some documents are irrelevant or weakly relevant to the research topic. In order to filter these documents, we used the maximum connected subgraph algorithm to get the largest connected subgraph. The largest connected subgraph is the core citation network for the target subject.

The maximum connected subgraph algorithm naturally realizes the screening and cleaning of the publications. In this study, the largest connected subgraph obtained contained 14,841 papers. Furthermore, 7825 papers did not generate a citation relationship with the largest connected subgraph.

(3) Expanded citation network

Additionally, using only the "Big Data" keyword query may miss some topic-related publications. Big data research generally involves distributed computing or parallel computing (i.e. Hadoop, MapReduce), cloud computing, deep learning, data analysis, etc. The research in

these fields is likely to be related to big data. Therefore, we used (TS = "Know discovery" OR "data-driven" OR "cloud Comput*" OR Hadoop OR Mapreduce OR "machine learn*" OR "deep learn*" OR "Data Science") NOT "Big Data" to retrieve expanded publications data from the Web of Science database, the search field is the subject, and the period is also from 2003 to April 27, 2021. A total of 144,483 big data-related documents and 4,558,574 reference records were obtained. The candidate expanded publications are achieved through the citation relationship between the current publications data set and the aforementioned core citation network. There are two categories of supplementary big data research publications. One category is the papers that cite at least five papers in the aforementioned core citation network, in which there are 467 papers in this category. The other category is the papers that have been cited at least five times by the papers in the aforementioned core citation network. There are 1,726 papers that meet the requirement in the candidate expanded publications. Furthermore, the two categories of publications actually overlap, so the actual number of expanded publications is 1,650. Thus, the total number of big data research-related papers obtained is 16,491.

Measuring and Mapping Country´s Research Contribution

How should one evaluate a country's research contribution and its distribution in science mapping of big data research? This study uses research output and highly-cited papers to evaluate and analyze the research contribution of a single country. Research output is measured by the number of papers published by a country, and highly-cited papers indicate that the citations of papers are in the top 5% in big data research. Since big data research involves different disciplines and research topics, a country may have advantages in specific research themes. This study tries to present a country's research contribution by science overlay mapping. Additionally, considering the importance of the first author and corresponding author, we extract the first author and corresponding author from the publication record, which are downloaded from the Web of Science database. In the study, only the first author and corresponding author are counted in the number of papers and highly-cited papers when calculating research contributions. In the TXT text format, which was downloaded from Web of Science, the "C1" field includes multiple lines, and the first line is the first author and his affiliation. The "RP" field is the corresponding author. Furthermore, the first author and corresponding author may belong to different countries in some cases. For such papers, our study calculates the country of the first author and the corresponding author separately.

Science overlay mapping is a kind of science mapping technology that draws local science mappings of institutions or research fields based on global science mapping (Rafols et al., 2010). Using overlay mapping can not only fully display the local knowledge structure but also reveal the local knowledge structure and its position and relationship in the global knowledge structure. Science overlay mapping is widely used in the field of knowledge visualization. Citespace and VOSViewer also provide the science overlay mapping functions. The construction of science overlay mapping first requires global science mapping as the base map and background. In our study, the knowledge structure of big data research is regarded as the big data global science mapping, and a country's big data research is regarded as local science mapping.

The construction process of science overlay mapping is as follows:

(1) Generate the big data research citation network based on big data research papers, use the Leiden algorithm (Traag et al., 2019) to cluster to form the knowledge structure of big

data research, and use VOSViewer software to generate science mapping on big data research. The Leiden algorithm is an improved version of the Louvain algorithm (Waltman & van Eck, 2013), which can overcome the defect of the Louvain algorithm generating bad clusters or scattered clusters.

(2) Export the map file of the big data science mapping from VOSViewer.

(3) Calculate the number of China's SCI/SSCI papers and highly-cited papers in each cluster and modify the map file of the big data science mapping according to the number of SCI/SSCI papers and highly cited papers in each cluster. Then use VOSViewer to generate the science overlay mapping of China.

Results

Big Data Science Mapping

Use the Leiden algorithm to cluster the big data citation network, set the adjustment parameter to 1, and finally form 33 clusters. There are 20 clusters that include more than 100 papers, the number of papers in other clusters is relatively small, and the largest cluster has only 28 papers. Figure 1 presents the big data knowledge structure mapping, where the node is a cluster, which represents a research topic in big data research. The figure is used as the global science mapping for big data, and it is the base map of overlay mapping for a country's research contribution. The upper-left area of Figure 1 represents social media and computational sociology, which mainly focuses on social media computing (computational sociology) (C5) and social issues (C0) related to big data, such as privacy and security. The upper-right area represents the Internet of Things (C3), Industry 4.0 (C8), and big data analysis (C1). The lower-right area is typical big data technology such as cloud computing, Mapreduce, Hadoop, etc. (C2). The lower-left area is mainly health information and precision medicine (C4), and the area on its right is bioinformatics (C12) and disease epidemiology (C16). The middle area represents the related topics of big data-related algorithm research, machine learning, smart grid, and digital earth (C6, C9, C10, C11). Therefore, from the perspective of science mapping, big data research mainly includes social media computing (computational sociology), big data analysis, big data processing technology (cloud computing, Hadoop, Mapreduce, etc.), the Internet of Things, Industry 4.0, big data health information, smart city, Digital Earth, etc. Furthermore, machine learning, deep learning, and artificial intelligence-related research are distributed in clusters in the middle of the science mapping (C5, C6, C7). These research topics are at the center of the big data knowledge structure, while the periphery is mainly the interdisciplinary application area between big data and different research fields.

Compared with previous big data review research (Chen et al., 2014; Khan et al., 2014; Yang et al. 2016; Yaqoob et al., 2016), the division and distribution of research content of big data are similar. These studies also reveal that big data processing and analysis are the core topic of big data research. Additionally, from the science mapping of big data research, it can be seen that the distribution of big data processing research is relatively concentrated, forming relatively stable clusters and regions. However, the research of big data analysis technology is relatively scattered, distributed in business data analysis related topics such as the Internet of Things, precision medicine, and Industry 4.0.

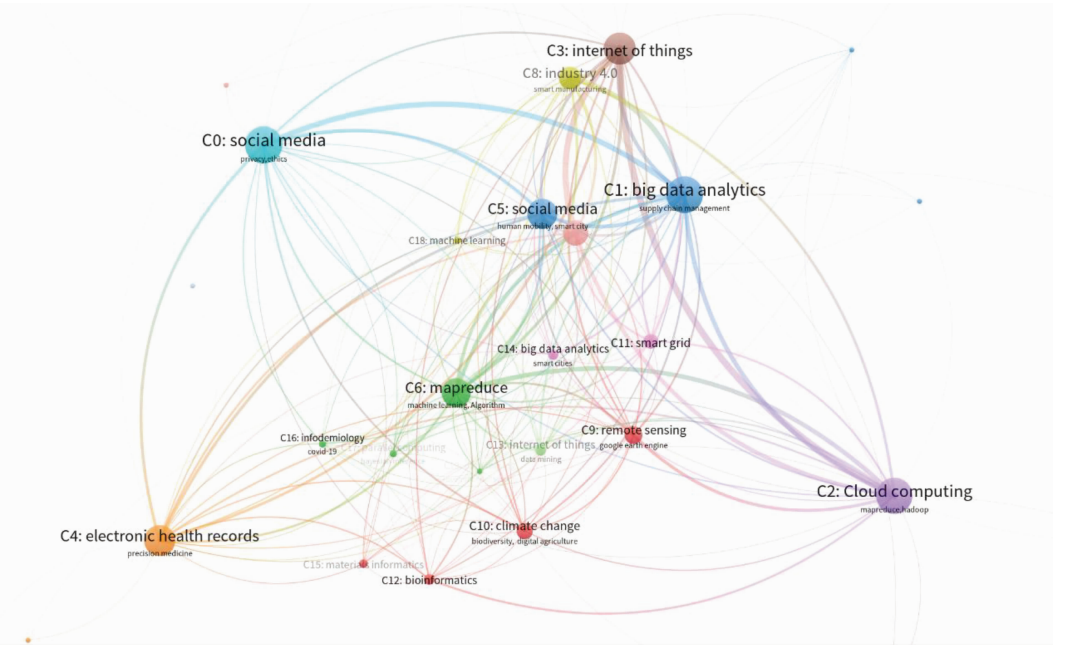

Figure 2 Science mapping on big data

China's Research Output and Distribution on Big Data

Table 1 presents the top 30 countries in terms of research output of big data. FAU indicates the number of papers based on the country of the first author. ORP is the number of papers based on the country of the corresponding author, which only includes the country of the corresponding author when the country of the first author is different. For papers where the corresponding author and the first author belong to different countries, we count the countries of the corresponding author and the first author separately. From Table 1, it is evident that the number of papers in China and the USA far exceeds that of other countries. The number of papers is 4,931 and 3,515, respectively. China and the USA are followed by England, South Korea, Australia, Spain, and India. The number of papers in each of these countries is more than 500. China and the USA are much more prolific in terms of big data research.

Table 1 Top 30 countries of Research output on Big Data

Country	FAU	ORP	Total
China	4797	134	4931
USA	3242	273	3515
England	785	113	898
South Korea	595	96	691
Australia	589	84	673
Spain	591	31	622
India	577	22	599
Italy	499	31	530

Country	FAU	ORP	Total
Canada	393	70	463
Germany	400	35	435
Taiwan,China	289	49	338
France	243	48	291
Netherlands	226	23	249
Brazil	174	19	193
Japan	156	28	184
Switzerland	152	19	171
Saudi Arabia	125	45	170
Pakistan	147	17	164
Iran	141	5	146
Singapore	119	23	142
Belgium	126	11	137
Malaysia	109	16	125
Turkey	118	4	122
Sweden	106	16	122
Greece	112	3	115
Finland	95	15	110
Scotland	93	11	104
Denmark	90	12	102
Norway	87	13	100

Figure 2 presents the position and distribution of China's research output on big data in terms of science overlay mapping. Almost all clusters contain China's papers. The clusters

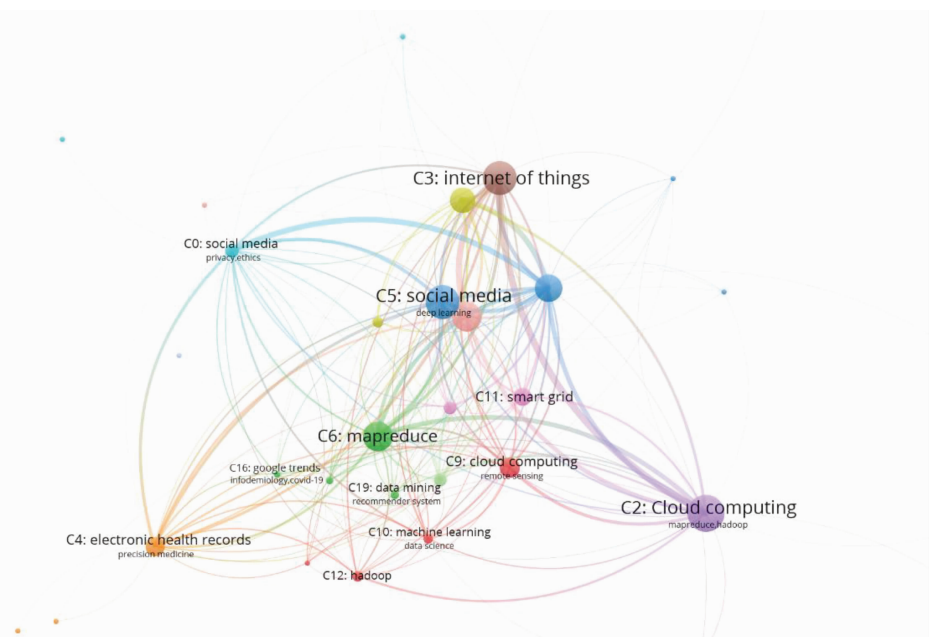

Figure 2 the overlay mapping of China research output on big data

with the most papers from China are C2 (754), C3 (610), and C5 (623). The research topics of these three clusters are cloud computing (C2), the Internet of Things (C3), and social media (C5). Secondly, there are also many papers from China which are distributed in C1, C6, C7, and C8. Table 2 lists keywords of some clusters which distributed many papers from China. This means that China has an important contribution in big data analysis, big data process analysis, social media, the Internet of Things, and cloud computing. From Figure 2, almost all important research topics have the distribution of China's papers.

Table 2 important clusters and their keyword lists

code	papers	Main keywords
C2	754	cloud computing, Mapreduce, Hadoop, security, cloud storage
C3	610	internet of things, cloud computing, edge computing, fog computing, smart city
C5	623	social media, deep learning, machine learning, human mobility,mobile phone data
C1	398	big data analytics, social media, supply chain management,business analytics
C6	473	Mapreduce, machine learning,data mining,feature selection,classification
C7	486	machine learning, deep learning, internet of things, big data analytics, security
C8	359	industry 4.0, internet of things, cloud computing, smart manufacturing, cloud manufacturing

China´s Distribution of Highly–Cited Papers in Big Data

Table 3 presents the top 30 countries in terms of the number of highly-cited papers on big data. PS is the number of papers for the country. HCP is the number of highly-cited papers listed with the country of the first author or corresponding author. PPtop 5% is the rate of highly-cited papers for the country. China and the USA stand out in terms of production over the other countries. Although the number of China's papers is more than that of the USA, both the number of highly-cited papers and the rate of highly cited papers (PPtop 5%) is not as good as those in the USA. In terms of the PPtop 5% indicator, the rate of China's highly-cited papers is less than 5% , which means that China' s research performance in big data is lower than the world average. In addition to the USA, England, Australia, Canada, and Germany have higher research output and research quality. In short, China has high research output, but the research performance is relatively low.

Table 3 top 30 countries or region of highly cited papers on Big Data

country	PS	HCP (top 5%)	PPtop 5%
China	4931	196	3.97%
USA	3515	299	8.51%
England	898	49	5.46%
South Korea	691	19	2.75%
Australia	673	51	7.58%
Spain	622	23	3.70%

country	PS	HCP (top 5%)	PPtop 5%
India.	599	9	1.50%
Italy	530	16	3.02%
Canada	463	25	5.40%
Germany	435	35	8.05%
Taiwan, China.	338	10	2.96%
France.	291	11	3.78%
Netherlands.	249	10	4.02%
Brazil.	193	5	2.59%
Japan.	184	9	4.89%
Switzerland	171	15	8.77%
Saudi Arabia.	170	2	1.18%
Pakistan.	164	1	0.61%
Iran.	146	5	3.42%
Singapore	142	22	15.49%
Belgium.	137	4	2.92%
Malaysia.	125	10	8.00%
Turkey.	122	4	3.28%
Sweden.	122	0	0
Greece.	115	8	6.96%
Finland.	110	2	1.82%
Scotland.	104	6	5.77%
Denmark.	102	1	0.98%
Norway.	100	4	4.00%

Figure 3 presents the distribution of highly-cited papers in big data science mapping. The highest number of highly-cited papers are the three clusters of C2, C3, and C8, followed by C5, C6, and C7. The number of highly-cited papers of cluster C2, C3, and C8 are more than 30 papers, and the rate of highly-cited papers of the three clusters are 4.6%, 5.3%, and 10.9%, respectively. C3 and C8 have a significant research quality, especially cluster C8. Table 2 has presented the main keywords of these important clusters. The main keyword of cluster C2 is cloud computing, the main keyword of C3 is the Internet of Things, and the main keywords of C8 are Industry 4.0 and Internet of Things. This means that these topics are important areas of China's attention and have certain research advantages. In addition to the three clusters, cluster C5, C6, and C7 also include a lot of highly-cited papers, but the rate of highly cited papers of these clusters are relatively low, and their rates are 2.4%, 3.8%, and 4.5%, respectively. These clusters' topics mainly involve social media and big data analysis, including machine learning, deep learning, data mining, etc.

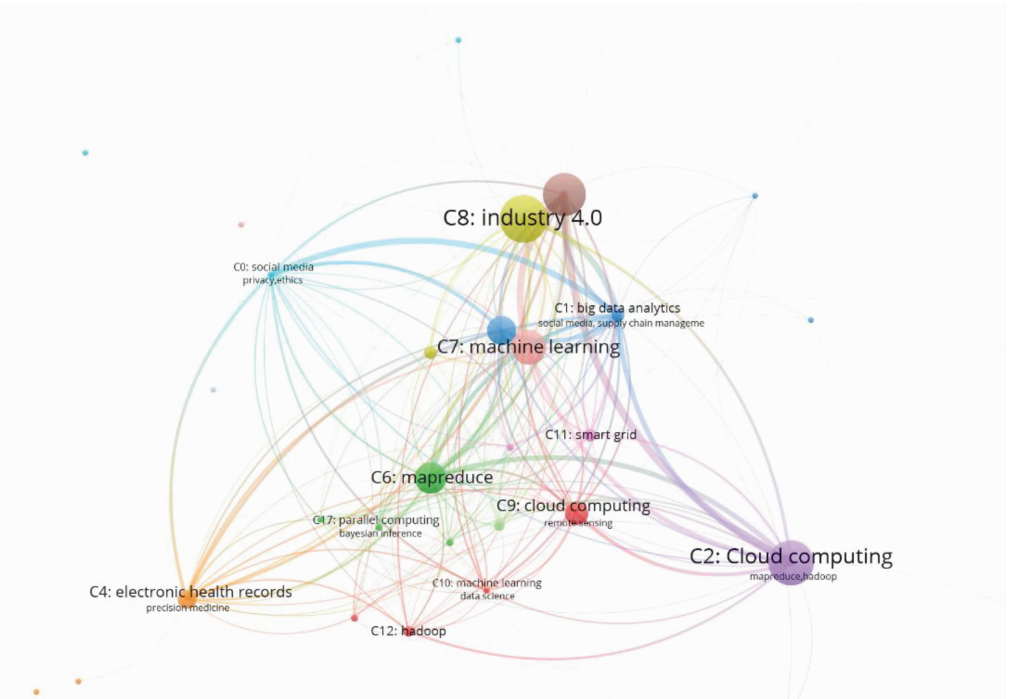

Figure 3 the overlay mapping of highly-cited papers on big data in China

Conclusion and Discussion

The study shows that China is one of the most important countries in big data research and China's research covers almost all areas of big data research. In terms of research output, China is the country that published the most papers in big data research. In particular, the number of papers in China and the USA far exceeds that of other countries. Furthermore, China and the USA are also the two countries with the highest number of highly-cited papers. Although the number of China's papers is more than that of the USA, both the number of highly-cited papers and the rate of highly-cited papers (PPtop 5%) is greater in the USA. In short, China has high research output, but the research performance is relatively low. In terms of knowledge structure with science overlay mapping, the topics with the most papers from China mainly focused on cloud computing, the Internet of Things, and social media. However, research topics with a higher rate of highly-cited papers are mainly distributed in cloud computing, big data medicine, and Industry 4.0.These topics are also the dominant areas of China's big data research.

The study discloses China's research contribution in combination with bibliometrics indicators and science overlay mapping. First, we propose a construction framework of highly-related publications data sets. This framework focuses on cleaning and filtering irrelevant or weakly relevant publications data via citation relationships and the maximum connected algorithm. Secondly, the Leiden algorithm is selected to cluster the citation network for big data, and VOSviewer is used to map the big data science structure. In order to better show the distribution of China's big data research, we utilized science overlay mapping to visualize the status quo of China's research in big data. This method can not only analyze China's performance in big data research as a whole, but more importantly, it

can also analyze China's research contributions in some important sub-fields of big data. Additionally, considering the importance of the first author and corresponding author, only the first author and corresponding author are counted in the number of papers and highly-cited papers when calculating research contributions. This approach can also reveal a country's research contributions more accurately.

Acknowledgements

This study is supported by the National Social Science Foundation of China (Grant No. 19ZDA348).

Reference

Anderson, C. (2008) . The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine, 16* (07) .

Chen, M., Mao, S. W., & Liu, Y. H. (2014) . Big Data: A Survey. *Mobile Networks & Applications, 19* (2) , 171–209. doi:10.1007/s11036–013–0489–0

Hey, T. (2012) . The Fourth Paradigm – Data–Intensive Scientific Discovery. In *E–Science and Information Management* (pp. 1–10) . Verlag Berlin Heidelberg: Springer.

Huang, C., Notten, A., & Rasters, N. (2011) . Nanoscience and technology publications and patents: a review of social science studies and search strategies. *Journal of Technology Transfer, 36* (2) , 145–172. doi: 10.1007/s10961–009–9149–8

Huang, Y., Schuehle, J., Porter, A. L., & Youtie, J. (2015) . A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for ´Big Data´. *Scientometrics, 105*, 2005–2022. doi:10.1007/s11192–015–1638–y

Kalantari, A., Kamsin, A., Kamaruddin, H. S., Ale Ebrahim, N., Gani, A., Ebrahimi, A., & Shamshirband, S. (2017) . A bibliometric approach to tracking big data research trends. *Journal of Big Data, 4* (1) . doi: 10.1186/s40537–017–0088–1

Khan, N., Yaqoob, I., Hashem, I. A., Inayat, Z., Ali, W. K., Alam, M., . . . Gani, A. (2014) . Big data: survey, technologies, opportunities, and challenges. *ScientificWorldJournal, 2014*, 712826. doi:10.1155/2014/712826

Lamba, K., & Singh, S. P. (2017) . Big data in operations and supply chain management: current trends and future perspectives. *Production Planning & Control, 28* (11–12) , 877–890. doi: 10. 1080/09537287. 2017.1336787

Liang, T.–P., & Liu, Y.–H. (2018) . Research Landscape of Business Intelligence and Big Data analytics: A bib–liometrics study. *Expert Systems with Applications, 111*, 2–10. doi:10.1016/j.eswa.2018.05.018

Liu, X. H., Sun, R. Q., Wang, S. Y., & Wu, Y. J. (2020) . The research landscape of big data: a bibliometric analysis. *Library Hi Tech, 38* (2) , 367–384. doi:10.1108/Lht–01–2019–0024

Mayer–Sch?nberger, V., & Cukier, K. (2014) . Big Data: *A Revolution That Will Transform How We Live, Work, and Think*. London, England: Mariner Books.

Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, S. J. (2018) . Big Data and supply chain manage–ment: a review and bibliometric analysis. *Annals of Operations Research, 270* (1–2) , 313–336. doi:10.1007/s10479–016–2236–y

Nobre, G. C., & Tavares, E. (2017) . Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study. *Scientometrics, 111* (1) , 463–492. doi:10.1007/s11192–017–2281–6

Raban, D. R., & Gordon, A. (2020) . The evolution of data science and big data research: *A bibliometric analy–sis. Scientometrics, 122* (3) , 1563–1581. doi:10.1007/s11192–020–03371–2

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010) . Science Overlay Maps: A New Tool for Research Policy and Library Management. *Journal of the American Society for Information Science and Technology, 61* (9) , 1871–1887. doi:10.1002/asi.21368

Sheng, J., Amankwah–Amoah, J., & Wang, X. (2017) . A multidisciplinary perspective of big data in manage–ment research. *International Journal of Production Economics, 191*, 97–112. doi:10.1016/j.ijpe.2017.06.006

Singh, V. K., Banshal, S. K., Singhal, K., & Uddin, A. (2015) . Scientometric mapping of research on ´Big Da–ta´. *Scientometrics, 105* (2) , 727–741. doi:10.1007/s11192–015–1729–9

Traag, V. A., Waltman, L., & van Eck, N. J. (2019) . From Louvain to Leiden: guaranteeing well–connected communities. *Sci Rep, 9* (1) , 5233. doi:10.1038/s41598–019–41695–z

Waltman, L., & van Eck, N. J. (2013) . A smart local moving algorithm for large–scale modularity–based com–munity detection. *The European Physical Journal B, 86* (11) . doi:10.1140/epjb/e2013–40829–0

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2016) . Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth, 10* (1) , 13–53. doi:10.1080/17538947.2016.1239771

Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016) . Big data: From beginning to future. *International Journal of Information Management, 36* (6) , 1231–1247. doi: 10.1016/j.ijinfomgt.2016.07.009

Zitt, M., & Bassecoulard, E. (2006) . Delineating complex scientific fields by an hybrid lexical–citation method: An application to nanosciences. *Information Processing &Management, 42* (6) , 1513–1531. doi:10.1016/j. ipm.2006.03.016