

Topic Detection Based on Weak Tie Analysis: A Case Study of LIS Research

Ling Wei^{1,2,3}, Haiyun Xu^{1†}, Zhenmeng Wang^{1,2}, Kun Dong^{1,2},
Chao Wang^{1,2} & Shu Fang¹

¹Chengdu Documentation and Information Center, Chinese Academy of Sciences,
Chengdu 610041, China

²University of the Chinese Academy of Sciences, Beijing 100049, China

³School of Information Management, Shanxi University of Finance & Economics,
Taiyuan 030006, China

Citation: Ling Wei, Haiyun Xu, Zhenmeng Wang, Kun Dong, Chao Wang & Shu Fang (2016). Topic Detection Based on Weak Tie Analysis: A Case Study of LIS Research.

Received: May 30, 2016

Revised: Aug. 23, 2016

Accepted: Sept. 12, 2016

Abstract

Purpose: Based on the weak tie theory, this paper proposes a series of connection indicators of weak tie subnets and weak tie nodes to detect research topics, recognize their connections, and understand their evolution.

Design/methodology/approach: First, keywords are extracted from article titles and pre-processed. Second, high-frequency keywords are selected to generate weak tie co-occurrence networks. By removing the internal lines of clustered sub-topic networks, we focus on the analysis of weak tie subnets' composition and functions and the weak tie nodes' roles.

Findings: The research topics' clusters and themes changed yearly; the subnets clustered with technique-related and methodology-related topics have been the core, important subnets for years; while close subnets are highly independent, research topics are generally concentrated and most topics are application-related; the roles and functions of nodes and weak ties are diversified.

Research limitations: The parameter values are somewhat inconsistent; the weak tie subnets and nodes are classified based on empirical observations, and the conclusions are not verified or compared to other methods.

Practical implications: The research is valuable for detecting important research topics as well as their roles, interrelations, and evolution trends.

Originality/value: To contribute to the strength of weak tie theory, the research translates weak and strong ties concepts to co-occurrence strength, and analyzes weak ties' functions. Also, the research proposes a quantitative method to classify and measure the topics' clusters and nodes.

Keywords Research topics; Weak tie network; Weak tie theory; Weak tie nodes; Library and Information Science (LIS)



JDIS

Journal of Data and
Information Science

Vol. 1 No. 4, 2016

pp 81–101

DOI: 10.20309/jdis.201626

<http://www.jdis.org>

[†] Corresponding author: Haiyun Xu (E-mail: xuhu@clas.ac.cn).

1 Introduction

Analyzing the current research status of a certain disciplinary subject can identify research directions and other implications for researchers in the field and promote the discipline's development. As a typical interdisciplinary discipline, library and information science (LIS) has been widely studied by scholars both in and outside China in relation to topic detection by using bibliometric methods such as word frequency statistics, co-word analysis, and knowledge mapping. Co-word analysis is a technique for discovering the linkages and associations among projects through the analysis of the co-occurrence frequency of pairs of word or noun phrases (Lee & Jeong, 2008). According to the co-occurrence strength, keywords are further classified to sum up the research focus, structure, and paradigm of a discipline by cluster analysis or other methods (Sedighi & Jalalimanesh, 2014). Co-word analysis has been used by many researchers to explore the research topics in different subject areas such as information retrieval (Ding, Chowdhury, & Foo, 2001), medical informatics (Wagner & Leydesdorff, 2005), international scientific studies, (Hou et al., 2006), management science (Yue, 2012), knowledge management (Sedighi & Jalalimanesh, 2014), and LIS (Chen et al., 2015; González-Alcaide et al., 2008; Guo et al., 2015; Jiang & Zhan, 2008; Liao, 2009; Qiu, & Lv, 2013; Xiao, Li, & Yuan, 2011).

Previous research has basically focused on the strong co-occurrence strength between keywords, but has paid little attention to the weak co-occurrence strength between keywords. The strong co-occurrence strength between two nodes reflects the close relationship of the topics. Such strong ties are important knowledge dissemination channels (Szulanski, 1996), which can efficiently promote the transfer of complex knowledge (Podolny & Baron, 1996). Recent research shows that strong ties are more important in internal knowledge sharing of knowledge-based subgroups (Poleacovschi & Javernickwill, 2015). From the view of interdisciplinary studies, strong ties are important in promoting knowledge dissemination in the same or related disciplines.

In this study we are more interested in the weak ties across disciplines. Weak ties theories (Granovetter, 1973 & 1983) describe how weak ties enable the flows of information between different groups, especially the flows of novel resources and information (Baer, 2010; Burt, 2004; Poleacovschi & Javernickwill, 2015). The weak co-occurrence strength between keywords stands for the weak ties between the topics. Theoretically, such weak ties are important for improving the breadth and depth of knowledge diffusion, especially the knowledge diffusion of interdisciplinary sciences. It is therefore meaningful to investigate the roles and functions of weak ties between topics to see how knowledge diffuses and combines, and how these combinations change.



In our previous study (Wei et al., 2015), we conducted a preliminary topic detection study based on weak tie analysis. While the weak ties between nodes are identified manually, partly, and qualitatively, the internal and external ties are not visualized clearly, nor are nodes and clusters discussed. As a follow-up to this research, our current study focuses on three questions: How do we pick out all the external weak ties between clusters? How do we quantitatively measure the roles and functions of nodes and clusters? What can we learn about interdisciplinary research based on the above discussion? This research contributes to the literature by offering a quantitative method to detect important research topics as well as their roles, interrelations, and evolution trends through translating weak and strong ties concepts to co-occurrence strength and analyzing the different types of weak ties' functions.

We begin by reviewing the principles behind tie strength and then discuss its proposed dimensions. Using the theory to support our definitions of weak subnets and weak nodes, we present a series of indicators to measure the roles and functions of the subnets and nodes. We end by discussing our main findings and summing up limitations and future work related to the research.

2 Weak Tie Theory

The weak tie theory, namely, the theory of the “strength of weak ties,” is a social network theory put forward by Granovetter (1973; 1983) and developed by Kavanaugh and Reese (2005) and Easley and Kleinberg (2010). It was used in its early stage to study interpersonal relations networks from the sociological perspective, and has been widely applied in recent years to topics such as social studies (Sharone, 2014; Zenou, 2015), economic management (Aubert, Léger, & Larocque, 2012; Takagi & Toyama, 2008), and computer science (Zhao, Wu, & Xu, 2010). Scholars worldwide have also conducted extensive research on LIS fields such as knowledge diffusion (Genius, 2005), scientific cooperation (Abbasi, Altmann, & Hossain, 2011; Bettoni & Bernhard, 2008; Yang, Morris, & Barden, 2009), frontier detection (Zhang, 2011), and open access (Li, Sheng & Wei, 2015; Pan & Sheng, 2014). So far, few studies have addressed topics detection based on weak ties, and prior research has generally provided only a qualitative description about weak ties and their possible functions, with little attempt to address quantitative analysis. This paper aims to bridge the gap, using a quantitative method to analyze research topics based on weak tie analysis.

Granovetter (1973) proposed four tie strength dimensions: amount of time, intimacy, intensity, and reciprocal service. Wellman and Wortley (1990) argued that providing emotional support, such as offering advice on family problems, indicates



a stronger tie. Burt (2004) proposed that structural factors such as network topology and informal social circles shape tie strength. Gilbert and Karahalios (2009) presented a predictive model that maps social media data to tie strength, and tested the seven dimensions of tie strength suggested by the existing literature. They found that intimacy makes the greatest contribution to tie strength. Gilbert (2009) also mentioned that threshold value can be used to define strong and weak ties. Sun et al. (2013) suggested using the link weight to measure the strength of social networks, where the links with higher weight means closer relationships, namely stronger ties, while the link with lower weight means weak ties.

Since the co-occurrence frequency of keywords in our research reflects their intensity, which is also the link weight of the co-occurrence network, it is reasonable to distinguish weak ties and strong ties by setting threshold value based on the co-occurrence frequency. Our work introduces a method to obtain a network consisting only of weak ties and nodes, and can quantitatively analyze the topics, roles, and functions of the weak ties and nodes.

3 Methodology

Before introducing the main steps of the research, it is necessary to clarify several terms used in the paper, noted below.

- 1) Strong tie and weak tie: According to the preamble analysis, we divide all the co-occurrence relationships of keywords into two classes by a threshold value, where those with frequency higher than the threshold are strong ties, and those with frequency lower than the threshold are weak ties;
- 2) Weak tie co-occurrence network and weak tie network: We define a network as a weak tie co-occurrence network obtained by filtering out all strong ties of the co-occurrence network generated through the keywords' co-occurrence matrix. Clusters and nodes included remain unchanged, and isolated nodes barely appear because of the internal links in each cluster. In order to focus on the weak ties between clusters, we remove all internal lines of each cluster to get a weak tie network. In the weak tie network, only the links between different clusters are left; if these lines are removed, the subnets will be independent from each other; and
- 3) Weak subnets and weak nodes: In the final weak tie network, all nodes are called weak nodes, and all subnets are called weak subnets.

The research ideas and main steps for the weak tie analysis on research topics detection are detailed below (Figure 1).



- 1) Selection of data and keywords: In order to detect the LIS topics, articles in LIS are collected, and keywords are extracted from article titles and preprocessed by the text analysis tool Thomson Data Analyzer (TDA);
- 2) Generation and clustering of co-occurrence networks: After data preprocessing, the top 300 high-frequency keywords are selected to generate a co-occurrence matrix and co-occurrence network using the social network analysis tools Ucinet and Gephi. When separating clusters, the Louvain community detection algorithm embedded in Gephi is applied, and the default value 1.0 is taken as the threshold;
- 3) Extraction of weak tie co-occurrence network: The high-frequency keyword co-occurrence network is filtered to a weak tie co-occurrence network on the premise that, the weak tie co-occurrence network should keep the basic characteristics of the original network, but not be too sparse. After several attempts, the nodes with degrees less than five, and the lines with weights below three or above 10 are removed;
- 4) Extraction of weak tie network: By removing all internal lines of each cluster, the weak tie network is extracted from the weak tie co-occurrence network;
- 5) Building of indicators: In order to analyze the roles and functions of weak subnets and weak nodes, a series of connection indicators are proposed; and
- 6) Analysis of subnets and nodes: In the last step, we try to find the answers to our research questions by analyzing the indicators of weak subnets and weak nodes.

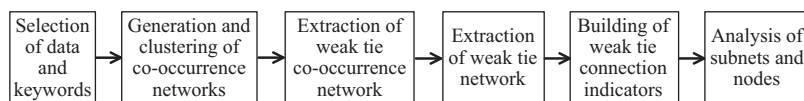


Figure 1. Research ideas and main steps for data analysis.

4 Data and Results

4.1 Data

As a comprehensive and general scientific research platform, Web of Science integrates a variety of databases that include a large amount of high-quality and multidisciplinary research literature. As a typical interdisciplinary field, library and information science (LIS) contains a wide variety of research topics that may create a large amount of weak ties. This paper takes LIS literature in SCI-EXPANDED, SSCI, CPCI-S, CCR-EXPANDED, and IC as data sources, and constructs the retrieval “WC = Information Science & Library Science” in selected “article” papers, creating a total of 37,769 records. The date of retrieval is July 25, 2014 and the time span is 2001–2014.



4.2 Indicators

4.2.1 Centrality Indicators

To understand networks and their participants, we evaluate the location of nodes in the networks. Measuring the network location requires determining the centrality of a node. There are three commonly used centrality measures that we focus on: degree centrality (Freeman, 1978; Wasserman & Faust, 1997), closeness centrality, and betweenness centrality (Brandes, 2004). In terms of this paper, degree centrality is the number of other nodes connected directly to a node, which is calculated by the number of that node's adjacent nodes. Closeness centrality is a measure of the degree to which a node is near all other nodes, defined as the "sum of reciprocal distance" of that node to any other nodes. The closer a node is to another node, the larger the measure is; the farther a node is to another node, the smaller the measure is. Betweenness centrality is an indicator of a node's centrality in a network, and is equal to the number of shortest paths between all vertices that pass through that node, and thus represents the degree of centralization of the node. A node with a high level of betweenness centrality strongly influences the transfer of items through the network, assuming the transfer follows the shortest paths (Freeman, 1977). In sociological terms, it measures the extent to which actors control resources.

In order to investigate the weak nodes' constitution and functions, this paper selects degree centrality as the main index and betweenness centrality as an auxiliary index. According to Gephi statistics, the two indices of most nodes have a positive correlation, and only a few betweenness centrality nodes have irregular changes. Nodes in a same subnet are displayed in the same color, where node size is consistent with degree centrality measure; the larger the value is, the bigger the node is. Links between different subnets are indicated in different colors, where the darker the color is, the more weights the lines have (Figures 2–5).

4.2.2 Weak Connection Indicators

In this study we define a series of connection indicators of weak subnets and nodes. Basically, the indicators are based on the degree centrality indicators. We also take the connection coverage of clusters and nodes into account to compare their connection strengths in the weak tie network. The types and indicator values of weak subnets and nodes are listed in Tables 1 and 2.

(i) Weak tie connection indicators of subnets: This section consists of two parts: subnet connection breadth indicator and subnet importance indicator (Table 1). The former measures how broadly one subnet connects the others, while the latter measures how important one subnet is in the whole weak tie network.



Indicator 1: subnet connection breadth (SCB) is the ratio of the sum of all edge nodes' degree centrality in the weak tie subnet to the sum of all nodes' degree centrality in the corresponding weak tie co-occurrence subnet. The higher the ratio is, the more other subnets that one subnet connects to, which means that the research topics in the subnet are relatively dispersive. The lower the ratio is, the more co-occurrence the internal nodes have, which means that the research topics in the subnet are relatively concentrated.

Indicator 2: subnet importance (SI) is the product of the subnet nodes' average connection strength and subnet connection density. The subnet nodes' average connection strength is the ratio of the sum of all edge nodes' degree centrality to the sum of all nodes' degree centrality in one subnet. It measures the average connectivity of a subnet's nodes. Subnet connection density is the ratio of the number of edges in one subnet to the whole network's edges. It measures the overall connectivity of the subnet. The higher the product is, the more important the subnet is in the whole network.

Table 1. Subnet types and indicators.

Indicator Subnet type	SI	SCB
Core subnet	High	High
Important subnet	Moderate	Moderate
Dense subnet	Low	Low

Note. SI refers to subnet importance and SCB refers to subject connection breadth.

According to the statistics and indicator values, taking indicator 2 as the main index and indicator 1 as an auxiliary index, we divide subnets into three types based on empirical observations:

- 1) Core subnets, which have high value in indicator 1 and relatively high value in indicator 2, where most nodes connect to external nodes. This type of subnet is in the core position of the whole network;
- 2) Important subnets, which have moderate value in both indicators 1 and 2. This type of subnet is a pivotal part of the whole network; and
- 3) Dense subnets, which have a lower value in both indicators 1 and 2. This type of subnet is at the edge of the entire network.

(ii) Weak tie connection indicators of nodes: We have indicator 3, the weak connection strength of nodes (WCS), which is defined as the ratio of one edge node's degree centrality value in the weak tie subnet to its degree centrality value in the weak tie co-occurrence subnet (Table 2). It measures the single node's connectivity, where the higher the ratio is, the more external nodes one node connects to.



Table 2. Node types and indicators.

Indicator \ Node type	Degree centrality	Betweenness centrality	WCS
Core node	Above 10	High	High
Important node	Between 5 and10	Moderate	Moderate
Common node	Below 5	Moderate/low	Low
Special node		Changed dramatically	

Note. WCS refers to the weak connection strength of nodes.

According to the statistics and index value, taking degree centrality as the main index and betweenness centrality and indicator 3 as auxiliary indices, we divide weak nodes into four types based on empirical observations:

- 1) Core weak nodes, which have high degree centrality (above 10) and relatively high WCS, and connect a large amount of nodes in and outside the subnet and thus play important roles in both the weak tie co-occurrence network and the weak tie network. Some links made up of these nodes are assigned more weight values, embodying the connection of important research topics;
- 2) Important weak tie nodes have relatively high degree centrality (between 5 and 10) and moderate WCS. Links between this kind of node are assigned lower weights, and are main components of the weak tie network;
- 3) Common weak tie nodes have the lowest degree centrality (below 5) and low WCS. Numerous common nodes interconnect with each other weakly, indicating special or novel research topics; and
- 4) Special weak tie nodes have betweenness centrality that decreases or increases dramatically. The nodes with decreased betweenness are more important in the weak tie co-occurrence network than in the weak tie network, meaning that these nodes are primarily connected internally, while nodes with increased betweenness are connected externally.

4.3 Results

Due to the space constraints and mass of data used for this paper, we do not analyze the results year by year. Because the retrieval date is July 25, 2014, the data of year 2014 are not complete. Besides, the differences between data of two years are small. As a result, we select year 2013 as the deadline, set the time span as two years, and focus on the years 2007, 2009, 2011, and 2013.

4.3.1 Weak Tie Analysis of Subnets and Nodes of 2007

The high-frequency keyword weak tie network of 2007 is clustered into five subnets, and includes 180 edges, 58 total nodes, 57 edge nodes, and 1 isolated node (Figure 2). The subnets are “user information seeking” (displayed in blue, with ID



0, 11 nodes), “bibliometric analysis” (displayed in red, with ID 1, 9 nodes), “communication techniques” (displayed in green, with ID 2, 12 nodes), “digital libraries” (displayed in grass green, with ID 3, 10 nodes), and “empirical investigation” (displayed in purple, with ID 4, 16 nodes). The biggest node’s label in each subnet indicates the topic of the subnet, and each node represents one specific subtopic. The indicator values of subnets and nodes are listed in Tables 3 and 4.

(i) Weak tie analysis of subnets: Subnets of “user information seeking,” “communication techniques,” and “empirical investigation” have more weak ties, while the other two subnets “digital library” and “bibliometric analysis” are at the edge of the entire network with much fewer weak ties (Figure 2).

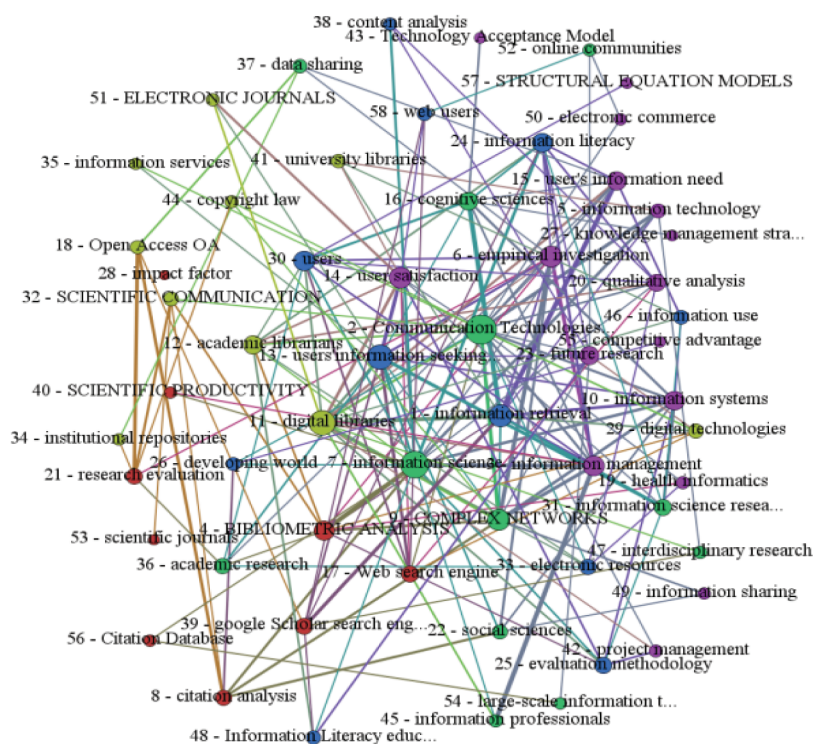


Figure 2. 2007 weak tie network of library and information science.

Subnets of “communication techniques” and “user information seeking” are core subnets, with a subnet importance of 4 and 3.48, respectively, and subnet connection breadth of 62.42% and 65.87%, respectively (Table 3). “Empirical investigation” is an important subnet, with a subnet importance of 2.81 and a connection breadth of

Research Paper

47.37%. The remaining two, “digital library” and “bibliometric analysis,” are dense subnets, and their connection breadth is both above 50%, with an importance of 1.56 for “digital library” and 1.17 for “bibliometric analysis.” The research topics of dense subnets cross and overlap to some degree, where the trend of concentration is obvious. Some topics of the “digital library” subnet are “academic library,” “scientific communication,” “open access,” and “institutional repository.” Typical nodes of the “bibliometric analysis” subnet are “academic information-seeking engines,” “citation analysis,” and “scientific output.”

Table 3. Subnet indices in 2007.

Subnet ID	Subnet label	SI	SCB	Subnet type
2	Communication techniques	4.00	62.42%	Core subnet
0	User information seeking	3.48	65.87%	Core subnet
4	Empirical investigation	2.81	47.37%	Important subnet
3	Digital library	1.56	55.41%	Close subnet
1	Bibliometric analysis	1.17	53.00%	Close subnet

Note. SI refers to subnet importance and SCB refers to subject connection breadth.

(ii) Weak tie analysis of nodes: This section focuses on the first type of weak tie connection nodes based on the weak tie network indices of 2007: core weak tie nodes (Table 4).

Table 4. Node indices in 2007 (partial list).

Node ID	Node label	Degree centrality	Betweenness centrality	SCB
2	Communication techniques	21	324.33	72.41%
7	Information science	18	224.33	69.23%
13	User information seeking	16	152.66	69.57%
11	Digital library	16	116.68	69.57%
1	Information retrieval	13	76.79	59.09%

Note. SCB refers to subject connection breadth.

Ranking the network indices of all nodes, core weak nodes are found at the top. The degree centrality and betweenness centrality of most nodes indicate a positive correlation. After making a detailed analysis of the top 10 nodes, we find that “communication techniques,” “information science,” and “complex network” are the most central nodes of the “communication techniques” subnet, in particular the former two nodes, ranking as the top two. The top three are “user information seeking behavior,” “information retrieval,” and “users,” with the latter two belonging to the same subnet. Note that “digital library” is among the top four and “empirical investigation,” “user satisfaction,” and “information management” are in the same subnet.



The core weak tie nodes of different subnets are frequently connected, where some weak links bear heavy weight. For example, “user information seeking” is linked to 16 nodes of the other four subnets, such as “Web seeking engine” and “Google scholar seeking engine” of the “bibliometric analysis” subnet, “information retrieval” and “complex network” of the “communication techniques” subnet, “digital library” of the “digital library” subnet, and “empirical investigation” and “user satisfaction” of the “empirical investigation” subnet. Among all the links, there are four heavy links whose weak tie weight is above 5.

Identifying these kinds of nodes and links can help detect the connections between primary research topics more clearly and intuitively.

4.3.2 Weak Tie Analysis of Subnets and Nodes of 2009

The high-frequency keywords weak tie network of 2009 is clustered into five subnets, and includes 216 edges, 93 nodes in all, 79 edge nodes, and 14 isolated nodes (Figure 3). The subnets are “information management” (displayed in purple,

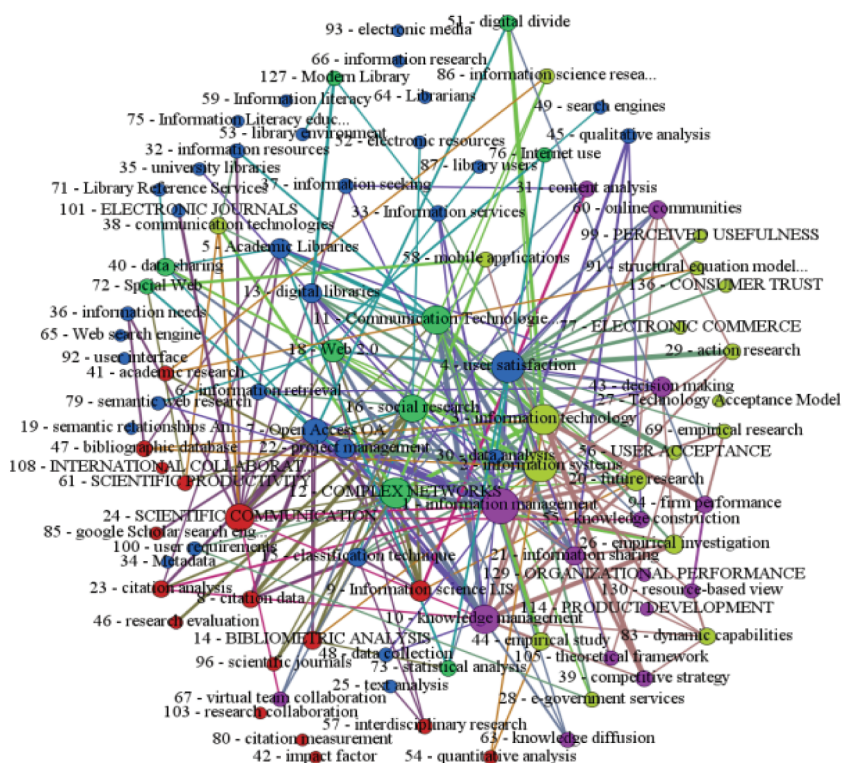


Figure 3. 2009 weak tie network of library and information science.



Research Paper

with ID 0, 15 nodes), “user satisfaction” (displayed in blue, with the ID 1, 33 nodes), “complex network” (displayed in green, with ID 2, 10 nodes), “information system” (displayed in gray yellow, with ID 3, 18 nodes), and “scientific communication” (displayed in red, with ID 4, 17 nodes). The indicator values of subnets and nodes are listed in Tables 5 and 6.

(i) Weak tie analysis of subnets: This section focuses on subnet indices for 2009 (Table 5). Subnets of “information management,” “user satisfaction,” and “information system” have many more weak links. “Complex network” and “information management” are core subnets, with a subnet importance of 3.42 and 2.67, respectively, and a subnet connection strength of 72.88% and 60%, respectively; “information system” is an important subnet, with a subnet importance and connection strength of 2.4 and 44.98%, respectively; “user satisfaction” and “scientific communication” are dense subnets, with a subnet importance in both below 2, and a subnet connection strength between 30% and 40%. Research topics of “user satisfaction” tend to focus on items such as “libraries,” “open access,” and “information sources and information services,” while “scientific communication” focuses on items such as “bibliometric analysis,” “citation data,” “citation analysis,” and “scientific output and evaluation.”

Table 5. Subnet indices in 2009.

Subnet ID	Subnet label	SI	SCB	Subnet type
2	Complex network	3.42	72.88%	Core subnet
0	Information management	2.67	60.00%	Core subnet
3	Information system	2.40	44.98%	Important subnet
1	User satisfaction	1.97	30.98%	Dense subnet
4	Scientific communication	1.28	40.27%	Dense subnet

Note. SI refers to subnet importance and SCB refers to subject connection breadth.

(ii) Weak tie analysis of nodes: This section focuses on the second type of weak connection nodes based on weak tie network indices of 2009: important weak nodes (Table 6).

Table 6. Node indices in 2009 (partial list).

Node ID	Node label	Degree centrality	Betweenness centrality	SCB
40	Knowledge sharing	6	88.45	75.00%
43	Decision support	8	32.88	67.00%
20	Future research	9	22.23	56.12%
5	Academic library	8	37.43	31.00%
9	Library and information Science	9	55.29	47.37%

Note. SCB refers to subject connection breadth.



In the rank of network indices, important nodes are behind the core nodes, with a relatively high degree centrality and betweenness centrality that not only link core nodes, but also a large amount of common nodes. While important nodes play a pivotal role in ensuring successful information communication in the network, most weak links are assigned a small weight. For example, “academic digital” of the “user satisfaction” subnet is linked to eight nodes of the other four subnets. Among these nodes, “communication techniques” and “scientific communication” are core nodes, “decision support” is an important node, and “bibliometric data” is a common node. Among the links, only the link with “communication techniques” has a high weight value (7), and the other links have a small weight value (3 or 4).

Identifying these kinds of nodes and links can help detect and summarize the connections between key research topics with more comprehension.

4.3.3 Weak Tie Analysis of Subnets and Nodes of 2011

The high-frequency keywords weak tie network of 2011 is clustered into six subnets, which includes 38 edges, 46 nodes in all, 29 edge nodes, and 17 isolated nodes (Figure 4). The subnets are “information technology” (displayed in sky-blue, with ID 0, 13 nodes), “information retrieval” (displayed in yellow, with ID 1, 8 nodes), “social network” (displayed in purple, with ID 2, 5 nodes), “digital library”

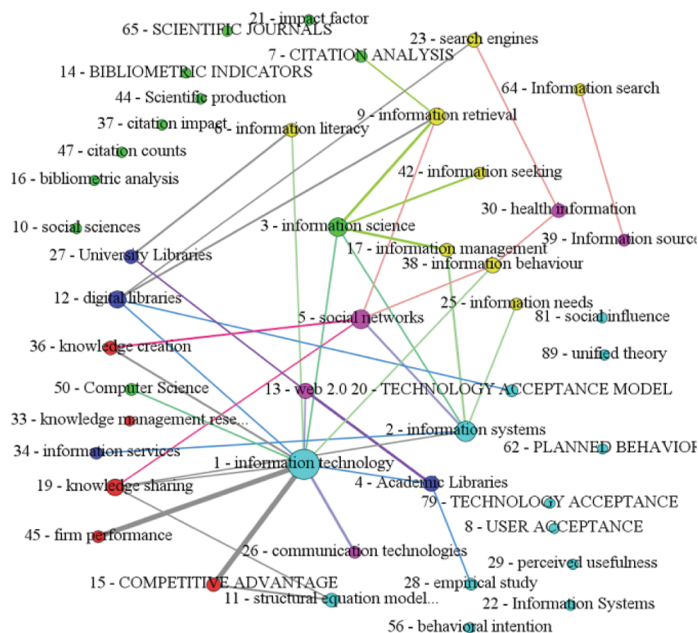


Figure 4. 2011 weak tie network of library and information science.

Research Paper

(displayed in blue, with ID 3, 4 nodes), “knowledge sharing” (displayed in red, with ID 4, 5 nodes) and “information science” (displayed in green, with ID 5, 11 nodes). The indicator values of subnets and nodes are listed in Tables 7 and 8.

(i) Weak tie analysis of subnets: The weak co-occurrence network of 2011 is a small network with a larger number of subnets, but a fewer number of nodes inside, where nodes co-occur frequently. We identify the weak tie network by moving the internal links of subnets. It is very sparse with 17 isolated nodes, and the network indicator values are quite small.

Among the six subnets in 2007 (Table 7), “information technology” stands out as a core subnet with an importance of 2.55, “information retrieval” and “social network” are important subnets with an importance of between 0.7 and 1, and the other three are dense subnets. “Digital library” focuses on the kinds of libraries, “knowledge sharing” focuses on “enterprise knowledge creation and management” as well as “comparative advantage,” while “information science” focuses on items such as “citation analysis,” “impact indices,” “bibliometric indices,” and “scientific output.”

Table 7. Subnet indices in 2007.

Subnet ID	Subnet label	SI	SCB	Subnet type
0	Information technology	2.55	38.00%	Core subnet
1	Information retrieval	0.84	33.33%	Important subnet
2	Social network	0.76	54.55%	Important subnet
3	Digital library	0.66	45.45%	Dense subnet
4	Knowledge sharing	0.53	56.25%	Dense subnet
5	Information science	0.32	21.88%	Dense subnet

Note. SI refers to subnet importance and SCB refers to subnet connection breadth.

(ii) Weak tie analysis of nodes: This section focuses on the third type of weak connection nodes based on weak tie network indices of 2011: common weak tie nodes (Table 8).

Table 8. Node indices in 2011 (partial list).

Node ID	Node label	Degree centrality	Betweenness centrality	SCB
11	Structural equation model	2	1.75	20.00%
6	Information literacy	2	11.00	40.00%
30	Heath information	2	2.10	50.00%
27	University library	2	0.50	40.00%
15	Comparative advantage	2	7.70	50.00%
7	Citation analysis	1	0.00	10.00%

Note. SCB refers to subject connection breadth.



Common weak tie nodes have the largest number, yet they have small degree centrality and betweenness centrality values. Identifying these nodes and links can help detect diverse connections between research topics and uncover special combinations. For example, the “structural equation model” links both “knowledge sharing” and “comparative advantage,” showing that the method has been used widely and intensively in the field.

4.3.4 Weak Tie Analysis of Subnets and Nodes of 2013

The high-frequency keywords weak tie network of 2013 is clustered into five subnets, which includes 103 edges, 84 nodes in all, 62 edge nodes, and 22 isolated nodes (Figure 5). The subnets are “information technology” (displayed in blue, with ID 0, 24 nodes), “information science” (displayed in red, with ID 1, 12 nodes),

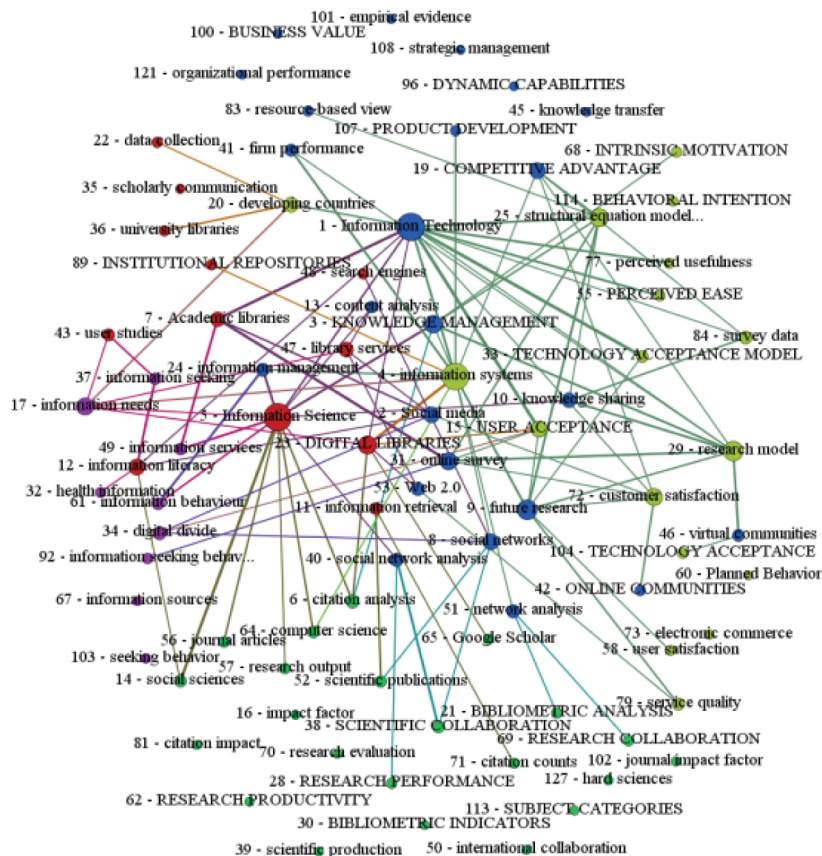


Figure 5. 2013 weak tie network of library and information science.

“citation analysis” (displayed in green, with ID 2, 22 nodes), “information need” (displayed in purple, with ID 3, 9 nodes), and “information science” (displayed in yellowish-green, with ID 4, 17 nodes). The indicator values of subnets and nodes are listed in Tables 9 and 10.

(i) Weak tie analysis of subnets: In this Section for 2013 (Table 9), “information technology” and “information system” are core subnets with more weak links, where the subnet importance for both is between 2 and 3. “Information science” is a pivotal subnet with an importance of 1.56. The other two, “information need” and “citation analysis,” are dense subnets with an importance below 0.4. “Information need” focuses on items such as “information service,” “information seeking,” and “information behavior,” while “citation analysis” focuses on items such as the kinds of libraries, “information literacy,” “institutional repository,” and “scientific communication.”

Table 9. Subnet indices in 2013.

Subnet ID	Subnet label	SI	SCB	Subnet type
0	Information technology	2.95	35.41%	Core subnet
4	Information system	2.10	34.81%	Core subnet
1	Information science	1.56	46.67%	Important subnet
3	Information need	0.40	41.46%	Dense subnet
2	Citation analysis	0.26	14.75%	Dense subnet

Note. SI refers to subnet importance and SCB refers to subject connection breadth.

(ii) Weak tie analysis of nodes: This Section focuses on the fourth type of weak tie connection nodes based on weak tie network indices of 2013: special weak tie nodes (Table 10).

Due to the definition of special weak tie nodes, they are divided into two types, those where the betweenness centrality either decreases or increases dramatically.

Nodes with decreased betweenness centrality include “knowledge management” and “knowledge sharing” of the “information technology” subnet, “citation analysis,” and “scientific performance” of the “citation analysis” subnet, and “technology acceptance model” of the “information system” subnet. These sorts of nodes may be the bridges in the subnet and thus play an important internal role.

Nodes with increased betweenness centrality include “future research” and “virtual community” of the “information technology” subnet, “scientific reports publication” of the “citation analysis” subnet, “digital divide” of the “information need” subnet, and “user satisfaction” and “developing countries” of the “information system” subnet. These sorts of nodes highlight research topics and their related subjects that are easily neglected or hidden in the normal keywords co-occurrence networks, where they may be emerging topics or frontiers that involve more interdisciplinary research.



Table 10. Node indices in 2013 (partial list).

Node ID	Node label	Betweenness centrality (weak co-occurrence)	Betweenness centrality (weak tie)
10	Knowledge sharing	145.27	35.94
33	Technology acceptance model	45.97	5.77
9	Future research	58.23	171.13
34	Digital divide	2.41	44.74
20	Developing countries	60.86	123.83

5 Discussion

Compared with the weak tie co-occurrence network, the weak tie network focuses on the analysis of weak links between subnets and nodes, by which we can first clearly observe the subnets' importance and connections between them from a micro level, and then judge the changing trends of the research topics.

5.1 Changing Law of Weak Subnets

Taking the “information science” subset, for example, by comparing the weak relation networks of years 2007, 2009, 2011, and 2013, we found that the numbers and topics of subnets changed yearly, even within the same subnet, yet the nodes and links could be different. There are three more outstanding features.

The subnets clustered with technique-related and methodology-related topics have been core and important subnets for years, including special prominent subnets such as “communication technique,” “information technology,” “information system,” “information retrieval,” and “information management.” These characteristics reflect the close relationships between information science and computer science, and indicate that as an application-oriented disciplinary subject, information science is heavily dependent on techniques and methodology.

Close subnets are highly independent, where research topics are generally concentrated and the majority topics are application-related. On the whole, the topics concentrate on “library construction and library service,” “bibliometric analysis,” “scientific communication and evaluation,” and “information need and information service.” This concentration trend reflects that the research objects of information science are still confined to relatively traditional and basic areas.

Combinations coming from technique-related topics, methodology-related topics, and application-related topics made up the four weak tie types summarized in our previous study (Wei et al., 2015): links between technique-related topics and application-related topics, links between methodology-related topics and application-related topics, links between application-related topics and application-related topics, and links between technique-related topics and methodology-related topics. Diverse combinations reveal the multifold interdisciplinary nature of information science.



5.2 Roles and Functions of Weak Nodes

In the weak tie network, core weak tie nodes and important weak tie nodes with high parameter values are responsible for connecting various research topics, playing the role of “strong bridge-nodes,” generating the “strong tie strength” that represents the combinations of important topics. Common weak nodes are playing the role of “weak bridge-nodes” and form a large amount of “weak tie strength” that represents the diversified linkages between topics. Special weak tie nodes are divided into two categories. The first is those with dramatically decreased betweenness centrality, which may be the key nodes inside the subnet that have a vital internal function. The second weak tie node category comprises those with dramatically increased betweenness centrality, considered as “special bridge-nodes” that can highlight the topics and their related subjects that are easily ignored or hidden in the normal co-occurrence network. These topics tend to focus on particular content, and are more likely to be interdisciplinary research.

6 Conclusion

Aiming at reducing the limitations of our previous study, this paper further analyzes the roles and functions of nodes and links by removing the internal links of subnets, omitting irrelevant nodes, and better visualizing the weak connections between the nodes of the weak tie co-occurrence network. The paper proposes a series of connection indicators of weak tie subnets and weak tie nodes to detect research topics, recognize important topics, and analyze topics evolution based on the weak tie theory. Taking “library and information science” as an example, this paper studies research topics by calculating and sorting the indicators, as well as using the social network analysis and time series analysis. The study finds that by using both weak tie connection indicators and social network degree indicators, we can reveal the features and changing trends of research topics’ clusters and summarize the roles and functions of different kinds of nodes and links.

According to the strong tie theory and the weak tie theory, accounting for longevity, stability, and meditation of strong co-occurrence ties between information research topics, strong ties enable the frequent knowledge exchange and stable cooperation, demonstrating a solid and consistent combination of topics. By contrast, because of the universality, heterogeneity, and intermediary of the weak ties between the information research topics, they make the process of knowledge exchange flexible and the cooperation more diversified. Those topics may contain some potentially emerging or frontier subjects that are not easily detected by analyzing the strong co-occurrence ties. Though the weak tie nodes can hardly represent the existing research foundation or the current research mainstream, they can be useful



compliments to the strong co-occurrence ties. It would therefore be better to study the topics by integrating analyses of the strong tie and weak tie relations between the keywords' co-occurrence.

This study is an effort to improve topics detection research based on the weak tie theory. It has two limitations. First, the parameter values are somewhat inconsistent, and it would be worthwhile to build up more scientific and reasonable indicators. Second, the weak tie subnets and weak tie nodes are classified based on empirical observations, and the conclusion is not verified and compared to other methods. Our future work will aim to detect the structural holes in the weak tie co-occurrence network and make detailed comparative analyses of the findings.

Acknowledgements

This work is funded by the National Social Science Youth Project “Study on the Interdisciplinary Subject Identification and Prediction” (Grant No.: 14CTQ033).

Author Contributions

L. Wei (weiling@mail.las.ac.cn) designed and performed the research and drafted the manuscript. H.Y. Xu (xuhy@clas.ac.cn, corresponding author) proposed the research idea and revised the manuscript. Z.M. Wang (wangzhenmeng@mail.las.ac.cn) wrote the program to process the data. K. Dong (dongkun@mail.las.ac.cn) and C. Wang (wangchao@mail.las.ac.cn) helped to analyze data. S. Fang (fangsh@clas.ac.cn) revised the final manuscript.

References

- Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measure and social network analysis measures. *Journal of Informetrics*, 5, 594–607.
- Aubert, B., Léger, P.M., & Larocque, D. (2012). Differentiating weak ties and strong ties among external sources of influences for enterprise resource planning (ERP) adoption. *Enterprise Information Systems*, 6(2), 215–235.
- Baer, M. (2010). The strength-of-weak-ties perspective on creativity: A comprehensive examination and extension. *Journal of Applied Psychology*, 95(3), 592–601.
- Bettoni, M., & Bernhard, W. (2008). Weak tie cooperation in the CoRe knowledge network. Retrieved on August 20, 2016, from <https://www.researchgate.net/publication/264840767>.
- Brandes, U. (2004). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2), 163–177.
- Burt, R.S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 349–399.
- Chen, G., Xiao, L., Hu, C.P., & Zhao, X.Q. (2015). Identifying the research focus of library and information science institutions in China with institution-specific keywords. *Scientometrics*, 103(2), 707–724.



Research Paper

- Ding, Y., Chowdhury, G.G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37(6), 817–842.
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge: Cambridge University Press.
- Freeman, L.C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40, 35–41.
- Freeman, L.C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215–239.
- Genius, S.K. (2005). Published literature and diffusion of medical innovation: Exploring innovation generation. *The Canadian Journal of Information and Library Science*, 1(29), 27–54.
- Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Sigchi Conference on Human Factors in Computing Systems* (pp. 211–220). New York: ACM.
- González-Alcaide, G., Castelló-Cogollos, L., Navarro-Molina, C., Aleixandre-Benavent, R., Valderrama-Zurián, J.C. (2008). Library and information science research areas: Analysis of journal articles in LISA. *Journal of the American Society for Information Science and Technology*, 59(1), 150–154.
- Granovetter, M.S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M.S. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1(1), 201–233.
- Guo, T., Xu, H.Y., Yue, Z.H., & Fang, S. (2015). Study on the interdisciplinary topics of information science based on TI index series (in Chinese). *Journal of the China Society for Scientific and Technical Information*, 34(10), 1067–1078.
- 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. *Science Research Management*, 22(3), 90–96.
- Jiang, Y.X., & Zhan, H.Q. (2008). Trend analysis of library and information sciences based on co-keyword statistics (in Chinese). *Library and Information Service*, 52(9), 28–31.
- Kavanaugh, A.L., & Reese, D.D. (2005). Weak ties in networked communities. *Information Society an International Journal*, 21(2), 119–131.
- Lee, B. & Jeong, Y.I. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics*, 77(1), 3–19.
- Li, J., Sheng, X.P., & Wei, C.M. (2015). Empirical research on open access resources sharing behavior from the perspective of weak ties (in Chinese). *Library Forum*, 2, 6–10, 87.
- Liao, S.J. (2009). Drawing and analysis on knowledge map of research frontiers of information science based on TDA (in Chinese). *Information Theory and Practice*, 32(11), 98–101.
- Pan, Y.F., & Sheng, X.P. (2014). Open access resource sharing behavior analysis based on weak ties (in Chinese). *Information Theory and Practice*, 37(7), 70–74, 80.
- Podolny, J.M., & Baron, J.N. (1996). Resources and relationships: Social networks and mobility in the workplace. *American Sociological Review*, 62(5), 673–693.
- Poleacovschi, C., & Javernickwill, A.N. (2015). Do strong or weak ties matter in knowledge networks? In the 5th International/11th Construction Specialty Conference (pp. 1–9). Vancouver, Canada: British Columbia.



- Qiu, J.P., & Lv, H. (2013). The hot domain, research fronts and knowledge base of international library and information visual analysis of 17 journals' knowledge map (in Chinese). *Document, Information & Knowledge*, 3, 4–15, 58.
- Sedighi, M. & Jalalimanesh, A. (2014). Mapping research trends in the field of knowledge management. *Malaysian Journal of Library & Information Science*, 19(1), 71–85.
- Sharone, O. (2014). Social capital activation and job searching: Embedding the use of weak ties in the American institutional context. *Work & Occupations*, 41(4), 409–439.
- Sun, Y., Liu, C., Zhang, C.X., & Zhang, Z.K. (2013) Epidemic spreading on weighted complex networks. *Physics Letters A*, 378(s 7–8), 635–640.
- Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17(S2), 27–43.
- Takagi, S., & Toyama, R. (2008). On growth of network and centrality's change analysis of co-inventors network in enterprise. *Communications in Computer & Information Science*, 19, 422–427.
- Wagner, C.S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, 34(10), 1608–1618.
- Wasserman, S., & Faust, K. (1997). *Social Network Analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Wei, L., Xu, H. Y, Guo, T., & Fang, S. (2015). Study on the interdisciplinary topics of information science based on weak co-occurrence and burst detecting (in Chinese). *Library and Information Service*, 59(21), 105–114.
- Wellman, B. & Wortley, S. (1990). Different strokes from different folks: Community ties and social support. *American Journal of Sociology*, 96(3), 558–588.
- Xiao, M., Li, G.J., & Yuan, H. (2011). Research fronts of international library and information visual analysis—based on bibliographic coupling analysis on JASIS&T (2000–2009) (in Chinese). *Library and Information Service Online*, 2, 1–5.
- Yang, L.Y., Morris, S.A., & Barden, E.M. (2009). Mapping institutions and their weak ties in a specialty: A case study of cystic fibrosis body composition research. *Scientometrics*, 79(2), 421–434.
- Yue, H. (2012). Mapping the intellectual structure by co-word: A case of international management science. *Web Information Systems and Mining*, 75(29), 621–628.
- Zenou, Y. (2015). A dynamic model of weak and strong ties in the labor market. *Journal of Labor Economics*, 33(4).
- Zhang, Y.J. (2011). Research on the scientific front detection by low-frequency occurrence phenomenon (in Chinese). Beijing. (Documentation and Information Centre, Chinese Academy of sciences Ph.D. dissertation)
- Zhao, J.C., Wu, J.J., & Xu, K. (2010). Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 82(1 Pt 2), 87–94.



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

Journal of Data and
Information Science

<http://www.jdis.org>