Doctoral dissertations of Library and Information Science in China: A co-word analysis

Qian-Jin Zong · Hong-Zhou Shen · Qin-Jian Yuan · Xiao-Wei Hu · Zhi-Ping Hou · Shun-Guo Deng

Abstract The aim of this paper is to map the intellectual structure of research in doctoral dissertations of Library and Information Science in China. By use of Co-word analysis, including cluster analysis, strategic diagram and social network analysis, we studied the internal and external structure and relationship of research fields in doctoral dissertations of Library and Information Science in China. Data was collected, during the period of 1994–2011, from six public dissertation databases and ten degree databases provided by the universities/institutes which have been authorized to grant doctoral degrees of Library and Information Science in China. The results show that Wuhan University is the most important institution of doctoral education in LIS in China. The focuses of researches, including information resource, ontology, semantic web, semantic search, electronic government, information resource management, knowledge management, knowledge innovation, knowledge sharing, knowledge organization, network, information service, information need and digital library. The research fields of LIS doctoral dissertations in China are varied. Many of these research fields are still immature; accordingly, the well-developed and core research fields are fewer.

Q.-J. Zong (✉) · H.-Z. Shen · Q.-J. Yuan · X.-W. Hu · Z.-P. Hou
School of Information Management, Nanjing University, Nanjing, China
e-mail: zongqianjin@gmail.com

S.-G. Deng
School of Economics and Management, South China Normal University, Guangzhou, China

Introduction

A dissertation shows proof of both an original contribution to knowledge and substantial subject knowledge in a discipline, and it also provides evidence of significant scholarly achievement (Kushkowski et al. 2003). Doctoral dissertation is not only a measure of research output, but is also a measure of the production of qualified manpower; a resource that is essential for contemporary knowledge societies (Andersen and Hammarfelt 2011). For these reasons, researches about doctoral dissertations in different disciplines and countries (Anwar 2004; Breimer 1996; Herubel 2007; Macauley et al. 2010; Villarroya et al. 2008; Yaman and Atay 2007) have been received a lot of attention. Doctoral dissertation of Library and Information Science (LIS) is also a valuable research subject, for example, Anwar (2004) examined the publication output of LIS graduates that was derived from their doctoral dissertations.

Since the ''opening-up'' of China to the West in 1978, Library and Information Science in China has developed rapidly along with the growth of Chinese economy (Wu and Yuan 1994). Peking University and Wuhan University were approved in 1990 authorized for the Library Science doctorate and Information Science doctorate respectively, and LIS doctor recruitment in China was started in 1991. Now, 20 years later, understanding the development of the discipline is particularly necessary. Doctoral dissertations offer unique insight into the field by revealing the foci of research and instruction within the institutions that produce LIS scholars (Finlay et al. 2012), and serve a critical function in the exploration of disciplinary development (Sugimoto et al. 2011). Thus, in this study, we map the intellectual structure of the research fields of LIS doctoral dissertations in China by using co-word analysis. This will contribute to better understanding the field and providing a basis for its future development. The remainder of this paper is organized as follows. First, we conducted a literature review. Next we introduced the methodology of this paper, included data collection, data process, and method of data analysis. Then descriptive statistic, cluster analysis, strategic diagram and social network analysis (SNA) was conducted to analyze the dataset, and the results were interpreted. Finally, we draw our conclusions in the last section.

Literature review

This section reviews the Co-word analysis and its latest development (especially Latent Dirichlet Allocation (LDA), a newly developed approach of topic modeling), and bibliometric research on dissertations (especially the LIS doctoral dissertations).

Co-word analysis and its latest development

Similar to the method of co-citation analysis (Small 1973; Small and Grifth 1974), co-word analysis rests on the assumption that a paper's keywords constitute an adequate description of its content. Two keywords co-occurring within the same paper are an indication of a link between the topics to which they refer (Cambrosio et al. 1993). The presence of many co-occurrences of pair of words within scientific papers reveals that they may correspond to a research theme. The main feature of co-word analysis is that it visualizes the intellectual structure of one specific discipline into maps of the conceptual space of this field, and that a time-series of such maps produces a trace of the changes in this conceptual space (Ding et al. 2001). Many researchers have used co-word analysis as an important method to

explore the concept network in different fields, for instance, biology (An and Wu 2011; Cambrosio et al. 1993; Rip and Courtial 1984), education (Ritzhaupt et al. 2010), Library and Information Science (Ding et al. 2000; Milojevic et al. 2011; Uzun 2002; Zhao and Zhang 2011), and so on. With the support from strategic diagram and SNA, co-word analysis, different from other co-occurrence methods, is able to visualize the intellectual structure of a specific discipline through measuring the association strength of keywords from publications in a research area (Liu et al. 2011).

Topic model is a kind of statistical model for discovering the scientific topics in a literature corpus, and can be considered as the latest development of co-word analysis. One of the newly developed approaches of topic model is LDA, which was proposed by Blei, Ng, and Jordan (Blei et al. 2003) as a generative probabilistic model for identifying the topics in a set of documents. LDA has been used to study the topical structure of a filed and it performs well. Griffiths and Steyvers (2004) presented a statistical inference algorithm for LDA to analyze abstracts from PNAS. The topics recovered by their algorithm picked out meaningful aspects of the structure of science and revealed some of the relationships between scientific papers in different disciplines. Zheng et al. (2006) extracted a set of major semantic concepts from a protein-related corpus of text words from MEDLINE titles and abstracts by applying the LDA model. They found that the identified concepts are semantically coherent, and most of them are biologically relevant. Expansions of LDA have also been used to understand (Sugimoto et al. 2011) correlations between topics (Blei and Lafferty 2007), authors (Rosen-Zvi et al. 2004), academic networks (Tang et al. 2008) and changes in topic overtime (Pruteanu-Malinici et al. 2010; Rzeszutek et al. 2010). Moreover, the topic modeling approaches are also used in field of community detection (Ding 2011), and some models (e.g., CTM, DCTM, etc.) are proposed to better understand the dynamic features of social networks and make improved personalized recommendations (Li et al. 2012).

Bibliometric research on LIS dissertations

As a bibliometric approach, citation analysis of LIS dissertations has been conducted to study the sources, rankings of disciplines and authors (Buttlar 1999; Gao et al. 2009; Sugimoto 2011). A few studies have investigated the topics of LIS dissertations. Schlater and Thomison investigated the methods used in Library Science dissertations (Schlater and Thomison 1974, 1982). Franklin and Jaeger (2007) examined the LIS doctoral dissertations written by African American women between 1993 and 2003, and the research fields were divided into five categories of research topics: information issues, library/librarianship issues, literature, and technology. Sugimoto et al. (2011) identified changes in dominant topics in LIS over time, by analyzing the 3,121 doctoral dissertations completed between 1930 and 2009 at North American Library and Information Science programs. In this study, core research areas (library history, citation analysis, and information-seeking behavior) was identified; meanwhile, one of the notable changes in the topics was the diminishing use of the word library (and related terms). Finlay et al. (2012) examined the topicality of LIS dissertations written between 1930 and 2009 at schools with American Library Association (ALA)-accredited university programs in North America. The results of this article indicated that the percentage of dissertations found to contain no instance of any of the selected library keywords had steadily risen since 1980; similarly, the percentage of dissertations found to contain instances of keywords in both the title and abstract had steadily declined.

Similarly, some researchers have analyzed the doctoral dissertations of Library Science or Information Science in China from the quantitative views. Gao et al. (2009) conducted a citation analysis of 14 doctoral dissertations in LIS at Wuhan University, the results revealed that the cited literatures came primarily from Chinese sources. Based on 110 doctoral dissertations of Information Science from China Doctoral Dissertations Full-text Database (CDFD), Wanfang Data, and National Science and Technology Library (NSTL), Wang et al. (2009) revealed the distribution of time and themes, as well as the focuses according to the frequency of keywords. They found that knowledge management and information services are the top two highest frequency keywords. Jin (2010) gathered 256 doctoral dissertations of LIS in China from 1994 to 2010. In this study, she investigated method systems of these dissertations, and results showed that researchers paid close attention to research methods, but ignored the methodology. Similarly, Yang (2011) surveyed research methods of 70 LIS doctoral dissertations of National Science Library of Chinese Academy of Sciences from 2000 to 2009, and the findings indicated that conventional methods were mainly used, such as investigation and experimental methods. Ye (2011) conducted a statistical analysis of doctoral dissertations of Library Science published in CDFD, Wanfang Data, and National Library of China (NLC) from 1994 to 2010, and analyzed the keywords with word frequency statistics method. The conclusion of this article indicated that publishing, digital libraries, library and information management whose frequency exceeded 10 and these keywords with high frequency can reflect the hot researches in Library Science.

However, until now, little attention has focused on the internal and external relationship of research fields in doctoral dissertations of LIS in China. Furthermore, the dataset of previous studies in China was only from public doctoral dissertations databases, such as CDFD, Wanfang data, and NLC. Thus, the data of these previous researches is still scarce. This paper aims to map the intellectual structure of the research fields of LIS doctoral dissertations in China. Compared with previous studies, in order to get more data, we not only gather doctoral dissertations from public degree databases, but also obtain doctoral dissertations from the degree databases provided by the universities/institutes which have been authorized to grant LIS doctoral degrees.

Methodology

Data collection and process

Keywords are the most important research elements in co-word analysis and should be exacted from publications when the research area is selected. We gathered data (1994–2011) from 16 databases, more specifically, six public degree databases and ten degree databases provided by the universities/institutes which have been authorized to grant LIS doctoral degrees.

In China, doctoral dissertations should be submitted to library/archive of the universities/institutes when the authors obtained their doctoral degrees. Until now, there are nine universities/institutes have been authorized to grant Library Science, or Information Science, or LIS doctoral degrees in China. They are Wuhan University (WHU), Nanjing University (NJU), Peking University (PKU), The National Science Library of Chinese Academy of Sciences (NSLC), Nankai University (NanKai), Jilin University (JLU), Renmin University of China (RUC), Central China Normal University (CCNU), and Sun

Yat-sen University (SYSU). Table 1 shows degree databases of the nine universities/institute.

Except for being submitted to library/archive of the degree-conferring universities/institutes, some doctoral dissertations (full text or bibliography) will be also submitted to the public degree databases. There are six common public degree databases, including CDFD, Wangfang Data, NCL, NSTL, China Academic Library and Information System (CALIS), and, Institute of Scientific and Technical Information of China (ISTIC). Table 2 shows the six public degree databases in China.

We gained data of doctoral dissertations by four steps. Firstly, we gathered bibliographies of LIS doctoral dissertations from the 16 degree databases. Secondly, we merged these data, and removed duplicated ones. Notice that, some bibliographies, especially the item "Keywords" is not correct in some public databases. Therefore, thirdly, we checked all keywords of the dataset one by one through reading full text or the first 16/24 pages in the databases above. Unfortunately, there were still some doctoral dissertations without full text (or full text of first 16/24 pages) in the databases or Internet. Thus, fourthly, we obtained these doctoral dissertations through document delivery services to fill the gap.

In this study, a program in RUBY was developed to processes the raw data. And then, BibExcel (Persson et al. 2009) was employed to calculate the frequencies that two keywords appeared together in the same doctoral dissertation. Subsequently, a symmetrical co-occurrence matrix based on the word co-occurrence was built. The value of the cell of two keywords was decided by the frequencies these two keywords both appear in the same dissertation. The higher co-occurrence frequency of the two keywords means the closer relationship between them (Ding et al. 2001). The symmetrical co-occurrence matrix was then transformed into a correlation matrix by using equivalence index (Cahlik 2000; Callon et al. 1991; Coulter et al. 1998). The equivalence index (E_{ij}) describes the strength of the association between words i and j in each word pair ij (Neff and Corley 2009; Callon et al. 1991):

$$E_{ij} = C_{ij}^2/C_{ii}C_{jj}.$$

Similarly, the value of the cell indicates the distance of two keywords; the higher value means the closer relationship between them.

We converted co-occurrence matrix to binary matrix by the program developed in RUBY. The average co-occurrence times between the high-frequency keywords (in the co-occurrence matrix) were 0.41. So, we set one as the threshold. If the value of the cell in co-occurrence was less than one, the value of the cell in binary matrix would be zero; otherwise, the value of the cell in binary matrix would be one.

Method of data analysis

Similar to other studies using co-word analysis, we chose hierarchical cluster analysis, strategic diagram and SNA. As we know, hierarchical cluster analysis and strategic diagram are commonly used in co-occurrence analysis.

Hierarchical cluster algorithm helps us find the clusters and the result of clustering can be graphically displayed as tree which shows the merging process and the intermediate clusters (Yang et al. 2012). Hierarchical clustering coupled with co-word analysis has been used widely in many studies (An and Wu 2011; Ding et al. 2001; Milojevic et al. 2011).

The strategic diagram is mainly used to describe the internal relations within a certain field and the interactions between fields (Law et al. 1988). This diagram (Bredillet 2009) is

Table 1 Doctoral dissertation database of LIS of the nine Universities/Institute in China

No.	Name	URL	Type
1	WHU	http://202.114.65.60/tpi_1/thesis/index.asp	Full text of first 16 pages, bibliography
2	NJU	http://202.119.38.70/js.asp	Bibliography
3	PKU	http://thesis.lib.pku.edu.cn/dlib	Full text, full text of first 16 pages, bibliography
4	NSLC	(1) http://www.las.ac.cn/index_others.jsp?subjectselect=ETD	Full text of first 16 pages, bibliography
		(2) http://search.gucas.ac.cn/SearchPage.aspx	
5	NanKai	http://www.lib.nankai.edu.cn/zycz/xwlw.html	Full text of first 16 pages, bibliography
6	JLU	http://202.198.25.8:8001/xwlw/user.htm	Full text of first 16 pages, bibliography
7	RUC	http://202.112.118.18/tpi/sysasp/share/login_1.asp?sysid=6	Full text of first 16 pages, bibliography
8	CCNU	http://202.114.34.37:8001/xwlw/index.jsp?agMode=1	Full text of first 16 pages, bibliography
9	SYSU	http://paper.sysu.edu.cn/xwlw/index.jsp	Full text of first 16 pages, bibliography

Table 2 Public database of doctoral dissertations in China

No.	Database name	URL	Type
1	CDFD	http://acad.cnki.net/Kns55/brief/result.aspx?dbPrefix=CDFD	Full text, bibliography
2	Wanfang Data	http://s.wanfangdata.com.cn/ThesisAdvancedSearch.aspx	Full text, bibliography
3	NCL	http://res4.nlc.gov.cn/home/index.trs?channelid=3	Full text of first 24 pages, Bibliography
4	CALIS	http://www.yidu.edu.cn/educhina/toAdvanceSearch.do	Bibliography
5	ISTIC	http://www.istic.ac.cn/dh/XueWei.aspx	Bibliography
6	NSTL	http://www.nstl.gov.cn/NSTL/facade/search/searchByDocType.do?subDocTypes=D01,D02&name_chi=%C2%DB%CE%C4	Bibliography

created by putting the strength of global context on the X axis (called *centrality*) and putting the strength of local context on the Y axis (called *density*). Two kinds of indexes (density and centrality) are used to measure the strength of local context and global context, respectively. Centrality is used to measure the strength of a subject area's inter-action with other subject areas. The value of the centrality of a given cluster can be the sum of all external link values (Courtial et al. 1993; Turner 1988) or the square root of the sum of the squares of all external link values (Coulter et al. 1998). In this study, we take the square root of the sum of the squares of all external link values (Coulter et al. 1998) as centrality. Density is used to measure the strength of the links that tie together the words making up the cluster; that is the internal strength of a cluster (He 1999). The density value can be the average value (mean) of internal links (Coulter et al. 1998; Turner 1988), the median value of internal links (Courtial et al. 1993), or the sum of the squares of the value of internal links (Bauin et al. 1991). In this paper, we take the average value (mean) of internal links as density (Coulter et al. 1998).

Social network analysis assesses the unique structure of interrelationships among individuals (Lurie et al. 2009), and has been extensively used in social science, manage-ment science, scientometrics, etc. SNA can also map the network by using methods of information visualization. We map the co-word network to show the relationships among research topics. Meanwhile, *k*-core analysis is commonly used in SNA. A *k*-core is a maximal group of nodes, all of which are connected to at least *k* other nodes in the group (Eschenfelder 1980; Maimon and Rokach 2005). By varying the value of *k* (that is, how many members of the group do you have to be connected to), different pictures can emerge. As the value of *k* becomes larger, group sizes will decrease, and the relationship among the members will be tighter (Yang et al. 2012). In bibliometrics, some studies have been investigated hot research topics though co-word analysis coupled with *k*-core analysis (Yang et al. 2012; Zhao and Zhang 2011).

In this study, the hierarchical cluster analysis and strategic diagram was conducted by using SPSS20. Simultaneously, the mapping and network were also obtained by analyzing original co-occurrence matrix and a binary matrix with Ucinet6.0 (Borgatti et al. 2002).

Result and discussion

Descriptive statistic of doctoral dissertations and keywords

We obtained 640 LIS doctoral dissertations in this study. Table 3 shows the distribution of institutions to which these dissertations belong. As shown in Table 3, Wuhan University has the largest number of LIS doctoral dissertations, indicating it is the most important institution of LIS doctoral education in China.

There were two dissertations which we could not get their keywords. We totally obtained 3,015 keywords (4.7 keywords per dissertation) from the 638 dissertations, and took the 3,015 keywords as the data sample of co-word analysis. Due to the lacking of unified indexing on keyword, we standardized these keywords by merging the synonyms (e.g., "Bibliometric analysis" is replaced by "Bibliometric"). Finally, 56 keywords with frequency more than six were selected as shown in Table 4. The frequencies of these 56 keywords are 612 times (about 20.3 % of the total), covering the main research topics of LIS doctoral dissertations in China. Notice that the keywords "library", "China", "countermeasure", "Information Science" and "information study" have very broad

Table 3 distributions of LIS doctoral dissertations in universities/institute in China

No.	Name	
1	WHU	256
2	NJU	111
3	PKU	97
4	NSLC	88
5	JLU	51
6	NanKai	16
7	RUC	12
8	CCNU	8
9	SYSU	1
Sum		640

meanings. In other words, this kind of keywords are meaningless for this study, and we excluded them in the below analysis.

The words with high frequency of occurrence and co-occurrence can reflect research focuses to some extent. The top ten keywords with high frequency of occurrence are knowledge management (39), digital library (34), network (31), ontology (22), information service (20), evaluation (20), electronic government (19), information resource (16), competitive intelligence (16), and library (15). The top ten keywords with high frequency of co-occurrence are network (32), knowledge management (32), digital library (30), information resource (26), ontology (25), knowledge organization (20), information resource management (19), electronic government (18), information retrieval (16), and evaluation (19). Notice that knowledge management, digital library, network, ontology, evaluation, electronic government, information resource have the higher frequency of occurrence and co-occurrence, and indicating that these research topics are major focuses and the bridges connecting other research topics (Liu et al. 2011) in the research of LIS doctoral dissertations in China.

Cluster analysis

We conducted the cluster analysis using hierarchical cluster analysis, with Ward's method (Ding et al. 2001; Gordon 1996; Neff and Corley 2009; Lee and Jeong 2008; Liu et al. 2011) and the distance measure is "Squared Euclidean distance" as recommended by Bacher (2002). The 51 keywords of LIS doctoral dissertations in China were divided into 15 clusters. It indicated that the research fields of LIS doctoral dissertations in China are varied. The dendrogram of the cluster analysis is shown in Fig. 1. Cluster names are given for each cluster as shown in Table 5.

As can be seen, cluster 10 has the largest number of keywords, indicating that the cluster 10 is the most centralized research fields. The keywords, that is, the research topics in cluster 10 are paid close attention to in LIS doctoral dissertations in China.

Drawing strategic diagram

Centrality and density, the two indicators of strategic diagram, could reflect the strength of relation within and between clusters. The strategic diagram can display the structure of

Table 4 The top 56 keywords

No.	Keyword	Abbreviation	Frequency
1	Knowledge management	KM	39
2	Digital Library	DL	34
3	Network	Net	31
4	Ontology	Onto	22
5	Information service	Info-Serv	20
6	Evaluation	Eval	20
7	Electronic government	EG	19
8	Information resource	IR	16
9	Competitive intelligence	CI	16
10	Library	Lib	15
11	Competitiveness	Comp	13
12	China	CN	13
13	Information resource management	IRM	13
14	Information retrieval	Info-Retr	12
15	Knowledge sharing	K-Shar	12
16	Countermeasure	Coun	12
17	Information industry	Info-Indu	11
18	Performance evaluation	Perf-Eval	10
19	Knowledge transfer	K-Tran	10
20	Resource sharing	RS	10
21	Information system	IS	10
22	Information safety	Info-Safe	9
23	Knowledge discovery	K-Disc	9
24	Search engine	SE	9
25	Information management	IM	9
26	Data mining	Data-Mini	9
27	Information Science	Info-Scie	8
28	Semantic web	Sema-Web	8
29	Knowledge organization	K-Orga	8
30	Information visualization	Info-Visu	8
31	Grid	Grid	8
32	Technological innovation	Tech-Inno	8
33	Knowledge service	K-Serv	8
34	Copyright	Copyright	7
35	Government information	Gov-Info	7
36	Electronic commerce	EC	7
37	Information study	Info-Stud	7
38	Intellectual property	IP	7
39	Information assurance	Info-Assu	7
40	Operating mechanism	OM	7
41	Network information resource	NIR	7
42	Information need	Info-Need	7
43	Library management	Lib-M	7

Table 4 continued

No.	Keyword	Abbreviation	Frequency
44	Information policy	Info-P	7
45	Personalized service	Pers-Serv	7
46	Digital reference service	DRS	7
47	Social network	Soci-Net	7
48	Citation analysis	Cita-Anal	7
49	Strategic management	Stra-Mana	6
50	Open access	OA	6
51	Humanities and social sciences	HASS	6
52	Resource allocation	Reso-allo	6
53	Academic journal	Acad-Jour	6
54	Bibliometrics	Bib	6
55	Semantic search	Sema-Sear	6
56	Knowledge innovation	K-inno	6

research fields, and it can also reveal the focuses and trends of research fields by dividing the clusters into four quadrants. We calculated the values of centralities and densities of all clusters (as shown in Table 6), and drew the strategic diagram (Fig. 2).

As shown in Fig. 2, clusters in quadrant I (upper right hand quadrant) include cluster 1 (information resource), cluster 2 (Ontology), cluster 3 (Electronic government), cluster 6 (Knowledge management) and cluster 14 (digital library). Both of the centrality and density of these clusters are high, indicating these clusters not only contain close internal connections but also are widely connected with other clusters. These fields are the research focuses in LIS doctoral dissertations of China and tends to be mature.

Clusters in quadrant II (upper left hand quadrant) only include cluster 4 (Information retrieval). This cluster has close internal connections, indicating that the research of this cluster has formed a relative stable scale. Contrary to the internal connections, connections between this cluster and the other clusters are not so close. That is to say, this field is located on the edge of the whole research network.

Clusters in quadrant III (lower left hand quadrant) include cluster 5 (social network), cluster 7 (evaluation of humanities and social sciences), cluster 8 (performance evaluation), cluster 9 (academic journal), cluster 10 (competitive intelligence), cluster 11 (library management), cluster 12 (bibliometrics) and cluster 15 (open access). These clusters have low centrality and density, thus, they have loose internal and external connections. The fields are still immature and located on the edge of research network.

Clusters in quadrant IV (lower right hand quadrant) only contain cluster 13 (information management). Although this field have loose internal connections, it has been attracted many researchers' attentions. Consequently, there is vast space for further development in this field. In other words, information management will become research trends in the future, and need to be further studied.

Social network analysis

We conduct two types of co-word networks through NetDraw. In each network, nodes represent keywords, and line between two nodes indicates that the two keywords have appeared in a same dissertation.
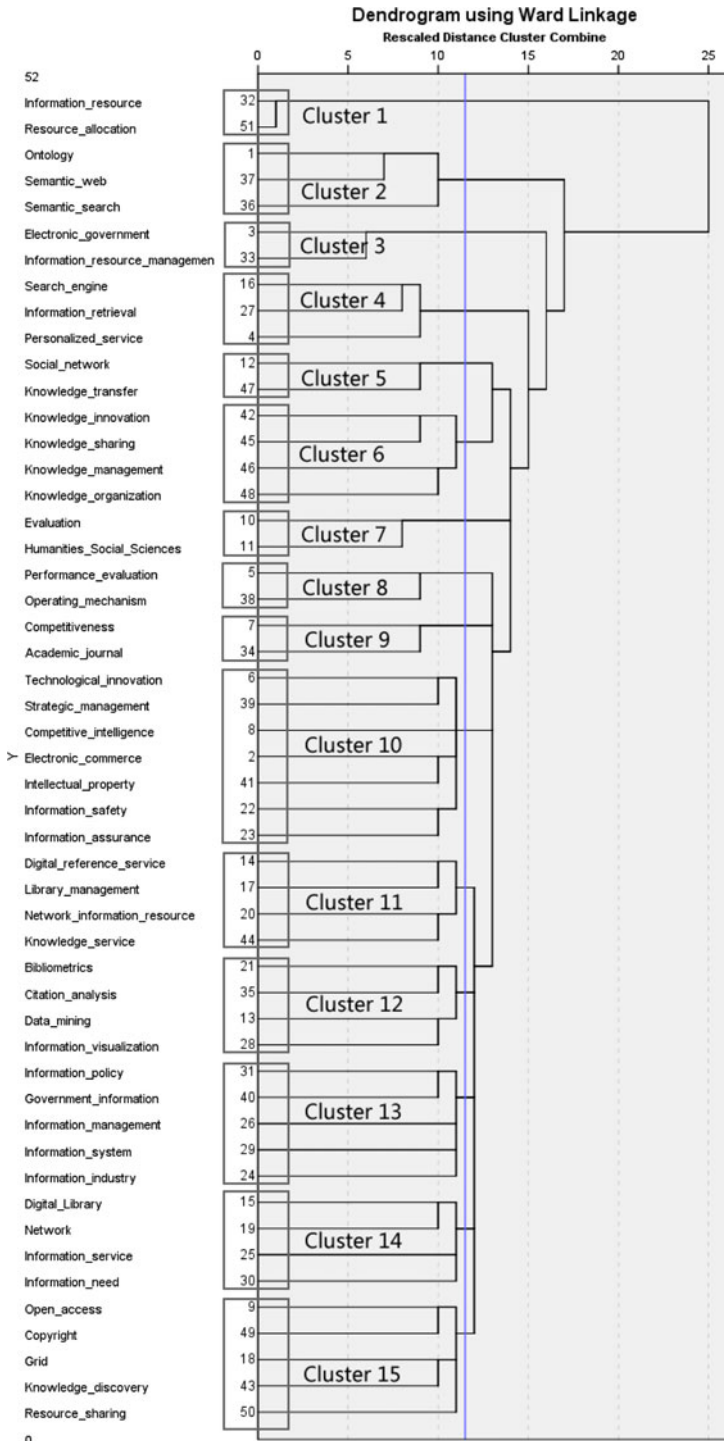
Dendrogram using Ward Linkage
Rescaled Distance Cluster Combine

0	5	10	15	20	25

52

Keyword	No.	Cluster
Information_resource	32	Cluster 1
Resource_allocation	51	
Ontology	1	
Semantic_web	37	Cluster 2
Semantic_search	36	
Electronic_government	3	Cluster 3
Information_resource_managemen	33	
Search_engine	16	Cluster 4
Information_retrieval	27	
Personalized_service	4	
Social_network	12	Cluster 5
Knowledge_transfer	47	
Knowledge_innovation	42	
Knowledge_sharing	45	Cluster 6
Knowledge_management	46	
Knowledge_organization	48	
Evaluation	10	Cluster 7
Humanities_Social_Sciences	11	
Performance_evaluation	5	Cluster 8
Operating_mechanism	38	
Competitiveness	7	Cluster 9
Academic_journal	34	
Technological_innovation	6	
Strategic_management	39	
Competitive_intelligence	8	
Electronic_commerce	2	Cluster 10
Intellectual_property	41	
Information_safety	22	
Information_assurance	23	
Digital_reference_service	14	
Library_management	17	
Network_information_resource	20	Cluster 11
Knowledge_service	44	
Bibliometrics	21	
Citation_analysis	35	
Data_mining	13	Cluster 12
Information_visualization	28	
Information_policy	31	
Government_information	40	
Information_management	26	Cluster 13
Information_system	29	
Information_industry	24	
Digital_Library	15	
Network	19	
Information_service	25	Cluster 14
Information_need	30	
Open_access	9	
Copyright	49	
Grid	18	
Knowledge_discovery	43	Cluster 15
Resource_sharing	50	

0

Fig. 1 Clusters of 51 keywords

Table 5 Fifteen clusters of research topics of LIS doctoral dissertations in China

| No.	Cluster name	Number of keywords	Member of cluster
1	Information resource	2	Information resource, resource allocation
2	Ontology	3	Ontology, semantic web, semantic search
3	Electronic government	2	Electronic government, information resource management
4	Information retrieval	3	Search engine, information retrieval, personalized service
5	Social network	2	Social network, knowledge transfer
6	Knowledge management	4	Knowledge innovation, knowledge sharing, knowledge management, knowledge organization
7	Evaluation of humanities and social sciences	2	Evaluation, humanities and social sciences
8	Performance evaluation	2	Performance evaluation, operating mechanism
9	Academic journal	2	Competitiveness, academic journal
10	Competitive intelligence	7	Technological innovation, strategic management, competitive intelligence, electronic commerce, intellectual property, information safety, information assurance
11	Library management	4	Digital reference service, library management, network information resource, knowledge service
12	Bibliometrics	4	Bibliometrics, citation analysis, data mining, information visualization
13	Information management	5	Information policy, government information, information management, information system, information industry
14	Digital Library	4	Digital Library, network, information service, information need
15	Open access	5	Open access, copyright, grid, knowledge discovery, resource sharing

The first network was generated by using original co-occurrence matrix (Fig. 3). It could intuitively show the relationship of research topics of LIS doctoral dissertations in China. The relative size of nodes is proportional to the frequency of keywords. Line thickness reflects the closeness of connections between two keywords, the thicker the line between two keywords, the closer the connection is. As shown in Fig. 3, the "KM (Knowledge management)" node has the biggest size, which represents it has the highest frequency of keyword. The thicker lines between two keywords, such as "IRM (Information Resource Management)" and "EG (Electronic Government)", "Onto (Ontology)" and "Sema-web (Semantic Web)", etc. represent closer relationships.

We conducted the second network by using the binary matrix which was converted from the original co-occurrence matrix. As show in Fig. 4, five cores are identified by k-cores analysis. In order to display the cores clearly, different shapes are configured: thirty up triangle nodes ($k = 5$) represent core themes of the network. Ten square nodes ($k = 4$) represent the secondary core themes. Six circle nodes ($k = 3$) are the themes which are located between core and periphery. Three down triangle nodes ($k = 2$) and two plus nodes ($k = 1$) are the periphery themes.

It should be noted that, Figs. 3, 4 are two different networks of the research topics of LIS dissertations in China. Figure 3 focuses on the relationships between research topics. Compared with Fig. 3, Fig. 4 focuses on finding core-verge research topics.

Table 6 Centrality and density of fifteen clusters

Cluster No.	Cluster name	Centrality	Density
1	Information resource	18.97	6.0
2	Ontology	17.80	6.0
3	Electronic government	16.97	6.0
4	Information retrieval	11.49	4.67
5	Social network	6.40	2.0
6	Knowledge management	26.32	6.0
7	Evaluation of humanities and social sciences	12.0	3.0
8	Performance evaluation	4.12	2.0
9	Academic journal	8.25	2.0
10	Competitive intelligence	9.27	3.43
11	Library management	6.48	2.0
12	Bibliometrics	8.37	2.0
13	Information management	15.20	2.0
14	Digital library	29.44	7.0
15	Open access	13.04	1.6
Average		13.61	3.71
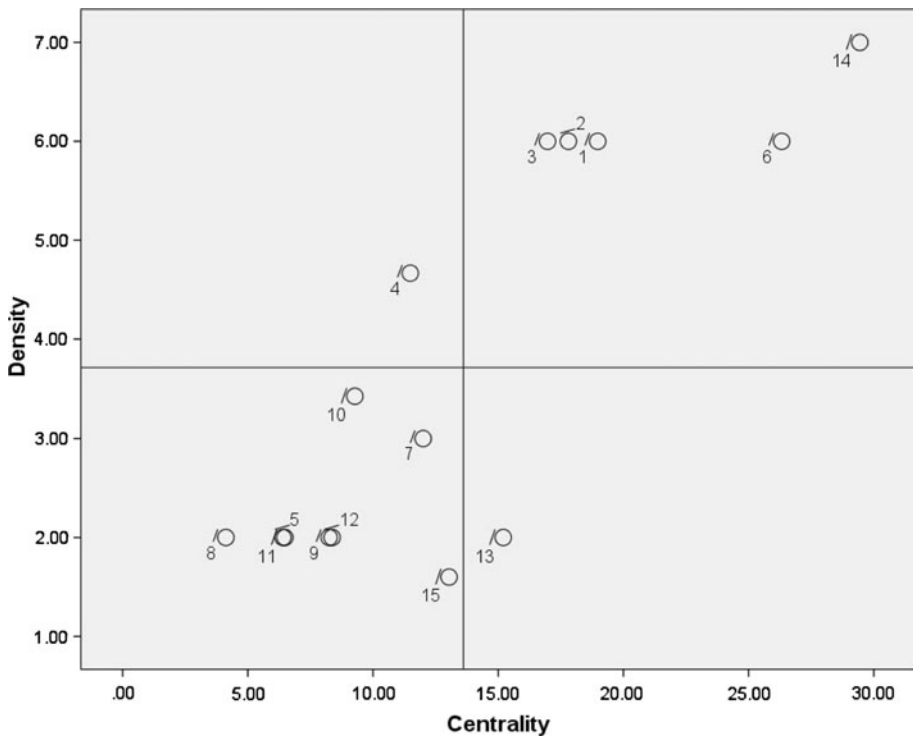
Fig. 2 Strategic diagram of Clusters
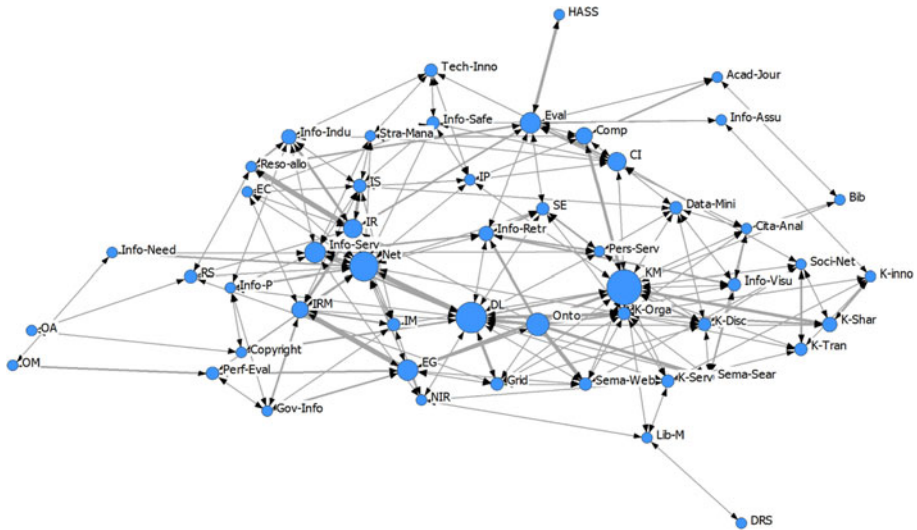
Fig. 3 Social network maps of original co-occurrence matrix
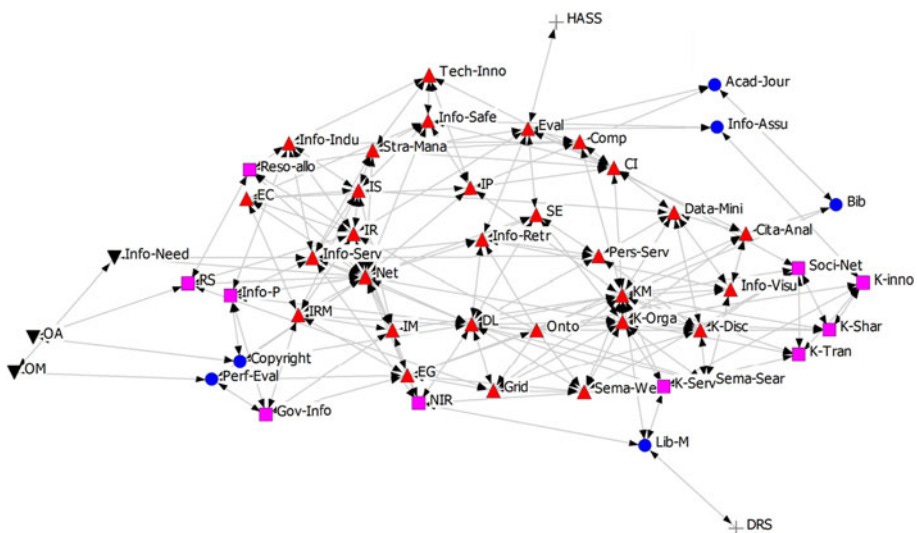
Fig. 4 *k*-cores analysis of binary matrix

Conclusions

In this paper, we investigated the intellectual structure of LIS doctoral dissertations in China by using co-word analysis, including hierarchical cluster analysis, strategic diagram and SNA. We obtain some clear and reasonable results about researches of LIS doctoral dissertations in China.

The distribution of LIS doctoral dissertations in universities/institutes implies that Wuhan University is the most important institution of doctoral education in LIS in China. School of Information Management of Wuhan University is the earliest LIS education institute in China. After 92 years of development, it has become a comprehensive and largest-scale research-oriented LIS education and research institute in China.

According to keyword frequency, strategic diagram and k-cores, we identify the focuses of researches in LIS doctoral dissertations in China, including information resource and allocation, ontology, semantic web, semantic search, electronic government, information resource management, knowledge management, knowledge innovation, knowledge sharing, knowledge organization, network, information service, information need and digital library. In these research focuses, there are only a small percentage of topics about library/library-related. This may be caused by many reasons, such as the researchers are no longer studying topics that are relevant to the practical field (Finlay et al. 2012). A lack of connections between research, education, and practice, is not only harmful to the development of LIS disciplines, but also to the future of the practice.

The hierarchical cluster analysis for top keywords with high frequency suggests that the research fields of LIS doctoral dissertations in China are varied. The evolution trends in strategic diagram reveal that many research fields in LIS doctoral dissertations in China are still immature; accordingly, the well-developed and core research fields are fewer, such as information resource, electronic government, ontology, digital library and knowledge management. In summary, it is notable that research fields of LIS doctoral dissertations in China are varied and many of them are still immature. This may be caused by two reasons. Firstly, the development of LIS disciplines (especially the discipline of Information Science) is still immature. Taking Information Science for example, there is no undergraduate education in Information Science in China. That is, there is only doctoral/master's education in Information Science in China. The researchers (e.g., Ph.D. candidate) are from different disciplines, such as mathematics, computer science, information systems, management science, economics, engineering, law, and so on. Thus, the research fields are varied due to the different backgrounds of disciplines. Secondly, the development of information technology and the growth of economics in China have motivated LIS schools to adjust their objectives, curriculum and knowledge structure in order to meet the various kinds of social needs (Dong 1997). Accordingly, the research fields of LIS doctoral dissertations become more specific and specialized. It should be noted that the doctoral education of LIS in China is only in a preliminary stage, thus, many research fields are still immature and need to be further studied.

As an exploratory study, this study also has some limitations. A little number of dissertations may contain some contents of confidentiality, such as patenting and/or commercial development possibilities. As a result, these dissertations through submission to degree databases may be delayed disclosure for years. We collected LIS doctoral dissertations from university/institute dissertation databases and public dissertation databases, therefore, the research may missed a small amount of dissertations.

Notwithstanding its limitation, this study has mapped an intellectual structure of the research fields of LIS doctoral dissertations in China. In future, we will do a longitudinal study in order to gain a full understanding of historical and contemporary developments in research. In addition, future studies may carry out a comparative study between China and oversea countries to integrate into the global trend.

Doctoral dissertation is an underdeveloped unit of analysis in contemporary bibliometric research. With the same view as the other researchers (Andersen and Hammarfelt

2011), we believe that the analysis of doctoral dissertations can provide valuable insights into the growth and structure of scientific fields and disciplines.

References

An, X. Y., & Wu, Q. Q. (2011). Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics, 88*(1), 133–144. doi:10.1007/s11192-011-0374-1.

Andersen, J. P., & Hammarfelt, B. (2011). Price revisited: on the growth of dissertations in eight research fields. *Scientometrics, 88*(2), 371–383. doi:10.1007/s11192-011-0408-8.

Anwar, M. A. (2004). From doctoral dissertation to publication: a study of 1995 American graduates in library and information sciences. *Journal of Librarianship and Information Science, 36*(4), 151–157. doi:10.1177/0961000604050565.

Bacher, J. (2002). *Cluster analysis.* Nuremberg: University of Erlangen—Nuremberg.

Bauin, S., Michelet, B., Schweighoffer, M. G., & Vermeulin, P. (1991). Using bibliometrics in strategic analysis: understanding chemical reactions at the CNRS. *Scientometrics, 22*(1), 113–137.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *Annals of Applied Statistics, 1*(1), 17–35. doi:10.1214/07-aoas114.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: software for social network analysis.* Harvard: Analytic Technologies.

Bredillet, C. N. (2009). Mapping the dynamics of the project management field: project management in action (part 2). *Project Management Journal, 40*(2), 2–6.

Breimer, L. H. (1996). Age, sex and standards of current doctoral theses by Swedish medical graduates. *Scientometrics, 37*(1), 171–176. doi:10.1007/bf02093493.

Buttlar, L. (1999). Information sources in library and information science doctoral research. *Library & Information Science Research, 21*(2), 227–245. doi:10.1016/s0740-8188(99)00005-5.

Cahlik, T. (2000). Comparison of the maps of science. *Scientometrics, 49*(3), 373–387. doi:10.1023/a:1010581421990.

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 chemistry. *Scientometrics, 22*(1), 155–205. doi:10.1007/bf02019280.

Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometrics—mapping over 70 years of biological safety research with co-word analysis. *Scientometrics, 27*(2), 119–143. doi:10.1007/bf02016546.

Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: a study in co-word analysis. *Journal of the American Society for Information Science, 49*(13), 1206–1223. doi:10.1002/(sici)1097-4571(1998)49:13<1206:aid-asi7>3.3.co;2-6.

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. doi:10.1007/bf02016216.

Ding, Y. (2011). Community detection: topological vs. topical. *Journal of Informetrics, 5*(4), 498–514. doi:10.1016/jjoi.2011.02.006.

Ding, Y., Chowdhury, G. G., & Foo, S. (2000). Incorporating the results of co-word analyses to increase search variety for information retrieval. *Journal of Information Science, 26*(6), 429–451. doi:10.1177/016555150002600606.

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. doi:10.1016/s0306-4573(00)00051-0.

Dong, X. Y. (1997). Transition of library and information science education in China: problems and perspective. *International Information & Library Review, 29*(1), 1–12.

Eschenfelder, A. H. (1980). *Magnetic bubble technology.* New York: Springer.

Finlay, C. S., Sugimoto, C. R., Li, D. F., & Russell, T. G. (2012). LIS dissertation titles and abstracts (1930–2009): where have all the librar* gone? *Library Quarterly, 82*(1), 29–46. doi:10.1086/662945.

Franklin, R. E., & Jaeger, P. T. (2007). A decade of doctorates: an examination of dissertations written by African American Women in library and information studies. *Journal of Education for Library and Information Science, 48*(3), 187–201.

Gao, S. J., Yu, W. Z., & Luo, F. P. (2009). Citation analysis of PhD thesis at Wuhan University, China. *Library Collections Acquisitions & Technical Services, 33*(1), 8–16. doi:10.1016/j.lcats.2009.03.001.

Gordon, A. D. (1996). Hierarchical classification. In P. Arabie, L. J. Hubert, & G. de Soete (Eds.), *Clustering and classification*. River Edge: World Scientific Publishing.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America, 101*, 5228–5235. doi:10.1073/pnas.0307752101.

He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends, 48*(1), 133–159.

Herubel, J. (2007). Pre 1990 French doctoral dissertations in philosophy: a Bibliometric profile of a canonical discipline (Proceedings of ISSI 2007: 11th International Conference of the International Society for Scientometrics and Informetrics, Vols I and II).

Jin, M. (2010). Survey research on method system of Ph.D. thesis of Library and Information Science in China. *Library and Information Service (in China), 54*(15), 81–85.

Kushkowski, J. D., Parsons, K. A., & Wiese, W. H. (2003). Master's and doctoral thesis citations: analysis and trends of a longitudinal study. *Portal-Libraries and the Academy, 3*(3), 459-479, doi:10.1353/pla.2003.0062.

Law, J., Bauin, S., Courtial, J. P., & Whittaker, J. (1988). Policy and the mapping of scientific change—a co-word analysis of research into environmental acidification. *Scientometrics, 14*(3–4), 251–264. doi:10.1007/bf02020078.

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. doi:10.1007/s11192-007-1819-4.

Li, D. F., Ding, Y., Shuai, X., Bollen, J., Tang, J., Chen, S. S., et al. (2012). Adding community and dynamic to topic models. *Journal of Informetrics, 6*(2), 237–253. doi:10.1016/j.joi.2011.11.004.

Liu, G. Y., Hu, J. M., & Wang, H. L. (2011). A co-word analysis of digital library field in China. *Scientometrics*, 1–15. doi:10.1007/s11192-011-0586-4.

Lurie, S. J., Fogg, T. T., & Dozier, A. M. (2009). Social network analysis as a method of assessing institutional culture: three case studies. *Academic Medicine, 84*(8), 1029–1035. doi:10.1097/ACM.0b013e3181ad16d3.

Macauley, P., Evans, T., & Pearson, M. (2010). Classifying Australian PhD theses: linking research and library practices. *Australian Academic & Research Libraries, 41*(1), 1–13.

Maimon, O., & Rokach, L. (2005). *Data mining and knowledge discovery handbook*. Heidelberg: Springer.

Milojevic, S., Sugimoto, C. R., Yan, E. J., & Ding, Y. (2011). The cognitive structure of Library and Information Science: analysis of article title words. *Journal of the American Society for Information Science and Technology, 62*(10), 1933–1953. doi:10.1002/asi.21602.

Neff, M. W., & Corley, E. A. (2009). 35 years and 160,000 articles: a bibliometric exploration of the evolution of ecology. *Scientometrics, 80*(3), 657–682. doi:10.1007/s11192-008-2099-3.

Persson, O., Danell, R., & Schneider, J. W. (2009). How to use Bibexcel for various types of bibliometric analysis. In F. Åström, R. Danell, B. Larsen, & J. W. Schneider (Eds.), *Celebrating scholarly communication studies: a Festschrift for Olle Persson at his 60th Birthday* (pp. 9–24). Leuven: International Society for Scientometrics and Informetrics.

Pruteanu-Malinici, I., Ren, L., Paisley, J., Wang, E., & Carin, L. (2010). Hierarchical Bayesian modeling of topics in time-stamped documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(6), 996–1011. doi:10.1109/tpami.2009.125.

Rip, A., & Courtial, J. P. (1984). Co-word maps of biotechnology—An example of cognitive scientometrics. *Scientometrics, 6*(6), 381–400. doi:10.1007/bf02025827.

Ritzhaupt, A. D., Stewart, M., Smith, P., & Barron, A. E. (2010). An investigation of distance education in North American research literature using co-word analysis. *International Review of Research in Open and Distance Learning, 11*(1), 37–60.

Rosen-Zvi, M., Grifftihs, T., Steyvers, M., & Smyth, P (2004). The author-topic model for authors and documents. In *20th conference on Uncertainty in artificial intelligence, Banff, Canada* (pp. 487–494).

Rzeszutek, R., Androutsos, D., & Kyan, M. (2010). Self-organizing maps for topic trend discovery. *IEEE Signal Processing Letters, 17*(6), 607–610. doi:10.1109/lsp.2010.2048940.

Schlater, G. A., & Thomison, D. (1974). *Library science dissertations, 1925–1972: an annotated bibliography*. Littleton: Libraries Unlimited.

Schlater, G. A., & Thomison, D. (1982). *Library science dissertations, 1973–1981: an annotated bibliography*. Littleton: Libraries Unlimited.

Small, H. (1973). Co-citation in the scientic literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.

Small, H., & Grifth, B. C. (1974). The structure of scientic literatures I: identifying and graphing specialties. *Science Studies, 4*(1), 17–40.

Sugimoto, C. R. (2011). Looking across communicative genres: a call for inclusive indicators of interdisciplinarity. *Scientometrics, 86*(2), 449–461. doi:10.1007/s11192-010-0275-8.

Sugimoto, C. R., Li, D. F., Russell, T. G., Finlay, S. C., & Ding, Y. (2011). The shifting sands of disciplinary development: analyzing North American Library and Information Science dissertations using latent dirichlet allocation. *Journal of the American Society for Information Science and Technology, 62*(1), 185–204. doi:10.1002/asi.21435.

Tang, J., Jin, R. M., & Zhang, J. (2008). A topic modeling approach and its integration into the random walk framework for academic search. In D. Gunopulos, F. Turini, C. Zaniolo, N. Ramakrishnan, & X. D. Wu (Eds.), *Icdm 2008: Eighth IEEE International Conference on Data Mining, Proceedings.* IEEE International Conference on Data Mining (pp. 1055–1060).

Turner, W. A. (1988). Packaging information for peer review: New co-word analysis techniques. In A. F. J. Van Raari (Ed.), *Handbook of quantitative studies of science and technohgy* (pp. 291–323). Netherlands: Elsevier Science Publishers.

Uzun, A. (2002). Library and information science research in developing countries and Eastern European countries: a brief bibliometric perspective. *International Information & Library Review, 34*(1), 21–33. doi:10.1006/iilr.2002.0182.

Villarroya, A., Barrios, M., Borrego, A., & Frias, A. (2008). PhD theses in Spain: a gender study covering the years 1990–2004. *Scientometrics*, *77*(3), 469–483. doi:10.1007/s11192-007-1965-8.

Wang, Z. J., Pan, Y. C., Xu, F., & Yan, B. N. (2009). The analysis of journal papers of Information Science doctors in China. *Library and Information (in China)*, *29*(2), 51–56.

Wu, Y. M., & Yuan, F. F. F. (1994). Library and Information education in the Peoples-Republic-of-China—the impact of Reform and Opening-Up. *Education for Information, 12*(2), 247–257.

Yaman, H., & Atay, E. (2007). PhD theses in Turkish sports sciences: a study covering the years 1988–2002. *Scientometrics, 71*(3), 415–421. doi:10.1007/s11192-007-1679-y.

Yang, Z. G. (2011). Survey on research methods of Library and Information Science PhD dissertation: a case study of National Science Library. *Journal of the National Library of China (in China)*, *20*(3), 82–88.

Yang, Y., Wu, M. Z., & Cui, L. (2012). Integration of three visualization methods based on co-word analysis. *Scientometrics, 90*(2), 659–673. doi:10.1007/s11192-011-0541-4.

Ye, S. S. (2011). Statistical analysis of Chinese doctoral dissertation of Library Science from 1994 to 2010. *New Century Library (in China)*, *32*(11), 6–9.

Zhao, L. M., & Zhang, Q. P. (2011). Mapping knowledge domains of Chinese digital library research output, 1994–2010. *Scientometrics, 89*(1), 51–87. doi:10.1007/s11192-011-0428-4.

Zheng, B., McLean, D. C., & Lu, X. H. (2006). Identifying biological concepts from a protein-related corpus with a probabilistic topic model. *Bmc Bioinformatics*, *7*, 58. doi:10.1186/1471-2105-7-58.