# Visualization analysis of health informatics research from 2001 to 2018

*Lu-Xi Zou and Ling Sun\**

*To analyse the scientific outputs of health informatics research and explore its frontiers and hotspots, we retrieve publications on health informatics research from the Web of Science Core Collection between 2001 and 2018. Geographic information system mapping and visualization techniques were used for analysing the countries, journals, institutions, authors, references and keywords of publications. Timelines of co-citation clusters and keywords with citation bursts were created to study the emerging trends and hotspots. Six thousand seventy-seven publications were identified in health informatics research; the annual number of publications has increased rapidly over the last two decades. Keywords analysis indicated that 'big data' for health informatics and clinical informatics could become a research frontier. This study reveals the hot spots, structure and trend of health informatics research. Nevertheless, further research and cooperation are needed.*

**Keywords:** Geographic information systems, health informatics, information networks, public health, visualization analysis.

HARNESSING the power of information technology is claiming unprecedented attention worldwide to improve health status, as well as deal with the challenge of global ageing, changes in the environment, economic and policy. Health informatics is developed to collect and analyse health information, improve the overall quality of healthcare in a cost-efficient manner and promote healthcare reforms[1]. It is a novel interdisciplinary field that involves information science, information technology, biological and medical knowledge for healthcare[2–4].

Articles on health informatics have been published in many academic journals, with an increasing tendency. The developments in health informatics are disparate in different regions all over the world, while few studies have been systematically conducted to analyse the current status of health informatics management and its scientific outputs. The present study aims to display the global visualizing information and identify the emerging trends and hotspots in scientific outputs of health informatics research using geographical information system (GIS) mapping and visualization tools.

## Materials and methods

We downloaded data on the publications from the Web of Science Core Collection (WoSCC) and set the period as 'January 2001 to December 2018'. On-line retrieval was performed using a keyword 'health informatics'. The present authors independently screened all titles and abstracts to select qualified articles based on the inclusion criteria (Figure 1), and only original and review articles were included in the study. We downloaded the following information for each publication: titles, authors, keywords, institutions and cited references.

GIS performs task such as the input, storage, management, maintenance, retrieval, synthesis, analysis and output of location-based or geographic information[5–7]. VOSviewerplays an important role in the construction and visualization of scientific landscapes, and contributes to easy-to-interpret analyses of co-citation and co-authorship[8]. Cite Space visualizes emerging trends and abrupt changes in particular fields within a specified period[9]. The spatial distributions of health informatics-related publications were displayed by ArcGIS version 10.3 (ESRI; Redlands, USA). The institution co-authorship, journal co-citation, authors co-citation as well as reference co-citation were displayed using VOSviewer 1.6.11. The timeline views for co-citation clusters and keywords were made using CiteSpace 5.3.

## Results

### Annual trends of publications

In total, 6077 publications were identified on health informatics research from 2001 to 2018. The number of publications has increased annually since 2001 (Figure 2),
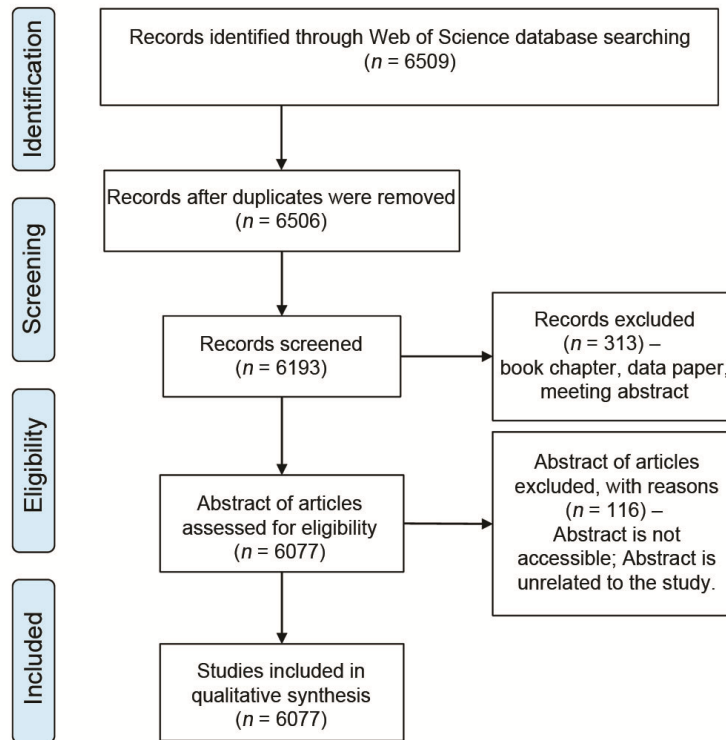
Lu-Xi Zou is in the School of Management, Zhejiang University, Hangzhou, Zhejiang, China and Ling Sun is in the Department of Nephrology, Xuzhou Central Hospital, Medical College of Southeast University, Xuzhou, Jiangsu, China.
*For correspondence. (e-mail: slpku@163.com)

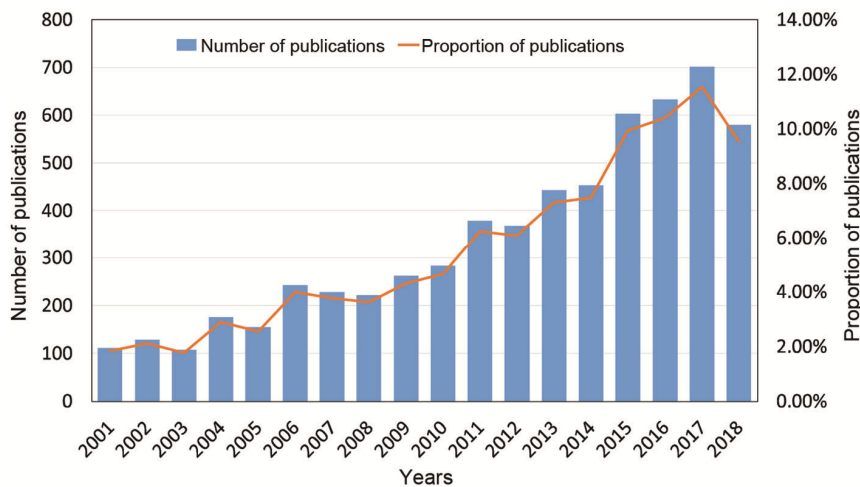**Figure 1.** Flow chart of health informatics research inclusion.



**Figure 2.** Annual trends in the number of academic health informatics-related publications from 2001 to 2018.

indicating that more researchers have focused on health informatics, especially after 2008.

*GIS mapping analysis*

These 6077 articles were from 115 countries or regions, of which the top 10 countries contributed to 4941 articles, accounting for 89.49% of the total number. USA published the largest number of articles (2716), followed by England (440), Canada (397), Australia (337) and Germany (297; Table 1).

Research projects need collaborations. Co-authorship analysis can evaluate the research status in a specific-field[10]; therefore, country co-authorship analysis reveals the influence as well as cooperation between the countries in this field. The countries were divided into developed and developing subgroups based on 'Standard Country or area codes for statistical use', published by the United Nations[11]. Figure 3 shows the network of country co-authorship in 2018. Blue node represents a paper and the links between the nodes reveal the cooperative relationships among countries or institutions, thicker

**Table 1.** Ranking of countries and institutions that published health informatics-related articles from 2001 to 2018

| Ranking | Country | Frequency | Institution (Country) | Frequency |
|---------|---------|-----------|-----------------------|-----------|
| 1 | USA | 2716 | Harvard University (USA) | 170 |
| 2 | England | 440 | University of Washington (USA) | 144 |
| 3 | Canada | 397 | Columbia University (USA) | 138 |
| 4 | Australia | 337 | Vanderbilt University (USA) | 126 |
| 5 | Germany | 297 | The University of Utah (USA) | 105 |
| 6 | Austria | 177 | University of Victoria (Canada) | 92 |
| 7 | Spain | 151 | University of Michigan (USA) | 90 |
| 8 | The Netherlands | 148 | Indiana University (USA) | 83 |
| 9 | Peoples' Republic of China | 142 | University of Pittsburgh (USA) | 82 |
| 10 | Sweden | 136 | Stanford University (USA) | 80 |
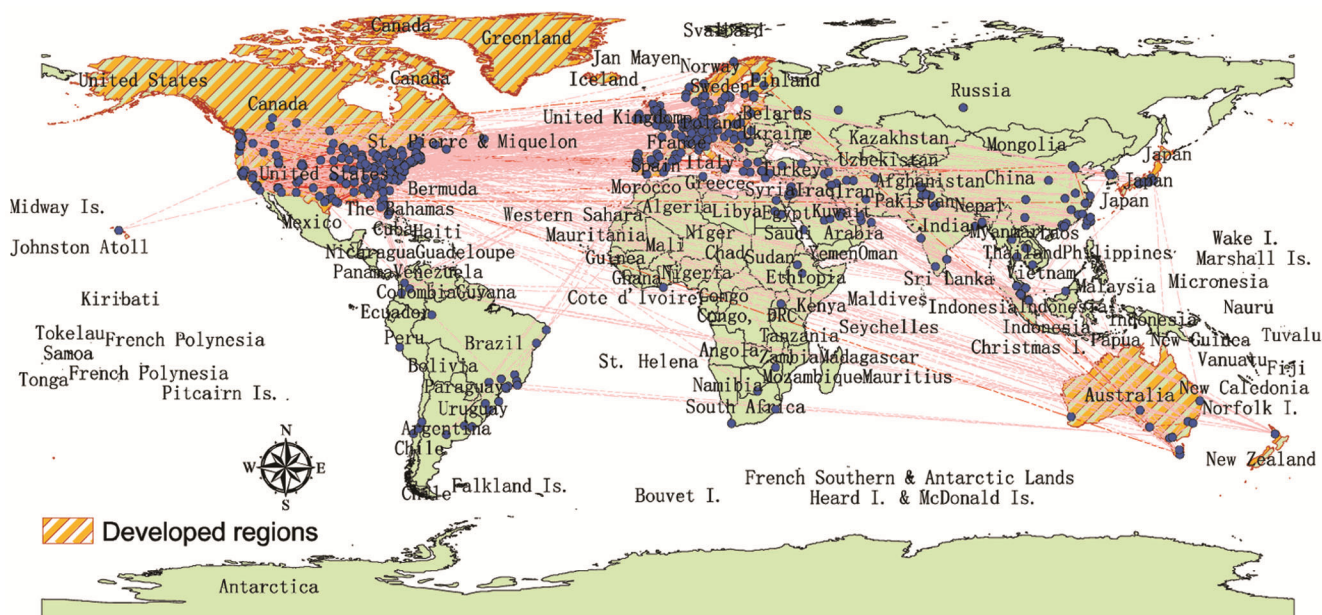


**Figure 3.** GIS mapping of health informatics-related publications; the country co-authorship network in 2018.

the link, stronger is the cooperation. Figure 3 shows that authors from developed countries published a total of 5893 articles, far outnumbering that published by developing countries (1123 articles). The top three nodes were USA, England and Canada, with total link strength values of 872, 490 and 311 respectively. In the field of health informatics research, USA was not only a research centre but also in close cooperation with many other countries, such as the developed countries in Europe and Oceania. This indicates that the research status of health informatics is unbalanced between developed and developing countries, in part due to the disparity of economic and technological levels between them. Furthermore, closer cooperation could result in more achievements that are advanced in scientific research.

*Network visualization analysis*

*Journal analysis:* More than 1000 academic journals contributed to these 6077 articles in health informatics

research. Table 2 presents the top 10 active journals. The highest number (275 articles, 4.98%) of articles was published by the *International Journal of Medical Informatics* (impact factor (IF) 2017, 2.957), followed by *Journal of the American Medical Informatics Association* (IF 2017, 4.27; 238 articles, 4.31%), *Methods of Information in Medicine* (IF 2017, 1.531; 174 articles, 3.15%), *Journal of Medical Internet Research* (IF 2017, 4.671; 139 articles, 2.52%) and *Journal of Biomedical Informatics* (IF 2017, 2.828; 129 articles, 2.34%).

IF of academic journals represents the average ratio of the number of citations to articles in a journal each year, and is used to measure the influence of the journal in its field. Among the top 10 active journals, 4 achieved 5-year IF > 3.0, with 781 articles accounting for 14.15% of all publications related to health informatics.

We used VOSviewer to draw the network of journal co-citation, wherein two items (say, authors, journals, references) are cited in one reference, which reflects the frequency and relationship between two co-cited items[12].

**Table 2.** Top 10 active journals that published health informatics-related articles from 2001 to 2018

| Ranking | Journal | Country | N1(%) | N2 | 2017 IF | Five-year IF |
|---|---|---|---|---|---|---|
| 1 | *International Journal of Medical Informatics* | Ireland | 275 (4.98%) | 7586 | 2.957 | 3.378 |
| 2 | *Journal of the American Medical Informatics Association* | England | 238 (4.31%) | 7318 | 4.27 | 4.337 |
| 3 | *Methods of Information in Medicine* | Germany | 174 (3.15%) | 2426 | 1.531 | 1.583 |
| 4 | *Journal of Medical Internet Research* | Canada | 139 (2.52%) | 3807 | 4.671 | 5.735 |
| 5 | *Journal of Biomedical Informatics* | USA | 129 (2.34%) | 2756 | 2.882 | 3.12 |
| 6 | *Applied Clinical Informatics* | Germany | 80 (1.45%) | 302 | 1.138 | 1.496 |
| 7 | *CIN-Computers Informatics Nursing* | USA | 78 (1.41%) | 754 | 1.056 | 1.13 |
| 8 | *BMC Medical Informatics and Decision Making* | England | 70 (1.27%) | 789 | 2.134 | 2.587 |
| 9 | *Journal of Medical Systems* | USA | 58 (1.05%) | 485 | 2.098 | 2.398 |
| 10 | *Journal of Public Health Management and Practice* | USA | 41 (0.74%) | 193 | 1.369 | 1.177 |

N1(%), Total number of health informatics-related articles in a journal from 2001 to 2018 (ratio of the total number of health informatics-related articles in a journal to the total number of articles from 2001 to 2018); N2, Number of citations of articles in a journal from 2001 to 2018; 2017 IF, Journal Impact Factor in 2017; Five-year IF, Five-year journal impact factor.

Journal co-citation can study the overall characteristics and structure of the academic journals network[13,14]. Figure 4 *a* shows the journal co-citation network with 165 nodes. The threshold was such that the minimum number of citations was more than 100. The size of nodes presents the number of article citations published in a journal, and reflects the activity of the journal. The distance between nodes represents their link strength; and shorter distance indicates higher citation frequency and stronger association.

The clustering technique is a parameterized and weighted variant of modularity-based clustering[15]. The VOSviewer classifies the nodes in one network to several clusters. A cluster is a set of closely related nodes, which is assigned by maximizing the function

$$V(c_1,...,c_n) = \sum_{i<j} \delta(c_i, c_j)(s_{ij} - \gamma),$$

where $c_i$ denotes the cluster to which node $i$ is assigned; $\delta(c_i, c_j)$ denotes a function that equals 1 if $c_i = c_j$, otherwise it is 0; $\gamma$ represents a resolution parameter that determines the clustering level, higher the $\gamma$ value, larger is the cluster[8]. Nodes of the same colour represent a cluster.

Here, the journals were divided into eight clusters. The red cluster contained *Journal of the American Medical Association* (2735 citations) and *British Medical Journal* (2091 citations), representing journals in general medicine. The green cluster contained *International Journal of Medical Internet Research* (1697 citations) and *Journal of Medical Informatics* (3707 citations), representing journals in information system and medical informatics. The yellow cluster contained *American Journal of Epidemiology* (237 citations) and *Lancet* (792 citations), representing journals in epidemiology and general medicine. The purple cluster contained *Nature* (591 citations) and *Science* (618 citations), representing journals in science and technology. The blue cluster contained *Journal of the American Medical Informatics Association* (7517 citations), representing journals in medicine, computer and interdisciplinary application. The baby-blue cluster contained *Journal of Digital Imaging* (277 citations), representing journals in medical imaging, nuclear medicine and radiology. The orange cluster contained *Journal of Methods of Information in Medicine* (2780 citations), representing journals in health care and service. The brown cluster contained *Journal of circulation* (373 citations), representing journals in the area of medical specialty. The top 10 active journals in Table 2 are also displayed in Figure 4 *a*; higher the IF, bigger is the node, indicating that publications in journals with higher IF received more citations and had a greater influence in health informatics.

*Institution analysis:* More than 1000 institutions had publications in health informatics research. The top 10 institutions published 1110 articles, of which Harvard University ranked first, followed by the University of Washington and Columbia University (Table 1).

Institution co-authorship shows the influence as well as communications between institutions in the research field[16,17]. The institutions were ranked based on the citation number of their publications. Figure 4 *b* shows the institution co-authorship network with the top 100 influential institutions. Harvard University, Columbia University, and Brigham and Women's Hospital were the top three influential institutions, with top three number of citations of publications in health informatics.

*Author co-citation analysis:* More than 1000 authors contributed to these 6077 articles. These authors were ranked according to the citation number of their published articles. The co-citation cited author network shows the top 100 authors (Figure 4 *c*). The largest node was found for Eysenbach, G. (419 citations), indicating his essential contribution to health informatics. The size of the circles represents their citation number; larger the circle, more are the citations. A shorter distance suggests a stronger and higher co-citation relationship
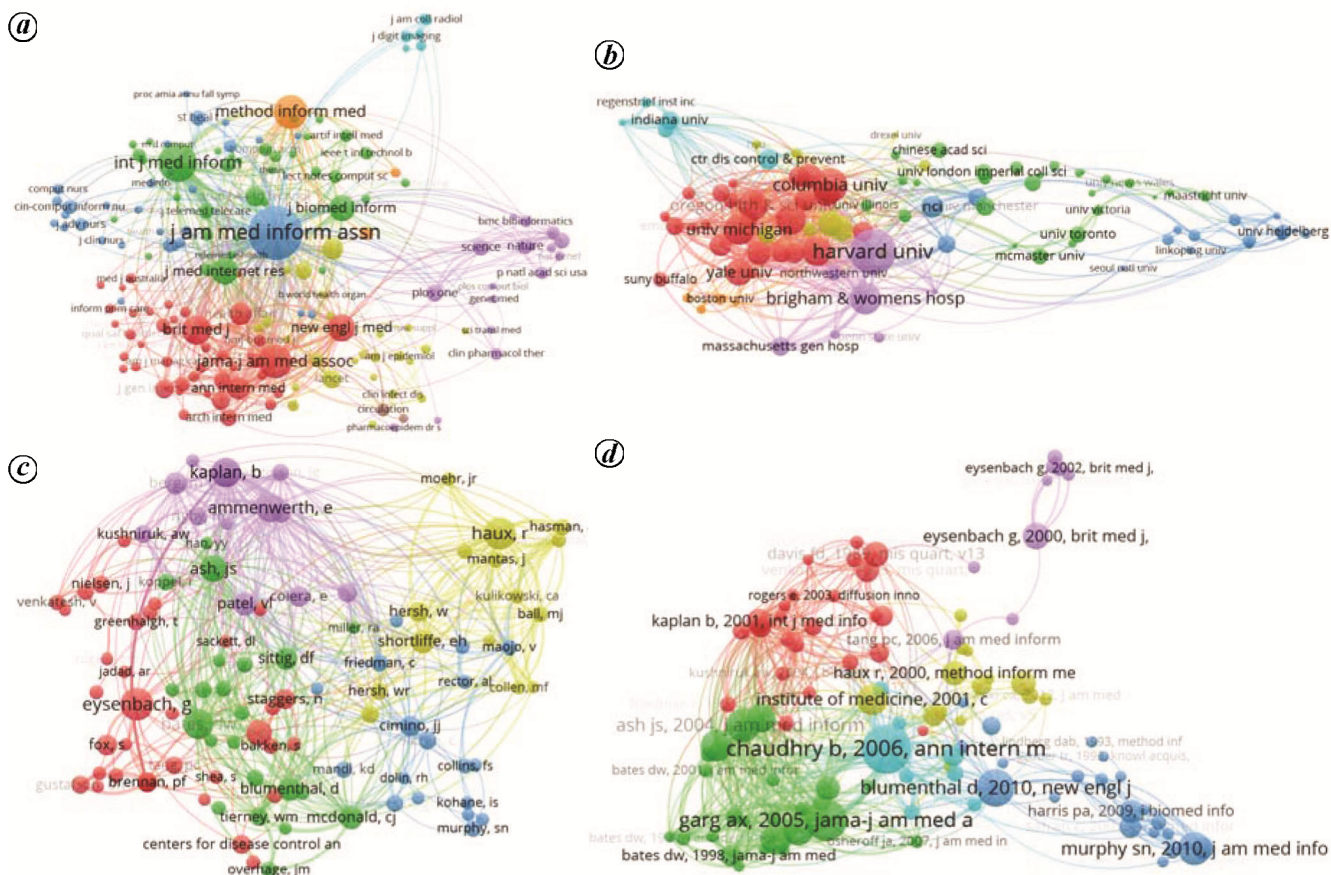
**Figure 4.** Network visualization analysis of health informatics-related publications from 2001 to 2018. *a*, Journal co-citation network. *b*, Institution co-authorship network. *c*, Co-citation cited author network. *d*, Reference co-citation network.

between two authors. Circles with the same colour were classified as similar field. In Figure 4 *c*, five clusters of authors have been distinguished. The leading researchers were Eysenbach, G. (373 citations), Haux, R. (408 citations), Bates, D. W. (351 citations), Ammenwerth, E. (335 citations) and Cimino, J. J. (182 citations) respectively in each cluster. Other researchers were directly or indirectly associated with one of these leading researchers.

*Reference co-citation analysis:* The co-citation of two papers can be established when they are cited simultaneously in a third paper[12]. The references ranked by their citations, the top 100 cited references made up a co-cited reference map with 2099 links (Figure 4 *d*). The size of nodes revealed not only their citation number, but also their association with health informatics. For instance, the most significant node in Figure 4 *d* was the article published in *Annals of Internal Medicine* 2006, by Basit Chaudhry, titled 'systematic review: impact of health information technology on quality, efficiency, and costs of medical care'[18]. This paper was cited 116 times by the 6077 health informatics-related articles, and more than 1300 times in the WoSCC.

## Emerging trends and research frontiers of health informatics

Citespace offers tools to analyse emerging trends and research frontiers[19]. We delineated the topics involved in health informatics-related articles by their keywords and titles. Figure 5 depicts a timeline visualization of publications in health informatics with 1423 nodes and 3854 connection lines using cluster labelling techniques, the log-likelihood ratio (LLR) tests of Citespace[9]. The Silhouette value (from –1 to 1) measures network homogeneity; greater the value, clearer is the theme for each cluster. The modularity value (from 0 to 1) reflects the network modularization, greater the value, better is the degree of the clustering network. In Figure 5, the network modularity value was 0.8983 and its silhouette value was 0.1969, indicating that the comprehensive landscape of health informatics was excellent for modularization and acceptable for homogeneity.

The network was decomposed into 20 clusters of co-cited references (Figure 5). Each cluster corresponded to a special field of health informatics research. Colour rings of cited references indicated that one or more citation bursts associated with this reference were detected.
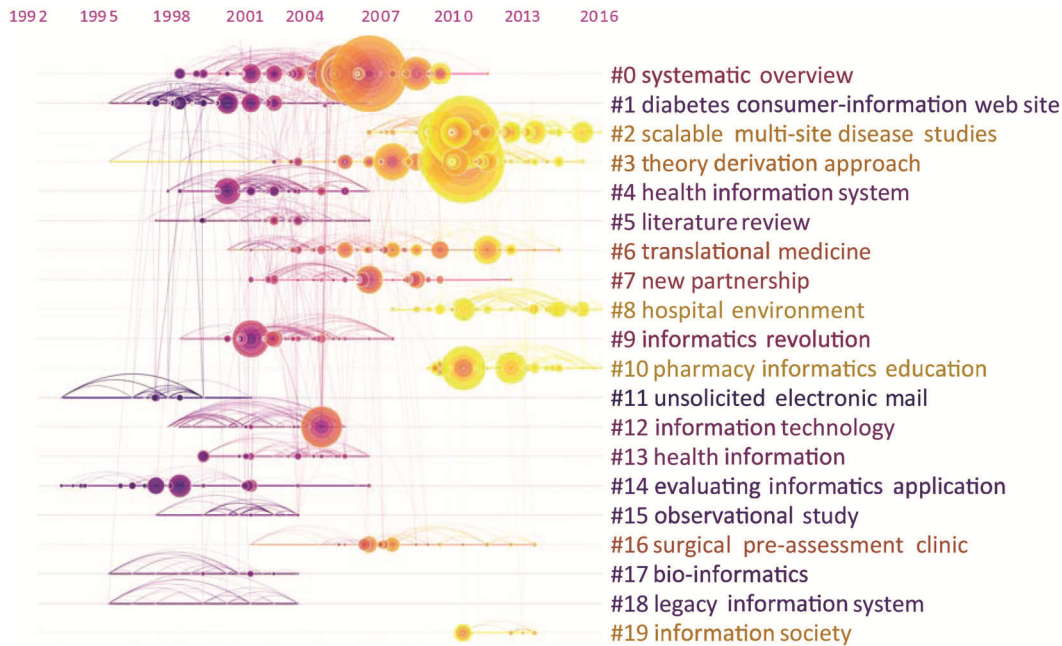
**Figure 5.** Reference co-citation time-view map of publications on health informatics from 2001 to 2018.

Citation bursts refer to abrupt increase of citation counts over a short period of time[9]. Since 2011, health informatics research has focused on four clusters: #2 scalable multi-site disease studies; #8 hospital environment; #10 pharmacy informatics education and #19 information society.

Keywords give sound descriptions of research hotspots, emerging trends and research frontiers[9,20]. Here, keywords with the strongest citation bursts could be identified as research frontiers. In Figure 6, we list the top 25 research frontiers during 2001–2018, plotted as green lines on the time intervals and highlight the periods of burst in red, which indicates the time interval of each burst from the beginning to the end.

## Discussion

In this study, visual analytic tools, including GIS mapping, VOSviewer and Citespace were applied in the analysis of current status and cooperation among countries and institutions, emerging trends and research frontiers in the field of health informatics. The results could be valuable in analysing critical developments in the specific field through newly published studies and will play an important role in a supplement to the reviews and articles in health informatics.

The timeline view of clusters of co-cited references in Figure 5 can be considered as the development of health informatics. For instance, since 2011, health informatics research has been focused on, as mentioned earlier. These results reveal the research frontiers in health informatics.

First, the cluster of scalable multi-site disease studies included 59 cited references. During the past decades, medical studies from a single centre usually reported opposite results, and some results could hardly be repeated. There could be confounders, such as the heterogeneity of patients with geographic and demographic boundaries, differences in the medical record system and treatment in different centres, which could influence the conclusions; therefore, studies have been conducted to establish a research system that could aggregate the information of patients from different hospitals in a unified way. Multi-site disease studies in health informatics have been developed and will be further improved by more and more researchers in the future.

Second, the cluster of hospital environment included 41 cited references. The ethical conflicts and ethical challenges in hospitalist systems have gained more attention in recent years, including utilization of medical resources, traditional models of hospital care versus a patients right to choose a hospitalist. Some studies have been designed to analyse and solve the above issues in multi-hospital environments. Third, the cluster of pharmacy informatics education had 28 cited references. As an important component of health informatics, pharmacy informatics ranges from pharmaceutical treatment, pharmacokinetics, to cost-effectiveness of medical management. It influences not only the choice of medication, but also pharmacy education. The last cluster, viz. information society had eight cited references. This burst was smaller than the others; however, the 21st century is an information society. We should recognize that continuing progress in the technology and methodology of communication and
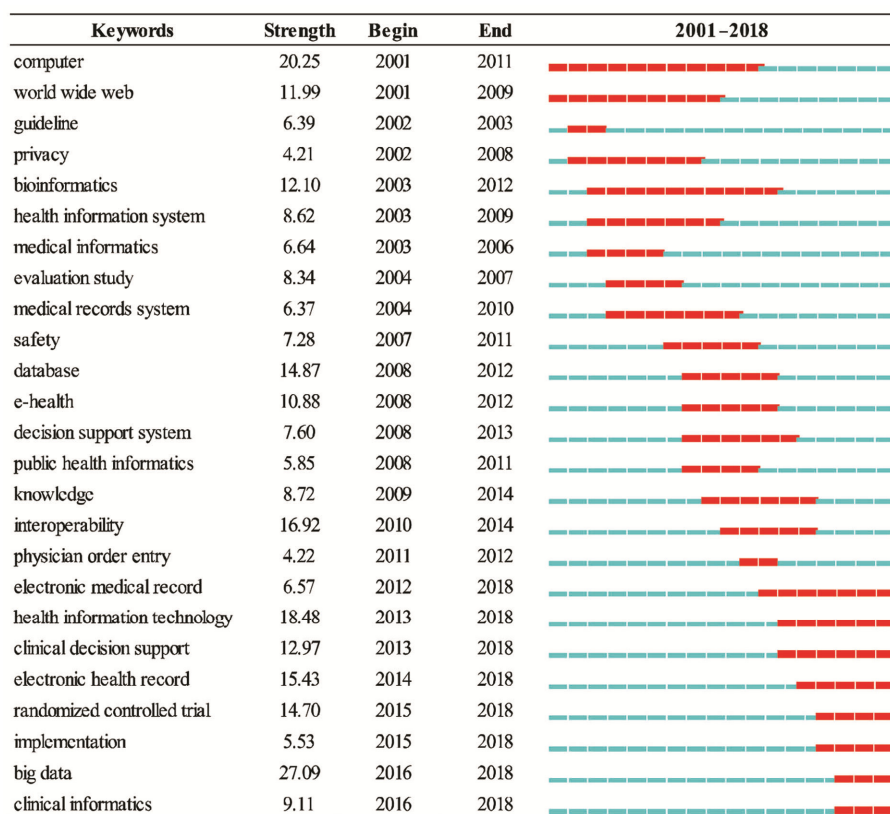
| Keywords | Strength | Begin | End | 2001–2018 |
|---|---|---|---|---|
| computer | 20.25 | 2001 | 2011 | |
| world wide web | 11.99 | 2001 | 2009 | |
| guideline | 6.39 | 2002 | 2003 | |
| privacy | 4.21 | 2002 | 2008 | |
| bioinformatics | 12.10 | 2003 | 2012 | |
| health information system | 8.62 | 2003 | 2009 | |
| medical informatics | 6.64 | 2003 | 2006 | |
| evaluation study | 8.34 | 2004 | 2007 | |
| medical records system | 6.37 | 2004 | 2010 | |
| safety | 7.28 | 2007 | 2011 | |
| database | 14.87 | 2008 | 2012 | |
| e-health | 10.88 | 2008 | 2012 | |
| decision support system | 7.60 | 2008 | 2013 | |
| public health informatics | 5.85 | 2008 | 2011 | |
| knowledge | 8.72 | 2009 | 2014 | |
| interoperability | 16.92 | 2010 | 2014 | |
| physician order entry | 4.22 | 2011 | 2012 | |
| electronic medical record | 6.57 | 2012 | 2018 | |
| health information technology | 18.48 | 2013 | 2018 | |
| clinical decision support | 12.97 | 2013 | 2018 | |
| electronic health record | 15.43 | 2014 | 2018 | |
| randomized controlled trial | 14.70 | 2015 | 2018 | |
| implementation | 5.53 | 2015 | 2018 | |
| big data | 27.09 | 2016 | 2018 | |
| clinical informatics | 9.11 | 2016 | 2018 | |

**Figure 6.** Keywords with the strongest citation bursts of health informatics-related publications from 2001 to 2018.

information is changing our communities, including the field of medical care.

Health informatics is the entire management of information in healthcare, and it focuses on the medical care of patients. The citation bursts of keywords in Figure 6 show that research hotspots in health informatics have switched to 'health information technology', 'clinical decision support', 'electronic health record', 'implementation', 'big data' and 'clinical informatics' in recent years. The imperfection of medical decision-making needs 'clinical informatics' for 'clinical decision support', through the 'implementation' of 'health information technology', 'electronic health record' and 'big data'. In the future, research on 'big data' should still be an important research direction.

The search method of retrieving information is based on 'health informatics' that includes all commonly known keywords related to the area of health informatics; however, this is a fast-developing field with emerging novel keywords, and some of them may not have been included in our analyses.

In addition, the developments in health informatics are disparate between developed and developing countries. There was only one developing country in the top 10, which contributed to the largest number of articles. No institution from developing countries ranked among the top 10, which had the largest number of articles or citations (Table 1 and Figure 4 b). This gap is in part due to the disparity of economic and technological levels between the two, as well as human resource[1]. With the progress of economy, science and technology, health informatics research will make great progress in the developing countries, considering their large populations.

This study has some limitations. The results are mainly from the WoSCC database with a collection of high-quality literature, most of them written in English; therefore, a linguistic bias might exist. Also, there were no data from other academic materials, such as patents and books on health informatics.

## Conclusion

The increasing annual number of health informatics-related publications indicate that the field has become a research hotspot in recent years. The *International Journal of Medical Informatics* published the largest number of articles, followed by *Journal of the American Medical Informatics Association*, *Methods of Information in Medicine*, and *Journal of Medical Internet Research*. The most influential country was USA, which not only contributed to the most number of articles, but also cooperated

closely with many other countries. The most influential institution was Harvard University, USA. Eysenbach, G., had the most number of cited articles. Recent research hotspots were 'big data' and 'clinical informatics'.

Our findings might provide valuable information to identify the perspectives of health informatics research. In the future, 'big data' for health informatics and clinical informatics could be a research frontier. The development of informatics technology and methodology is expected to promote continuous quality improvement in medical care in an ageing society. Health informatics will enhance the opportunities for global access to medical knowledge and health services, and help achieve better and cost-effective medical care for patients all over the world. Nevertheless, further studies and collaborations are needed.

*Competing interests:* The authors declare no competing interests.

*Data availability:* The raw data can be obtained from the Web of Science Core Collection of Thomson Reuters.

1. Fridsma, D. B., Health informatics: a required skill for 21st century clinicians. *Br. Med. J.*, 2018, **362**, k3043.
2. Novak, L. L., Kuziemsky, C. and Kaplan, B., Samantha Adams Festschrift: Sam Adams and the social construction of technology and health-implications for biomedical informatics. *Appl. Clin. Inform.*, 2018, **9**, 496–499.
3. Harrisa, S. *et al.*, Critical care health informatics collaborative (CCHIC): data, tools and methods for reproducible research: a multi-centre UK intensive care database. *Int. J. Med. Inf.*, 2018, **112**, 82–89.
4. Carrion, J., Improving the patient–clinician interface of clinical trials through health informatics technologies. *J. Med. Syst.*, 2018, **42**, 120.
5. Walshe, N., Geographical information systems for school geography. *Geography*, 2018, **103**, 46–49.
6. Sun, L. and Zou, L.-X., Spatiotemporal analysis and forecasting model of hemorrhagic fever with renal syndrome in Mainland China. *Epidemiol. Inf.*, 2018, **146**, 1680–1688.
7. Kose, E., Erbas, M. and Ersen, E., An integrated approach based on game theory and geographical information systems to solve decision problems. *Appl. Math, Comput.*, 2017, **308**, 105–114.
8. van Eck, N. J. and Waltman, L., Software survey: Vosviewer, a computer program for bibliometric mapping. *Scientometrics*, 2010, **84**, 523–538.
9. Chen, C. M., Eugene Garfield's scholarly impact: a scientometric review. *Scientometrics*, 2018, **114**, 489–516.
10. Samitas, A. and Kampouris, E., Empirical investigation of co-authorship in the field of finance: a network perspective. *Int. Rev. Financ. Anal.*, 2018, **58**, 235–246.
11. Primarily, U. N. S., Standard country or area codes for statistical use (m49), Statistics Division of the United Nations Secretariat, 1999, p. 1.
12. Boyack, K. W. and Klavans, R., Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? *J. Am. Soc. Inf. Sci. Technol.*, 2010, **61**, 2389–2404.
13. Muhuri, P. K., Shukla, A. K., Janmaijaya, M. and Basu, A., Applied soft computing: a bibliometric analysis of the publications and citations during (2004–2016). *Appl. Soft Comput.*, 2018, **69**, 381–392.
14. Zhou, H., Tan, W., Qiu, Z., Song, Y. and Gao, S., A bibliometric analysis in gene research of myocardial infarction from 2001 to 2015. *PeerJ*, 2018, **6**, e4354.
15. Saha, S. and Das, R., Exploring differential evolution and particle swarm optimization to develop some symmetry-based automatic clustering techniques: application to gene clustering. *Neural Comput. Appl.*, 2018, **30**, 735–757.
16. Xing, D., Zhao, Y., Dong, S. J. and Lin, J. H., Global research trends in stem cells for osteoarthritis: a bibliometric and visualized study. *Int. J. Rheum. Dis.*, 2018, **21**, 1372–1384.
17. Hou, D., Bi, X., Mao, Z., Fan, Y., Hu, X. and Li, X., Biomaterials research of China from 2013 to 2017 based on bibliometrics and visualization analysis. *PeerJ*, 2019, **7**, e6859.
18. Chaudhry, B. *et al.*, Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann. Int. Med.*, 2006, **144**, 742–752.
19. Lin, G., Hu, Z. and Hou, H., Research preferences of the G20 countries: a bibliometrics and visualization analysis. *Curr. Sci.*, 2018, **115**, 1477–1485.
20. Yu, D., Wang, W., Zhang, W. and Zhang, S., A bibliometric analysis of research on multiple criteria decision making. *Curr. Sci.*, 2018, **114**, 747–758.