DE GRUYTER
OPEN

Research Paper

Detecting Dynamics of Hot Topics with Alluvial Diagrams: A Timeline Visualization

Wenjing Ruan, Haiyan Hou & Zhigang Hu†

WISE Lab, Dalian University of Technology, Dalian 116024, China

Citation: Wenjing Ruan, Haiyan Hou & Zhigang Hu (2017). Detecting Dynamics of Hot Topics with Alluvial Diagrams: A Timeline Visualization.
Vol. 2 No. 3, 2017
pp 37–48
DOI: 10.1515/jdis-2017-0013
Received: Oct. 15, 2016
Revised: Jan. 10, 2017
Accepted: Feb. 7, 2017

Abstract

Purpose: In this paper, we combined the method of co-word analysis and alluvial diagram to detect hot topics and illustrate their dynamics.

Design/methodology/approach: Articles in the field of scientometrics were chosen as research cases in this study. A time-sliced co-word network was generated and then clustered. Afterwards, we generated an alluvial diagram to show dynamic changes of hot topics, including their merges and splits over time.

Findings: After analyzing the dynamic changes in the field of scientometrics from 2011 to 2015, we found that two clusters being merged did not mean that the old topics had disappeared and a totally new one had emerged. The topics were possibly still active the following year, but the newer topics had drawn more attention. The changes of hot topics reflected the shift in researchers' interests. Research topics in scientometrics were constantly subdivided and re-merged. For example, a cluster involving "industry" was divided into several topics as research progressed.

Research limitations: When examining longer time periods, we encounter the problem of dealing with bigger data sets. Analyzing data year by year would be tedious, but if we combine, e.g. two years into one time slice, important details would be missed.

Practical implications: This method can be applied to any research field to illustrate the dynamics of hot topics. It can indicate the promising directions for researchers and provide guidance to decision makers.

Originality/value: The use of alluvial diagrams is a distinctive and meaningful approach to detecting hot topics and especially to illustrating their dynamics.

Keywords Dynamics; Alluvial diagram; Hot topics; Timeline approach

† Corresponding author: Zhigang Hu (E-mail: huzhigang@dlut.edu.cn).

Research Paper

1 Introduction

In the scientific community, literature resulting from academic research forms complex networks that include multiple information streams and huge data sets. As the keywords of literary works directly represent the main content or topics of research, we use keywords provided by authors to detect the hot topics in their fields. We can therefore discern the main focus of research fields and grasp the direction of their development, and such understanding can contribute to scientific advancement.

The term "hot topics" refers to subjects that attract extensive attention from researchers in a short period of time and was addressed frequently in literature (Wan et al., 2015). As the development of knowledge is continuous and fluid, changing focus in hot topics reflects the constant change in knowledge and information gathering. This idea drives us to find patterns in these changes and to detect hot topics from previous research in order to predict the future research direction.

2 Literature Review

Detecting dynamics of hot topics in a research field is central to discerning the core content of that research field. To detect hot topics based on scientific articles, researchers have employed various approaches such as word frequency analysis, co-word analysis, co-citation analysis, and social network analysis.

Word frequency analysis is a basic method of hot topic detection and often combined with other methods. For example, Gong and Ye (2006) analyzed research hotspots using index terms that captured the essence of a topic because they believed that keywords with a high frequency represented academic hotspots in a given field. Xiao (2011) employed the *h*-index and keyword frequency to identify hot topics. In his study, high-frequency keywords above *h*-index corresponded to research hotspots. Co-word analysis is also a common method used to detect hot topics. Courtial, Callon, and Sigogneau (1993) extracted keywords from the titles of food patents and generated the co-word network, allowing for the detection of technology hotspots in the field. Co-citation analysis is another method frequently employed by researchers to identity hot topics. Pan and Qiu (2015) conducted a bibliometric analysis based on co-citation to reflect the most popular topics on student learning within the literature.

Social network analysis (SNA), a popular approach at present, is a quantitative analysis approach based on mathematics and graph theory. It is able to build the social network model from complex literature networks and detect research hotspots, and it is widely used in Sociology, Information Science, Economics, and Management (Otte & Rousseau, 2002). Ding (2011) applied the topology- and

topic-based community detection approaches to the coauthorship networks of information retrieval areas. She suggested that in the future the community detection approach should be used to identify dynamic changes of topics rather than emphasizing the relationships between communities and topics. Li et al. (2015) used classical word frequency analysis and co-word analysis along with centrality analysis and cohesive subgroups analysis to reveal the hot topics of international economic disciplines from 1999 to 2013. Song (2011) applied social network analysis to explore hot research topics and enhance the objectivity of measurements by drawing a global graph of co-citation networks and visualizing the graph's components, bridges, cut-points, *k*-cores, and clusters.

Researchers have used different ways to track topic changes. Tang and Hu (2013) developed an integrative approach to tracking and visualizing the changes of research streams. Using research cohesion score, whose value is determined by the summation of shared keywords, they measured similarities in the focus of research of a pair of articles and found that the triggered research streams diffused via extended co-authorships.

With the development of visual technology in recent years, it is possible to have a more vivid display of hot topics' dynamics. Many kinds of visualization software have been developed to deal with complex data. Among these visualization software, CiteSpace (Chen, 2006), VosViewer (van Eck & Waltman, 2010), and Sci^2 (Sci2 Team, 2009) received the broadest acceptance. Hou et al. (2006) used knowledge maps instead of word frequency to identify research hotspots and trends of research fronts of international science studies. The alluvial diagram is an emerging visualization to reveal the process of mergers and splits of clusters over time. It was designed by Rosvall and Bergstrom in 2010[①], and unfortunately, has not been well known until now. In this study, we used this alluvial diagram to show its functions and features.

3 Data and Methodology

3.1 Data Collection

Publications in the research field of scientometrics were chosen for this study. Data was retrieved from Thomson Reuters's (presently Clarivate Analytics) Web of Science (WoS) on May 16, 2016. To collect the most relevant publications in the field of scientometrics during the past five years, we conducted the literature retrieval in the following steps. First, we searched for articles published in *Scientometrics* since it was founded in 1978, and retrieved a total of 3,901 publications. Second, we collected all data that had cited at least one of these 3,901

① The alluvial diagram tool is available on http://www.mapequation.org/apps/MapGenerator.html.

articles and were published in the period of 2011–2015. Third, these citing articles were restricted to three WoS categories: (1) Information Science & Library Science, (2) Computer Science, Interdisciplinary Applications, and (3) Computer Science, Information Systems. The eligible articles were combined with 1,397 articles that were originally published in *Scientometrics* between 2011 and 2015, and finally we obtained a total of 3,368 articles in the field of scientometrics.

3.2 Alluvial Diagram and PageRank

There are many popular timeline designs to fit different information streams, such as the three-dimensional spiral timeline, chessboard timeline, interaction timeline, relationship timeline, Gantt timeline, and complex timeline. Choosing an appropriate timeline approach depends on user demand. We used the alluvial diagram to demonstrate the dynamics of hot topics and to identify the structural changes of research. Specifically, main clusters in a scientific network at a given time occupy a column in the diagram and are horizontally connected to significant preceding and succeeding clusters by stream fields. We generated a co-word network before importing and generating the alluvial diagram. We then clustered the network and gave a label to each cluster. This approach relies on word profiles derived from articles citing a cluster of co-word articles, based on the assumption that the word profiles characterize the nature of a co-word cluster.

In the alluvial diagram, PageRank is used to reflect the importance of each cluster and word. We not only consider the frequency of keywords, but each keyword's weight using PageRank (Page et al., 1998). PageRank is a link-analysis algorithm which is designed and used by Google to measure the importance of a webpage in the first place. After years of development, this algorithm has been used everywhere in network analysis. Using PageRank, we are able to identify hot topics more accurately than the traditional approach that only considers keyword frequency. PageRank gives each keyword a value of weight. The words used by a paper with high influence will be set a higher weight than those used by an ordinary paper. This idea was actually initiated from citation analysis (Pinski & Narin, 1976), which posits that the number of citations a paper receives can reflect the influence and importance of its research. PageRank algorithm extends this approach by not counting inbound links of all pages equally, but normalizing the number of outbound links and importance of neighboring pages, as shown in Equation (1):

$$PR(p) = (1-d) + d\sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)}, \quad (1)$$

where $PR(p)$ means the PageRank value of page p; T_i ($i=1,2...$) means the inbound links of page p; $C(T_i)$ denotes the number of outbound links on page T_i; $PR(T_i)/C(T_i)$ means the PageRank value that page T_i (the inbound link of page p) gives to page

p; *d* is the damping factor which can be set between 0 and 1 and reflects the probability that a user reaches a page randomly; and (1 – *d*) gives the total PageRank value of all pages as 1. In this study, we set *d* equal to 0.85.

3.3 Data Analysis Process

We first calculated the frequency of keywords and chose the top 100 keywords with frequencies beyond 48. There were 28 significant keywords that needed to be normalized. Concentrating on hyphenated words and singular and plural nouns, we then replaced these keywords with normalized keywords in the original data set. This provided unitive data for further analysis.

4 Results

4.1 Alluvial Diagram of Hot Topic Dynamics

After choosing the top 5% of frequent keywords from the articles, we generated five co-word networks with CiteSpace for years 2011 through 2015. By conducting cluster analysis, we detected hot topics for each year. After that, we drew an alluvial diagram between years, which displayed the dynamics of every cluster. In the alluvial diagram as shown in Figure 1, each block represents a cluster. The height of a block represents the cluster's PageRank value. Blocks were ranked in descending order by PageRank value from the bottom to the top. To do this in a simple way, only the dynamics of the top 13 clusters with the highest PageRank value was shown, such as the cluster of *Industry*, *Google Scholar*, and *Italy*.

With the alluvial diagram in Figure 1, we can intuitively observe the mergers and splits of the top clusters from 2011 to 2015. The height of a stream field represents the flowing nodes' total PageRank value. From this diagram, scientometrics has undergone constant development and changes in recent years. The *Industry* cluster had the highest PageRank value in 2011 and also the highest PageRank value in the whole five-year period. The main research contents of the *Industry* cluster is study of network collaboration among authors, organizations, or nations using scientometric methods, where both the characteristics and influencing factors of collaboration can be identified. This research aims to promote collaboration and drive knowledge flow between industries and universities, and help researchers anticipate and address the demands of industry to develop new technologies. From this hot topic of *Industry*, we are able to confirm that scientometric research has played a significant role in scientific and technological development by assisting in the decision-making issues in recent years. This promotes the research focus on the industry-related research using scientometric methods. Obviously, application-oriented research is at the center of scientometric research, where its scope is still constantly expanding.

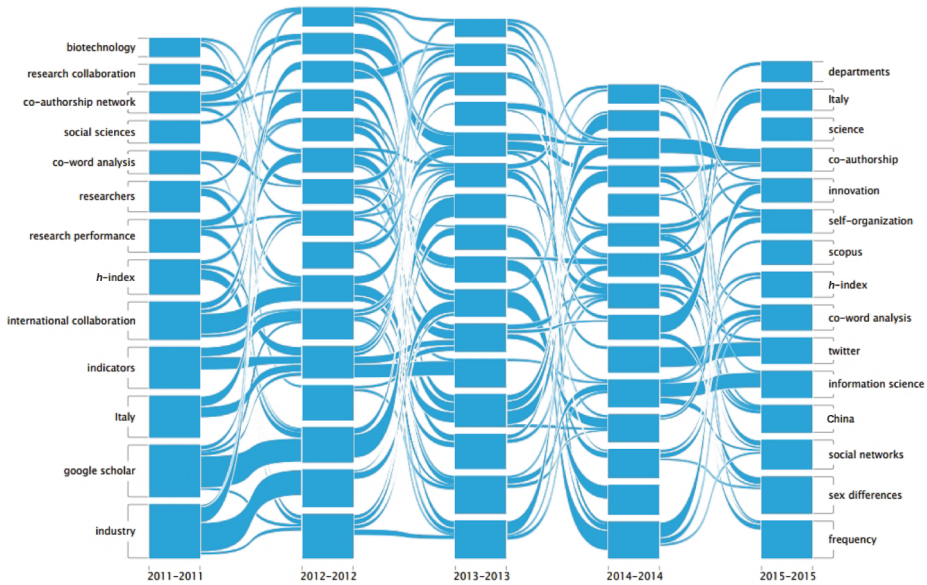

Figure 1. Alluvial diagram of the dynamics of hot topics in the field of scientometrics.

To analyze the importance of every cluster, we listed the name of every cluster and its PageRank value from 2011 to 2015 in Table 1.

As seen in Figure 1 and Table 1, we found that the cluster *Industry* and *Triple helix* had high PageRank values every year from 2011 to 2014. *Triple helix* refers to the three-way, innovative government-industry-university cooperation in this field, originally based on the "three-screw" structure in biology. It implies that research related to industry and government-industry-university collaboration was a hot topic in scientometrics during these years. Furthermore, there were some emerging clusters that showed up in 2015, such as *Sex differences* and *Self-organization*. The cluster of *Sex differences* was an interesting topic because researchers of different sexes had different performances in some scientific fields. Another cluster of *Self-organization* belongs to the scope of complex networks with big data backgrounds. Along with the rapid development of information science, research about complex networks will attract more attention, even in the field of scientometrics, because information science can promote its development.

4.2 Alluvial Diagram of the *Industry*: A Specific Case

The dynamics of the *Industry* cluster, which has the highest PageRank value over the five years, are marked in red in Figure 2, in which all of its topic splits and mergers are presented.

Table 1. PageRank values of the clusters from 2011 to 2015.

	2011		2012		2013		2014		2015	
	Clusters	PR	Clusters	PR	Clusters	PR	Clusters	PR	Clusters	PR
1	#INDUSTRY	11.0%	#PATENT CITATIONS	8.7%	#TERMS	7.4%	#INNOVATION	7.2%	#FREQUENCY	7.3%
2	#GOOGLE SCHOLAR	10.0%	#TRIPLE HELIX	7.3%	#CITATION NETWORKS	7.3%	#PUBLISHERS	5.8%	#SEX DIFFERENCES	7.3%
3	#ITALY	8.2%	#SCOPUS	6.9%	#SCALES	6.8%	#PERCENTILES	5.6%	#SOCIAL NETWORKS	5.8%
4	#INDICATORS	8.2%	#ACADEMIC WEB	6.8%	#COOPERATION	6.4%	#WEBOMETRICS	5.4%	#CHINA	5.4%
5	#INTERNATIONAL COLLABORATION	7.5%	#SCIENCE	6.2%	#SCIENCE	5.5%	#INFORMATION SCIENCE	5.4%	#INFORMATION SCIENCE	5.2%
6	#*H*-INDEX	6.8%	#CLASSIFICATION	5.9%	#RESEARCH PRODUCTIVITY	5.4%	#TWITTER	5.0%	#TWITTER	5.1%
7	#RESEARCH PERFORMANCE	6.5%	#SCIENTIFIC COLLABORATION	5.1%	#INNOVATION	5.3%	#ITALY	4.7%	#CO-WORD ANALYSIS	5.0%
8	#RESEARCHERS	6.1%	#MANAGEMENT	5.0%	#RESEARCH PERFORMANCE	5.0%	#RANKING	4.7%	#*H*-INDEX	5.0%
9	#CO-WORD ANALYSIS	4.5%	#PROXIMITY	4.7%	#MAPS	4.7%	#SCIENTIFIC PRODUCTIVITY	4.5%	#SCOPUS	4.7%
10	#SOCIAL SCIENCES	4.4%	#LIBRARY	4.7%	#SCOPUS	4.6%	#INTERDISCIPLINARY	4.5%	#SELF-ORGANIZATION	4.6%
11	#CO-AUTHORSHIP NETWORK	4.2%	#COOPERATION	4.7%	#KNOWLEDGE	4.6%	#SOCIOLOGY	4.3%	#INNOVATION	4.6%
12	#RESEARCH COLLABORATION	3.9%	#INDEX	4.4%	#CO-AUTHORSHIP NETWORK	4.6%	#PATTERNS	4.1%	#CO-AUTHORSHIP	4.6%
13	#BIOTECHNOLOGY	3.7%	#NANOTECHNOLOGY	4.2%	#*H*-INDEX	4.3%	#CO-AUTHORSHIP	4.0%	#SCIENCE	4.5%
14	#WORD ANALYSIS	3.4%	#DIFFUSION	4.1%	#KNOWLEDGE MANAGEMENT	4.2%	#NANOSCIENCE	3.9%	#ITALY	4.2%
15	#INFORMATION SCIENCE	3.2%	#SOCIAL NETWORK ANALYSIS	4.0%	#CLASSIFICATION	4.0%	#TRIPLE HELIX	3.6%	#DEPARTMENTS	3.9%

Note. PR refers to PageRank value.

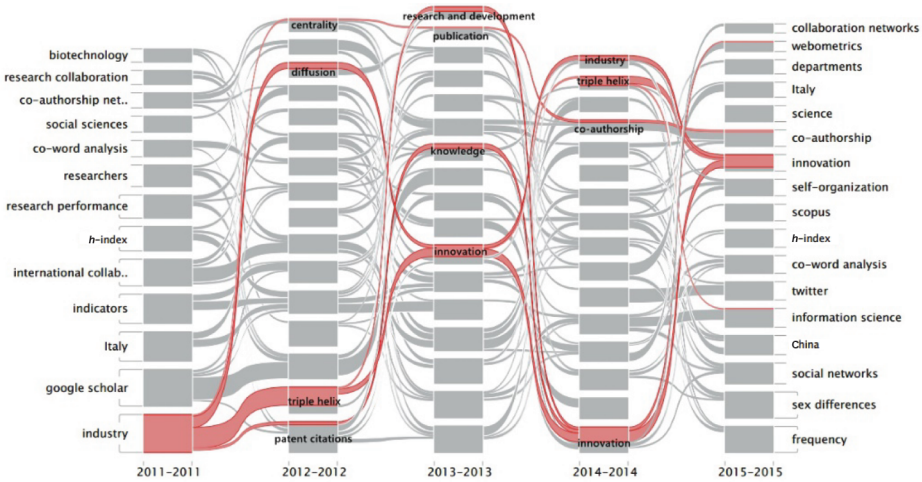

Figure 2. Alluvial diagram of the dynamics of the hottest topic *Industry*.

The *Industry* cluster split into four clusters in 2012 and maintained this status for the next three years only. The reason for this is the cluster was labeled by the keyword within it with the highest PageRank value. The research topics of *Industry* in 2012 became more intensive than in 2011. First, the *Patent citations* cluster in 2012 analyzed the state of technology development based on patent citations. Second, the *Triple helix* cluster in 2014 investigated collaboration among government, industry, and universities, where it became one of the main inflow clusters of *Industry*. Third, the *Diffusion* cluster in 2012 mainly refers to knowledge diffusion on the networks, such as institutional collaboration networks and national collaboration networks. Last, the *Centrality* cluster in 2012 was an important measure in the social network analysis. It had significant implications for identifying the author or institution that occupied the core place of the network. Based on this change, the research contents of the *Industry* cluster became more specific. By analyzing its dynamic changes from 2012 to 2015, we found the steady cluster *Innovation*, which split off from the *Industry* cluster. This finding is in accordance with the current demands placed on technology development and reflects researchers' attention paid to innovation.

In order to see the details of dynamic changes of the *Industry* cluster, we listed the node of every cluster and highlighted its nodes in blue in Figure 3. The keywords of every cluster were listed in descending order based on its PageRank value.

If the keyword is used frequently in 2011 as well as 2012, it will show up in the diagram. Every node represents a dynamic change, and there are no keywords with a high PageRank value during all five years. Although most keywords do show up

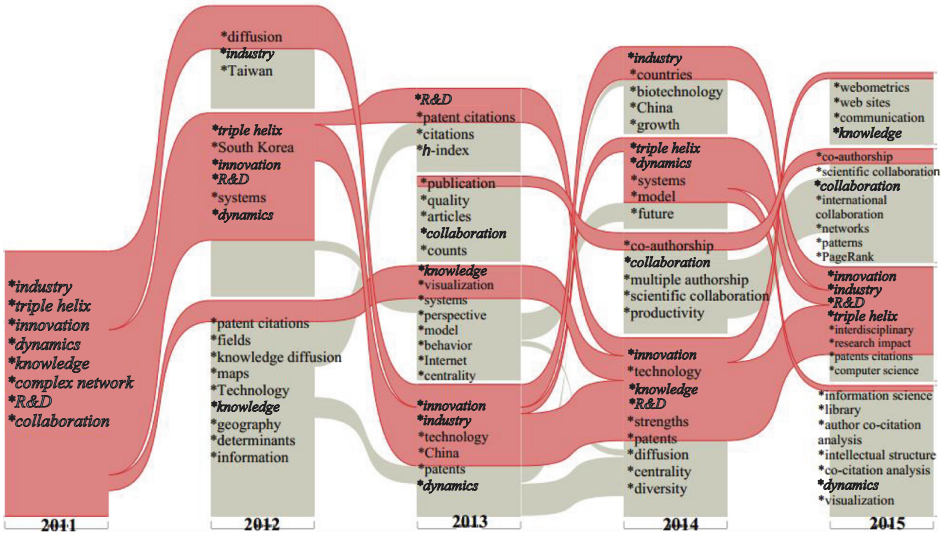

Figure 3. Alluvial diagram of the details of every flow through clusters of the hottest topic *Industry*.

in the following year, they do not have equal importance. In Figure 3, the keyword *industry* did exist in a cluster in 2012, but its PageRank value was lower than that of the keyword *diffusion*. Of the top three keywords with a high PageRank value in 2011, *industry*, *triple helix*, and *innovation*, the latter two were also significant in the following four years, even exceeding *industry*—to become the top two hottest research topics in 2015. In 2014 *Innovation* became the most significant cluster with the highest PageRank value, and in 2015 it split into four keywords from the *Industry* cluster and came to occupy the top four positions. This suggests that research on innovation was a hot topic in 2015, yet its research content was a little different from 2011. The topic *interdisciplinary* also belongs to this *Innovation* cluster, reflecting that interdisciplinary study became a new and hot research topic this year. The three keywords, knowledge, collaboration, and dynamics, flowed into three different clusters in 2015 (Figure 3) and were ranked in a low position in their respective cluster, where they did not cluster with other keywords. This reflects their weak relationship with other keywords, in that they could not represent the main research content of their respective cluster. The *Innovation* cluster thus became the final, significant cluster of the *Industry* cluster, which drew attention of many researchers in 2015.

5 Discussions and Conclusions

After analyzing the dynamic changes in scientometrics from 2011 to 2015, we revealed the pattern of how a cluster was divided and merged over time. The topic

could still be hot the following year, but other topics may have become more popular. As the alluvial diagram revealed, we were able to track a field of interest and which cluster it belonged to the following year. The changes in a research field reflect shifts in researchers' interests and changing research objectives. Because scientometrics is a long-standing disciplinary field in Library and Information Science, a variety of methods can be applied to any discipline to solve research challenges and problems. While classic analytical methods continue to be important to scientometrics, methods coming from Information Science, which are more automated and efficient, are becoming more popular, as they can satisfy the increasing demands of big data processing.

Detecting the dynamics of hot topics means identifying their mergers and splits as research progresses. For example, the *Industry* cluster is subdivided into several topics in the alluvial diagram. Theoretical research attracts a smaller portion of researchers' attention, while the demand for practice research, especially that involves research policy, is continually increasing.

Many researchers have applied various methods to find hot topics, including visualization technology that can highlight the hot fields. But as hot topics are only popular during a certain period of time, the time factor should be taken into consideration when gauging the topics value or influence. In networks, time is not easy to display clearly. The alluvial diagram can show the clusters clearly, while also adding the time slice to every cluster. By transforming the static network to a dynamic alluvial diagram, we can easily figure out the clusters' contexts. Detecting the dynamics of hot topics in a field with the alluvial diagram can thus help researchers quickly ascertain the state of development of a field at the macro-level. Also, we can extract the keywords in each cluster. This allows us to better understand the internal knowledge structure of this field at the meso-level. Finally, we are able to track the changing path of the hot topics, which are reflected by the keywords, at the micro-level, and this can provide direction for researchers and suggestions for decision makers.

There are limitations to this study. The study period was only five years. When examining longer time periods, we encounter the problem of dealing with bigger data sets. Analyzing data year by year would be tedious, but if we combine e.g. two or more years into one time slice, important details would be missed. In our future studies, we aim to find a solution to this problem.

Acknowledgements

This work is supported by the National Social Science Foundation of China (Grant No.: 14BTQ030).

Author Contributions

H.Y. Hou (htieshan@dlut.edu.cn) proposed the research idea, and planned and designed the first outline. W.J. Ruan (vikki0608@163.com) wrote the first draft and performed data analysis. H.Y. Hou, Z.G. Hu (huzhigang@dlut.edu.cn, corresponding author), and W.J. Ruan revised the paper, joined discussion of the findings, and contributed to writing the paper.

References

Chen, C.M. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology, 57(3), 359–377.

Courtial, J.P., Callon, M., & Sigogneau, A. (1993). The use of patent titles for identifying the topics of invention and forecasting trends. Scientometrics, 26(2), 231–242.

Ding, Y. (2011). Community detection: Topological *vs*. topical. Journal of Informetrics, 5(4), 498–514.

Gong, F., & Ye, B. (2006). The hot topics and key words of educational research in China from the year 2000 to 2004. Based on the statistical analysis of CSSCI (in Chinese). Journal of Higher Education, 27(9), 1–9.

Hou, H.Y., Liu, Z.Y., Chen, Y., Jiang, C.L., Yin, L.C., & Pang, J. (2006). Mapping of science studies: The trend of research fronts (in Chinese). Science Research Management, 27(3), 90–96.

Li, M., Zhang, J.L.,Yin, S.Q., & Wang, D.W. (2015). Empirical study on subject research hotspots based on social network analysis: Taking economic subject as an example (in Chinese). Sci-Tech Information Development & Economy, 25(22), 119–122.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. Journal of Information Science, 28(6), 441–453.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web. Technical Report. Stanford InfoLab. Retrieved on January 10, 2017, from http://ilpubs.stanford.edu:8090/422/.

Pan, L., & Qiu, S. (2015). Hot topics and development trends involving researches on student learning on an international scale since 21st Century. Based on analysis of document co-citation of the theme of "student learning" from WOS 2000–2014 (in Chinese). Educational Research, 37(7), 126–135.

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with applications to the literature of physics. Information Processing and Management, 12(5), 297–312.

Rosvall, M., & Bergstrom, C.T. (2010). Mapping change in large networks. Proceedings of the National Academy of Sciences, 5(1), e8694.

Sci2 Team. (2009). Science of Science (Sci2) tool. Indiana University and SciTech Strategies. Retrieved on January 10, 2017, from https://sci2.cns.iu.edu.

Song, G. (2011). Exploring method of measuring and visualizing focuses based on SNA. Journal of Modern Information, 31(5), 46–54.

Tang, L., & Hu, G.Y. (2013). Tracing the footprint of knowledge spillover: Evidence from U.S.–China collaboration in nanotechnology. Journal of the American Society for Information Science and Technology, 64(9), 1791–1801.

Wan, H., Tan, Z.Y., Lu, J.J., & Zhu, X.L. (2015). Summary of the evolution of citation analysis research: 2001–2014 (in Chinese). Library and Information Service, 59(6), 120–136.

van Eck, N.J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538.

Xiao, K. (2011). Application of *h*-index in focus analysis of subject research based on library and information science (in Chinese). Journal of Intelligence, 30(3), 69–73.

This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).