The Present of Education Big Data Research in China: Base on the Bibliometric Analysis and Knowledge Mapping

Cheng Jiang1, Qi Wang2, Wu Qing 2,*, LeiYe Zhu1, Si Cheng1, HaiBin Wang3,1, *

1 School of Educational Science, Huangshan University, Huangshan, 245041, China
2 Student Office, Huangshan University, Huangshan, 245041, China
3 School of Business Administration, Zhejiang Gongshang University, Hangzhou 310018, China
*Corresponding author. Wu Qing, qingwu@hsu.edu.cn or HaiBin Wang, wanghaibin@hsu.edu.cn

Abstract. Big data is a term used to refer to data sets that are too large or complex for traditional data-processing application software to adequately deal with. With the deep integration between information technology and education, researchers at home and abroad have made great progress in the field of big data in education. Nevertheless, a comprehensive quantitative analysis of the emerging research trends and topics has not yet been discovered. To reveal the research characteristics and current status of educational big data, 2052 related papers from the China National Knowledge Infrastructure (CNKI) were analyzed by CiteSpace V software. The results display that: (1) the domain of education big data was launched in 2013 and has experienced a rapid growth in the past five years, reaching an all-time high in 2017. XianMing Yang and Central China Normal University are at the forefront of contributing authors and organization respectively. (2) “Education in ideology and politics” is the most frequently-used keywords and “electronic schoolbag”, “educational reform” and “MOOC” have the longest outbreak range. “Big data era”, “educational innovation”, “big data”, “educational evaluation”, “educational management”, “educational application”, “higher education” and “educational informatization” are the top eight largest clusters. Those are the topical research topics in the field of educational big data.

1. Introduction
Big data is a term used to refer to data sets that are too large or complex for traditional data-processing application software to adequately deal with, which has the characteristics of sheer volume, wide variety, high velocity and extreme veracity [1]. With the deep integration between information technology and education, great progress has been made by researchers at home and abroad, in areas of big data in education. Thereafter, western researchers from different disciplines began to the study of big data in education, and obtained important achievements, including the construction of education big data technology system, the educational application of big data, the teaching innovation with big data, and so on [2-3].

The 13th Five-Year Plan for the Development of National Education issued by The State Council in 2017, clearly proposed "accelerating the construction of education big data and open sharing" [4]. Although the focus on education big data starts relatively late in China [5], China's educational big data research has achieved a rapid development in the context of the country’s emphasis on education big data. Chinese researchers pay more attention to the technology and application research of
education big data among these achievements. In the technology research of education big data, researchers explore the techniques of collecting, mining and analyzing educational big data, such as with the adoption of the Python [6] and a map-based visual analysis method for big data in education [7]. In the application research of education big data, researchers start with the significance and function of the education big data construction [8-9], and the educational reform by education big data [10]. As has been shown, Chinese researchers have acquired rich achievement in the area of big data in education in the past 5 years. Nevertheless, a comprehensive quantitative analysis to objective and effective authors and organizations have not been completed, along with the burgeoning research trends and topics. For the sake of better promoting the development of Chinese education big data, this paper has conducted a quantitative review of education big data research in China based on the bibliometric analysis and knowledge mapping with the support of the CiteSpace V.

2. Methodology

2.1. data sources
The source and reference analytical papers come from the China National Knowledge Infrastructure (CKNI). There are 2052 papers of all types published conducting in the CNKI collection, under the headings “big data” and “education” or “educational big data”. The data sources contain 1991 journal articles and 61 Dissertations.

2.2 data processing
2052 papers were saved in RIS format which includes the crucial information, such as titles, authors’ names and affiliations, abstract and keywords. The bibliometric analyses and knowledge mapping were conducted with Java CiteSpace V software [11], which supports the construction and visualization of bibliographic record networks.

3. Results

3.1 Bibliometric analysis results

3.1.1 Chronology statistics. In view of annual analysis, the number of annual studies from 2013 to 2018 in China is revealed in Fig. 1. Chinese research of education big data began in 2013 and showed a year-on-year growth trend, reaching an all-time high in 2017.

![Fig 1. The annual research of education big data published in China.](image-url)
3.1.2. Author statistics. Table 1 shows the Chinese top 10 outstanding authors and the number of their published papers. As shown, XianMing Yang tops the list with a total of eleven articles, followed by KeYun Zhao and Hai Zhang, each of which has published six articles.

| Rank | Author         | Frequency |
|------|----------------|-----------|
| 1    | XianMing Yang  | 11        |
| 2    | KeYun Zhao     | 6         |
| 3    | Hai Zhang      | 6         |
| 4    | PengGao Zhang  | 4         |
| 5    | NanZhong Wu    | 4         |
| 6    | HongHai Yu     | 3         |
| 7    | YanNan Zhang   | 3         |
| 8    | FuXiang Hu     | 3         |
| 9    | HuaiJie Li     | 3         |
| 10   | FanGang Hu     | 3         |

3.1.3. Affiliation statistics. Table 2 depicts the Chinese top 10 organizations facilitating the studies on education big data. As indicated, Central China Normal University has the best performing, followed by East China Normal University and Southwest University. Not surprising, nine of the top 10 outstanding organizations are normal university, proving that normal universities play a totally important role in education big data research.

| Rank | Author                                      | Frequency |
|------|---------------------------------------------|-----------|
| 1    | Central China Normal University             | 30        |
| 2    | East China Normal University                | 29        |
| 3    | Southwest University                        | 29        |
| 4    | Northeast Normal University                 | 23        |
| 5    | Beijing Normal University                   | 23        |
| 6    | Qufu Normal University                      | 18        |
| 7    | Jiangsu Normal University                   | 17        |
| 8    | University of Electronic Science and Technology of China | 16 |
| 9    | Shaanxi Normal University                   | 16        |
| 10   | Fujian Normal University                    | 14        |

3.2. Knowledge mapping results

3.2.1. High frequency keywords. Table 3 lists the top 10 keywords of studies on Chinese education big data, apart from their frequencies. Obviously, the most frequently cited keywords are “education in ideology and politics” in addition to the subject keywords “big data”. In addition, “colleges and universities”, “innovation”, “undergraduate” are positioned in the 2nd, 3rd and 4th places respectively, except the subject keywords. It indicates that the current researchers focus on the educational innovation with the help of the big data technology.

| Rank | Keywords                                           | Frequency |
|------|----------------------------------------------------|-----------|
| 1    | big data                                           | 2161      |
| 2    | big data era                                       | 732       |
| 3    | education in ideology and politics                 | 563       |
| 4    | big data in education                              | 264       |
| 5    | colleges and universities                          | 234       |
| 6    | innovation                                         | 177       |
| 7    | undergraduate                                      | 152       |
| 8    | university's ideological and political education    | 144       |
| 9    | education                                          | 111       |
| 10   | higher education                                   | 111       |
3.2.2. Popular research trends. To seek the popular research trends, the keywords with citation burst are analyzed in this paper. Fig 2 presents the top ninth keywords with the strongest citation. As showed, “electronic schoolbag”, “educational reform” and “MOOC” have the longest outbreak range, each of which exceeds more than two years. Furthermore, the topic of “data processing” is one of the latest research frontiers, indicating that the processing technology of education big data can be the popular research trends in future.

![Top ninth keywords with the strongest citation bursts](image)

Fig 2. Top ninth keywords with the strongest citation bursts.

3.2.3. Popular research topics. In order to explore popular research topics, this paper analyzes the keywords with clustering. With running CiteSpace, it presents 39 clusters whose Modularity Q is 0.898 and the Mean Silhouette is 0.901. Fig 3 presents the top eight largest clusters. They were big data era, educational innovation, big data, educational evaluation, educational management, educational application, higher education and educational informatization. As presents, big data era (#0) is the largest cluster. It contains 32 articles and its Silhouette is 0.972. The cluster's high-frequency keywords include “big data era”, “information innovation” and “digital campus”. Educational innovation is the second largest cluster (#1), which includes 27 articles and its Silhouette is 0.864. The high frequency keywords of this cluster contain “innovation”, “big data technology”, “education mode” and “educational innovation”. Big data is the third largest cluster (#2), which comprises 25 articles and its Silhouette is 0.983. The high-frequency keywords in this cluster contain “big data”, “ideological and political education”, “undergraduate” and “data mining”.

![Clusters of the current researches](image)

Fig 3. Clusters of the current researches.

4. Conclusions and further research
In the past five years, there are a large number of papers on education big data having published in China. Nevertheless, a comprehensive quantitative analysis to objectively identifying influential researchers and organizations as well as the emerging research trends and topics has not yet been accomplished. To order to fill this vacancy, with the help of the CiteSpace V software, this paper
quantitatively reviews China’s education big data research on account of the bibliometric analysis and knowledge mapping. It was presented the chronology, author and affiliation statistics in the bibliometric analysis section. The results show that research of education big data in China started since 2013 then showed the year-on-year growth trend since 2013, and reached its highest point in history in 2017 with the annual number 615. Xianing Yang tops the list with a total of eleven articles and Central China Normal University has the best performing among the contributing organizations. Furthermore, it is discovered that nine of the top 10 outstanding organizations are normal university, proving that normal universities play a totally important role in education big data research.

In the section of Knowledge mapping analysis, the high frequency keywords are obtained. As shown in table 3, the most frequently cited keywords were “education in ideology and politics”, “colleges and universities”, “innovation”, “undergraduate”, “university's ideological and political education” and “higher education” besides the subject keywords “big data” or “education”, illustrating that Chinese researchers pay more attention to the application and educational innovation with the help of the big data technology, especially the education in ideology and politics among undergraduates. It is particularly important to discover the major popular research trends and topics in the field of Chinese education big data through keywords bursting and clustering. Fig 2 shows the top nine keywords with the strongest citation bursts. “Electronic schoolbag”, “educational reform” and “MOOC” have the longest outbreak range, illustrating that the three topics are the primary popular research trends in last years. In addition, the topic of “data processing” is one of the typical frontiers of recent research, indicating that the processing technology of education big data can be the fashionable research trends in future. The top eight largest clusters are depicted in Fig 3. Big data era (#0) is the largest cluster, which includes 32 articles. And “big data era”, “information innovation” and “digital campus” were the high frequency keywords. Educational innovation (#1) is the second largest cluster, which includes 27 articles. And the high frequency keywords were “innovation”, “big data technology”, “education mode” and “educational innovation”. Big data (#2) is the third largest cluster, which contains 25 articles, and the high frequency keywords were “big data”, “ideological and political education”, “undergraduate” and “data mining”. It implies that the current study focuses on the education application and reform with the support of the big data technology.

Obviously, there are still some limitations in this research. The database used is only from the China National Knowledge Infrastructure (CKNI). As a matter of fact, there are generous excellent articles about education big data in abroad. Future research needs to further expand the scope of the search and to further explore the current status quo of the education big data research.

Acknowledgement
This research was supported by the Key Project of Humanities and Social Sciences in Anhui Province of China (Grant NO. SK2014A390 and SK2016A0882) and the Teaching research project of Anhui Province in China (Grant NO. 2017zhkt414 and 2017jyxm0445).

References
[1] Breur, T. (2016) Statistical power analysis and the contemporary “crisis” in social sciences. Journal of Marketing Analytics, 4:1-5.
[2] ShiJing, L., Rong, Z. (2018) A comparative analysis of research hot spots and trend on educational big data at home and abroad. China medical education technology, 32: 240-245.
[3] RenFeng, H., Yao, L. (2018) Analysis on the hotspots of educational big data research. Journal of Yanbian Institute of Education, 32: 58-60
[4] The State Council. (2017) Notice of the State Council on Printing and Distributing the 13th Five-Year Plan for the Development of National Education. http://www.moe.gov.cn/jyb_xwgh/moe_1777/moe_1778/201701/t20170119_295319.html
[5] XiaoJun, W. (2013) The future of education big data. Primary and secondary school information technology education, 2: 23.
[6] ShiChun, W., XinHua, X., HongChun, Z., JiaChen, H. (2018) Research on the influence of
individual differences on python crawler's access to educational big data. China Education Informationization, 9:79-81.
[7] SanNvYa, L., DongBo, Z., Hao, L., JianWen, S., Jie, Y. (2018) A study of map-based visual analysis method for big data in education. Audio-visual education research, 7:49-56.
[8] XiaoMin, L. (2017) Analysis of the current situation of the construction of educational big data in China and the initial establishment of the framework. China's electrification education, 8:128-131.
[9] XiaoFeng, Y. (2016) Education in the era of big data: prospect and action. Higher education research, 37:7-12.
[10] Ying, P. (2018) Problems and challenges in the study of education big data in China. Modern trade industry, 39:58-59.
[11] Chen, C. (2017) Expert Review. Science Mapping: A Systematic Review of the Literature. Journal of Data & Information Science, 2: 1-40.