Association for Information Systems
AIS Electronic Library (AISeL)

Winter 12-4-2017

A Co-Citation Bibliometric Analysis of Crowdsourcing Research

Xiaobei Liang
Tongji University, Shanghai, China, liangxiaobei@tongji.edu.cn

Minghua He
Tongji University and Jinggangshan University, China, hmhminy@163.com

Follow this and additional works at: http://aisel.aisnet.org/iceb2017

A Co-Citation Bibliometric Analysis of Crowdsourcing Research

Xiaobei Liang, Tongji University, Shanghai, China, liangxiaobei@tongji.edu.cn
Minghua He*, Tongji University and Jinggangshan University, China, hmhminy@163.com

ABSTRACT

Crowdsourcing has gained increasing attention from scholars in a broad range of fields such as computer science, engineering, information system, business, and economics. However, few crowdsourcing studies are conducted from the bibliometric perspective. This paper conducts document co-citation, author co-citation, journal co-citation, and keyword co-word analysis in the crowdsourcing research field by using CiteSpace and *Web of Science TM Core Collection* database, aiming to identify highly cited articles and journals and influential authors in the crowdsourcing research field during the time span from 2008 to 2016 and to find out current hot research topics and future directions in the crowdsourcing research field.

Keywords: Crowdsourcing; co-citation; bibliometric analysis

_____
*Corresponding author

INTRODUCTION

Since Howe (2006) coined the term "crowdsourcing" in *Wired Magazine*, crowdsourcing has gained increasing attention from scholars in fields such as computer science, business, economics, science and marketing. A lot of scholars have conducted a broad range of studies on crowdsourcing. Previous studies paid a lot of attention to solving the questions such as what exactly crowdsourcing is, how the crowdsourcing platform operates, and what difference exists between an outsourcing task and a crowdsourcing task. Several authors, especially Chinese authors have conducted a literature review on crowdsourcing research (Feng and Li, 2015; Zhang *et al.*, 2012). A crowdsourcing literature review was also conducted from the human resource management perspective (Buettner, 2015), and a systematic mapping study on crowdsourcing was conducted (Hosseini *et al.*, 2015). However, few crowdsourcing studies are conducted from the bibliometric perspective.

Bibliometric analysis is a statistical analysis of written publications, such as books or articles. The purpose of our study is to conduct a bibliometric analysis by using CiteSpace to detect and visualize important articles, influential authors, current hot research topics, as well as evolving trends in the crowdsourcing research. The bibliometric analysis results will be presented in a quantitative and visualized manner by using CiteSpace. Bibliographic techniques are distinctive functions of CiteSpace, and are usually used to conduct co-citation analysis and keyword co-word analysis in a research field (Liu *et al.*, 2015). In this paper, we will apply the following bibliographic techniques provided by CiteSpace:

- document co-citation analysis (Chen, 2009 ; Leydesdorff, 2005)
- author co-citation analysis (Chen and Carr, 1999 ; Nerur *et al.*, 2008; White & Mccain, 1998)
- journal co-citation analysis (Hu *et al.*, 2011; Liu, 2005; Tsay *et al.*, 2003)
- keyword co-word analysis (Callon *et al.*, 1991; He, 1999; Lee & Su, 2010)

Document co-citation analysis is intended to find out important articles with a high citation frequency in the crowdsourcing research field; author co-citation network is intended to find out the important and influential authors in the crowdsourcing research field; journal co-citation network is intended to find out which journals are dominant sources of articles regarding crowdsourcing; and keyword co-word analysis measures the frequency of keywords, and tries to identify current research front and evolving trends in the crowdsourcing research field.

In summary, this paper contributes to identify which articles, authors, and journals play an important role in the evolution of the crowdsourcing research and to find out hot topics in the research field in recent years, so that interesting scholars and practitioners in this field can develop a good insight into the current hot research topics and future research directions of the crowdsourcing research field.

This paper is structured as follows. First, we discuss the method and data for bibliometric analysis. Second, a basic description of crowdsourcing research is provided, which includes the number of articles, the number of citations, top productive authors, and top journals in the crowdsourcing research field during the time span from 2008 to 2016. Third, we describe empirical results of the bibliometric analysis in detail, which include results of document co-citation analysis, author co-citation analysis, disciplinary distribution analysis, journal co-citation analysis, and keyword co-word analysis in the crowdsourcing research field. Lastly, we conclude our analysis results, analyze limitations of our study, and discuss our future research work.

METHOD AND DATA

In our study, we employed the Java-based CiteSpace program as the analysis tool. The version of CiteSpace used in our study is CiteSpace 5.0. R1 SE (hereinafter referred to as CiteSpace V).

Though Howe (Howe, 2006) coined the term "crowdsourcing", it was in 2008 that the first scholarly research using the word crowdsourcing was published. Therefore, we retrieved literature records from the Web of Science ™ Core Collection database by using crowdsourcing as the keyword and setting the time span from 2008 to 2016 (as of October 27, 2016). A total of 1237 publications were retrieved and classified into eleven document types. Among all the 1237 publications, there were 590 proceedings papers, accounting for 47.81 % of the total, followed by journal articles ((498, 40.11%), editorial materials (54, 4.38 %), meeting abstract (29, 2.35 %) and review (25, 2.03 %). We adopted the 498 journal articles as the input data for further analysis.

INITIAL DATA STATISTICS

Since the year of 2008, the number of studies on crowdsourcing has been on the rise (except a slight fall in the year 2010). In the year of 2016, the number of papers in the crowdsourcing research field reached 135, accounting for 27.1% of all the 498 articles. During the time span from 2014 to 2016, crowdsourcing related studies have become increasingly hotter (with the number of articles accounting for over 70%; see Figure 1).
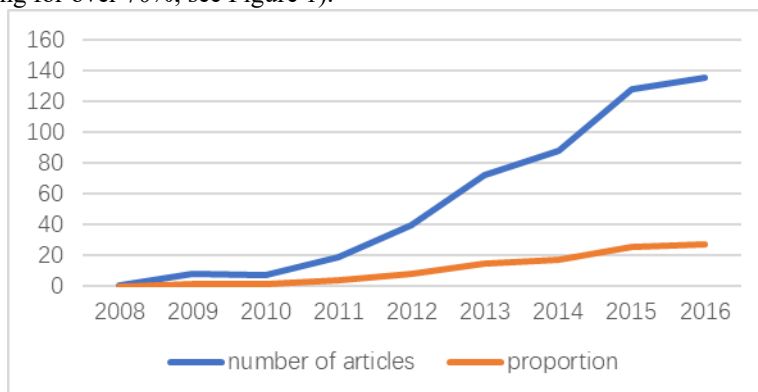

Figure 1: Number of Articles Regarding Crowdsourcing Research

According to our search results, the 498 articles were cited 3457 times during the time space from 2008 to 2016, and the number of citation counts amounted to 2885 times if self-citations were excluded, with the average of citations being 6.94. As shown in Figure 2, the number of citations has witnessed continuous growth since the year of 2009, indicating that previous studies have exerted a larger impact on current crowdsourcing research. Details of citations will be further analyzed in the following sections.
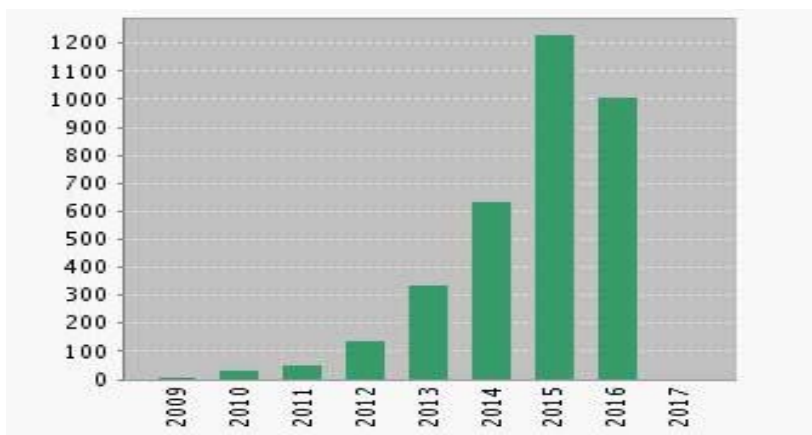

Figure 2: Number of Citations of Articles Each Year

Table 1 lists the top 20 productive authors in the crowdsourcing research field during the time span from 2008 to 2016. In this list, the top five productive authors are Wu, Chen, Fritz, Brabham, and Yang, each of whom has published at least four articles in the crowdsourcing research field. Judging from Table 1, we could find that several productive authors are from China, showing that Chinese authors are very active in this research field. To a certain degree, Chinese authors have paid close attention to hot research topics in this filed, thereby contributing to the evolution of crowdsourcing research.

Table 1: Top 20 Productive Authors in The Crowdsourcing Research Field

Author	Number	Proportion	Author	Number	Proportion
WU	6	1.212	HUI	4	0.808
CHEN	6	1.212	DUSTDAR	4	0.808
FRITZ	5	1.01	ZHUANG	3	0.606
BRABHAM	5	1.01	ZHAO	3	0.606
YANG	4	0.808	ZHANG	3	0.606

WANG	4	0.808	ZHANG	3	0.606
STOLOVITZKY	4	0.808	ZHANG	3	0.606
SEE	4	0.808	WANG	3	0.606
PERGER	4	0.808	SYED	3	0.606
LU	4	0.808	SU	3	0.606

All the 498 journal articles were published on 309 journals, which were contributed by 500 scholars from 59 countries. USA was the top productive country, publishing a total of 247 articles, followed by the People's Republic of China (publishing 76 articles), England (publishing 59 articles) and Germany (publishing 37 articles). The top four productive institutions all came from the USA, namely, University of Michigan, University of North Carolina, Sandford University, and Harvard University. It is worth noting that China's Tsinghua University and Wuhan University were all among the list of the ten most productive institutions.

Table 2 lists top 20 journals publishing articles in the crowdsourcing research field and their respective impact factor in 2015. The PLOS ONE journal published the most papers in the field of crowdsourcing (18 in total). This journal is the world's first multidisciplinary open access journal, covering primary research from any discipline within science and medicine. Because the concept of crowdsourcing was initially defined as an online problem-solving model (Howe, 2006), the crowdsourcing field is usually concerned with web technology, computer science, engineering, and information system. Therefore, it is not surprising that related articles are usually published on journals such as Journal of Medical Internet Research (11 in total), IEEE Transactions on Knowledge and Data Engineering Science (9 in total), and IEEE Internet Computing (8 in total).

Table 2: Top 20 Journals in The Crowdsourcing Research Field

Name	Number	Impact factor
PLOS ONE	18	3.057
JOURNAL OF MEDICAL INTERNET RESEARCH	11	4.532
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	9	2.476
IEEE INTERNET COMPUTING	8	1.4
ACM TRANSACTIONS ON INTELLIGENT SYSTEMS AND TECHNOLOGY	8	2.414
IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS	7	2.534
DECISION SUPPORT SYSTEMS	7	2.604
COMPUTER NETWORKS	7	1.446
INFORMATION RETRIEVAL	6	0.917
IEEE COMMUNICATIONS MAGAZINE	6	5.125
DATABASE THE JOURNAL OF BIOLOGICAL DATABASES AND CURATION	6	2.627
SENSORS	5	2.033
JOURNAL OF INFORMATION SCIENCE	5	0.878
IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	5	2.661
CHEMICAL ENGINEERING NEWS	5	0.33
SCIENTIFIC REPORTS	4	5.228
INFORMATION COMMUNICATION SOCIETY	4	2.109
IEEE TRANSACTIONS ON MOBILE COMPUTING	4	2.456
EXPERT SYSTEMS WITH APPLICATIONS	4	2.981
TRANSPORTATION RESEARCH RECORD	3	0.522

CO-CITATION ANALYSIS RESULTS

Document Co-Citation Analysis

We analyzed references of the retrieved papers by using CiteSpace V, and chose the top 20 references in each time slice for detailed analysis. By using CiteSpace V, we obtained the top 10 cited references in terms of citation frequency during the time span from 2006 to 2016 (see Table 3). Among all the cited references, Brabham (2008) article was cited the most in our database. In fact, Brabham (2008) article was also the first scholarly research using the word crowdsourcing in this field. Brabham (2008) introduced crowdsourcing from the perspective of theoretical grounding and exemplar cases and explored the possibilities for the crowdsourcing model in detail. It is not surprising that Brabham (2008) article served as an important basis for all subsequent researches in the crowdsourcing research field. Other papers in the top 10 were also linked closely to Brabham (2008) article.

Howe (2008) book was cited 46 times in total. In his book, Howe explained how companies had managed to exploit the global pool of talent, and proposed that crowds would always perform better than internal employees, which further developed the crowdsourcing concept and advocated the advantages of crowdsourcing. In Estellés-Arolas (2012) article, current crowdsourcing related definitions were deeply analyzed to retrieve general elements and establish eight basic characteristics of any crowdsourcing initiative. Knowing these basic characteristics is a prerequisite for carrying out research on crowdsourcing applications and initiatives. In Doan *et al.* (2011) article, various online crowdsourcing systems were discussed, a variety of crowdsourcing applications were explored, and a lot of technological challenges were raised in the process of applying crowdsourcing means to solving problems. In Paolacci *et al.* (2010) article, a survey was conducted to get new demographic data about the Mechanical Turk subject population, aiming to make a difference between the quality of data acquired through crowdsourcing-based labor markets and the data acquired through traditional methods. In Kittur and Kraut (2008) article, Wikipedia-a forerunner employing the crowdsourcing model, was taken as an example to explain how the coordination methods affected article quality, concluding that it was of critical importance to effectively harness the "wisdom of the crowds" in online production environments. In Snow *et al.* (2008) article, Amazon's Mechanical Turk system was proved to be a highly efficient method for collecting wisdoms of crowds over the Web, further emphasizing the advantages of crowdsourcing model. In Jeppesen and Lakhani (2010) article, the significance and value of open innovation in the process of recruiting labor forces online were examined, thereby contributing to the literature research on connections between crowdsourcing and open innovation. In Leimeister *et al.* (2009) article, users' incentives and motives of participation in crowdsourcing idea competitions were explored by using a two-step model, which provided support for the development of promising innovative ideas in crowdsourcing idea competitions. In Howe (2006) article, the term "crowdsourcing" was coined for the first time, thereby serving as a cornerstone for subsequent studies in the crowdsourcing research field.

Through above analysis of the top 10 cited papers, we concluded that these papers mainly discussed the basic concepts, definitions, advantages, and applications of crowdsourcing, and were therefore basic and important articles in the crowdsourcing research field. Almost all scholars who want to carry out basic research into crowdsourcing may cite these references.

Table 3: Top 10 Cited References in The Crowdsourcing Research Field

Study	Citation counts in our dataset	Type of source
Brabham (2008)	55	Journal Article
Howe (2008)	46	Book
Estelles-arolas (2012)	45	Journal Article
Doan *et al.* (2011)	42	Journal Article
Paolacci *et al.* (2010)	29	Journal Article
Kittur and Kraut (2008)	24	Journal Article
Snow *et al.* (2008)	21	Conference proceedings
Jeppesen and Lakhani (2010)	19	Journal Article
Leimeister *et al.* (2009)	19	Journal Article
Howe (2006)	18	Journal Article

Referring to the visualized map of crowdsourcing references (see Figure 3), we could find that this network is formed by 128 nodes and 208 links, with each node representing one cited reference (Chen, 2009). It should be noted that only the references with a threshold of at least ten citations were labelled in the visualized map (see Figure 3), indicating that they are the important references in the crowdsourcing research field. In Figure 3, several large clusters are developed, indicating that those researches were combined sufficiently. It is worth noting that nodes in the blue cluster at the bottom of the map are closely linked to Allio (2004) article. Through further literature analysis, we found that Allio (2004) article was an interview with Carroll, President and Chief Executive Officer of InnoCentive—a famous crowdsourcing platform, during which Carroll explained in detail the chances, challenges, applications and future development of the InnoCentive model for open innovation.

As shown in Figure 3, besides nodes labeled with Brabham (2008), Howe (2008), Doan (2011) and Allio (2004), other nodes with purple rings include Brabham (2009), Brabham (2010), Cooper *et al.* (2010), Hippel (2005), and Albors *et al.* (2008), indicating that these articles are also very important in the crowdsourcing research field. In Brabham (2009) article, dominant theories relevant to crowdsourcing model were systematically explained to prove that the crowdsourcing model was a practical and effective online problem solving model for business settings and that it was also applicable to public sectors. In a sense, Brabham (2009) article has extended the application scope of crowdsourcing from business settings to public sectors. In Brabham (2010) article, Threadless was taken as an example to empirically explore motivations of crowd participation in crowdsourcing applications. Since crowds' motivations are crucial factors for applying the crowdsourcing model to new forms of problem-solving, at present more and more researches are conducted, by referring to Brabham (2010) article, to examine the effect of crowd's motivations on crowd participation. In Cooper *et al.* (2010) article, interactive multiplayer games were taken as an example to demonstrate that crowdsourcing model can even be used to solve more complex scientific problems. This further highlighted the advantages of applying crowdsourcing model to scientific problems. In Hippel (2005) article, Hippel

explained in detail how the emerging process of user-centered, democratized innovation worked and how user innovation provided a very necessary complement to and feedstock for manufacturer innovation. This provided us a major paradigm shift in innovation, that is, users were taking the place of manufacturers to become the leading force of innovation in many fields. In Albors *et al.* (2008) article, new learning network collaboration practices, namely, crowdsourcing, wikis, and open source, were deeply analyzed and their motivation, consequences, as well as factors driving their development were also discussed, with an aim to define these practices clearly and apply these practices more efficiently.

In summary, all the above papers are also representative literatures in the crowdsourcing research field, and may be considered important and classical literatures in this field.
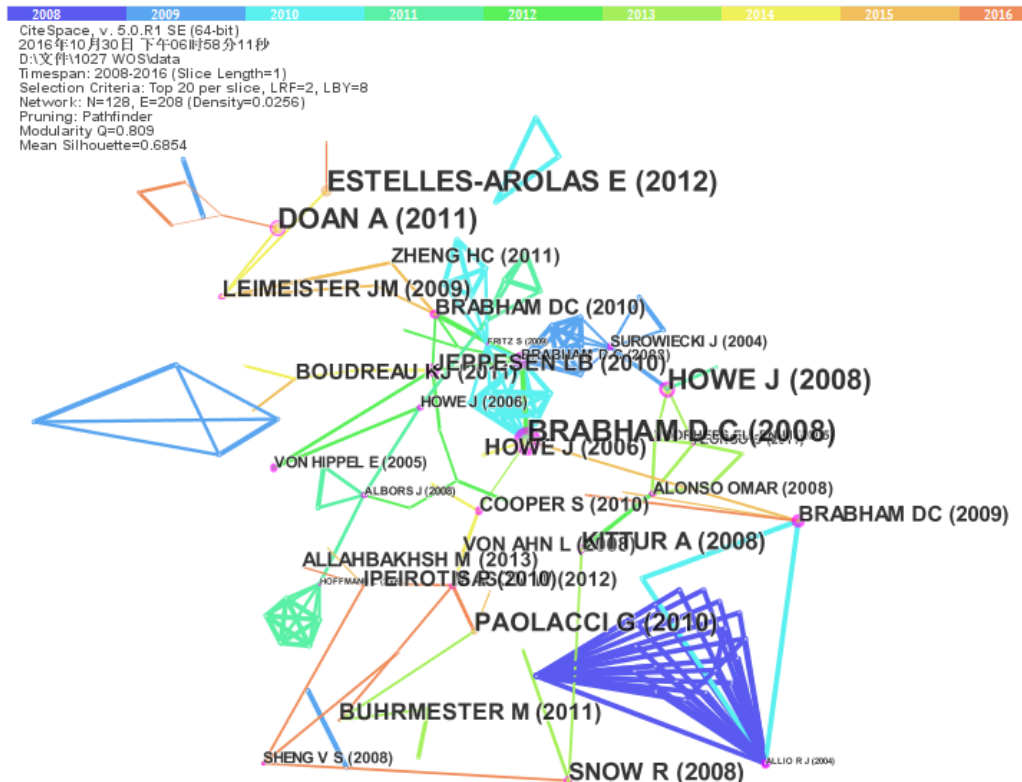

Figure 3: Document Co-Citation Network

Author Co-Citation Analysis

We still employed CiteSpace V in this part to generate a narrative table and a map of author co-citation networks. Table 4 lists co-cited authors in the crowdsourcing research field. It goes beyond doubt that Howe was the pioneering author in the crowdsourcing research field. Howe was cited a total of 173 times, with a centrality of 1.11, indicating that he was the most influential author in this field. Brabham was cited 114 times, with a centrality of 1.15, indicating that he was the most important author in this field. This was consistent with our earlier document co-citation analysis results, that is, Brabham has published several important papers in this field.

Through comparison between Table 3 and Table 4, we found that only four authors' names (namely, Ipeirotis, Surowiecki, Mason, and Von) did not appear in Table 3. On the one hand, all these authors are frequently cited in this field, proving that they are important and influential authors in the crowdsourcing research field. On the other hand, only few authors other than those in the list have carried out ground breaking work in this research field during the past decade. Through literature analysis, we found that the four authors not appeared in Table 4 also had made great contributions to the crowdsourcing research.

Besides the paper co-authored with Paolacci, Ipeirotis has published several papers regarding Amazon Mechanical Turk. For example, Ipeirotis *et al.* (2010) article proposed an algorithm for managing data quality of crowdsourcing services including Amazon Mechanical Turk and carried out an experiment to determine whether the algorithm was effective in various scenarios. Therefore, Ipeirotis's previous research findings were very useful and could help us better understand how crowdsourcing platform was operated, laying a foundation for subsequent studies in the crowdsourcing research field. Surowiecki is also one of influential authors in the crowdsourcing field. In Surowiecki (2006) article, Surowiecki pointed out that decisions made by a group were often better than those made by experts or a single member in the group, attaching great importance to wisdoms of crowds. In addition, Surowiecki (2006) article provided a list of four criteria for defining a wise crowd and summarized three types of main crowd wisdoms. Mason co-authored a paper regarding Amazon Mechanical Turk with Suri in 2012. In Mason and Suri (2012) article, they summarized benefits of the Amazon Mechanical Turk platform for crowdsourcing purposes and illustrated in detail how to conduct crowdsourcing on the Amazon Mechanical Turk platform. It is worth nothing that Mason and Suri (2012) article also explored how to carry out behavioral researches by using the crowdsourcing platform. In a sense, Mason and Suri (2012) article has extended the application scope of crowdsourcing from business scenario to behavioral research domain. Von is one of crowdsourcing pioneers, and he has been committed to research on human computation and

CAPTCHAs. Von Ahn *et al.* (2008) explored how to implement human-based character recognition through CAPTCHAs, aiming to demonstrate that CAPTCHAs are practical web security measures.

TABLE 4: Top 10 Co-Cited Authors in The Crowdsourcing Research

Author	Frequency	Centrality	Institution
Howe	173	1.11	Northeastern University, USA
Brabham	114	1.15	University of Southern California, USA
Kittur	60	0.03	Carnegie Mellon University, USA
Ipeirotis	55	0.62	New York University, USA
Surowiecki	52	0.52	The New Yorker, USA
Doan	45	0.09	University of Wisconsin, USA
Mason	44	0.03	Facebook, USA
Estellés-arolas	43	0	Technical University of Valencia, Spain
Von Ahn	39	0	Carnegie Mellon University, USA
Snow	39	0.31	Stanford University, USA

Looking at Table 3, nine of the top 10 cited authors were from USA, showing that American authors played the dominant role in the crowdsourcing research field. Only Estellés-arolas was from Europe, who is now a PhD crowdsourcing researcher in the Technical University of Valencia, Spain. It is worth noting that though there are a lot of productive Chinese scholars in the crowdsourcing research field, none of them is ranked in the top 10 lists in terms of citation frequency, showing that only few Chinese scholars are influential ones in this field.

By searching for their educational backgrounds, we found that all the top ten co-cited authors have got a doctorate degree, indicating that they have received a series of strict and systematic academic training. In addition, most of these authors have been dedicated to research in the fields of computer science and crowdsourcing for many years.
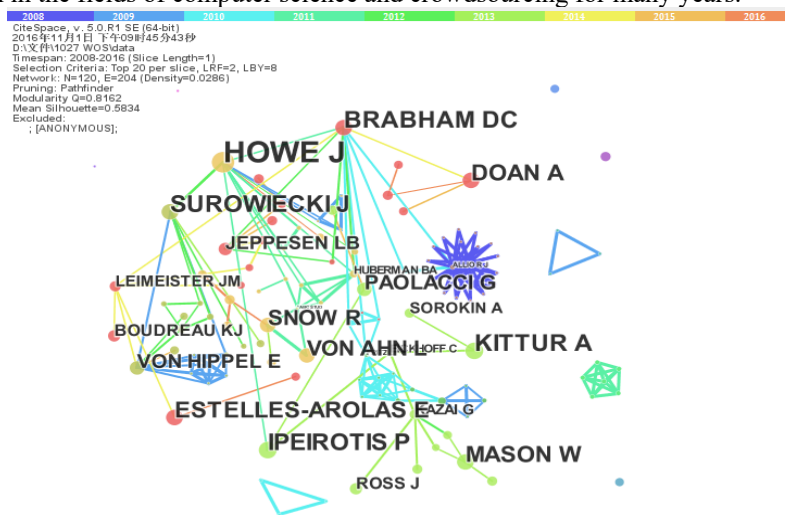

Figure 4: Author Co-Citation Network

Figure4 illustrates a map of the author co-citation network. Excluding the anonymous authors, this network comprises 120 nodes and 204 links, but only authors with a threshold of at least ten citation counts are labeled in the map.

Referring to the map in Figure 4, we could find that most authors are linked closely to Howe and Brabham, showing that Howe and Brabham were the most influential authors in the crowdsourcing research field. In addition, nodes Howe, Brabham, and Surowiecki are linked each other, showing that their research results might have been co-cited during past years. When comparing Figure 3 and Figure 4, we found that there was a blue cluster where nodes were closely linked to Hippel. Hippel has been committed to research on free innovation and user innovation. Hippel (2005) article has been introduced in the document co-citation analysis. As we all know, crowdsourcing is a form of user-driven innovation. In this sense, Hippel, being the initiator of user innovation, is one of important and influential contributors in the crowdsourcing research field. Other authors with at least ten citation counts include Ross, Boudreau, Alonso, Quinn, Raykar, and Afuah. Ross *et al.* (2010) was a proceedings paper regarding crowdworkers. In Ross *et al.* (2010) paper, the demographics of Amazon Mechanical Turk workers were investigated and it was found that population of Amazon Mechanical Turk workers had become more globalized and shifted from a previous US-dominant moderate-income workforce to an India-dominant young and well-educated international workforce. Their research findings indicated that the crowd workforce became more international and the crowdsourcing model would change people's future work and living styles in developing countries. Boudreau co-authored a crowdsourcing related paper with Lakhani in 2013. In Boudreau and Lakhani (2013) article, they identified when crowds were inclined to perform better than employees in the internal organization, and also provided instructions for selecting the best

form of crowdsourcing for a given setting. Their research findings may help enterprises improve efficiency in handling crowdsourcing takes, and are thus of practical significance. In addition, Boudreau has published several papers regarding how to manage outside innovation in an empirical manner. Alonso co-authored a proceedings paper regarding crowdsourcing with Lease in 2011. In Alonso and Lease (2011) paper, opportunities and challenges of the crowdsourcing model and crowdsourcing platforms were analyzed, and a foundation was provided for those who began to apply crowdsourcing in the context of their own particular tasks. Quinn co-authored a proceedings paper regarding human computation with Bederson in 2011. In Quinn and Bederson (2011) paper, they made a comparison between human computation and other related terms including crowdsourcing. As we all know, crowdsourcing and human computation are emerging fields that are intersected in economics and computer science. Therefore, their research findings may facilitate the use of crowdsourcing in solving human intelligence tasks. Raykar is a research scientist in IBM Research, India. Raykar *et al.* published a paper titled *Leaning From Crowds* (Raykar *et al.*, 2010) in Journal of Machine Learning Research. In Raykar *et al.* (2010) article, a probabilistic approach for supervised learning was proposed, and this probabilistic approach was proved by experiment to be better than commonly used methods. In this sense, Raykar *et al.* (2010) article has extended the application scope of crowdsourcing to the field of machine learning. Afuah co-authored an article regarding crowdsourcing with Tucci in 2012. In Afuah and Tucci (2012) article, they proposed certain conditions under which a firm may choose crowdsourcing model to seek solutions to its internal problems.

Based on the above literature analysis results, the foregoing six authors with at least ten citations also have made some contributions to the crowdsourcing research, and were important contributors in this research field.

Disciplinary Distribution Analysis
The purpose of disciplinary distribution analysis is to classify subject categories having a close relationship with the crowdsourcing research field. The function of subject category co-occurrence analysis in CiteSpace program can be used to implement disciplinary distribution analysis. Figure 5 shows the disciplines involved in the crowdsourcing research. For convenience of description, only the disciplines with a frequency threshold of at least 45 were labeled in the map of Figure 5.

Looking at the map in Figure 5, we could find that the disciplinary distribution network comprises 55 nodes and 71 links, indicating that 55 disciplines are distributed in the crowdsourcing research field. The largest node is Computer Science, with a frequency of 215. This result is consistent with our earlier author co-citation analysis results, that is, most of authors in the crowdsourcing research field have been dedicated to the research on computer science for many years. Other disciplines with high frequency include Computer Science and Information Systems (124), Engineering (103), Engineering and Electrical& Electronic (66), Business and Economics (57), Computer Science and Artificial Intelligence (55), Computer Science and Software Engineering (48), Information Science & Library Science (46), and Telecommunications (45). This indicates that crowdsourcing is also an interdisciplinary research field which covers disciplines ranging from computer science, business, management, engineering to information science.
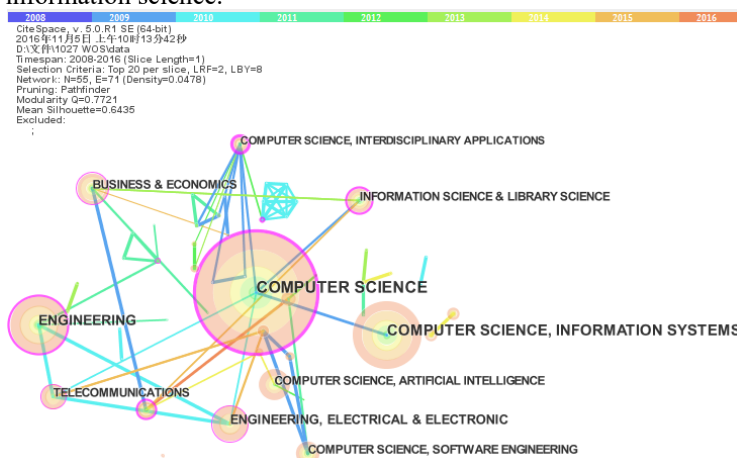

Figure 5: Disciplines Involved in the Crowdsourcing Research

Journal Co-Citation Analysis
Unlike the disciplinary distribution analysis, journal co-citation analysis is conducted to identify cited journals in the crowdsourcing research field. Therefore, through journal co-citation analysis, we can obtain important insights into the journals that jointly constitutes the knowledge base of a research field (Liu *et al.*, 2015).

Table 5 lists the top 10 cited-journals in the crowdsourcing research field. Through careful analysis of this list, we found that *Wired* magazine was the most frequently cited journal for articles in the crowdsourcing research field. This result may be largely attributed to the fact that the term "crowdsourcing" was first proposed in *Wired* magazine. Though the journal *Communications of the ACM* was ranked the third place in the list, it has the highest centrality of 0.86, showing that this journal was the most important source of cited articles in the crowdsourcing research field. In addition, the world's top

academic journals *Science* and *Nature* were also in the top ten list, which demonstrated that scholars in the crowdsourcing research field paid close attention to the world's latest academic trends and tried to establish a link between their own researches and mainstream researches in other fields.

TABLE 5: Top 10 Cited-Journals in The Crowdsourcing Research Field

Frequency	Centrality	Journal	IF in 2015
115	0.85	Wired Magazine	N/A
106	0.28	Lecture Notes in Computer Science	0.8
99	0.86	Communications of the ACM	3.301
75	0.36	Science	34.661
62	0.13	Nature	38.138
58	0.03	Management Science	2.741
56	0	Journal of Information Science	0.878
55	0	Convergence-US	0.845
45	0	PLOS ONE	3.057
42	0	Behavior Research Methods	3.048

By using CiteSpace V, we developed a journal co-citation network comprising 110 nodes and 183 links (see Figure 6). As shown in Figure 6, there are two blue triangular clusters, one of which is linked to *Administrative Science Quarterly* (with an Impact factor of 5.316 in 2015) and the other is linked to *Academy of Management Review* (with an impact factor of 7.288 in 2015). These two clusters comprise a lot of nodes, and are linked to journal *Communications of the ACM* respectively, indicating that these two journals are also the dominant sources of cited articles for crowdsourcing scholars.

When taking a closer look at the map in Figure 6, we also found that besides the nodes labelled with *Wired Magazine*, *Lecture Notes in Computer Science*, *Communications of the ACM*, and *Science*, other nodes with purple rings include *Academy of Management Review*, *Organization Science*, *Journal of Management Information System*, and *Administrative Science Quarterly*, showing that these highly cited journals are also very important sources of cited articles for crowdsourcing scholars. In addition, these four journals are involved in the management discipline, showing that crowdsourcing scholars have kept a close eye on the latest research results in the management discipline.

Furthermore, there are several nodes with a red core (see Figure 6), indicating that these nodes have a strong citation bursts. By using the burst detection function of CiteSpace V, we obtained the top five journals with the strongest citation bursts (see Figure 7). As shown in Figure 7, the journal *International Journal of Electronic Commerce* has the strongest strength (4.4185), with the citation burst beginning from 2014 and ending in 2016, showing that its citation frequency reached 11 times in 2015 but dropped dramatically to zero in 2016. The journal *Organization Science* has the second strongest strength (3.659), with the citation burst beginning from 2014 and ending in 2016, showing that its citation frequency reached 14 times in 2015 but dropped dramatically to zero too. The third to the fifth place are *GeoJournal* (with the burst strength of 3.2635), *Journal of Management Information System* (with the burst strength of 3.1179), and *Administrative Science Quarterly* (with the burst strength of 2.5462).
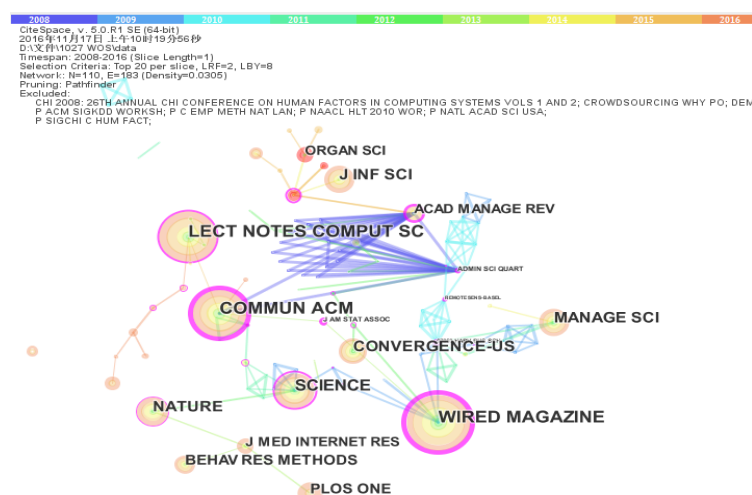

Figure 6: Journal Co-Citation Network

Top 5 Cited Journals with the Strongest Citation Bursts

Cited Journals	Year	Strength	Begin	End	2008 - 2016
ADMIN SCI QUART	2008	2.5462	**2008**	2012	
GEOJOURNAL	2008	3.2636	**2009**	2010	
J MANAGE INFORM SYST	2008	3.1179	**2013**	2016	
ORGAN SCI	2008	3.659	**2014**	2016	
INT J ELECTRON COMM	2008	4.4185	**2014**	2016	

Figure 7: Top 5 Cited Journals with the Strongest Citation Bursts

Keyword Co-Word Analysis

In general, keywords of a published article can reflect core content of the article, and are thus reliable indicators for analyzing hot topics and topic distribution in a research field. A keyword co-word network analysis can help to find out hot research topics and the future directions of a knowledge domain (Liu *et al*., 2015).

We obtained the top 20 keywords (see Table 6) by using CiteSpace V. Since the term ''crowdsourcing'' is used as the search keyword, it is not surprising that the keyword ''crowdsourcing'' appeared most frequently (250 times). Other keywords with a higher appearance frequency include "system" (32), "innovation" (27), "information" (27), "design" (23), "model" (22), "web" (21) and "knowledge" (20). In terms of centrality, there is an interesting finding that centralities of the keywords "Amazon mechanical turk" and "Algorithm" are zero. This indicates that previous hot research topics related to concept's antecedents and applications of the crowdsourcing platform have been gradually losing importance in recent years. In contrast, such keywords as "knowledge", ''participation'', "performance", "motivation", and "citizen science" have a higher centrality, which are mainly related to studies on customers' participation in crowdsourcing. This indicates that the research focus in the crowdsourcing field have shifted from mere applications to customer behaviors and that the application scenario of crowdsourcing has extended from the original business settings to public sectors.

TABLE 6: Top 20 Keywords in the Crowdsourcing Research Field

Frequency	Centrality	Keyword	Frequency	Centrality	Keyword
250	0.13	Crowdsourcing	15	0.11	Performance
32	0.14	System	14	0.11	Citizen science
27	0.1	Innovation	14	0.34	Motivation
27	0.18	Information	13	0.12	Quality
23	0.09	Design	13	0.02	Open innovation
22	0.1	Model	13	0.09	Management
21	0.05	Web	12	0.51	Community
20	0.43	Knowledge	11	0	Algorithm
19	0.09	Network	11	0.07	Internet
16	0.21	Participation	10	0	Amazon mechanical turk

Figure 8 illustrates the keyword co-word network comprising 164 nodes and 313 links. It is notable that the important keywords in the crowdsourcing field were inter-linked closely. As shown in Figure 8, there are three large notable clusters: one cluster is linked to the keyword "community", one cluster is linked to keyword "motivation", and another one is linked to "knowledge", indicating that motivation of users as well as knowledge sharing in the crowdsourcing community has become hotter research topics in recent years. Some small clusters including such keywords as network economy, behavior, human translation, and crowcasting are located far from the center of the map, indicating that some new topics began to appear in the crowdsourcing research field, but did not play a dominant role in the current research domain.
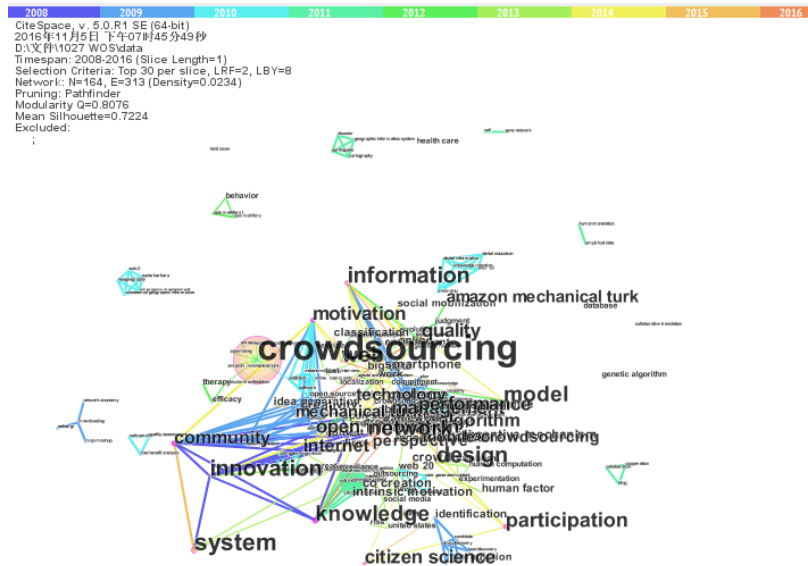
Figure 8: Keyword Co-Word Network

To further detect the evolving trends of the crowdsourcing field, we queried for how often the main keywords in the crowdsourcing research field appeared each year during the time span from 2008 to 2016. For brief description, we chose the keyword "crowdsourcing" as an example. Figure 8 illustrated the history of appearance of the keyword "crowdsourcing".
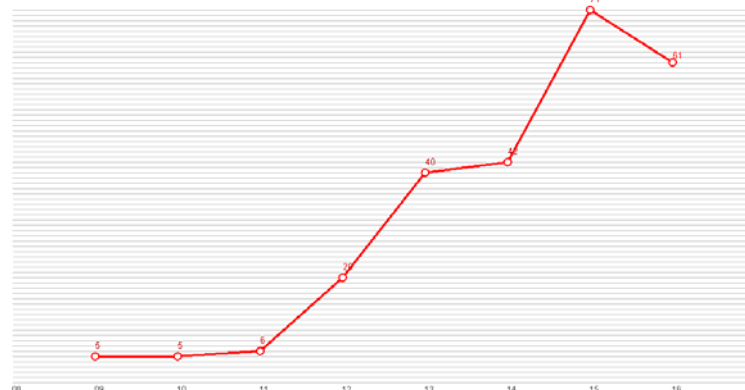

Figure 9: History of Appearance of The Keyword "Crowdsourcing"

As shown in Figure 9, the word "crowdsourcing" began to appear as a keyword in journal articles in 2009. The frequency of its appearance has been on the rise (growing from five times to 71 times) between 2009 and 2015, indicating that research topics regarding crowdsourcing has become increasingly hotter during this period. As of October 27, 2016, the keyword "crowdsourcing" have appeared 250 times totally, indicating that the research on crowdsourcing is more mature compared to the year 2008 when the first research paper was published.

CONCLUSIONS

In this paper, we have conducted a bibliometric analysis on the crowdsourcing studies during the time span from 2008 to 2016 by using CiteSpace V program and *Web of Science ™ Core Collection* database. The bibliometric analysis is mainly conducted on 498 journal articles during the time span from 2008 to 2016, from the perspective of document co-citation, author co-citation, journal co-citation, keyword co-word occurrence, as well as disciplinary distribution, where document co-citation, author co-citation, and journal co-citation analysis helped us identify the important articles, major contributors, and important journals in the crowdsourcing field, while keyword co-word analysis helped us find out hot research topics and the future directions in the crowdsourcing field.

According to our document co-citation analysis results, the top ten important works in the crowdsourcing research field were Brabham (2008) article, Howe (2008) book, Estelles-arolas (2012) article, Doan *et al.* (2011) article, Paolacci *et al.* (2010) article, Kittur and Kraut (2008) article, Snow *et al.* (2008) article, Jeppesen and Lakhani (2010) article, Leimeister *et al.* (2009) article, and Howe (2006) article. Other important papers included Allio (2004) article, Brabham (2009) article, Brabham (2010) article, Cooper *et al.* (2010) article, Hippel (2005) article, and Albors *et al.* (2008) article.

According to our author co-citation analysis results, the most influential authors in the crowdsourcing research field are Howe (who coined the term crowdsourcing for the first time) and Brabham (who was the first scholar to publish scholarly research paper using the term crowdsourcing). Other important authors include Kittur, Doan, Ipeirotis, Surowiecki, Doan, Mason, Estellés-arolas, Von, Snow, Hippel, Ross, Boudreau, Alonso, Quinn, Raykar, and Afuah. Through disciplinary distribution analysis, we found that the top three popular subject categories related to the crowdsourcing research field are Computer

Science, Computer Science and Information Systems, and Engineering and that crowdsourcing was an interdisciplinary research field which covers disciplines ranging from computer science, business, management, engineering to information science. The journal co-citation analysis results shew that *Wired magazine*, *Lecture Notes in Computer Science*, and *Communications of the ACM* were the three dominant sources for cited articles in the crowdsourcing research field. The keyword co-word analysis results provided us insights into the evolving trends of the crowdsourcing research field, showing that the research on crowdsourcing has become more mature and that the research focus and application scope of crowdsourcing have become more extensive and diversified. Most of our analysis results were presented in a visualized and quantitative manner by using the functions of CiteSpace V, which may provide a valuable reference for subsequent researches in the crowdsourcing field.

However, it is worth noting that our research has some limitations. First, our data comprises only journal articles without considering other document types, which may slightly affect the overall analysis results. Second, our analysis results are mainly acquired on the basis of metrics such as citation frequency and centrality, without considering sigma, another important indicator in the CiteSpace. Third, for a better visual effect, we labelled only nodes with a higher threshold in the maps made by CiteSpace V, which may result in loss of some minor details. In our future research, we will continue to focus on research topics regarding crowdsourcing, since this research field is a fairly new area and a lot of research topics can be further explored. In addition, we will conduct bibliometric analysis in other research fields by using more functions of CiteSpace V.

ACKNOWLEDGMENT

REFERENCES

[1] Afuah A. & Tucci C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review*, 37(3), 355-375.

[2] Albors J., Ramos J. C., & Hervas J. L. (2008). New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source. *International Journal of Information Management*, 28(3), 194-202.

[3] Allio R. J. (2004). Ceo interview: The innocentive model of open innovation. *Strategy & Leadership*, 32(4), 4-9.

[4] Alonso O. & Lease M. (2011). Crowdsourcing 101: Putting the wsdm of crowds to work for you. In *Proceedings of the ACM International Conference on Web Search and Data Mining* (pp. 1-2). ACM, Hong Kong, China, February 9-12.

[5] Boudreau K. J. & Lakhani K. R. (2013). Using the crowd as an innovation partner. *Harvard Business Review*, 91(4), 60.

[6] Brabham D. C. (2008). Crowdsourcing as a model for problem solving: An introduction andcases. *Convergence the International Journal of Research Into New Media Technologies*, 14(1), 75-90.

[7] Brabham D. C. (2009). Crowdsourcing the public participation process for planning projects. *Planning Theory*, 8(3), 242-262.

[8] Brabham D. C. (2010). Moving the crowd at threadless. *Information Communication & Society*, 13(8), 1122-1145.

[9] Buettner R. (2015). A systematic literature review of crowdsourcing research from a human resource management perspective. In *Proceedings of the Hawaii International Conference on System Sciences* (pp. 4609-4618). HICSC, Hawaii, USA, January 5-8.

[10] Callon M., Courtial J. P., & Laville F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, 22(1), 155-205.

[11] Chen C. (2009). Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the China Society for Scientific & Technical Information*, 57(3), 359-377.

[12] Chen C. & Carr L. (1999). Trailblazing the literature of hypertext: Author co-citation analysis (1989–1998). In *Proceedings of the Tenth ACM Conference on Hypertext and Hypermedia: Returning To Our Diverse Roots: Returning To Our Diverse Roots* (pp. 51-60). ACM, Darmstadt, Germany, February 21-25.

[13] Cooper S., Khatib F., Treuille A., Barbero J., Lee J., Beenen M., Leaverfay A., Baker D., Popović Z., & Players F. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466(7307), 756-60.

[14] Doan A. H., Ramakrishnan R., & Halevy A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the Acm*, 54(4), 86-96.

[15] Estellés-Arolas E. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189-200.

[16] Feng J. H. & Li G. L. (2015). A survey on crowdsourcing. *Chinese Journal of Computers*, 38(9), 1713-1726.

[17] He Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, 48(1), 133-59.

[18] Hippel E. V. (2005). Democratizing innovation: The evolving phenomenon of user innovation. *Journal Für Betriebswirtschaft*, 55(1), 63-78.

[19] Hosseini M., Shahri A., Phalp K., Taylor J., & Ali R. (2015). Crowdsourcing: A taxonomy and systematic mapping study. *Computer Science Review*, 17, 43-69.

[20] Howe J. (2008) *Crowdsourcing: Why the power of the crowd is driving the future of business*, Crown Business, N.Y.

[21] Howe J. (2006). The rise of crowdsourcing. *Wired*, 14(6), 176-183.

[22] Hu C. P., Hu J. M., Gao Y., & Zhang Y. K. (2011). A journal co-citation analysis of library and information science in china. *Scientometrics*, 86(3), 657-670.

[23] Ipeirotis P. G. (2010). Analyzing the amazon mechanical turk marketplace. *XRDS*, 17(2), 16-21.

[24] Ipeirotis P. G., Provost F., & Wang J. (2010). Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 64-67). ACM, Washington DC, USA, July 25-25.

[25] Jeppesen L. B. & Lakhani K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization Science*, Vol .21(5), 1016-1033.

[26] Kittur A. & Kraut R. E. (2008). Harnessing the wisdom of crowds in wikipedia:Quality through coordination. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 37-46). ACM, San Diego, USA, November 8-12.

[27] Lee P. C. & Su H. N. (2010). Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. *Innovation Management Policy & Practice*, 12(1), 26-40.

[28] Leimeister J. M., Huber M., Bretschneider U., & Krcmar H. (2009). Leveraging crowdsourcing: Activation-supporting components for it-based ideas competition. *Journal of Management Information Systems*, 26(1), 197-224.

[29] Leydesdorff L. (2005). Similarity measures, author cocitation analysis, and information theory. *Journal of the American Society for Information Science & Technology*, 56(7), 769-772.

[30] Liu Z. (2005). Visualizing the intellectual structure in urban studies: A journal co-citation analysis (1992-2002). *Scientometrics*, 62(3), 385-402.

[31] Liu Z., Yin Y., Liu W., & Dunford M. (2015). Visualizing the intellectual structure and evolution of innovation systems research: A bibliometric analysis. *Scientometrics*, 103(1), 135-158.

[32] Mason W. & Suri S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, 44(1), 1-23.

[33] Nerur S. P., Rasheed A. A., & Natarajan V. (2008). The intellectual structure of the strategic management field: An author co-citation analysis. *Strategic Management Journal*, 29(3), 319-336.

[34] Paolacci G., Chandler J., & Ipeirotis P. G. (2010). Running experiments using amazon mechanical turk. *Judgment & Decision Making*, 5(5), 411-419.

[35] Quinn A. J. & Bederson B. B. (2011). Human computation:A survey and taxonomy of a growing field. In *Proceedings of the International Conference on Human Factors in Computing Systems*, (pp. 1403-1412). ACM, Vancouver, Canada, May 7-12.

[36] Raykar V. C., Yu S., Zhao L. H., Valadez G. H., Florin C., Bogoni L., & Moy L. (2010). Learning from crowds. *Journal of Machine Learning Research*, 11(2), 1297-1322.

[37] Ross J., Irani L., Silberman M. S., Zaldivar A., & Tomlinson B. (2010). Who are the crowdworkers?: Shifting demographics in mechanical turk. In *Proceedings of the International Conference on Human Factors in Computing Systems* (pp. 2863-2872). ACM, Georgia, USA, April 10-15.

[38] Snow R., O'Connor B., Jurafsky D., & Ng A. Y. (2008). Cheap and fast---but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 254-263). ACL, Hawaii, USA, October 25-27.

[39] Surowiecki J. (2006). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. *Personnel Psychology*, 59(4), 982-985.

[40] Tsay M. Y., Xu H., & Wu C. W. (2003). Journal co-citation analysis of semiconductor literature. *Scientometrics*, 57(1), 7-25.

[41] Von A. L., Maurer B., Mcmillen C., Abraham D., & Blum M. (2008). Recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895), 1465.

[42] White H. D. & Mccain K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the Association for Information Science and Technology*, 49(4), 327-355.

[43] Zhang L., Zhong F., & Tu H. (2012). A research review on crowdsourcing. *Science & Technology Progress & Policy*, 29(6), 154-160.