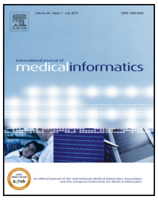
International Journal of Medical Informatics 98 (2017) 22–32

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.ijmijournal.com



Review article

Visualizing the knowledge structure and evolution of big data research in healthcare informatics

Dongxiao Gu[a], Jingjing Li[a], Xingguo Li[a,*], Changyong Liang[a,b]

[a] School of Management, Hefei University of Technology, 193 Tunxi Road, Hefei, Anhui 230009, China
[b] National Joint Engineering Research Center for Intelligent Decision and Information Systems, 193 Tunxi Road, Hefei, Anhui 230009, China

ARTICLE INFO

Article history:
Received 31 July 2016
Received in revised form 8 November 2016
Accepted 21 November 2016

Keywords:
Big data
Healthcare informatics
Bibliometrics
Knowledge structure
Knowledge management

ABSTRACT

Background: In recent years, the literature associated with healthcare big data has grown rapidly, but few studies have used bibliometrics and a visualization approach to conduct deep mining and reveal a panorama of the healthcare big data field.

Methods: To explore the foundational knowledge and research hotspots of big data research in the field of healthcare informatics, this study conducted a series of bibliometric analyses on the related literature, including papers' production trends in the field and the trend of each paper's co-author number, the distribution of core institutions and countries, the core literature distribution, the related information of prolific authors and innovation paths in the field, a keyword co-occurrence analysis, and research hotspots and trends for the future.

Results: By conducting a literature content analysis and structure analysis, we found the following: (a) In the early stage, researchers from the United States, the People's Republic of China, the United Kingdom, and Germany made the most contributions to the literature associated with healthcare big data research and the innovation path in this field. (b) The innovation path in healthcare big data consists of three stages: the disease early detection, diagnosis, treatment, and prognosis phase, the life and health promotion phase, and the nursing phase. (c) Research hotspots are mainly concentrated in three dimensions: the disease dimension (e.g., epidemiology, breast cancer, obesity, and diabetes), the technical dimension (e.g., data mining and machine learning), and the health service dimension (e.g., customized service and elderly nursing).

Conclusion: This study will provide scholars in the healthcare informatics community with panoramic knowledge of healthcare big data research, as well as research hotspots and future research directions.

1. Introduction

With the rapid development of emerging information technologies, including cloud computing, social networks, mobile commerce, and the Internet of Things, data are rapidly growing and the advent of big data has raised myriad opportunities and challenges in a variety of research fields. The term "big data" mainly refers to volumes of large, complex, linkable data [1], but in the existing attempts of explicit definitions for big data, there is still not an agreement in the scientific community. One of the earliest descriptions of big data was given by John R. Masey, chief data scientist at Silicon Graphics, Inc. in 1997[2]. In the early 2000′s, Laney presented a framework expressing the 3-dimensional increase in data volume, velocity, and variety. Although he did not mention big data explicitly in this work, the model, later nicknamed as "the 3V's", was used as its definition [3]. In 2003, big data (analytics) started to emerge when Google, Yahoo, and some other high technology firms started to use big data for business data analysis [4]. Big data became "bigger"– more volume, variety, and velocity – and organizations started focusing on the data-driven economy. Organizations of all types are now developing data-based offerings to remain competitive [5]. Big data is a term for any collection of data sets that is too large and complex for traditional database management systems and data processing applications to deal with [6,7]. Big data has five typical features, also called five V's: volume, value, velocity, variety, and veracity [8].

Big data research has been the subject of extensive attention in both academia and industry worldwide. Moreover, it has become a hot topic of discussion in both the industrial and the academic

* Corresponding author.
E-mail addresses: gudongxiao@hfut.edu.cn (D. Gu), 847840274@qq.com (J. Li), mikehfut0551@163.com (X. Li), cyliang@hfut.edu.cn (C. Liang).

http://dx.doi.org/10.1016/j.ijmedinf.2016.11.006
1386-5056/© 2016 Elsevier Ireland Ltd. All rights reserved.

research communities. First, *Nature Publishing Group* published a special journal, *Big Data,* in 2008 [9]. Then, *Science Publishing Group* followed by also publishing a special journal, *Dealing with Data*, in 2011 [10]. In March 2012, U.S. President Barack Obama declared that the United States government would invest $200 million to launch *The Big Data Research and Development Initiative* [11]. In practice, big data techniques and analyses have been successfully applied to many fields, including healthcare informatics, business analytics, Internet finance, social media user behavior analysis, Internet public opinion analysis, e-business, e-health, and manufacturing [12]. The information storm resulting from the advent of big data is not only changing people's lives, careers, and ways of thinking but also initiating great transformations [13].

More recently, big data research and development has received unprecedented attention in various fields. In the healthcare field, the development of the Internet of Things (IOT) and sensor networks has greatly propelled the growth of medical and healthcare big data. Many countries are advancing robustly in terms of medical informationization and in various emerging information technology applications in the healthcare field. Moreover, an increasing number of medical institutions are obtaining strategic and funding support to work on big data research and analysis, which significantly contributes to endeavors made by most countries [14]. The application of big data in medicine has a bright prospect, as well as a profound significance. Healthcare big data can make big impact and values [15]. Through big data analysis, certain diseases may be detected in their early stages, allowing patients to obtain more effective treatments [16]. Big data analysis and its results will also help manage special patients, promote public health, and detect social medical insurance fraud more quickly and efficiently [17]. In recent years, healthcare big data research has enjoyed rapid growth, and endless related research issues have emerged, including big data applications for medical service quality improvement [18,19], the outlook for healthcare big data development [20,21] as well as healthcare big data organization, modeling, and knowledge discovery.

With the rapid increasing of publications associated with big data, some scholars began to find and aggregate relevant existing literature and provide knowledge support for other researchers in the healthcare (medical) informatics[1] community. For example, White (2014) conducted a literature review related to the challenges and opportunities presented by medical care big data [22], and Luo (2016) completed a literature review on big data applications in biomedical research and health care [17]. However, both of these investigations are qualitative in nature and lack large-scale literature-based analysis. It is still difficult for readers to find core literature, journals, and researchers, and to understand the overall situation and hotspots of the healthcare big data field. There are also some literature reviews in the big data domain using systematic review and metadata analysis [23]. As a methodology commonly used in historical literature analysis, systematic review generally selects relevant primary studies and uses methods of synthesis [24,25]. Researchers generally use systematic review to find and aggregate relevant existing evidence about a specific research issue of interest, such as cardiovascular diseases using big-data by Hadoop, or Mhealth technologies for chronic diseases and elders [26,27], but such systematic review does not help readers understand the overall situation, hotspots, and prominent future directions of healthcare big data research. However, there are few studies of the overall developmental trend of the big data literature, particularly those reviewing articles based on a vast literature and bibliometrics, a methodology for the review of scientific literatures. Scientific and technical literatures have important advantages in analysing the general research structure and development of disciplines [28]. Many prominent scholars contribute their findings to the scientific literature.

Bibliometrics is a special type of quantitative analysis in knowledge fields that examines large amounts of scientific literature as its objects of analysis. It comprehensively uses mathematics, statistics, philology, and other professional knowledge and methods to analyze the research achievement distribution of a subject, subject development, and research trends, and finally intuitively displays results from the analysis by visualization. In contrast to a systematic review method, which focuses on a specific research issue and analyses only a part of the literature, bibliometrics generally obtains large amounts of literature data via literature retrieval under specific retrieval conditions. It performs data analysis based on almost all the retrieved articles and can alleviate incomplete analysis caused by insufficient knowledge and partial literature coverage. It generally also uses various literature analysis software programs for visualization analysis and presentation of results. For example, it uses various visual mapping methods and intuitively shows the overall knowledge structure, research framework, and development trends of a discipline. This is very helpful for researchers to rapidly comprehend the overall research status and hotspots.

In recent years, medicine has become one of the focuses in big data research. It would be of great significance to understand the knowledge structure, development trends, and research focuses of healthcare big data through a bibliometric analysis. This study conducted a comprehensive analysis on the healthcare big data research using bibliometrics. By analysing 2398 journal articles indexed in the Web of Science (WOS) database, we performed a bibliometric analysis on the core data collection ranging from 2003 to 2016, and identified the knowledge structure and study focus in healthcare big data research [29]. The results of our study will provide scholars in the healthcare informatics community with knowledge support, and help them quickly understand the research status, structure, and hotspots, as well as development trends of healthcare big data. It will also be helpful to promote the research, development, and applications of healthcare big data in the future.

The remainder of this paper is organized as follows. The next section introduces the methodology we used for this study, including data sources and research toolkits. Section 3 introduces the knowledge map of the time-and-space analysis process and results, including a time distribution map and a space distribution map of healthcare big data research outcomes. Section 4 presents the knowledge base analysis process and the results. Section5 presents the research focus of healthcare big data in detail, including keyword extraction and frequency counting, network construction, and the research focus analysis. The last section concludes the paper and suggests issues, challenges, and trends for future research.

2. Methodology

2.1. Data source

As a general matter, academic journals function as a channel for scholars to publish, spread, accumulate, comment on, and assume the lead in scientific research [30]. Compared with studies published in books or reports, articles published in journals tend to be more direct and consistent in addressing critical issues in a research field [31]. According to Bradford's law, most key studies are published in core international journals. We therefore collected data on articles published in core international journals between January 1st, 2003 and April 16th, 2016 from the Web of Science

[1] According to Wikipedia: "Health informatics (also called health care informatics, healthcare informatics, medical informatics, nursing informatics, clinical informatics, or biomedical informatics) is informatics in health care." (https://en.wikipedia.org/wiki/Health_informatics)

(SCI-EPANDED, CPCI-S, CCR-EXPANDED, and IC (Index Chemicus) included) for our literature review [32]. We selected the year of 2003 as the starting year for our study, because big data as analytics tools started emerging then when high technology firms began using big data in practice, leading to more and more big data research achievements and corresponding publications, according to Larson and Chang' research [4].

We performed a literature search on healthcare big data. The search strategy we used is *TS = (big data AND #̈) AND the literature type: (Article),* in which the symbol # represents such keywords as "health", "health care", "medical", "diseases" and seventeen other words associated with medicine and healthcare. We connected these twenty words with "OR". Two reviewers screened all titles and abstracts for potentially eligible studies. Full texts of all potentially eligible studies were then assessed by the same two reviewers. We eventually obtained 2398 journal articles.

2.2. Toolkits

The concept of mapping knowledge domains originated from a symposium organized by the National Academy of Sciences (NAS) of the United States in 2003. With the development of information visualization, an increasing number of tools were exploited to assist in mapping knowledge domains [33]. CiteSpace is a commonly used information visualization tool developed by Professor Chen Chaomei's team at Drexel University, USA [34]. The tool can calculate the number of articles in a specific field and explore the critical path of the field's evolution and knowledge turning points by relying on co-citation analysis theory and PFNET through a series of visualized maps. This study analyzed the subject evolution potential dynamic mechanism and explored subject development [35]. CiteSpace III big data processing has been undertaken to analyze the knowledge structure and basis of healthcare big data research, aiming to help researchers understand the knowledge structure in this field with the assistance of various knowledge mapping domains. To analyze and identify critical issues, we adopted SATI3.2 [36] to build a keyword co-occurrence matrix; and converted the data format with Ucinet 6.0 to finally achieve keyword co-occurrence mapping. Excel 2010 and Histcite were also used in the study.

3. Knowledge map of time-and-space analysis

3.1. Time distribution map of healthcare big data research

To evaluate the outcomes of big data research in the medicine/healthcare field, we collected journal articles published from 2003 to 2016 and tracked annual publishing trends and changes (shown in Fig. 1). These articles were among the first to combine computer technology with field research, accumulating a large database with a relatively good information technology basis [37,38]. In 2003, there were 53 articles on big data research in healthcare. Fig. 1 shows the growth trend between 2003 and 2016. Prior to 2013, the number of articles on healthcare big data rose consistently; however, from 2013 to 2015, this figure increased sharply and much more quickly than what is forecasted by the predicting index curve. Although the total number of articles published in 2016 is unknown yet, it is likely that the trend will continue. Overall, it is clear that healthcare big data research is developing vigorously.

The annual total number of authors of the articles is shown in Fig. 2. The trend of the number of authors is consistent with that of the number of articles. There were 237 authors in 2003; by 2015, the number of authors grew to 2711, an increase of 10 times the number 12 years earlier. This increase marked a rapid growth of authors.

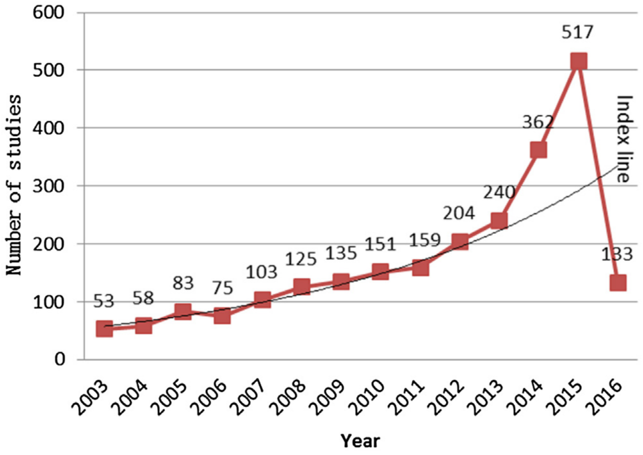


Fig. 1. Annual number of published articles on healthcare big data.

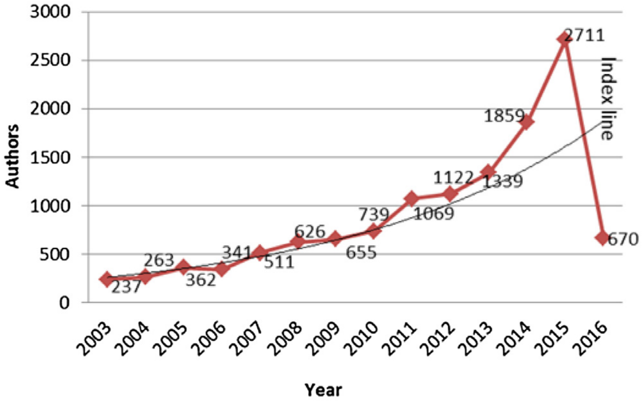


Fig. 2. Annual number of authors.

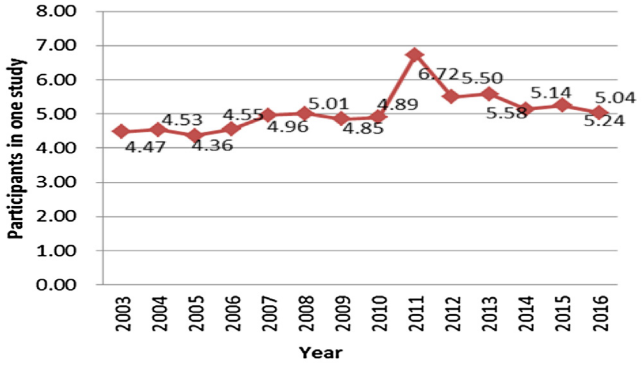


Fig. 3. Average number of co-authors per article.

To evaluate the input-output ratio of researchers in the field of healthcare big data, we summarized the average number of co-authors per article (shown in Fig. 3). In 2013, the average number of co-authors per article was 4.47. This number continued to rise, reaching a maximum of 6.72 in 2011, and then started to decrease, ending at 5.24 in 2015. Fig. 3 reveals a trend of collaboration among authors in the field of healthcare big data research. Collaboration, to some extent, ensures the quality of the published articles and reflects the emphasis on healthcare big data research in the big data era.

Table 1
Institutions and the number of articles published (20 or more).

Institution	No. of published articles	TLCS	TGCS
Harvard University	55	28	1711
University of CA Los Angeles	30	4	292
Stanford University	23	2	186
Chinese Academy of Sciences	22	0	116
Columbia University	22	1	445
Oxford University	22	2	246
University of CA San Francisco	21	7	165
University of Toronto	21	2	135
University College London	20	1	241
University of CA San Diego	20	15	174

3.2. Space distribution map of healthcare big data research outcomes

3.2.1. Institutional distribution

Table 1 lists the top ten academic groups and institutions in terms of number of published journal articles on healthcare big data. Harvard University ranks first, with 55 articles. The University of California, Los Angeles ranks second, with 30 articles, narrowly outnumbering Stanford University by seven more. The other eight institutions from multiple countries are comparable, which is a sign of international interest and emphasis on healthcare big data research.

There are two divisions of citation frequency in the Histcite system: LCS and GCS. LCS (Local Citation Score) refers to the citation frequency of an article in the local database, whereas GCS (Global Citation Score) refers to the citation frequency of an article in the Web of Science (WOS) database. TLCS and TGCS are used to represent the total citation frequency in terms of LCS and GCS, respectively [39].

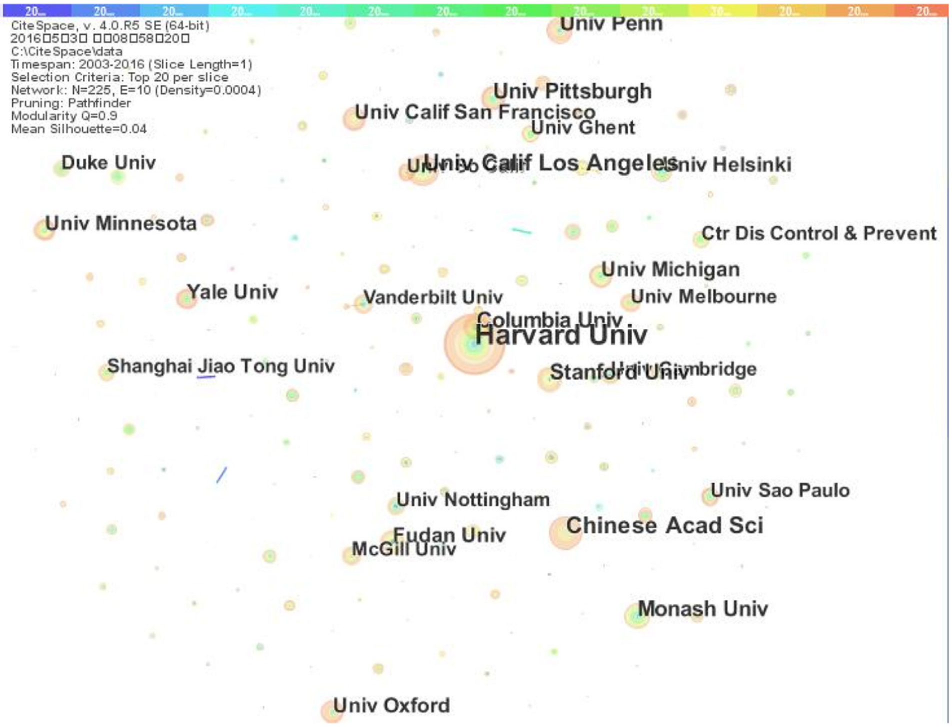
Research collaboration is an important means of enhancing overall research strength, allowing researchers to complement one another's advantages and share information [40]. The level of research collaboration is one of the indexes employed to evaluate the state of research in a specific field. To study institutional collaboration in healthcare big data research, we set relevant parameters with CiteSpace and drew an institutional collaboration network, as shown in Fig. 4. A node represents an institution, the size of the node represents the quantity of its total published articles, the circles represent years, the label font size represents centrality, and an edge represents institutional collaboration [41]. Fig. 4 shows that the number of nodes is 225, the number of edges is 10, and the density is 0.0004, indicating that there was little collaboration across the institutions.

3.2.2. National distribution

To study the national collaboration distribution in healthcare big data research, we set the relevant parameters using CiteSpace III and formed a national collaboration network (shown in Fig. 5). Altogether, there are 109 countries with 217 national collaboration edges. Based on the national collaboration network analysis, we extracted the relevant information for those countries with more than 40 published articles, as shown in Table 2. In terms of the total number of articles published, the USA tops the list with 662 articles, followed by China and the United Kingdom with 235 and 191 articles, respectively. In terms of centrality, the USA ranks first with a ratio of 0.44, the United Kingdom second with a ratio of 0.24, and Germany third with a ratio of 0.12; these countries have been marked with purple circles. Centrality can describe the importance of a node among other nodes. Judging from the index, the USA and the United Kingdom have comparative advantages in terms of the number of articles and important influence. Most countries began publishing in 2003, with the Netherlands (2004), Japan (2004), India (2005), and Turkey (2005) starting a little later.

4. Knowledge base analysis of healthcare big data research

A knowledge base consists of the previous research content and structure of a specific field. It is a means for better understanding



```
CiteSpace, v. 4.0.R5 SE (64-bit)
2016□5□3□ □□08□58□20□
C:\CiteSpace\data
Timespan: 2003-2016 (Slice Length=1)
Selection Criteria: Top 20 per slice
Network: N=225, E=10 (Density=0.0004)
Pruning: Pathfinder
Modularity Q=0.9
Mean Silhouette=0.04
```

Fig. 4. Institutional collaboration network.

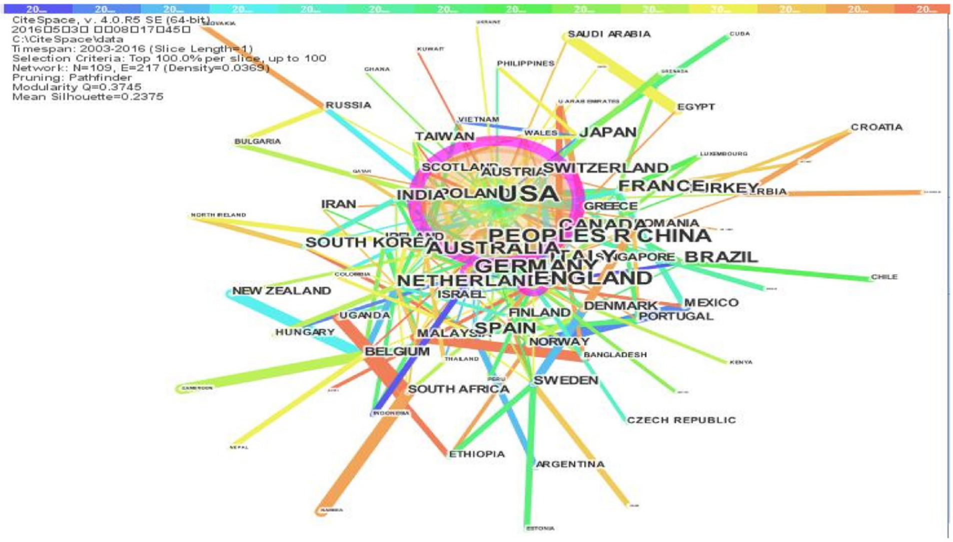


Fig. 5. National collaboration network.

Table 2
Countries with 40 or more published articles in healthcare big data research.

Nations	No. of published articles	Centrality	Year of first articles published
USA	662	0.44	2003
People's R China	235	0.03	2003
United Kingdom	191	0.24	2003
Germany	174	0.12	2003
Australia	102	0.05	2003
Italy	93	0.04	2003
Canada	84	0.07	2003
Spain	70	0.04	2003
Netherlands	63	0.08	2004
France	57	0.06	2003
Brazil	57	0.03	2003
India	53	0	2005
Switzerland	50	0.01	2003
Turkey	49	0.01	2005
South Korea	49	0.01	2003
Japan	49	0.03	2004
Poland	42	0.02	2003

the essence of a research front [42]. The research front refers to the development in a research field; therefore, the introduction of a research front constitutes the related knowledge base. Chen [43] redefined the knowledge base (intellectual base) of a research front as the citation trajectory of the research front in the literature.

To identify the knowledge base and its research innovation path in healthcare big data research, we conducted a literature co-citation analysis and drew the co-citation networks. Co-citation networks are knowledge networks formed in specific situations when two articles are cited by a third article or by several different articles concurrently [44]. Co-citation analysis focuses on the frequency of co-citation by other articles to describe the relationship between the articles; for example, the more likely that two articles are cited in a third article, the closer relationship they have. In other words, the two articles share a similar research background [45]. When articles, journals, or institutions are repeatedly cited by their peers, those co-cited studies would gradually be accepted by the scientific community and then develop into a scientific paradigm. By analyzing the literature co-citation network, this type of paradigm can be visualized [46]. According to Kuhn's theory of scientific paradigms, scientific paradigms refer to the common and important beliefs held in common by a specific discipline community [47]. A scientific paradigm consists of a series of concept systems and analytical methods that are commonly accepted and used in a specific subject area [48]. Therefore, co-citation networks could represent the knowledge base of a research field.

Fig. 6 shows the article co-citation network. Each node represents a cited article, the line between two nodes represents the co-citation relationship, and the thickness of the line represents the frequency of co-citation [45]. The figure shows that an article titled "The Inevitable Application of Big Data to Healthcare," which was published by Mourdoch, TB (2013) in the *JAMA-J AM MED ASSOC*, was cited 28 times and cross-connected with Gottesman (2013), Weber GM (2014), Groves P (2013), Altshuler DM (2012) and Blumentald (2010). The lines connecting it with Gottesman (2013) and Weber GM (2014) are thicker, indicating stronger co-citation relationships. The figure also shows a strong correlation and similar literature theme with articles published by Mourdoch, TB (2013), Gottesman (2013) and Weber GM (2014). An article published in *Nature* by Ginsberg J (2009), "Detecting influenza epidemics using search engine query data", was cited 18 times and shares a thicker line with Butler D (2013) and Lazer D (2014). This line indicates that this article has a topic similar to the topics of the articles published by BUTLER D (2013) and LAZER D (2014). An article published by LAZER D (2014) in SCIENCE was cited 15 times, which was similar to that published by COOK S (2011), CARNEIRO HA (2009) and GINSBERG J (2009). Judging from this overall pattern, the co-citation networks of medicine and healthcare research are distributed in a dispersed form because big data was newly introduced to the field of Medical Treatment and Public Health, and the co-citation relationship was not completely formed, lacking sufficient integration and development. Fig. 7 shows a co-citation time chart.

This study used intermediary centrality to represent important scientific studies. The important scientific studies on medicine and healthcare research in each period have been marked on the map, and they constitute the innovation path of this field. In addition, the Citation Bursts show that those scientific studies marked by red spots recorded the historical trace of the scientific research in this field. Before 2003, BAUM M (2002) published an article in LANCET, an authorized academic journal in the world medical community. Information regarding scientific literature with high intermediary centrality published after 2003 is shown in Table 3. Based on a detailed analysis of the innovation path of this field, we divide its development into three stages: the disease research stage, which focuses on studying cancer, obesity, and hypertension; the

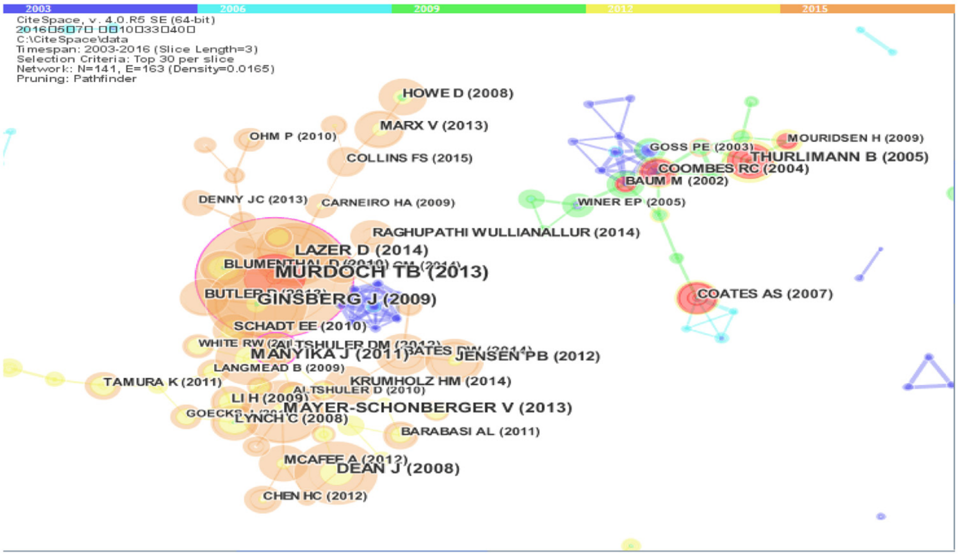


Fig. 6. Article co-citation network.

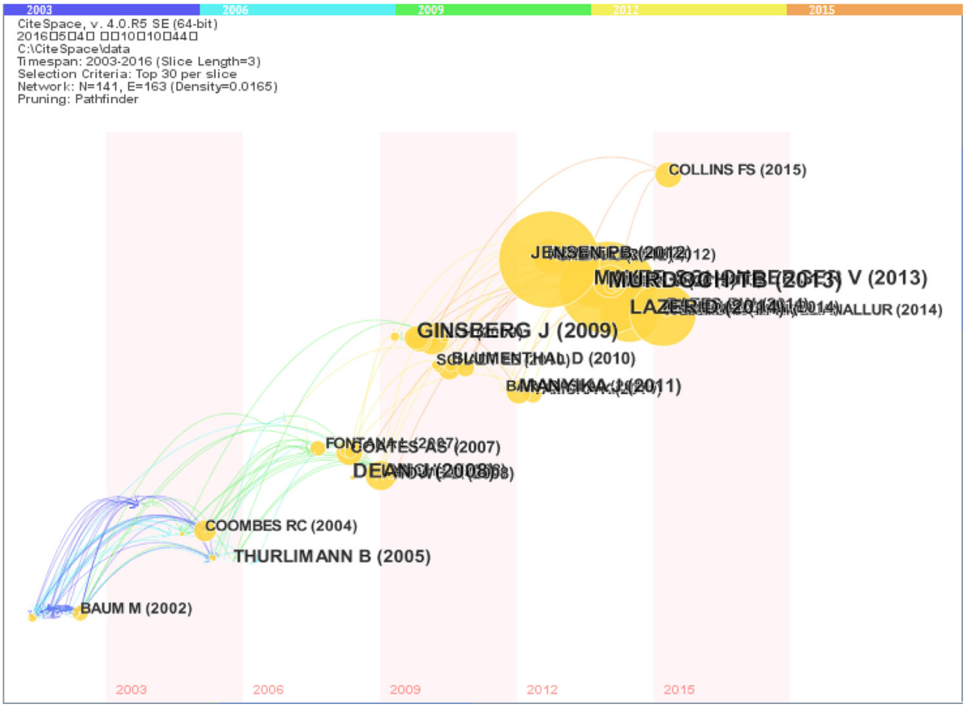


Fig. 7. Co-citation time chart.

Table 3
Information nodes with high intermediary centrality.

Year	Author	Study name	Journal
2004	COOMBES RC	A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer	NEW ENGL J MED
2008	BAHAR I	Comparison of Three Different Techniques of Corneal Transplantation for Keratoconus	AM J OPHTHALMIOL
2009	ARAIN MB	Determination of arsenic levels in lake water, sediment, and foodstuff from selected areas of Sindh, Pakistan: Estimation of daily dietary intake	FOOD CHEM TOXICOL
2011	ALTSHULER DM	An integrated map of genetic variation from 1092 human genomes	NATURE
2013	Mourdoch, TB	The Inevitable Application of Big Data to Healthcare	JAMA-J AM MED ASSOC
2015	COLLINS	A New Initiative on Precision Medicine	NEW ENGL J MED

Table 4
Nodes in the literature co-citation networks (Centrality > 0.06).

Authors of nodes	Articles	Year	Journal	Citation frequency	Centrality
Mourdoch, TB	The Inevitable Application of Big Data to Healthcare	2013	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	28	0.11
Altshuler, David M	An integrated map of genetic variation from 1092 human genomes	2012	NATURE	8	0.11
WEBER GM	Finding the Missing Link for Big Biomedical Data	2014	JAMA-J AM MED ASSOC	7	0.07

life and health research stage, studying food, hygiene, and human genomes; and the nursing research stage, focusing on personalized medicine and related medical information technologies. The three stages are shown in Table 3. The scientific literature constitutes the knowledge base of healthcare big data research, providing a path for researchers to study critical issues and to further explore the innovation path.

Table 4 provides information about nodes in the literature co-citation networks with centrality greater than 0.6. The articles published by TB Mourdoch, David Altshuler, and GM Weber all enjoyed a comparatively important position in healthcare big data research. TB Mourdoch, who worked at the University of Calgary, specialized in Transplantation, General Internal Medicine, Surgery, Gastroenterology, Hepatology, and Cell Biology. He excels at transplantation research. David M Altshuler worked at Harvard University via Boston Healthcare and System Massachusetts General Hospital; his specialties are Genetics Heredity, Endocrinology Metabolism, Science Technology (other Topics), Biochemistry, Molecular Biology and general Internal Medicine. He has published more than 140 articles on Genetics Heredity and 53 articles on Endocrinology Metabolism and was noted as a productive researcher, with 200 articles published overall. GM Weber, who worked at Harvard University, specialized in Medical Informatics, Computer Science, Information and Library Science, Healthcare Sciences Services, and Chemistry. He has published 11 articles in Medical Informatics. Based on all these researchers' consistent efforts, the knowledge base of healthcare big data research has been developed and completed by their successors.

In summary, the analysis of important researchers and important articles in the scientific literature has revealed and promoted the development of healthcare big data research, playing a vital role in developing related theories.

5. The healthcare big data research focus analysis

Research focus refers to the academic focus and highlights the accumulated factors of a discipline during a specific period, such as the number of academic articles, the emergence of academic tides and the growth of the research community in a specific research field or discipline [49]. Kuhn emphasized that the development of science is alternately propelled by normal science and scientific revolution. This statement indicates that the scientific revolution is an agent of change; the new and old paradigms are incommensurable [50]. With this incommensurability, the vocabulary systems for the new and old paradigms differ. Therefore, whether a scientific revolution would break out might be determined from the vocabulary changes that occurred during the period. By calculating the co-occurrence frequency of the keywords in scientific literature, we can predict the correlation between the keyword unit and the research focus of a specific academic field during the period. Therefore, co-term analysis of keywords can reveal the research structure and focus on a specific field. Callon [51] first proposed co-term analysis, which was later introduced to information science and spread widely in this field. Co-term analysis originated from the concept of bibliographic coupling and co-citation in bibliometrics, and refers to the situation in which two terms expressing a research theme or orientation in a specific field appear in one article, and it is thus certain that they are correlated with one another. The higher the likelihood that they appear together, the closer their relationship and distance [52]. Co-term analysis is more commonly used in bibliometrics than citation analysis or co-author analysis.

To detect the research focus in healthcare big data research, this study has adopted the co-term analysis approach, which contains three stages: keyword extraction, co-term matrix construction, and data analysis [53].

5.1. *Keyword extraction and frequency counting*

Keywords embody the soul of a scientific literature, summarizing researchers' major content, academic thoughts, and principal research methods. Keywords can reflect an article's research orientation and range, serving as an important index in scientific and quantitative studies [54]. To construct a reasonable keyword co-occurrence network, this study adopted SATI3.2 (Statistical Analysis Tool for Infometrics) to assess the gathered articles, to extract 1820 keywords, and then pre-process the selected keywords. We initially deleted unnecessary keywords: data, prevalence, depression, and analysis. We then grouped synonyms, for example, with health and healthcare in the same column and aging and elderly in the same column. Keywords with high frequency can fully reflect the research focus and trends in a specific field [55]. Therefore, this study extracted 56 keywords, more than 7 times the frequency of those in the research samples, as shown in Table 5. Certain high frequency keywords, such as epidemiology, breast cancer, and obesity, are revealed in the table, all of which are concerned with issues of high interest in healthcare big data research; data mining and machine learning also have become important technological issues. Moreover, researchers in the healthcare big data field are concerned with personality, health, and mortality. Because quality of life has been greatly improved, healthcare, quality of life, and the aging society are critical issues in this field.

5.2. *Network construction and the analysis of research focus*

Based on the 56-keyword samples selected in Table 4, this study employed SATI3.2 to draw a 56 × 56 keyword co-term matrix, as shown in Table 6. We used Ucinet 6.0 to convert the format of the co-term matrix, then visualized the matrix with Netdraw and formed the keyword co-occurrence network as shown in Fig. 8. In the chart, the nodes represent the keywords and the size of the nodes marks different betweennesses. The larger the nodes, the greater the betweenness of the keywords; likewise, the higher the central position the nodes have, the more likely it is that the keywords would become the focus in healthcare big data research. The lines connecting the nodes indicate the co-occurrence frequency of two different keywords. The thicker the lines, the more co-occurrence they have – and the closer the relationship between the keywords [56].

Fig. 8 shows that keywords such as big data, epidemiology, obesity, public health, and diabetes mellitus form larger nodes; they

Table 5
Keyword frequency (7 plus).

	keywords	frequency		keywords	frequency		keywords	frequency
1	big data	139	20	treatment	13	39	genetics	9
2	epidemiology	46	21	public health	13	40	endothelin-1	9
3	personality	32	22	genomics	12	41	privacy	9
4	breast cancer	30	23	prevention	12	42	mental health	9
5	data mining	30	24	nursing	12	43	stress	9
6	health	25	25	endothelin	12	44	primary care	8
7	mortality	24	26	incidence	11	45	smoking	8
8	obesity	23	27	meta-analysis	11	46	metabolic syndrome	8
9	cancer	21	28	risk factors	11	47	electronic health records	8
10	machine learning	20	29	adolescents	11	48	risk assessment	8
11	healthcare	20	30	disease	11	49	MapReduce	8
12	children	19	31	clinical trials	11	50	climate change	8
13	quality of life	19	32	stroke	10	51	monitoring	8
14	elderly	19	33	Alzheimer's disease	10	52	social media	8
15	diabetes mellitus	19	34	cloud computing	10	53	control	8
16	survival	18	35	hypertension	10	54	tuberculosis	8
17	physical activity	16	36	policy	10	55	medicine	8
18	bioinformatics	14	37	HIV	9	56	precision medicine	8
19	prognosis	14	38	personalized medicine	9	i i	i i	i i

Table 6
Selected keywords co-term matrix.

	big data	epidemiology	personality	breast cancer	data mining	health	mortality	...
big data	139	3	0	0	15	3	1	...
epidemiology	3	46	2	1	1	0	3	...
personality	0	2	32	0	0	1	1	...
breast cancer	0	1	0	30	0	0	1	...
data mining	15	1	0	0	30	0	0	...
health	3	0	1	0	0	25	1	...
mortality	1	3	1	1	0	1	24	...
obesity	0	1	0	0	0	1	1	...
cancer	0	1	1	0	1	0	1	...
machine learning	15	0	0	0	6	2	0	...
children	1	1	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...

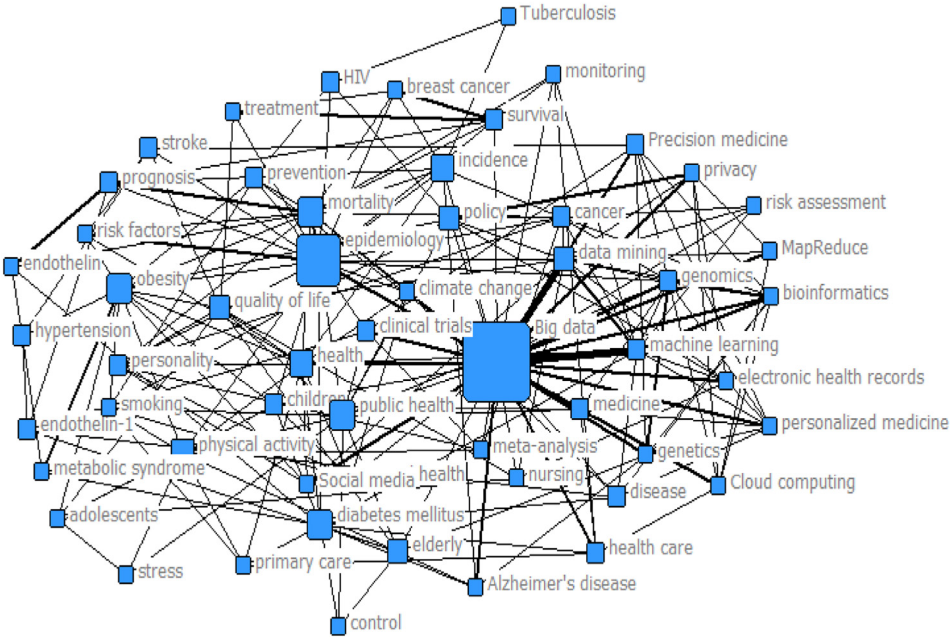


Fig. 8. Co-word network of keywords.

are the research focus of international healthcare big data research. Epidemics, a basic subject of public hygiene, has a long history dating back to ancient Greece. In recent years, changes in people's lifestyles, unhealthy diets, smoking, drinking, less physical activity, greater psychological pressures, environmental pollution, and an aging population [57] have led to an ever-increasing incidence of chronic epidemic diseases. With the increase and development of quality of life and medical technology, research on epidemics has expanded from being concerned with dangerous factors to being concerned with the entire life process, propelling the formation and development of life course epidemiology[58]. Traditional methods of data collection and analysis technology can no longer meet increasing needs; the introduction of big data processing into healthcare research has alleviated the situation. Obesity, public health, and diabetes are closely related to today's society and people's lifestyles in today's world, which are of course the research focus of healthcare big data research.

The big data nodes share thicker lines with data mining, machine learning, and medicine, indicating a tight relationship among them. The tight relationship between big data and data mining, machine learning, and medicine shows their internal relationships because the application of big data in medical hygiene should receive related technological support; however, data digging and machine learning provide the required support. With ever-increasing requirements for life quality and healthcare services, healthcare big data analysis is becoming increasingly important. Refined healthcare, personalized health services and improved endowment quality are critical issues in the academic communities of healthcare, information technology management, and decision sciences.

6. Concluding remarks, challenges, and future trends

6.1. Concluding remarks

In this study, we completed a bibliometric analysis of healthcare big data. The main work and findings of this study are as follows: (a) We found the trend of research output as well as the changes of co-author numbers in each research paper in the healthcare big data research field using a time distribution map analysis; (b) The core and productive institutions, as well as their distribution in the world, were found by agency collaboration network analysis and national collaboration network map analysis; (c) We found the core literature and core authors in the healthcare big data research field, the analysis of literature co-citation, and the innovation path of this research field by means of a literature co-citation time chart; and (d) We provided scholars in healthcare informatics with an overview of healthcare big data research and hotspots by a co-word network analysis of keywords. Our results will provide researchers with an important background for their future research in the broad healthcare big data field.

Generally, we examined the knowledge base and innovation path of healthcare big data research and analyzed critical issues, aiming to provide an important reference for future researchers to understand the overall development of this field. In terms of the time distribution map of research, the number of researchers working in healthcare big data research domain is substantial, with an average of five co-authors per article. This study found that collaborations among authors are numerous but that the cooperative community is dispersed and lacks significant collaboration. It is strongly recommended that authors from different institutions strengthen their collaboration with one another. Additionally, collaboration between institutions or organizations should also be enhanced to use resources more completely. In terms of national distribution, the USA outnumbered other countries by an overwhelming proportion. China should focus on improving its quality; despite a large number of published articles, the country has very low centrality. Throughout the world, the proportion of multi-national and multi-regional cooperative arrangements in healthcare big data research has increased consistently, indicating that such research has become much more highly proficient, comprehensive, and well-rounded [59].

In terms of knowledge base, this study examined the prominent researchers and core scientific literature in this field during the research period. These researchers and scientific outcomes made great contributions to the construction of the knowledge base in this field. This study categorized healthcare big data research into three stages: the disease research stage, the life-and-health research stage, and the nursing research stage. The research content in each stage is related to the social background and technology of that period.

In terms of research focus, healthcare big data research has a diversified focus, ranging from diseases such as epidemics, breast cancer, obesity, and diabetes, to technology such as data mining, machine learning, and healthcare, including personalization, aging, and death rates. These keywords reflected the major social concerns of global healthcare research. This study primarily analyzed core data from the Web of Science; although the data samples were not comprehensive, future study is planned that will focus on collecting more data from various databases to enable a more comprehensive analysis.

6.2. Challenges and future trends

As an emerging technology, big data is changing the lives of human beings, driving the changes in scientific research paradigms, and promoting the development of science. However, as an emerging discipline, big data still faces many issues in its development. As an important component of big data research, healthcare big data has certain typical characteristics: (a) large data size, complexity, multi-source heterogeneity that characterizes not only healthcare databases (such as electronic health records databases and health examination records databases) but also Internet databases (such as health consultation and service websites and mobile medical apps) in which large-scale data have been generated and reached the level of petabyte(PB) [16]; (b) rich data types, including both structured data, such as basic patient information, and unstructured data, such as medical images; (c) rapidly growing speed in data size; and (d) data quality problems, such as noise and incompleteness. The above features of healthcare big data pose several challenges regarding appropriate storage, transfer, analysis, visualization, and aggregation of data.

Firstly, high-throughput methods yield massive data and allow the fast, efficient, broad, and unbiased investigation of diseases on more detailed level, providing a global view on medical systems. Nevertheless, inferring causality from healthcare big data alone is often an unsolved issue. Due to the exponentially growing amount of data in the healthcare domain, as well as the diversity of data sources, healthcare big data analysis, integration, and development of models is today one of the bottlenecks in healthcare informatics research. One of the biggest challenges for big data is on how to develop effective data mining models and tools for fast finding the right knowledge from complex medical data, as well as supporting medical decision making.

Secondly, the criteria governing healthcare big data collections and use are not unified, which can lead to difficulties in data sharing and patients' privacy protection. Patient privacy and rights protection is a challenging issue in data collection and use. Data sanitation does not yield personal privacy. Since identifying data can still be revealed by intersecting multiple databases or complex statistical analysis, special care on privacy protection must be taken. There exist solutions such as differential privacy to the problem, but they are often too restrictive to be practically implemented.

Another challenge is associated with data storage technology. Although Hadoop and NoSQL ease database storage problems, data storage technology still needs further breakthroughs to meet the requirements of rapid storage and effective processing. In addition, data-sharing platforms are generally scattered, which might be inconvenient for researchers in acquiring data for experiments.

Cloud computing and service technology could be helpful to solve this problem.

In spite of the above-mentioned challenges, with the large volumes, healthcare big data have the tremendous advantage of being beneficial for lots of, very diverse research issues and are a great shared resource for the healthcare informatics research community. It is necessary to accelerate organizational collaboration among research agencies, hospitals, governments, and community health service agencies, as well as companies in medicine or health care domain. And scientists worldwide also should actively carry out international collaboration working together on core and frontier issues of healthcare big data.

Current and future trends for healthcare big data research and development include:

(a) Integrated analysis of heterogeneous health big data. An integrated analysis to support diagnoses, therapy, and health promotion based on heterogeneous data is becoming an emerging trend. The data for integrated analysis come from both online and offline databases related to healthcare such as electronic medical record, electronic health records data, the data of population, as well as healthcare on the Internet. One of important basis to realize the integrated analysis of heterogeneous health big data is develop an effective mechanism to assure patient-related data accessible by authorized organizations or individuals and preserve patient privacy the best way possible by legal means.

(b) The design of interpretation technique for medical big data mining. Designing an interpretation technique to address potentially highly nonlinear prediction functions for medical knowledge discovery is also an emerging research trend. Various data analysis tools are indispensable to the development of precision medicine. For some complex highly nonlinear prediction issues in medicine, it is very important to develop an interpretation technique to meet the requirement of the solution of objective problems.

(c) *Big data meeting public health and aging research*. One trend in the application domain is to collect aging-related “big data” from myriad fields (e.g., clinical medicine, Chinese traditional medicine and health science, psychology, nutrition sciences, and social science) in public health and aging research. Big data can improve people’s health by providing insights into the causes and outcomes of various diseases, better drug targets for precision medicine, as well as enhanced early detection, prediction and prevention of disease [1] [17]. Moreover, it can improve our understanding of various health behaviors (such as smoking, drinking, and sleeping late) and accelerate the knowledge-to-diffusion cycle. Accordingly, the big data-driven precision health administration and aging research (such as big data-driven elderly health assessment, personalized elderly health service recommendation, and elderly health promotion) are attracting research interest of an increasing number of researchers.

(d) *Cloud-based healthcare big data research*. Lots of biostatistics software packages have been used to handle large clinical datasets, which enabled the features of cloud computing [60–62]. Cloud computing provides users with shared processing resources as a service [4]. Big Data in the cloud offers excellent opportunities for the scientists in healthcare informatics communities to process and analyze different types of big data (medical images included) quickly [63]. Such related research topics as cloud-based medical big data service models, collaborative mechanism of service, security and intellectual property rights, and the corresponding systems (medical or healthcare apps included) design, applications, as well as value research, are becoming a new trend.

Summary points

- Literature research of healthcare big data possessed a diversified focus
- From 2003 to 2016, healthcare big data research has been vigorously developing worldwide.
- In terms of the total number of articles published, the USA tops the list with 662 articles, followed by China and the United Kingdom.
- In terms of centrality, the USA ranks first with a ratio of 0.44, and the United Kingdom and Germany rank second and third.
- Throughout the world, the proportion of multi-national, multi-regional cooperation in healthcare big data research has increased consistently.

Author contributions

Study conception and design: Prof. Gu, Prof. X. Li.

Manuscript writing and preparation: Prof. Gu, Mr. J. Li, Prof. Liang.

Data collection and analysis: Ms. J. Li, Prof. Liang.

Critical revision: Prof. Liang.

Gu and X. Li conceived and designed the study. J. Li and Liang performed the scale development, data collection and analysis. Gu and J. Li wrote the first full draft. Based on J. Li’s work, Gu and X. Li conducted further improvement and completed the revised version of the paper. Liang reviewed and edited the manuscript. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The data collection, analysis, and interpretation of this research were partially supported by the National Natural Science Foundation of China (NSFC) under grant Nos. 71331002, 71273125, 71301040, 71271072 and 71573071, and by the China Social Science Foundation (major project program) under grant No. 12&ZD221.

References

[1] M.J. Khoury, J.P. Ioannidis, Big data meets public health, Science 346 (6213) (2014) 1054–1055.

[2] J.R. Mashey, Big Data and the Next Wave of Infrastress. In Computer Science Division Seminar, University of California, Berkeley, 1997.

[3] A.D. Mauro, M. Greco, M. Grimaldi, What is big data? a consensual definition and a review of key research topics, AIP Conf. Proc. 1644 (1) (2015) 97–104.

[4] D. Larson, V. Chang, A review and future direction of agile, business intelligence, analytics and data science, Int. J. Inf. Manage. 36 (2016) 700–710.

[5] T.H. Davenport, Big data at work: dispelling the myths, Uncovering the Opportunities Big Data Bootcamp (2014) 49–62.

[6] M. Ebrahimi, A. Mohan, A. Kashlev, S. Lu, BDAP: a big data placement strategy for cloud-Based scientific workflows, in: IEEE First International Conference on Big Data Computing Service and Applications, IEEE Computer Society, 2015, pp. 105–114.

[7] E. Baro, S. Degoul, R. Beuscart, E. Chazard, Toward a literature-driven definition of big data in healthcare, BioMed Res. Int. 15 (4) (2015) 1–9.

[8] M.H. Kuo, T. Sahama, A.W. Kushniruk, E.M. Borycki, D. Grunwell, Health big data analytics: current perspectives, challenges and potential solutions, Int. J. Big Data Intell. 2 (2014) 114–126.

[9] C. Lynch, Big data: how do your data grow? Nature 455 (2008) 28–29.

[10] H.C. Kum, S. Ahalt, T.M. Carsey, Dealing with data: governments’ records, Science 332 (2011) 1263.

[11] G. Li, Research status and scientific thinking of big data, Bull. Chin. Acad. Sci. 27 (6) (2012) 647–657.

[12] G.H. Kim, S. Trimi, J.H. Chung, Big data applications in the government sector: a comparative analysis among leading countries, Commun. ACM 57 (3) (2014) 78–85.

[13] B.C. Medeiros, S. Satram-Hoang, D. Hurst, K.Q. Hoang, F. Momin, C. Reyes, Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States, Ann. Hematol. 94 (7) (2015) 1127–1138.

[14] Y. Zhang, S.L. Guo, L.N. Han, et al., Application and exploration of big data mining in clinical medicine, Chin. Med. J. (Engl.) 129 (6) (2015) 731–738.

[15] S.F. Wamba, S. Akter, A. Edwards, G. Chopin, D. Gnanzou, How 'big data' can make big impact: findings from a systematic review and a longitudinal case study, Int. J. Prod. Econ. 165 (2015) 234–246.

[16] W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Health Inf. Sci. Syst. 2 (1) (2014) 1–10.

[17] J. Luo, M. Wu, D. Gopukumar, Y. Zhao, Big data application in biomedical research and health care: a literature review, Biomed. Inf. Insights 8 (2016) 1–10.

[18] D.W. Bates, S. Saria, L. Ohnomachado, A. Shah, G. Escobar, Big data in health care: using analytics to identify and manage high-risk and high-cost patients, Health Aff. (Millwood) 33 (7) (2014) 1123–1131.

[19] M.J. Khoury, J.P.A. Ioannidis, Big data meets public health, N. Z. Med. J. 93 (676) (2014) 1054–1055.

[20] J. Roski, G.W. Bolinn, T.A. Andrews, Creating value in health care through big data: opportunities and policy implications, Health Aff. (Millwood) 33 (7) (2014) 1115–1122.

[21] M. Craven, C.D. Page, Big data in healthcare: opportunities and challenges, Big Data 3 (4) (2015) 209–210.

[22] S.E. White, A review of big data in healthcare: challenges and opportunities, Open Access Bioinform. (6) (2014) 13–18.

[23] P. Gaitanou, E. Garoufallou, P. Balatsoukas, The Effectiveness of Big Data in Health Care: A Systematic Review. Metadata and Semantics Research, Springer International Publishing, 2014.

[24] K.R. Felizardo, S.G. Macdonell, E. Mendes, J.C. Maldonado, A systematic mapping on the use of visual data mining to support the conduct of systematic literature reviews, J. Softw. 7 (2) (2012) 450–461.

[25] M.P. Gagnon, M. Desmartis, M. Labrecque, et al., Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals, J. Med. Syst. 36 (1) (2012) 241–277.

[26] S. Thakur, M. Ramzan, A systematic review on cardiovascular diseases using big-data by Hadoop, in: International Conference – Cloud System and Big Data Engineering, IEEE, 2016.

[27] G. Chiarini, P. Ray, S. Akter, C. Masella, Mhealth technologies for chronic diseases and elders: a systematic review, IEEE J. Sel. Areas Commun. 31 (9) (2013) 6–18.

[28] D.J. Good, C.M. McIntyre, Use of journal clubs within senior capstone courses: analysis of perceived gains in reviewing scientific literature, J. Nutr. Educ. Behav. 47 (5) (2015) 477–479 (e1).

[29] A. Balaid, M.Z.A. Rozan, S.N. Hikmi, J. Memon, Knowledge maps: a systematic literature review and directions for future research, Int. J. Inf. Manage. 36 (3) (2016) 451–475.

[30] H.J. No, Y. An, Y. Park, A structured approach to explore knowledge flows through technology-based business methods by integrating patent citation analysis and text mining, Technol. Forecast. Soc. Change 97 (2014) 181–192.

[31] L. Pan, S. Wang, A bibliometrics analysis on Chinese education research hotspots based on literature keywords co-occurrence knowledge map, Educ. Res. Exp. 6 (2011) 20–24.

[32] K.S. Condic, Citation analysis of student dissertations and faculty publications in reading and educational leadership at oakland university, J. Acad. Librariansh. 41 (5) (2015) 548–557.

[33] X. Zhang, Y. Gao, X. Yan, P.O.D. Pablos, Y. Sun, X. Cao, From e-learning to social-learning: mapping development of studies on social media-supported knowledge management, Comput. Hum. Behav. 51 (2015) 803–811.

[34] Z. Lin, C. Wu, W. Hong, Visualization analysis of ecological assets/values research by knowledge mapping, Acta Ecol. Sin. 35 (5) (2015) 142–154.

[35] Q. Xu, W. Zhang, L. Hu, J. Wang, J. Jin, The development and research of bioinformatics in neuroscience, Aasri Procedia 1 (3) (2012) 359–364.

[36] Q. Liu, A study on mining bibliographic records by designed software sati: case study on library and information science, J. Inf. Resour. Manage. 2 (1) (2012) 50–58.

[37] F.M. Afendi, N. Ono, Y. Nakamura, K. Nakamura, L.K. Darusman, N. Kibinge, et al., Data mining methods for omics and knowledge of crude medicinal plants toward big data biology, Computat. Struct. Biotechnol. J. 4 (5) (2013) 1–14.

[38] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, et al., Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey & Company, New York, 2011.

[39] L. Bornmann, W. Marx, Histcite analysis of papers constituting the h, index research front, J. Informetr. 6 (2) (2012) 285–288.

[40] A. Ebadi, A. Schiffauerova, How to become an important player in scientific collaboration networks? J. Informetr. 9 (4) (2015) 809–825.

[41] Z. Lin, C. Wu, W. Hong, Visualization analysis of ecological assets/values research by knowledge mapping, Acta Ecol. Sin. 35 (5) (2015) 142–154.

[42] H. Choe, D.H. Lee, H.D. Kim, I.W. Seo, Structural properties and inter-organizational knowledge, flows of patent citation network: the case of organic solar cells, Renew. Sustain. Energy Rev. 55 (2016) 361–370.

[43] C.M. Chen, Citespace ii: detecting and visualizing emerging trends and transient patterns in scientific literature, J. Am. Soc. Inf. Sci. Technol. 57 (3) (2009) 359–377.

[44] H. Small, Co-citation in the scientific literature: a new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. 24 (24) (1973) 265–269.

[45] M. Navonil, B. Nik, J.E.T. Simon, S. Stelios, Exploring the e-science knowledge base through co-citation analysis ☆, Procedia Comput. Sci. 19 (2013) 586–593.

[46] C.H. Hsiao, C. Yang, The intellectual development of the technology acceptance model: a co-citation analysis, Int. J. Inf. Manage. 31 (2) (2011) 128–136.

[47] K. Backhaus, L. Kai, M. Koch, The structure and evolution of business-to-business marketing: a citation and co-citation analysis, Ind. Mark. Manage. 40 (6) (2011) 940–951.

[48] R.M. Chang, R.J. Kauffman, Y.O. Kwon, Understanding the paradigm shift to computational social science in the presence of big data, Decis. Support Syst. 63 (3) (2014) 67–80.

[49] H. Özçınar, Mapping teacher education domain: a document co-citation analysis from 1992 to 2012, Teach. Teach. Educ. 47 (2015) 42–61.

[50] H. Sankey, Scientific realism and the semantic incommensurability thesis, Stud. Hist. Philos. Sci. Part A 40 (2) (2009) 196–202.

[51] M. Callon, J.J.P. Courtial, W.A. Turner, S. Bauin, From translations to problematic networks—an introduction to co-word analysis. soc sci inf sur les sci soc, Soc. Sci. Inf. 22 (2) (1983) 191–235.

[52] U. Sandouk, K. Chen, Learning contextualized semantics from co-occurring terms via a siamese architecture, Neural Netw. 76 (C) (2015) 65–96.

[53] C.C. Wu, H.J. Leu, Examining the trends of technological, development in hydrogen energy using patent co-word map analysis, Int. J. Hydrogen Energy 39 (33) (2014) 19262–19269.

[54] H.J. Li, H.Z. An, Y. Wang, J.C. Huang, X.Y. Gao, Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: based on two-mode affiliation network, Phys. A Stat. Mech. Appl. 450 (2016) 657–669.

[55] J. Huenteler, J. Ossenbrink, T.S. Schmidt, V.H. Hoffmann, How a product's design hierarchy shapes, the evolution of technological knowledge-evidence from patent-citation networks in wind power, Soc. Sci. Electron. Publ. 45 (6) (2014) 1195–1217.

[56] Z. Xie, Z.Z. Ouyang, Q.M. Liu, J.P. Li, A geometric graph model for citation networks of exponentially growing scientific papers, Phys. A-Stat. Mechan. Appl. 456 (2016) 167–175.

[57] L.S. Katie, B. Thomas, P. Andrew, Bordetella pertussis epidemiology and evolution in the light of pertussis resurgence, Infect. Genet. Evol. 40 (2016) 136–143.

[58] J. Lynch, G.D. Smith, A life course approach to chronic disease epidemiology, Annu. Rev. Public Health 26 (2005) 1–35.

[59] S. Krätke, Regional knowledge networks: a network analysis approach to the interlinking of knowledge resources, Eur. Urban Reg. Stud. 17 (1) (2010) 83–97.

[60] N.V. Chawla, D.A. Davis, Bringing big data to personalized healthcare: a patient-centered framework, J. Gen. Intern. Med. 28 (3) (2013) 660–665.

[61] E.E. Schadt, M.D. Linderman, J. Sorenson, L. Lee, G.P. Nolan, Cloud and heterogeneous computing solutions exist today for the emerging big data problems in biology, Nat. Rev. Genet. 12 (3) (2011) 224.

[62] A. O'Driscoll, J. Daugelaite, R.D. Sleator, 'Big data', Hadoop and cloud computing in genomics, J. Biomed. Inform. 46 (5) (2013) 774–781.

[63] S.S. Sahoo, C. Jayapandian, G. Garg, F. Kaffashi, S. Chung, A. Bozorgi, et al., Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data' using cloud computing for epilepsy clinical research, J. Am. Med. Inform. Assoc. 21 (2) (2014) 263–271.