DSI DATA SCIENCE AND INFORMETRICS
VOL.3, NO.1, FEB. 2023

RESEARCH ARTICLE

Big data in education: Themes and trends

Teng Zhao[a], Shiji Chen[b,c]*, Junping Qiu[a,b,c], Yunlong Yu[b,c]

a. Zhejiang Academy of Higher Education, Hangzhou Dianzi University, Hangzhou, China

b. Chinese Academy of Science and Education Evaluation, Hangzhou Dianzi University, Hangzhou, China

c. Academy of Data Science and Informetrics, Hangzhou Dianzi University, Hangzhou, China

ABSTRACT

With the rapid growth of big data research, the existing research has investigated the research themes and trends of big data in different disciplines, but less has paid attentions on the field of education. Using bibliographic data from the Web of Science (WoS), we conduct bibliometric analysis and science mapping to explore the research themes and trends of big data in education. The results show that though education is not the major producer of big data research, it does have a positive development trend. In addition, we find that big data in education mainly serves as a tool to facilitate educational outcomes. Implications, limitations, and future directions are discussed.

KEYWORDS

Big data; Education; Themes and trends; Bibliometric analysis; Science mapping

1 Introduction

With the rapid emergence of innovative technologies, there is a significant increase in big data demand in a variety of industries such as insurance, healthcare, and e-commerce (Baig et al., 2020). Education—as one of the most crucial sectors in promoting the development of society—is not an exception, especially in the present digital era. It has been tied closely with big data, not only because it produces big data through various educational activities, such as the applications of smart devices (Singh & Miah, 2020) and the development of online courses (Baig et al., 2020), but also because it in turn relies heavily on these data. Undoubtedly, effectively utilizing these big data helps educational stakeholders to make better informed decisions and improve educational effectiveness (Fischer et al., 2020).

According to Laney (2001), big data has three typical features, namely three Vs: Volume, Velocity, and Variety. In the current education setting, there are substantial amounts (volume) of data generating over time (velocity) at different levels (variety), which is of significant value. Consequently, it has gained attention from many countries, and some, such as the United States and China, have placed big data in a priority (Chen et al., 2021). The National Center for Educational Statistics (NCES) in the U.S. have conducted multiple nationally representative studies for researchers to assess different educational outcomes in different periods and at different levels. As examples, the Education Longitudinal Study of 2002 (ELS:02) and

* Corresponding Author: chensj@hdu.edu.cn

the High School Longitudinal Study of 2009 (HSLS:09) are both high school-based longitudinal studies, but they are conducted with separate cohorts (different periods), whereas the Integrated Postsecondary Education Data System (IPEDS) primarily focus on postsecondary institutions (different levels). Though not as abundant as these in the U.S., China has also collected several nationally representative education-related big data, such as China Family Panel Studies (CFPS) by Peking University. It should be noted that in education, scholars also regard large-scale datasets as "big data" (Sorensen, 2019).

Noteworthy, since educational research has valued and encouraged interdisciplinarity increasingly (Chettiparamb, 2007), advanced sources (e.g., big data) and analytical approaches are inevitably required to process and interpret (Fischer et al., 2020). This leads educational researchers to shift considerable attention to leveraging big data to assess educational outcomes from different perspectives. For instance, Zhao and Perez-Felkner (2022) utilized the HSLS:09 to examine the effects of perceived mathematical and scientific abilities and interests on STEM major choices, with approximately 23,000 students across the country. Wang (2013) analyzed 14,000 high school graduates in the ELS:02 to understand the influential factors of their entrance to STEM majors in colleges. Focusing on 1,6000 Chinese middle school students, Lei (2021) examined the effect of community socioeconomic status on high school attendance.

However, limited research has systematically and quantitatively trimmed the patterns of research themes and trends in educational research related to big data. With the rapid growth of big data research, the existing research (e.g., Aboelmaged & Mouakket, 2020; Liu et al., 2020; Singh et al., 2015) has applied bibliometric methods and science mapping to investigate the research themes and trends of big data. In addition, many studies are exploring and analyzing the application of big data in different disciplines. Focusing on medical big data, Liao et al. (2018) studied the status of medical big data. Concentrating on the field of business and management, Ardito et al. (2019) depicted the literature links between big data and management phenomena. Yet, to our knowledge, limited research has explored the research outputs of educational big data research, including thematic distributions and trends.

The present study aims to fill this research gap by using bibliometric analysis and science mapping method. Guided by Chen et al. (2021) and Huang et al. (2015), we adopt the core lexical query, the maximum connected subgraph algorithm, and the expanded lexical query to construct a reliable data set for big data-related papers, using the Web of Science (WoS) databases. Then, we limit the field to education by only focusing on the *Education and Educational Research* category, thereby obtaining big data-related papers in education. By analyzing and visualizing these papers, this study explores the research themes and trends of big data in education. It not only helps to shed light on the research outputs of big data in education and offers educational research promise, but also paves the way to the intersection between big data and the education field.

2 Materials and Methods

2.1 Data Source

The source of bibliometric data is from the WoS, which is currently owned and maintained by Clarivate Analytics. WoS databases are often used by scholars to conduct bibliometric analysis (AlRyalat et al., 2019; Archambault et al., 2009; Huang et al., 2022). Reference tracing and citation reporting are two major advantages of the WoS datasets, which could help to identify themes and trends in a given research area, with leading academic journals and citation networks (Huang et al., 2022). WoS core collection consists of Science Citation Index (SCI), Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI), and the former two were selected for the present study. We set the search field as the subject and limited the time period from 2003 to April 27, 2021. In addition, we limited paper types to Article and Proceeding papers.

2.2 Data Set Construction

2.2.1 Overall Process of Data Set Construction

The lexical query has been widely used to construct the data set for subject-based bibliometric analysis and/or science mapping (Chen et al., 2021; Yalcin & Daim, 2021). However, in some cases, papers obtained through lexical query may be irrelevant or weakly relevant to the research topic, which leads to interference or deviations. Scholars (e.g., Mejia et al., 2021) use citation networks that are measured by citation relationships to identify whether papers are related to a topic or not. Thus, a paper appears to have little relevance to the topic if no citation relationship is observed between this paper and the topic-related citation network.

With these in mind, our study adopted three steps to construct our analytic data, constructing the initial citation network, core citation network, and expanded citation network, respectively. More detailedly, we constructed the citation network by matching DOIs of published papers and their references. Then, we applied the maximum connected graph algorithm to extract the core citation network from the initial citation network. At last, we established the expanded citation network by expanding the lexical query to retrieve more papers that related to the intended research. The process of data set construction is presented as Figure 1.

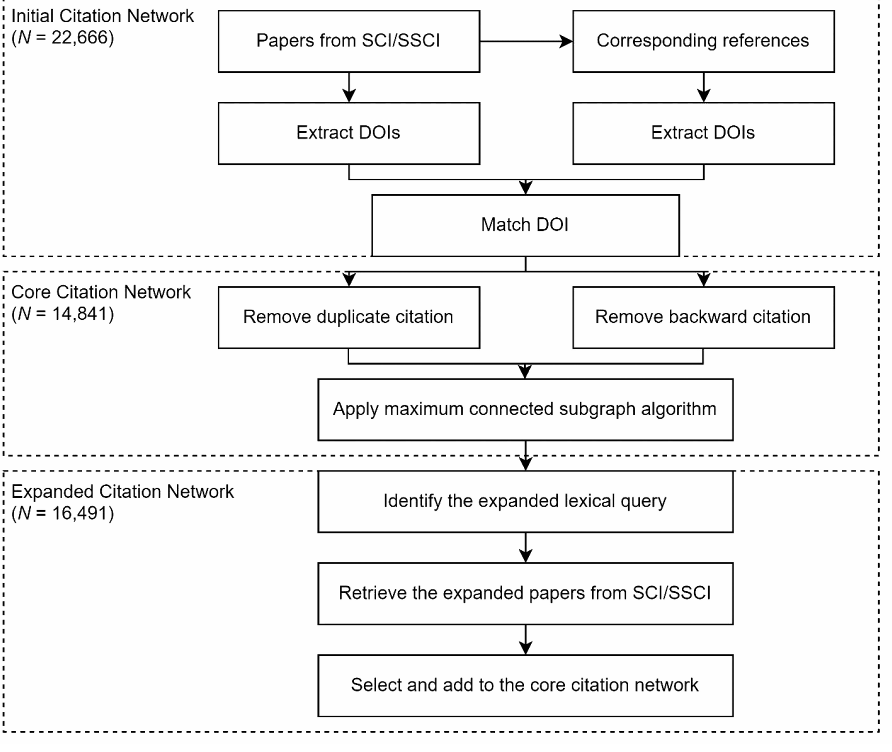

Figure 1 The Process of the Analytic Data Set Construction.

2.2.2 Constructing the Initial Citation Network

To retrieve big data-related research from SCI and SSCI databases, we used "big data" and "bigdata" as the core lexical query. The full record format of papers along with the corresponding references was derived. These records were analyzed and processed using a *Java* program in order to increase the efficiency of data analysis. Then, we extracted the DOIs from both retrieved papers and their corresponding references. Finally, we matched these two sets of DOIs to construct the initial citation network between papers, obtaining 22,666 papers and 740,982 references.

2.2.3 Generating the Core Citation Network

Repeated citations and backward citations may exist when constructing citation networks (Small, 1999), which would hurt the accuracy of citation networks. Repeated citations may be caused by errors in data, while backward citations may occur when papers were these online first without an issue and page number. To ensure the precision of our citation network, we removed these repeated and backward citations.

In addition, not all the papers were strongly relevant to the research topic, therefore, the maximum connected subgraph algorithm was adopted to obtain the largest connected subgraph (Hu et al., 2016). The algorithm was designed to monitor and clean the target papers, thereby forming the core citation network for big data-related research. After performing the algorithm, 7,825 papers were dropped due to the lack of a citation relationship, and left 14,841 papers in the core citation network for the present study.

2.2.4 Constructing the Expanded Citation Network

As known, big data involves a variety of fields, such as computing (e.g., cloud computing), machine learning, and data analysis. To obtain papers related to big data as complete as possible, we further performed a lexical query with TS = "comput*" OR "machine learn*" OR "deep learn*" OR "data*", in order to expand our data of big data-related papers. The asterisk was used as a truncation symbol to accommodate variations of the search term (Mejia et al., 2021). For example, "comput*" could include "computing, computer, computers". It should be noted that this query process was the same to the initial query process, that is, both the search field and the search period were constant.

However, in this expanded query, these retrieved papers were likely to be irrelevant or weakly relevant to the research topic. Hence, we set two additional selection criteria to filter and clean the papers, resulting in a more precise citation network. Specifically, we limited papers to those that either cite or have been cited by at least five papers in the core citation network mentioned above. After dropping the overlap papers extracted by the two selection criteria, we had 1,650 expanded big data-related papers. Adding these expanded papers into the core citation network, the final 16,491 big data-related papers in the expanded citation network were obtained.

2.3 Measuring and Mapping Big Data in Education

In order to understand the research outputs of big data in education, we further limited all the generated big data-related papers to the education field. We derived the big data-related papers from the *Education and Educational Research* category, and sorted them as the research outputs of big data in education. The total number of papers on big data in education was 100. Using these 100 papers, the present study attempted to present the research themes of big data in education by science overlay maps.

Science overlay maps have been widely used to visually map scientific outputs locally and globally (Carley et al., 2017; Rafols et al., 2010). It contributes to visualize local knowledge structure in a variety of scientific fields including education, as well as its positions and relationships in the global knowledge structure (Chen et al., 2021). We used VOSViewer software to generate science overlay maps, treating the whole big data-related research as the global science maps and educational big-data-related research as the local science maps.

Specifically, we generated the expanded citation network based on the whole big data-related papers. Then, we adopted Leiden algorithm suggested by Traag et al. (2019) to cluster the knowledge structure of these big data-related papers and use VOSViewer to create science overlay maps on big data. Last, we modified the big data science overlay maps based on the number of big data-related papers in education in each cluster, generating the science overlay maps of educational research output on big data.

3 Results

3.1 Science Overlay Maps on Big Data

Figure 2 displays the knowledge structure maps of big data. The nodes in the figure represent a cluster of research topics in big data research. The size of the nodes represents the weights of the nodes, which means the bigger the node, the larger the weight is. There were 32 clusters after we adopted the Leiden algorithm and set the adjustment parameter to "1" . However, we did not include all of them in Figure 2 in order to better display the map, and truncated a partial map only including major clusters.

As shown in Figure 2, in the upper-left corner, the node represents social media (C0), which primarily concentrates on social media computing and social media issues related to big data like privacy and ethics. Big data analytics (C1) and Internet of Things (C3) were the two major nodes in the upper-right corner. Looking at the middle area of Figure 2, Mapreduce (C6), smart grid (C11), industry 4.0 (C8), google trends (C16), Internet of Things (C13), machine learning (C10), and Hadoop (C12) were the main nodes. In the lower-left corner, electronic health records (C4) was the main node. Focusing on the lower-right corner, cloud computing (C2) had a relatively large node size. Therefore, according to the science overlay map on big data, big data research mainly concentrated on social media, big data analytics, Internet of Things, industry 4.0, big data processing technology such as Mapreduce, Hadoop, and cloud computing, etc.

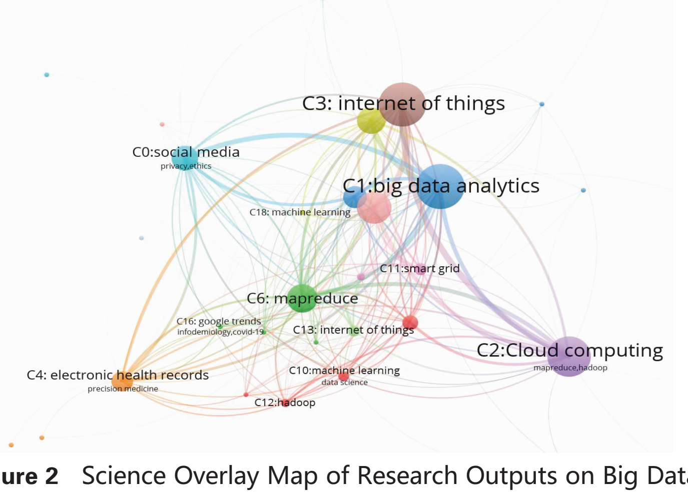

Figure 2 Science Overlay Map of Research Outputs on Big Data.

3.2 Research Output and Trends on Big Data in Education

Figure 3 displays the positions and distributions of research outputs on big data in education. Unsurprisingly, only 12 of 33 clusters contained educational papers. Looking at Figure 3, social media (C0), Internet of Things (C13), and big data analytics (C1) were the largest three clusters on the map, indicating most educational big data research was in these three research topics. Table 1 presents the numbers of educational big data-related papers as well as the whole big data-related papers, by research topics. The largest three clusters—C0, C13, and C1—had 63, 10, and 6 papers, respectively. In addition, the share of educational big data-related papers in the whole big data-related papers ranged from 0.13% to 3.16%, indicating education may not substantially contribute to big data research.

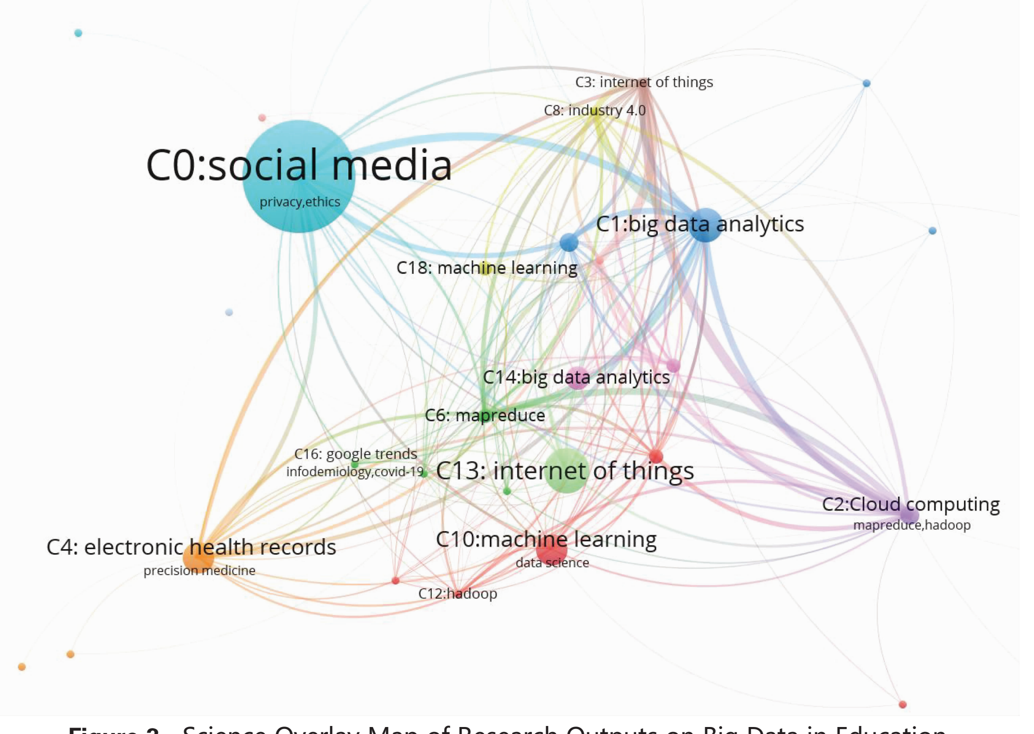

Figure 3 Science Overlay Map of Research Outputs on Big Data in Education.

Table 1 Key Research Clusters of Big Data in Education.

Code	Research cluster	Big data papers in education	Big data papers	Percentage
C0	Social media	63	1992	3.16%
C13	Internet of things	10	626	1.60%
C1	Big data analytics	6	4533	0.13%
C10	Machine learning	5	587	0.85%
C4	Electronic health records	5	1480	0.34%

Furthermore, we sorted the number of educational big data-related papers by year. Figure 4 displays that there was a substantial increase in the number of educational big data-related papers from 2013 to 2017. From 2017 to 2020, the numbers ranged from 16 to 21. It should be noted that there were only 7 papers in 2021 because we limited the time period of data to April, 2021. These results indicated that there is a development trend of big data research in the field of education.

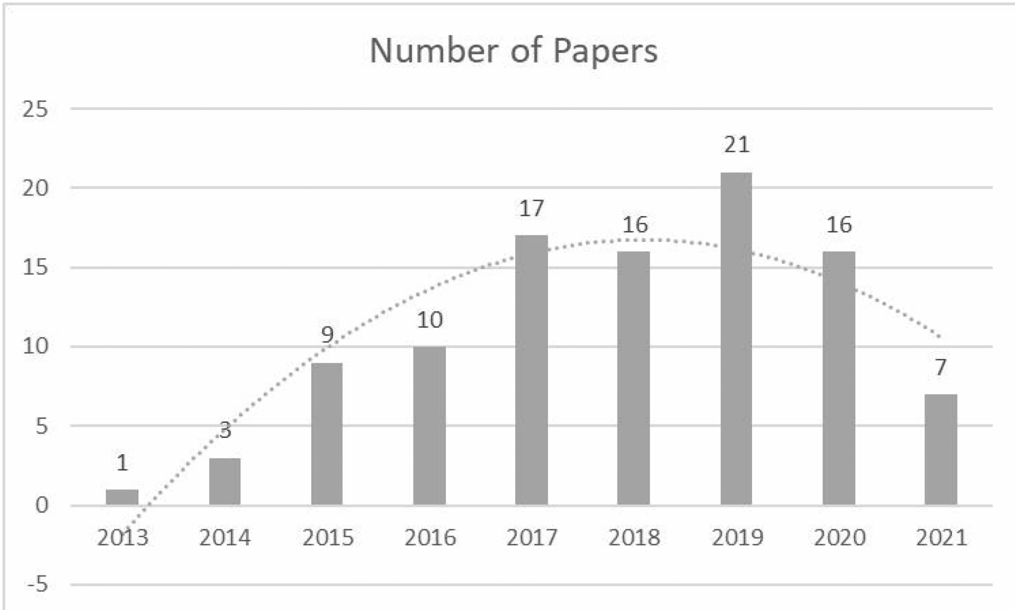

Figure 4 The Numbers of Big Data-related Papers in Education, by Year.

3.3 Content Analysis of Research on Big Data in Education

Further, we were particularly interested in the research content of big data in education, thus we carefully reviewed these 100 papers on big data in education and sorted them by e-ducational levels and keywords, as presented in Table 2. With respect to educational level, results showed that, among them, there were 17 papers clearly specified their research contents were related to higher education including engineering education (4), health professional education (2), and medical education (2), while only 4 papers were to K-12. Other 79 papers did not specify the educational level and thus we sorted them to the overall education including pre-school, K-12, and higher education.

Looking at Top 5 research contents derived from these 100 papers, there were 30 papers relating big data to learning, such as learning experiences, learning challenges, learning outcomes, and learning tools. Also, there were 14 papers focusing on big data analytics, which were used to identify problems, conduct performance assessments, and so on. It should be noted that this number was different from the number in C1 because here we did not use "keywords" to sort research contents, instead, we carefully reviewed all the 100 papers. In addition, other big data papers in education focused on teachers & teaching (9), courses & curriculum (5), and education governance (5).

Table 2 Research Contents of Big Data-related Papers in Education, by educational level and keywords.

Categories	Number of papers	Percentages
Education level		
K–12	4	4%
Higher education	23	23%
Education–overall	79	79%
Top 5 research contents		
Learning	30	30%
Analytics	14	14%
Teachers & teaching	9	9%
Courses & curriculum	5	5%
Education governance	5	5%

4 Discussions and Conclusions

The present study seeks to understand the research themes and trends of big data in education through big data-related papers between 2003 and 2021 which are included in the SCI and SSCI databases. The inner patterns of big data research and big data research in the field of education, as well as the trend of big data research in education have been studied.

First, our findings reveal that big data analysis and big data processing are the core themes of big data research, which is consistent with Chen et al. (2014) and Xu and Yu (2019). Chen et al. (2014) conducted a systematic review and focused on four stages of big data: data generation, data acquisition, data storage, and data analysis, including computing, Internet of Things, and Hadoop, which are also been extracted from our bibliometric analysis. Xu and Yu (2019) conducted a bibliometric and visual analysis to investigate all the big data publications from 2009 to 2018, and found that authors frequently used keywords such as “big data analytics”, “MapReduce” and “social media” in big data research. These words also appear as main research clusters in big data-related papers. Moreover, big data analysis research is relatively scattered in business data analysis-related topics, including precision medicine, industrial automation, and the internet of things. In contrast, big data processing research is largely concentrated, forming relatively stable clusters and regions, as can be seen from the science overlay map of big data research.

Second, our results show that though education is not the major producer of big data research, it does have a positive development trend. This is consistent with Eynon (2013), which also pointed out that the use of big data was still relatively niche in education, but it appears to be on the rise. Our finding of the trend of educational big data-related papers also reveals that from the year 2013, the number of big data-related papers started to increase. In addition, we find that most educational papers related to big data were linked to social media. Paul et al. (2017) pointed out that social media analytics could be applied to the field of education since it is the process of collecting various information to make informed decisions. Ababneh et al. (2020) illustrated that in the current era, education is one of the most common formal fields that frequently use social media as an educational tool to improve student outcomes. These could help to explain why a majority of educational big data-related papers are on the theme of social media.

Third, our results of the distributions of research contents of big data-related papers in education indicate that big data in education mainly serves as a tool to facilitate educational outcomes. For example, for the aspect of student learning, Saqr et al. (2017) used 133 students' online activity data to predict their risk of under-achieving in a blended medical education course. For the aspect of education governance, Sorensen (2019) used large-scale administrative data and applied machine-learning techniques to predict the risk of school dropout for Grade 3 to 8 students in North Carolina. For the aspect of teacher and teaching, Dixon and Moxley (2013) analyzed the 118,611 comments from instructors on student writing, and found the patterns of instructor higher-order concerns and lower-order concerns. Other than the role of a tool to facilitate education, our results also reveal that education, in turn also plays a role in producing big data. For example, Cope and Kalantzis (2016) systematically reviewed the implications of big data in education, focusing on student writing, and pointed out that it could generate unprecedented amounts of data. In addition, with wider applications of innovative tools in education, massive data could also be generated and explored (Aljawarneh, 2020; Cox, 2021).

Two limitations of this study should be noted. First, despite we observed important results through the bibliometric analysis and visualization on big data-related papers, with a focus on the field of education, we only retrieved papers from SCI and SSCI databases. Future research could follow Mejia et al. (2021) to include more databases such as the Arts & Humanities Citation Index and the Emerging Sources Citation Index. Second, all the papers that analyzed in the present study are written in English, however, papers written by other languages also should be taken into account. Future research could additionally analyze big data-related papers in other languages. This could not only help to make the bibliometric analysis more comprehensive but also to understand the regional differences in research outputs of big data, especially in education.

Statements and Declarations

Ethics: Not applicable.

Consent to participate and for publication: Not applicable.

Conflict of interest: The authors declare no competing interests.

Funding: This work was supported by Zhejiang Provincial Philosophy and Social Sciences Planning Project (23NDJC161YB).

Authors' contributions: T.Z., S.C., and J.Q. conceptualized the manuscript; T.Z. led the writing; S.C. performed data collection, conducted the statistical analyses, and led the review; Y.Y. provided review. All authors have read and agreed to the published version of the manuscript.

References

Ababneh, M., Al–Jarrah, A., & Cavus, N. (2020). *Social media usage in education: Big data perspective.* Pro–ceedings of the 14th International Technology, Education and Development Conference, Valencia, Spain.

Aboelmaged, M., & Mouakket, S. (2020). Influencing models and determinants in big data analytics research: A bibliometric analysis. *Information Processing & Management, 57* (4), 102234. doi: 10.1016/j.ipm.2020.102234

Aljawarneh, S. A. (2020). Reviewing and exploring innovative ubiquitous learning tools in higher education. *Journal of Computing in Higher Education, 32* (1), 57–73. doi: 10.1007/s12528–019–09207–0

AlRyalat, S. A. S., Malkawi, L. W., & Momani, S. M. (2019). Comparing bibliometric analysis using pubmed, scopus, and web of science databases. *Journal of Visualized Experiments, 152*, e58494. doi:10.3791/58494

Archambault, é., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the web of science and scopus. *Journal of the American Society for Information Science and Technology, 60* (7), 1320–1326. doi: 10.1002/asi.21062

Ardito, L., Scuotto, V., Del Giudice, M., & Petruzzelli, A. M. (2019). A bibliometric analysis of research on big data analytics for business and management. *Management Decision, 57* (8), 1993–2009. doi: 10.1108/MD-07-2018-0754

Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2020). Big data in education: A state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education, 17* (1), 44. doi: 10.1186/s41239-020-00223-0

Carley, S., Porter, A. L., Rafols, I., & Leydesdorff, L. (2017). Visualization of disciplinary profiles: Enhanced science overlay maps. *Journal of Data and Information Science, 2* (3), 68–111. doi:10.1515/jdis-2017-0015

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19* (2), 171–209. doi: 10.1007/s11036-013-0489-0

Chen, S., Qiu, J., & Yu, B. (2021). China' s research contribution in big data. *Data Science and Informetrics, 1* (3), 1–13.

Chettiparamb, A. (2007). *Interdisciplinarity: A literature review*. Report, Interdisciplinary Teaching and Learning Group, University of Southampton.

Cope, B., & Kalantzis, M. (2016). Big data comes to school: Implications for learning, assessment, and research. *AERA Open, 2* (2), 2332858416641907. doi: 10.1177/2332858416641907

Cox, A. M. (2021). Exploring the impact of artificial intelligence and robots on higher education through literature-based design fictions. *International Journal of Educational Technology in Higher Education, 18* (1), 3. doi: 10.1186/s41239-020-00237-8

Dixon, Z., & Moxley, J. (2013). Everything is illuminated: What big data can tell us about teacher commentary. *Assessing Writing, 18* (4), 241–256. doi: 10.1016/j.asw.2013.08.002

Eynon, R. (2013). The rise of big data: What does it mean for education, technology, and media research?. *Learning, Media and Technology, 38* (3), 237–240. doi: 10.1080/17439884.2013.771783

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education, 44 (1), 130–160. doi: 10.3102/0091732x20903304*

Hu, J., Wu, X., Cheng, R., Luo, S., & Fang, Y. (2016). *Querying minimal steiner maximum-connected subgraphs in large graphs.* Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, USA.

Huang, X., Yang, Z., Zhang, J., Wang, R., Fan, J., Zhang, H., Xu, R., Li, X., Yu, S., Long, L., & Huang, H. (2022). A bibliometric analysis based on web of science: Current perspectives and potential trends of smad7 in oncology. *Frontiers in Cell and Developmental Biology, 9*. doi: 10.3389/fcell.2021.712732

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* (3), 2005–2022. doi: 10.1007/s11192-015-1638-y

Laney, D. (2001). 3d data management: Controlling data volume, velocity and variety. *META Group Research Note, 6* (70), 1.

Lei, L. (2021). The effect of community socioeconomic context on high school attendance in China: A generalized propensity score approach. *Sociology of Education, 95* (1), 61–88. doi: 10.1177/00380407211057305

Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., & Zeng, X.-J. (2018). A bibliometric analysis and visualization of medical big data research. *Sustainability, 10* (1), 166. doi: https://www.mdpi.com/2071-1050/10/1/166

Liu, X., Sun, R., Wang, S., & 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

Mejia, C., Wu, M., Zhang, Y., & Kajikawa, Y. (2021). Exploring topics in bibliometric research through citation networks and semantic analysis. *Front Res Metr Anal, 6*, 742311. doi: 10.3389/frma.2021.742311

Paul, P. V., Monica, K., & Trishanka, M.(2017). *A survey on big data analytics using social media data.* 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India. doi: 10.1109/IPACT.2017.8245092.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and li–brary management. *Journal of the American Society for Information Science and Technology, 61* (9), 1871–1887. doi: 10.1002/asi.21368

Saqr, M., Fors, U., & Tedre, M. (2017). How learning analytics can early predict under–achieving students in a blended medical education course. *Med Teach, 39* (7), 757–767. doi: 10.1080/0142159x.2017.1309376

Singh, H., & Miah, S. J. (2020). Smart education literature: A theoretical analysis. *Education and Information Technologies, 25* (4), 3299–3328. doi: 10.1007/s10639–020–10116–4

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

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Sci–ence, 50* (9), 799–813. doi: 10.1002/(SICI)1097–4571(1999)50:9<799::AID–ASI9>3.0.CO;2–G

Sorensen, L. C. (2019). "Big data" in educational administration: An application for predicting school dropout risk. *Educational Administration Quarterly, 55* (3), 404–446. DOI: 10.1177/0013161x18799439

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From louvain to leiden: Guaranteeing well–connected com–munities. *Scientific Reports, 9* (1), 5233. doi: 10.1038/s41598–019–41695–z

Wang, X. (2013). Why students choose stem majors: Motivation, high school learning, and postsecondary con–text of support. *American Educational Research Journal, 50* (5), 1081–1121. doi: 10.3102/0002831213488622

Xu, Z., & Yu, D. (2019). A bibliometrics analysis on big data research (2009– 2018). *Journal of Data, Informa–tion and Management, 1* (1), 3–15. doi: 10.1007/s42488–019–00001–2

Yalcin, H., & Daim, T. (2021). A scientometric review of technology capability research. *Journal of Engineering and Technology Management, 62*, 101658. doi: 10.1016/j.jengtecman.2021.101658

Zhao, T., & Perez–Felkner, L.(2022). Perceived abilities or academic interests? Longitudinal high school science and mathematics effects on postsecondary stem outcomes by gender and race. *International Journal of STEM Education, 9* (1), 42. doi: 10.1186/s40594–022–00356–w