

Crowdsourcing Applied in the Field of Public Health: A Bibliometric Perspective

Enjun XIA, Lingling WANG, Wei WANG

School of Management and Economics, Beijing Institute of Technology, P.R. China,
linglingbbt@qq.com

Abstract Crowdsourcing has been widely adopted in many fields, but there is no study on summarizing how crowdsourcing applies to public health. Applying bibliometric method, this paper explores the time-scope, country and institution, published journals, emerging trends, and citation cluster analysis of crowdsourcing applying to public health. The results show that American scholars lead in the crowdsourcing application research; research object of crowdsourcing application has been gradually diversified and crowdsourcing is widely used to analyze the issues such as health communication, preventive medicine and epidemiology, etc.

Key words Crowdsourcing; Public health; Bibliometric

1 Introduction

With increasing globalization and 21st-century trends, the flattening effect of Globalization 3.0 creates an environment that encourages the growth of crowdsourcing. Crowdsourcing can extend the innovation activities of enterprises to an infinite and vast network space; it also explores, utilizes and integrates the innovation resources of the whole society and the wisdom of the society through the Internet. From the field of innovative design to hygiene, positioning services, and new product development, crowdsourcing quietly subverts business models and traditional social structures. In recent years, application of crowdsourcing in public health has been of scholarly interest.

The Internet has greatly reduced the cost of information transfer and the boundaries of participating activities. Liu Bin (2015) believes that everyone can create, arrange and spread information. The original top-down communication path has become a “fission-like spread” in today's situation ^[1]. In this context, the concept of health 2.0 has been proposed with the practice of crowdsourcing in the field of health communication ^[2]. Like Internet 2.0, Terry (2009) ^[3] argues that Health 2.0 promotes self-creation, sharing, community concepts, and user self-empowerment, all of which coincide with the aim of crowdsourcing. Liu Chuncheng (2015) ^[4] believes that crowdsourcing can effectively compensate for the limitations of traditional health communication methods in terms of poor pertinence, lack of innovation, high cost, etc.

We retrieve 1957 papers in Web of Science database from 2006 to 2017. After removing 132 articles published in 2017, Figure 1 shows that the number of papers is growing rapidly after 2010. Among them, 16 high-cited crowdsourcing papers cover four application domains: event warning, location services, marketing management, and public health. Among them, public health areas include infectious disease prevention, health management, and Internet medical care. Application of crowdsourcing in the public health field has become a research hotspot. At present, no scholars in China have summarized the research on crowdsourcing in specific fields, especially in the field of public health.

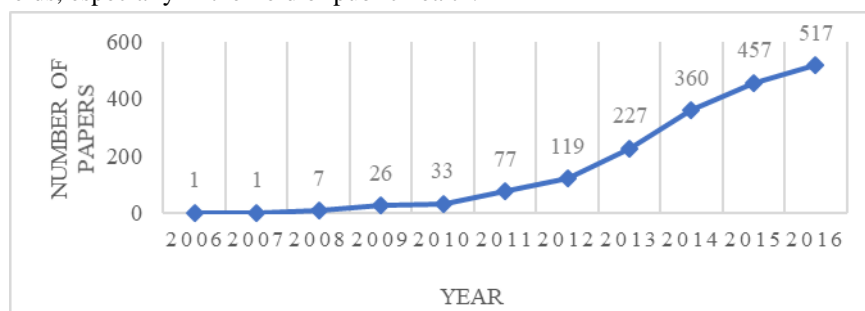


Figure 1 The number of papers about crowdsourcing research from 2006-2016

Using the Web of Science as a data source, this paper conducts bibliometric analysis of papers from 2007-2017 for the application research of crowdsourcing in the field of public health. Based on the research of spatial-temporal analysis and journals clustering research, this paper deeply explores the research frontiers and hotspots, and clarifies the evolution path and development trends of crowdsourcing in this field. The innovation of this study is to summarize and summarize the application research of crowdsourcing in the field of public health, which is conducive to the hotspots and frontiers of open innovation and crowdsourcing for Chinese scholars, thus promoting innovation management research.

2 Methodology and data

This study uses bibliometric and knowledge graph to analyze the relationship of extant literature and publications in given field. Through co-citation analysis and co-occurrence analysis, knowledge graph can present the research hotspots, evolution history, and development trends of a research field in a visual form in a certain period. From an objective and quantitative perspective, bibliometrics use the information of authors, key words, references, journals, countries, institutions to report emerging trends. Hence, this method can reflect the overall research and development trends of crowdsourcing in the field of public health.

Citespace III software was used to conduct knowledge graph analysis in this article. Citespace III is a visual analysis tool developed by Professor Chaomei Chen (Drexel University, Dalian University of Technology, Changjiang Scholar) ^[5] based on JAVA platform, which can realize the co-citation analysis, keyword co-occurrence analysis, collaborative analysis of institutional authors, etc. Compared with other visualization software, Citespace III has the advantages of more convenient data processing, better visualization and easier interpretation. Therefore, it can meet the requirements of the literature co-citation and keyword co-occurrence analysis of large samples.

This study retrieves data from the core collection database of Web of Science (WOS), limiting search to the Science Citation Index (SCI), Social Science Citation Index (SSCI), Conference Proceedings Citation Index- Science (CPCI-S), and Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH). Firstly, we used crowdsourcing, crowd sourcing, and crowd-sourcing that are the similar meanings as keywords. Secondly, the time search span was set from January 2006 to June 2017 because the “crowdsourcing” was proposed by Howe in 2006. Thirdly, the search scope included existing research results of article, proceeding papers and reviews. After refining, 175 documents were retrieved. The search design process is shown in Table 1.

Table 1 Summary of Search Details

Settings	Contents
Index	SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH
Search rules	TS= (“crowdsourcing” and “health”) OR TS= (“crowd sourcing” and “health”) OR TS= (“crowd-sourcing” and “health”)
Literature types	Article, proceeding papers, review
Time span	2006.01 – 2017.06
Total	175

3 Results

3.1 Time-scope analysis

As shown in Figure 2, the orange line indicates the literature statistics of crowdsourcing in the field of public health, and the blue bar indicates the total amount of literature in given field. Since the concept of crowdsourcing was introduced in 2006, it was not until 2009 that applied research was started in the field of public health. The number of literatures has a significantly increase from 2011 to 2013. After the steady stage in 2013-2014, the number of literatures in 2015-2016 has risen sharply. Especially in 2016, the number has exceeded 60. Since we count the literature for June 2017, it is expected that the number of literatures will continue to grow in the future. At the same time, we also find that the growth of applied research in the field of public health is basically consistent with the growth trend of the whole crowdsourcing field. Absolutely, crowdsourcing research in the field of public health is the hot spot.

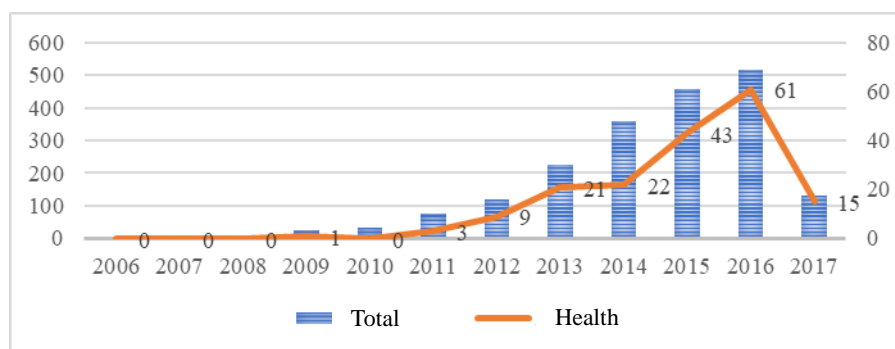


Figure 2 Time span of research literature in the field of crowdsourcing

3.2 Country and institution analysis

Citespace III can display the number of papers, cooperation and centrality published by countries in the form of annual rings. The size of the annual ring shows the number of papers issued, and the outermost purple circle of the annual ring represents the centrality^[5]. We select country and institution as network nodes, with the data extraction object of Top 30, and the path data is visually analyzed by pathfinder algorithm. According to the visual network map shown in Figure 3, the papers in this field are mainly composed of two clusters of the United States and the United Kingdom. Combined with Figure 1, from the number of papers issued by each node, the United States has the largest contribution to the literature output in this field (108 articles), accounting for 61.71% of the total number of the world, much higher than other countries, followed by the United Kingdom, the Netherlands, and China, Switzerland, Germany and Australia. In this entire network, the United States has the most centrality, which means that he has direct or indirect cooperation with many countries in the co-existing network, such as Germany and China. In recent years, China has issued a total of 8 papers. Although the research started late (the first time of publication was 2015), the number of papers has gradually increased year by year.

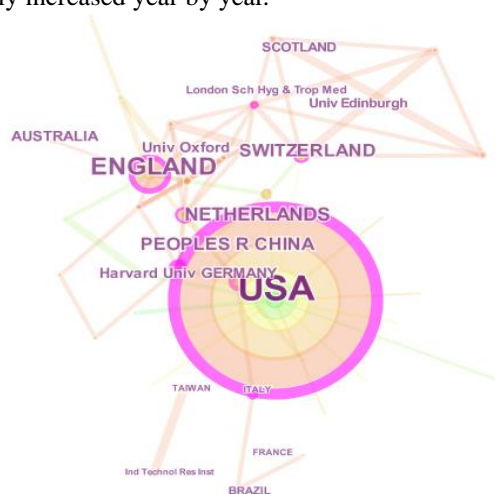


Figure 3 Network graph of country and institution

As the country with the largest output of academic papers in this field, the United States has its research institutes mainly located in universities, such as Harvard University, University of Pittsburgh, New York University, and the University of North Carolina. Similarly, representative research institutions in the UK include Cambridge University, Edinburgh University, and Imperial College. Some medical schools have also begun to study crowdsourcing in the public health field, such as The London School of Hygiene & Tropical Medicine, the Guangdong Provincial Center for Disease Control and Prevention, and Boston Children's Hospital, etc. Among them, Cambridge University, Harvard University, London School of Hygiene and Tropical Medicine are also the core global institutions that drive crowdsourcing research in the field of public health.

3.3 Journals analysis

when identifying core journals in a field, it is necessary not only to pay attention to the number of posts, but also to focus on journals with high centrality and high frequency of citations. We take a visualization of co-citation journals analysis of sample data using Citespace III and the results are shown in Table 2. From the number of posts, as top journals in medical science, *Journal of Medical Internet Research*, *New England Journal of Medicine*, *Lancet*, *Jama-Journal of the American Medical Association*, and *American Journal of Public Health* Magazine have led the research of crowdsourcing in the field of public health. The related papers posted on these journals represent the research frontiers and hotspots in this field. In addition, *Nature* and *Science* are comprehensive top-level journals that provide a knowledge base for medical, biological, economics, management, and psychology in the field of public health research.

Table 2 List of high-profile and high-cited journals

No.	Name of journal	Publication	Impact factor	Country
1	Journal of Medical Internet Research	56	5.175	Canada
2	PLoS One	48	2.806	US
3	Nature	35	40.137	UK
4	New England Journal of Medical	24	72.406	US
5	Journal of General Internal Medicine	23	3.701	US
6	Science	23	37.205	US
7	Lancet	23	47.831	UK
8	Jama-Journal of the American Medical Association	23	44.405	US
9	American Journal of Public Health	22	3.858	US
10	Journal of the American Medical Informatics Association	21	3.698	US

3.4 Analysis of emerging trends in crowdsourcing research in the field of public health

Citespace III can generate a research evolution graph based on co-cited literature, showing the knowledge base and frontier of a domain research in the form of a time zone view. In this way, we can grasp the key achievements of crowdsourcing research in the field of public health and the development of the whole field [5]. The time span is set to 2006-2017, and cited reference is selected as the network node. We adopt the path search algorithm (pathfinder) and use the time zone view (time zone) to display. The result after running the software is shown in Figure 4. From Figure 4, we can identify key results of the application of crowdsourcing in the field of public health.

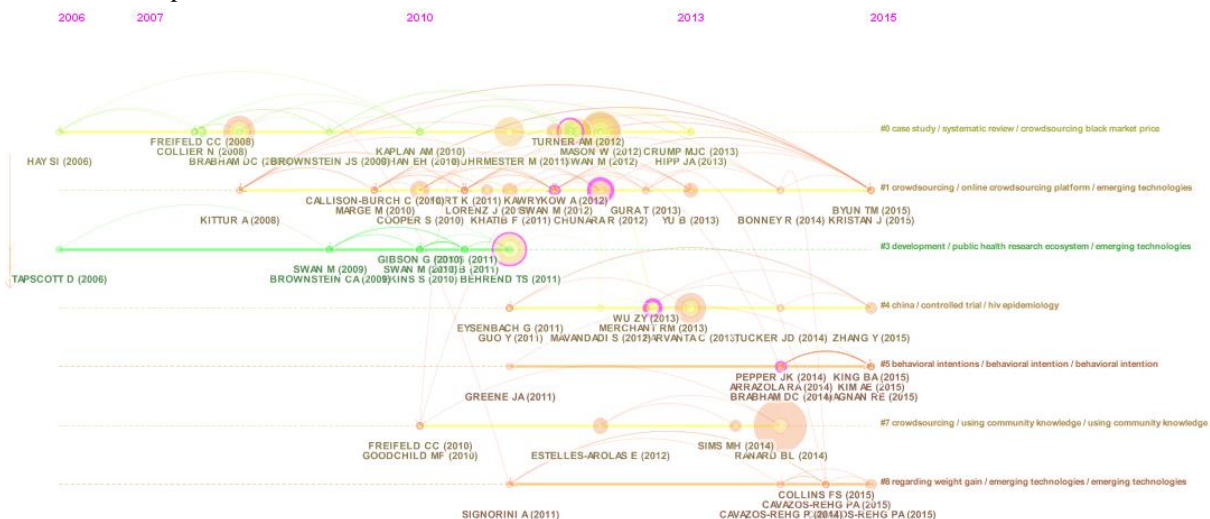


Figure 4 Evolutionary path knowledge graph of crowdsourcing in the field of public health

As shown in Table 3, the years, authors, article titles, and cited numbers of key results are listed. The two highly cited articles by DC Brabham (2008) [6] and W Mason (2012) [7] are both exploratory studies of crowdsourcing in the public domain and online crowdsourcing, laying the foundation for later research. Since 2012, crowdsourcing has gained attention in the field of public health. Many scholars have begun to study the application of crowdsourcing in specific fields such as personal care, biological agents, sexual health communication, etc. Chinese scholar Y Zhang has conducted

research on the innovation contribution competition in the field of promoting sexual health, providing a research foundation for Chinese scholars. The detailed research results are stated as follows:

Table 3 List of key research

Year	Author	Title	Frequency
2008	DC Brabham ^[6]	Crowdsourcing as a model for problem solving: Leveraging the collective intelligence of online communities for public good	1104
2011	TS Behrend ^[8]	The viability of crowdsourcing for survey research	259
2012	M Swan ^[9]	Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen	104
2012	W Mason ^[7]	Conducting behavioral research on Amazon's Mechanical Turk.	1124
2013	C Parvanta ^[10]	Crowdsourcing 101 A Few Basics to Make You the Leader of the Pack	37
2014	BL Ranard ^[11]	Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review	63
2015	Y Zhang ^[12]	Creative Contributory Contests to Spur Innovation in Sexual Health: 2 Cases and a Guide for Implementation	11

3.5 Reference cluster analysis

The title words and keywords condense the focus of the author's research, which are the core and essence of a paper. Therefore, through the co-occurrence analysis of the title words and keywords, we can find research hotspots in the scientific field. We run Citespace III software to conduct citation analysis of papers and keywords, and generate knowledge maps. These results are shown in Figure 5 and Figure 6.

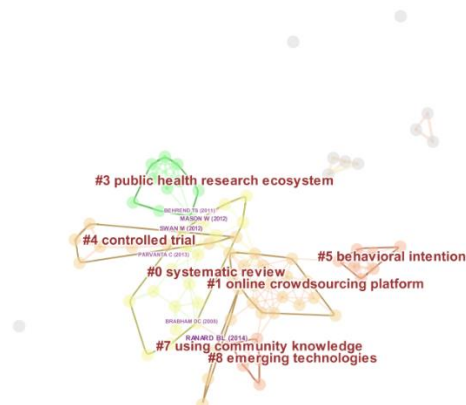


Figure 5 Literature co-citation network graph

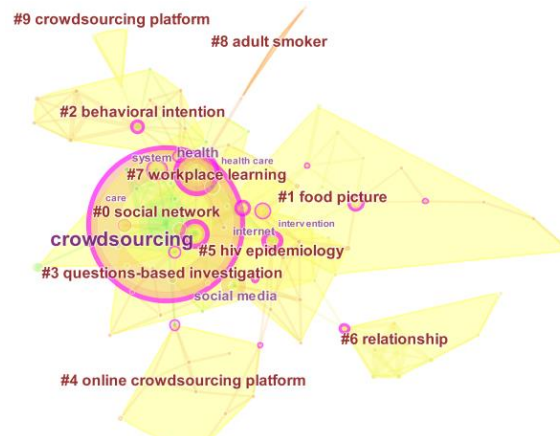


Figure 6 Keyword clustering network graph

We extract the high-frequency vocabulary in Table 4, with the high-frequency title words listed on the left and the high-frequency keywords on the right. Before the keyword clustering, some scattered words appeared, such as crowdsourcing, health, social media, Internet, health care, disease, intervention, mental health, prevention, etc. This phenomenon indicates that crowdsourcing was applied to public health, mental health, preventive medicine, interventional therapy and other fields, with the tools of social media, the Internet, etc. Comparing the clustered title words and keywords, we can find that in addition to high-frequency word crowdsourcing and health, public health ecosystem, community knowledge, systematic evaluation and other words appear in the title words. And the keywords indicate specific subjects such as AIDS epidemiology, adult smokers, and problem investigations.

Table 4 List of high frequency title words and keywords

No.	High frequency title words	No.	High frequency key words
1	Systematic review	1	Social network
2	Online crowdsourcing platform	2	Food picture
3	Public health research ecosystem	3	Behavioral intention
4	Controlled trial	4	Question-based investigation
5	Behavioral intention	5	Online crowdsourcing platform
6	Using community knowledge	6	HIV epidemiology
7	Emerging technologies	7	Relationship
		8	Workplace learning
		9	Adult smoker

4 Conclusion

From a multi-dimensional, time-sharing and dynamic perspective, this paper rationalizes the development trends and hotspots of the research, and provides empirical evidence for the application research of network crowdsourcing in China to keep up with the international frontier. The main conclusions are as follows.

Firstly, the number of crowdsourcing research literature has proliferated, and the field of network crowdsourcing has been more abundant, especially in public health and health communication. At present, the United States leads the crowdsourcing research in the field of public health. Scholars from the Guangdong Provincial Center for Disease Control and Prevention cooperate with American universities to apply crowdsourcing in sexual health communication and HIV prevention. They have made outstanding contributions to the application of crowdsourcing in the field of public health in China.

Secondly, the application research of crowdsourcing has been gradually diversified, for example, from the initial labor crowdsourcing to the current network crowdsourcing health communication, from participant behavior research to preventive medicine, mental health and personalized prevention in the field of health, etc. The network crowdsourcing model not only promotes the application of healthy big data and the construction of intelligent health platform but also promotes the process of social medical crowdsourcing.

Finally, crowdsourcing research in this field focuses on four domains: study crowdsourcing as an Internet health media; application of crowdsourcing in the field of prevention and treatment; the role of crowdsourcing in the public health care ecosystem; applied research of crowdsourcing competitions in infectious diseases and epidemiology. In addition, it is found that the application research of crowdsourcing in China's public health field is only carried out in Guangdong Province, and there are no significant achievements in other regions.

References

- [1] Liu Bin. Reconstruction and Innovation of Health Communication under the Network Environment[J]. *News Affection*, 2015(07):86-88.
- [2] Hughes B, Joshi I, Wareham J. Health 2.0 and Medicine 2.0: tension and controversies in the field[J]. *Journal of Medical Internet Research*, 2008, 10(3): e23.
- [3] Terry M. Twittering healthcare: social media and medicine[J]. *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association*, 2009, 15(6): 507-510.
- [4] Liu Chuncheng, Tang Songyuan, Chongyi Wei, Joseph Tucker. From the masses to the masses: crowdsourcing and health communication [A]. 2015 China Health Communication Conference excellent papers [C]. 2015
- [5] Chen Chaomei, Chen Yue, Hou Jianhua, et al. CiteSpace II: Identification and Visualization of New Trends and New Dynamics in Scientific Literature[J]. *Journal of the China Society for Scientific Information*, 2009, 28(3): 401-421.
- [6] Brabham D C. Crowdsourcing as a model for problem solving: Leveraging the collective intelligence of online communities for public good[J]. *Convergence the International Journal of Research into New Media Technologies*, 2008, 14(1):75-90.
- [7] Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. [J]. *Behavior Research Methods*, 2012, 44(1):1-23.
- [8] Behrend T S, Sharek D J, Meade A W, et al. The viability of crowdsourcing for survey research[J]. *Behavior Research Methods*, 2011, 43(3):800.
- [9] Swan M. Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen[J]. *Journal of Personalized Medicine*, 2012, 2(3):93.
- [10] Parvanta C, Roth Y, Keller H. Crowdsourcing 101 A Few Basics to Make You the Leader of the Pack[J]. *Health Promot Pract*, 2013, 14(2):163-7.
- [11] Ranard B L, Ha Y P, Meisel Z F, et al. Crowdsourcing--harnessing the masses to advance health and medicine, a systematic review[J]. *Journal of General Internal Medicine*, 2014, 29(1):187.
- [12] Zhang Y, Kim J A, Liu F, et al. Creative Contributory Contests to Spur Innovation in Sexual Health: 2 Cases and a Guide for Implementation[J]. *Sexually Transmitted Diseases*, 2015, 42(11):625.