Identifying the research focus of Library and Information Science institutions in China with institution-specific keywords

Guo Chen · Lu Xiao · Chang-ping Hu · Xue-qin Zhao

Abstract In order to distinguish the research focus between different Library and Information Science (LIS) research institutions in China, we use the Keyword Activity Index (KAI) to identify their institution-specific keywords. The KAI, whose idea is borrowed from the Activity Index, measures whether an institution has alternatively comparative advantage in a particular topic according to its share in publications. In this study, a total of 65,653 papers from 19 core LIS journals in China during the period of 2000–2013 are collected. The top 8 most prolific LIS research institutions in China are selected for further investigation of the utility of KAI. Their institution-specific keywords are extracted based on the KAI values to represent their research focus and then clustered using co-word analysis; the research advantages of each institution are analyzed and compared according to these clusters. The reasons of their research advantages are analyzed based on their research function and research background.

Keywords Research institution · Research focus · Keyword analysis · Keyword Activity Index · Library and Information Science in China

Introduction

Research institutions are crucial to scientific innovation and the competiveness of nations (Abramo and D'Angelo 2014; An et al. 2014). Yan and Sugimoto (2011) stated that the research institution is "a stable and representative element of studying the production,

G. Chen · L. Xiao (✉) · C. Hu
Center for Studies of Information Resources, Wuhan University, Wuhan 430072,
People's Republic of China
e-mail: ahjk_xiaolu@163.com

X. Zhao
Humanities and Communications College, Shanghai Normal University, Shanghai 200234,
People's Republic of China

diffusion and consumption of knowledge". An analysis of the research profiles of important institutions not only help identify their weakness and find collaborators, but also provide information for making scientific policies at several levels (Hicks 2012). Institutions within the same field may focus on different research emphasis in terms of specialization and diversification (Glänzel et al. 2009). The prominent institutions usually stand leading positions in hot research topics and keep their competitive advantages (An et al. 2014). As a result, it is necessary to identify the research focus of different institutions and compare their specializations.

The Library and Information Science (LIS) research in China has developed rapidly in the past decade, which motivates researchers to map the intellective structures of LIS research in China based on the analysis of co-author (Yan et al. 2010), co-citation (Hu et al. 2011), and co-word (Zong et al. 2013; Hu et al. 2013). However, the deep analysis of research topics of LIS institutions in China is seldom studied (An et al. 2014). In China, the LIS research institutions are diverse in research functions and research backgrounds. For example, some institutions are established by government-leading with specific research functions, others belong to different schools of universities with diverse research backgrounds. Therefore, their research focuses are significant different from each other, which should not be ignored when studying the overall researches of LIS in China. The study of research profiles of LIS institutions in China can be beneficial to recognize the LIS research in China more clearly as well as improving research collaboration based on their research advantages.

In this paper, we analyze the publication keywords of institutions from a new perspective—instead of identifying research focus using hot topics; our approach investigates the research topics that can better represent the specialization of an institution, aiming to reveal the comparative advantage of different institutions. The goal of this paper is to (1) establish the quantitative method of identifying institution-specific topics of the main LIS institutions in China; (2) reveal the research focus and comparative advantages of these institutions based on their institution-specific topics. The key innovation of our method is to calculate the Activity Index of an institution for particular keywords, so that we can select their institution-specific topics.

Related work

Bibliometrics analysis of research institutions

Nowadays, the academic ranking has become a very popular way of institution assessment, especially for universities (STolZ et al. 2010; Aguillo et al. 2010; Hazelkorn 2014). Although the peer review is believed to be the best way to evaluate research output (Bornmann 2011), its major limits are the costs and time (Abramo et al. 2013). Therefore, the bibliometric indicators were widely utilized to capture the quantity and quality of academic research output. Related indicators are mainly focus on macro analysis of research activity and performance of institutions based on their publication and citation data (Bornmann et al. 2012; Waltman et al. 2012). For example, the most common method is to use the number of publications as the basic indicator of research activity, the citation impact is used as a basic indicator of research performance (Moed et al. 2011), and there are many improved approaches based on these basic indicators. Zhu et al. (2014) incorporated the international collaboration as a new dimension besides the quantity and quality dimensions when evaluating the research institutions of China. Recently, Bornmann et al. (2014) has

introduced a web application to visualize institution performance within specific subject areas based on the basic indicators we mentioned above.

As Thijs and Glänzel (2008) pointed out that "the usage of bibliometrics indicators cannot disguise that comparing institution remains often like comparing apples with pears". Institutions usually have different subject profiles in the context of specialization and diversification, so it is necessary to utilize more sophisticated techniques in the performance evaluation at the institutional level (Glänzel et al. 2009). Leta et al. (2006) developed a method to classify and map the European and Brazilian institutional landscaped on the basis of their research profiles. Miguel et al. (2008) used 53 subject categories to describe the disciplinary profile of an institution so as to reveal its intellectual structure and main research fronts. Thijs and Glänzel (2010) constructed the field vectors based on the field assignment of journals to represent the research profile of institutions; they also believed that even the groups within the same field differ in particular profile (Thijs and Glänzel 2008). So it's necessary to distinct their difference at a more subtle level. Belter and Sen (2014) recommend topics as one metric for the evaluation of NOAA R&D. An et al. (2014) find that early researches at the institution level in a certain research field lack of the details of the specific research topics, so they defined the salient research topics as those high-frequency controlled terms and then compared the main LIS institutions in America and China.

Identifying research topic using keywords

In bibliometrics, researchers believed that publication keywords are the basic knowledge elements which can represent the publication's key concepts (Ding et al. 2001; Yi and Choi 2012). Many researchers have utilized the keywords to reveal the intelligent structure of research entities in the form of co-word clustering (Rip and Courtial 1984; Callon et al. 1991) or keyword network (Choi et al. 2011; Assefa and Rorissa 2013).When identifying the research topics of research groups, the high-frequency keywords assigned by more authors or indexers in publications are usually considered as more important. For example, Zhao and Wang (2011) used keywords with frequency above 60 times to analyze the research foci of pervasive and ubiquitous computing. Niu et al. (2014) used the top 30 high-frequency keywords to find significant differences between geosciences, multidisciplinary and environmental sciences. In some other databases, high frequency keywords are used in the analysis of hotspots and developing trends of research fields (Su et al. 2014).

However, the use of high-frequency keywords as research topics has long been questioned. Quoniam et al. (1998) argued that classical bibliometric techniques ignored the innovative aspects which are often described by the long-tail keywords; Milojević et al. (2011) pointed out that some high-frequency keywords are non-specific words and therefore are of limited values in the bibliometric analysis; Liu and Ma (2013) argued that the high-frequency keywords cannot be used for distinguishing knowledge content. In the analysis of research institutions, Moed et al. (2011) found that "an institution may show internally a low publication activity in a field compare to its output in other filed, but externally, compare to other institutions in the same field, be among the most productive ones." Normally, a research institution may have special preference on some topics due to their development history and research orientation (Huang et al. 2006). We regard such topics as "institution-specific" topics, and the identification of them can help us recognizing the strategic positioning of a research institution as well as its comparative advantage.

Methodology

The Keyword Activity Index

The institution-specific topics are usually more concentrated on publications of the given institution and can reflect its research focus. To compute these topics, we borrow the idea of the Activity Index from Frame ([1977](#)). The Activity Index comes from the Revealed Comparative Advantage Index in economics (Balassa [1965](#)). It measures whether a country has alternatively comparative advantage in a particular research field. The definition of it is:

$$\text{AI} = \frac{\text{the share of the given country in publications in the given filed}}{\text{the share of the given country in publications in all science fileds}} \quad (1)$$

In formula ([1](#)), $\text{AI} > 1$ indicates that the country emphasizes the given science field as compared to its overall research production while $\text{AI} < 1$ indicates that it has a relative poor research on the field as compared to its overall research production.

AI has been widely used to represent the research profiles of countries/regions. Thijs and Glänzel ([2008](#)) used AI to describe the national profile of 8 research fields in European countries. López-Illescas et al. ([2011](#)) used AI as an indicator to investigate the university's research performance with considerations on discipline specializations. Pouris and Ho ([2014](#)) used AI to identify the emphasized and underemphasized research fields of Africa. Harzing and Giroud ([2014](#)) used AI to discover the competitive advantages of nations for different academic disciplines.

The AI was often used at the macro level (to reveal the research field emphasis of countries or regions); the similar idea can also be applied at the micro level to measure whether or not a research institution emphasizes on a given research topic. As the research topics are usually represented by the publication keywords, we can extend the AI to the Keyword Activity Index (KAI):

$$\text{KAI} = \frac{\text{the share of the institution in publications containing the given keyword}}{\text{the share of the institution in all publications}} \quad (2)$$

The idea of KAI resembles to the Revealed Comparative Advantage Index in economics. We can think of an institution's research on a given topic as a country's export of a given product. In formula ([2](#)), the denominator is similar to a country's share of world exports; the numerator is similar to the country's share of world exports on a given product. Accordingly, the KAI can help us to measure whether an institution has comparative advantage in studying a given topic. $\text{KAI} > 1$ indicates that the topic is emphasized in the institution above its average level; $\text{KAI} < 1$ indicates that the topic is underemphasized in the institution.

In bibliometrics, an institution's share of publications can be calculated as dividing its publication number by the total publication number of all institutions; its share of publications on a given topic can be calculated as dividing the number of its publications containing the given keyword by the number of all publications containing the given keyword. Let $n(i, j)$ denotes the number of publications containing the keyword $K(i)$ in the given institution $I(j)$; $n(i, \text{all})$ denotes the total number of publications containing the keyword $K(i)$ in all institutions; $n(\text{all}, j)$ denotes the number of publications contributed by the given institution $I(j)$; $n(\text{all}, \text{all})$ denotes the total number of publications of the given research field. Accordingly, formula ([2](#)) can be rewritten as:

$$\text{KAI} = \frac{n(i,j)/n(i,\text{all})}{n(\text{all},j)/n(\text{all},\text{all})} \qquad (3)$$

Procedures

The procedures of identifying the research focus of institutions based on their institution-specific keywords can be demonstrated as Fig. 1.

Step 1: Constructing the background corpus and institutional corpuses based on the publications collected from an academic database. The background corpus consists of all publications from the core LIS journals in China. The number of publications for each given keyword is calculated and then listed in <keyword, frequency> format, so that we can get the value of $n(i, \text{all})$ from the background corpus when calculating the KAI values according to Eq. (3). Each institution corpus consists of all publications contributed by the given institution in the background corpus; the keywords are calculated using the same method as in the background corpus, so that we can get the value of $n(i, j)$ in Eq. (3) from the related institution corpus.

Step 2: Constructing the keyword frequency matrix of institutions, in which the rows, columns, and cells indicate institutions, keywords, and the frequency of keywords (which equals the number of publications containing the keywords) within the institution's publications.

Step 3: For each institution, we compute the KAI of each keyword for a given institution corpus according to Eq. (3), and then we select the keywords with the highest KAI values as its institution-specific keywords. Note that, there exist some keywords with very low frequency for a given institute but also have extremely high KAI values because they were rarely assigned by authors in other institutions. We believe that they are not good representations of the institution's research focus. Therefore, we select keywords with high KAI values from each institution's high-frequency keywords as its institution-specific keywords. To be specific, in our data corpus, the high-frequency keywords are defined as the top 200 keywords of each institution, which neither
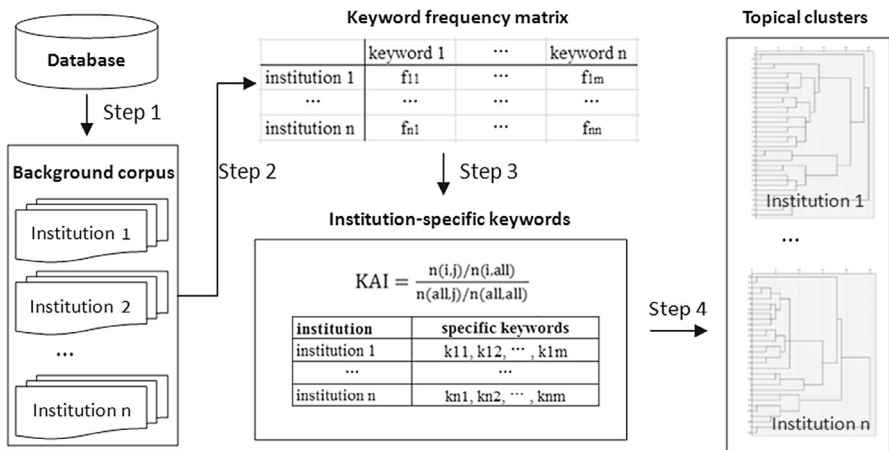

Fig. 1 Procedures of mapping research focus of institutions based on their specific keywords

eliminate too many hot topics for the relatively large institutions nor retain too many unimportant topics for the relatively small institutions. The KAI threshold is set to 1.5 because the keywords with KAI values between 1 and 1.5 are mostly words with general meanings (such as "information", "library") which cannot be used for better distinguishing different institutions.

Step 4: Clustering the institution-specific keywords of each institution based on the co-word clustering method and identifying the comparative advantage research focus of each institution based on the clusters.

Data collection and process

Constructing the background corpus of LIS research in China

The Chinese Journal Full-Text Database (CJFTD) was used as the data source. Like other researches (Hu et al. 2013), we use the published papers from all core journals of LIS in China to resemble the research field of LIS, we collected the related papers by setting: (1) the category of data source as "core journals" in "Library and Information Science" which contains all 19 core LIS journals in China; and (2) the time span from 2000 to 2013. In total, we retrieved 65,653 publications. Note that, one of the core journals, Journal of the China Society for Scientific and Technical Information, has not been indexed by CJFTD since 2003. So we download the journal publications during 2003–2013 from the Wanfang Data (http://librarian.wanfangdata.com.cn/) in China. The data is easily merged using the NoteExpress format, which are available in both two databases. We extract all author-provided keywords of those publications as the representation of research topics, because these keywords are believed to be carefully selected to identify distinctive research focus of scientific papers (Abrahamson 1996; McCloskey 1998), and they are easy to attain in our database.

Next, we have to manually remove the keywords which are mistakenly assigned by indexers (Law and Whittaker 1992), and eliminate the domain stop words (He 1999) such as "research", "counter measure", "problem" and so on. Since different authors may use various keywords when describing the same concept, we normalize keywords with the same meaning into a standard form, for example, "electronic library" is normalized into "digital library". As there are so many keywords, we have only scanned keywords with frequency higher than 1 (one is excluded). After the data preprocessing, 67,786 keywords are kept, and their total frequency is 277,721. The background corpus of LIS in China is constructed based on these keywords.

Identifying the top institutions and constructing institutional corpuses

As Van Raan (2005) pointed out, identifying the attribution of publications, i.e., find the specific organization of a publication from the provided affiliation information is an extremely important and challenging technical problem in bibliometric analysis. Taking into account the problems that Van Raan mentioned as well as the naming conventions of Chinese organizations, we firstly select those institution addresses that have at least 200 papers and extract their main organization names. After that, the main organization names are searched in the background corpus one by one, so that we can manually normalize the results. An example is demonstrated in Fig. 2, in which the Wuhan University Library will

Fig. 2 An example of institution name normalization

not be merged because it is an independent research institute with the School of Information Management.

After the normalization of the major research institutions, we select several top institutions as our analysis target. The distribution of publication number over number of institutions is shown in Fig. 3. As we can see, there is a quantity gap between the 9th most prolific institution and the 8th one. In addition, the top 8 institutions have been widely studied in many bibliometric researches and universally recognized as the most important LIS research institutions in China. So the top 8 most prolific institutions are chosen for our analysis. Their institutional corpuses are constructed in the same way as the background corpus. The basic information of all corpuses is shown in Table 1.

Clustering the institution-specific keywords

According to the discussion in the step 3 of "Procedures" section, the KAI value of top 200 high-frequency keywords in each institution corpus are calculated based on Eq. (3). The keywords with KAI > 1.5 in each institution are selected as their specific keywords. Note that author keywords are unsystematic and some of them are idiosyncratic. The effects are
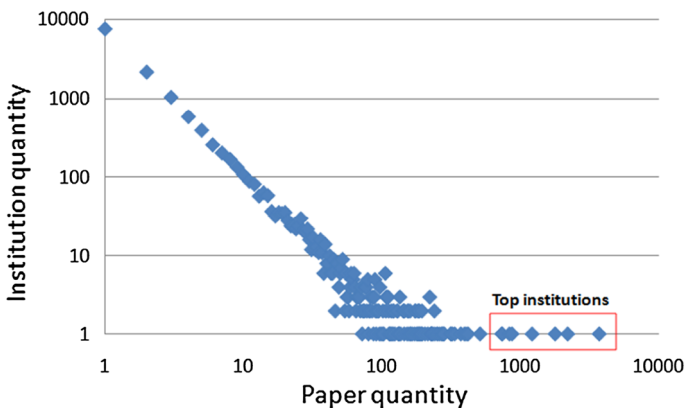

Fig. 3 The distribution of papers over institutions

Table 1 Basic information of eight institutional corpuses and the background corpus of LIS in China

Institutions	Abbreviations	Number of papers	Output scale (%)	Number of keywords	Total frequency of keywords
School of Information Management, Wuhan University	WHU	3,722	5.67	7,168	15,839
The National Science Library of China	NSLC	2,202	3.35	5,327	10,209
School of Information Management, Nanjing University	NJU	1,787	2.72	4,537	7,923
Department of Information Management, Peking University	PKU	1,217	1.85	3,147	5,312
Department of Information Management, Jilin University	JLU	872	1.33	2,162	3,698
School of Information Management, Sun Yat-sen University	SYU	835	1.27	2,120	3,531
School of Information Management, Central China Normal University	CNU	741	1.13	2,058	3,250
Institute of Scientific and Technical Information of China	ISTIC	735	1.12	2,157	3,470
All institutions in the LIS field	–	65,653	100	67,786	277,721

more noticeable when identifying the institution-specific keywords. Since the core LIS journals in China have not yet provided options for the author to select keywords from controlled thesaurus provided by the manuscript submission system, we mapped these specific keywords into established thesaurus. The online "Library, Information Science & Technology Thesaurus" in the EBSCO HOST Database and the printed "Great Dictionary of Library and Information Science" is used as references.

For better analysis and comparison, we utilized the co-word analysis method to cluster keywords based on their co-occurrence in the publications of each institution. Firstly, the co-word frequency of each two institution-specific keywords is counted as the number of publications containing both of them. Secondly, the symmetric co-word matrixes of each institution are built, in which the rows and columns represent the specific keywords of the institution, and the cells is set as the co-word frequency of keyword pairs. Then those original matrixes are transformed into Pearson's correlation matrixes to indicate the similarity and dissimilarity of each keyword pair (Van Eck and Waltman 2009). Thirdly, the co-word matrixes are imported into the SPSS 19.0, in which the cluster analysis is conducted using the hierarchical clustering with Ward's method, and the distance measure is set as "Square Euclidean distance". Note that there are few high-frequency keywords in the institution-specific keywords, so the relationships among them are much looser, leading to sparser co-word matrixes. To attain a better clustering result, we invited 3 experts in LIS to adjust the co-word clusters results and assign the subject label of each cluster.

Result analysis

The clustering results of institution-specific keywords of the top 8 LIS institutions in China are listed in Table 2. A keyword may belong to different cluster subjects in different

Table 2 Clusters of institution-specific keywords of the top eight LIS institutions in China

Institutions	Clusters	Institution-specific keywords (keyword frequency in the institution, the KAI value)
WHU	Information management: government	Value-added exploitation (23, 3.2); information reuse (11, 3.9); public sector (21, 4.2); government information publicity (28, 2.0); archives (11, 2.5); information policy (35, 1.9); information resource allocation (14, 1.9); government information (45, 1.7)
	Information analysis: informetrics	Multidimensional Scaling (10, 3.7); Informetrics (65, 2.0); scientific evaluation (16, 2.1); Webometrics (27, 3.1); information visualization (52, 1.6); content analysis (32, 1.6); Social Network Analysis (32, 1.6); Mapping Knowledge Domain (25, 1.6); co-word analysis (29, 1.5)
	Information society: economics	Information economy (20, 3.3); information product (25, 2.8); online marketing (10, 1.9); information market (15, 1.7); game theory (10, 1.5)
	User and service	User interaction (45, 3.3); user experience (20, 2.1); user behavior (14, 1.6); customer relation management (11, 1.5); subject information gateway (18, 2.3); cloud computing service (15, 2.5)
	Publishing	Publishers & publishing (31, 3.5); internet publishing (34, 2.5)
	LIS education	LIS education (11, 3.5); library education (24, 1.8); Boone Library School (17, 3.5)
	Others	Wikipedia (16, 2.3)
NSLC	Information management: scientific research	National Science Library of China (29, 11.1); Chinese Academy of Sciences (32, 10.3); e-Science(8, 7.9); Research libraries (8, 5.3); scientific researchers (18, 2.7); subject service (21, 3.4); scientific data (13, 3.3); academic libraries (30, 2.8); subject information gateway (11, 2.4); S&T journals (18, 1.5); S&T literature (18, 1.5); graduate student service (8, 1.7)
	Information management: Open Access	Open access publishing (76, 3.8); long-term preservation (49, 4.4); DSpace (8, 3.2); institutional repository (44, 2.2)
	Knowledge organization: classification	Classification methodology (62, 1.8); classification system (14, 2.4); literature classification (14, 4.6)
	Knowledge organization: technology	Knowledge organization systems (16, 5.3); Mashups (17, 14.5); Linked Data (26, 3.4); information extraction (17, 2.6); interoperability (18, 2.5); semantic web (26, 1.9); text mining (12, 1.6); ontology (68, 1.7); metadata (54, 1.5)
	Competitive intelligence: strategic	Strategic competitive intelligence (15, 6.6); industrial competitive intelligence (8, 5.8); intelligence analysis (52, 2.8); strategic planning (15, 2.5); patent analysis (31, 2.1)
	Others	Co-word analysis (23, 2.1); user behavior (12, 2.3); citation analysis (37, 1.5)
NJU	Research evaluation: Humanity and Social Science	Chinese Social Sciences Citation Index (37, 7.7); citation indexes (17, 3.5); academic evaluation (22, 3.3); Humanities and Social Sciences (13, 4.8); Web Impact Factor (8, 3.7); journal evaluation (23, 3.2); Social Sciences (13, 3.0); scholarly periodicals (11, 1.7); evaluation system (50, 1.6)

Table 2 continued

Institutions	Clusters	Institution-specific keywords (keyword frequency in the institution, the KAI value)
	Information analysis: informetrics	Conditional Random Field (8, 7.2); link analysis (16, 3.6); relevance (15, 2.9); quantitative research (23, 3.0); Social Network Analysis (20, 1.8); Mapping Knowledge Domain (16, 1.6); Analytic Hierarchy Process (17, 1.6); citation analysis (49, 2.4)
	Competitive intelligence: theory	Counterintelligence (8, 3.2); competitive intelligence system (14, 1.9); intelligence analysis (26, 1.7); competitive intelligence (85, 1.6); patent analysis(18, 1.5)
	Information society: social informatization	Digital campus (12, 9.5); social media (8, 8.0); informatization measurement (13, 5.7); digital divide (12, 1.6); society informatization (13, 6.3); information society (8, 1.6)
Others	Data warehousing (12, 1.6)	
PKU	Public library	Knowledge capability (5, 13.8); American public libraries (6, 6.0); knowledge assistance (6, 5.0); American Library Association (6, 2.9); local library (13, 2.5); network reading (5, 1.6); library spirit (8, 1.5)
	Public library: policy	Library legislation (45, 6.0); right of privacy (8, 5.1); information policy (13, 2.2); Internet communication (5, 3.1); library right (10, 2.0); Right of communication of information on networks (5, 1.6)
	Basic theory	Philologist (5, 13.8); Wang Zhongmin (7, 9.7); information theory (5, 3.5); library science history (8, 3.2); bibliographies (13, 2.2); philology (11, 2.2); librarianship (51, 1.5)
	Information industry	Digital content industry (9, 1.8); culture industry (5, 4.5); information service industry (8, 1.6)
	Information retrieval	Natural language retrieval (7, 2.5); cross-language information retrieval (6, 3.2)
	LIS education	Discipline system (8, 3.6); library education (11, 2.5)
	Others	Domain analysis (6, 10.7); complex network (7, 4.1)
JLU	Information management: enterprise	Technology absorptivity (11, 16.5); business website (10, 16.3); information supply chain (6, 26.1); technology market (6, 23.2); industrial cluster (7, 4.9); enterprise informatization (27, 1.6); competitive advantage (9, 2.6); information service industry (18, 2.7); information technology (41, 2.0); information environment (10, 2.1); information system (29, 3.0); crisis management (16, 2.3); technological innovations (16, 2.9)
	Knowledge management: enterprise	Knowledge acquisition (8, 3.9); knowledge transfer (15, 1.8); knowledge map (6, 1.8); knowledge sharing (20, 1.5); Data Warehouse (9, 2.5)
	Knowledge organization: technology	Formal Concept Analysis (12, 22.0); concept lattice (16, 20.6); semantic grid (16, 11.2); grid (20, 3.9); domain ontology (13, 3.1); knowledge organization (24, 1.7)
	Information society: ecology	Information ecochain (13, 15.6); information ecology (23, 11.0); information ecosystem (20, 8.6)

Table 2 continued

Institutions	Clusters	Institution-specific keywords (keyword frequency in the institution, the KAI value)
SYU	Public library	Library spirit (12, 3.3); community services (9, 8.7); International Federation of Library Associations (12, 7.1); American Library Association (6, 4.2); library right (13, 3.8); intellectual freedom (12, 4.3); Library Science history (5, 2.9); disadvantaged groups (6, 1.7); library history (5, 1.5); community library (10, 1.7)
	User and service	Usability (17, 7.9); recommended system (13, 3.3); reference service (22, 1.7); user satisfaction (10, 1.8); user research (10, 2.2); library orientation (11, 1.5)
	LIS education	Library education (12, 4.0); educational reform (5, 1.9)
	Information analysis: web	Webometrics (22, 11.3); web impact factor (8, 7.9); blogs (13, 1.9); online public opinion (12, 2.1)
CNU	Information society: ecology	Information ecological niche (8, 19.3); information person (7, 14.3); information ecology (17, 9.6); information ecochain (12, 16.9); information ecosystem (12, 6.1)
	Knowledge management: knowledge community	Virtual communities (18, 13.0); academic blog (5, 12.4); knowledge sharing (33, 2.9); knowledge transfer (16, 2.3); tacit knowledge (12, 1.7)
	Information society: economics	Information consumers (5, 22.7); information consumption (19, 13.4); information product (13, 3.7); e-business (34, 2.6); enterprise (36, 1.5)
	Others	Click stream (6, 20.5); Zhang Shunhui (5, 20.5); topic map (13, 13.0); reading rate (5, 12.8); user satisfaction (8, 1.6); tags(17, 3.5)
ISTIC	Knowledge organization: thesaurus	Term interrelations (13, 27.1);Chinese thesauri (5, 17.2); thesauri (50, 12.6); terminology services (6, 20.38); notion (9, 3.5); co-word analysis (10, 2.6)
	Information retrieval	Syntactic analysis (14, 15.3); machine translation (6, 12.2); natural language processing (17, 5.2); cross-language information retrieval (8, 8.3); term frequency (9, 7.3); keywords (12, 2.5)
	Research evaluation: S&T	S&T evaluation (9, 23.7); S&T literature (23, 5.7); S&T information (18, 6.5); S&T journals (16, 4.0); impact factor (8, 2.4); journal evaluation (8, 2.7)
	Competitive intelligence: industrial	Industrial competitive intelligence (13, 28.2); patent analysis (19, 3.8); small and medium-sized enterprises (9, 4.0); intelligence analysis (14, 2.3); competitive intelligence (68, 3.1)
	Knowledge organization: technology	Knowledge extraction (14, 16.5); knowledge element linkage (9, 12.5); Wikipedia (11, 8.0); content analysis (9, 2.3); text mining (6, 2.4); linked data (7, 2.2); ontology (36, 2.7); knowledge organization (17, 1.5); interoperate (7, 2.0); Semantic Web (7, 1.5); knowledge engineering (5, 7.6)
	Others	Information quality control (19, 9.6);technical control (5, 22.9); National S&T Library (11, 17.5)

institutions, because different institutions may focus on different aspects of the same topic. From Table 2, we can see that, on the one hand, different institutions have their own focus on research topics; on the other hand, some institutions share the same research focus, but

we are still able to find their differences in detail by comparing their institution-specific keywords.

For a better result visualization, we displayed the hierarchical structure of clusters and the relationship between clusters and institutions in the EXTRAVIS software (Holten et al. 2007) (see Fig. 4). In Fig. 4, the clusters are organized hierarchically according to Table 2, for example, both "S&T evaluation" and "Humanity & Social Science evaluation" are two sub-clusters of "research evaluation"; the link between an institution and a cluster indicates that the institution focus on the research topics in the cluster.

In the following sections, we will treat the cluster which belongs to only one institution as its "unique advantage subject"; and other clusters as its "advantage subjects". As Huang et al. (2006) pointed out, institutions have different focus on research topics mainly because of their development history and research orientation. When analyzing the result in Table 2, we are looking for the reasons of why these institutions focus on those topic clusters from their research background and organizational function.

1.   School of Information Management, Wuhan University

This institute dates back to the Boone Library School in the early twentieth century, which is first modern Library school in China. Since then, Wuhan University has always
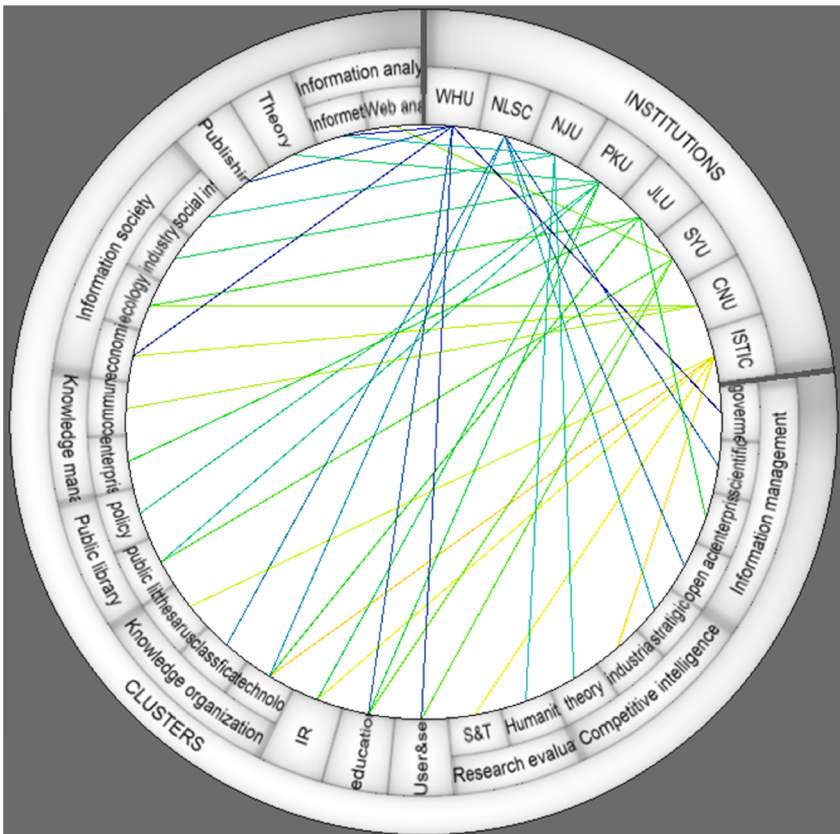

Fig. 4 A visualization of relationships between clusters and institutions

stood in the first place of the evaluations of LIS research and education in China. Nowadays, it is the biggest and the most comprehensive LIS School in China with 5 majors including the Library Science, Information Management and System, Archives, Editing and Publishing Science, E-Commerce.

As we can see in Table 2, Wuhan University has unique advantages in public information resource management and publishing. The advantages of public information resource management, particularly the development and utilization of government information, may originate from the integration of its traditional advantages in information sciences and archives. The comparative advantages of publishing may come from the comprehensiveness of its major coverage (five majors). Besides, informetrics, information economics, user and service, and LIS education are also the advantage subjects of Wuhan University. Based on Table 2, we find that (1) comparing to Nanjing University, the informetrics research of Wuhan University focus more on the research contents (such as content analysis and co-word analysis) and information visualization (such as information visualization and Multidimensional Scaling); (2) comparing to Sun Yat-sen University, the research of user and service focuses more on the user interaction and user experience; and (3) it has gotten superiority on the cloud computing service research, which is an emerging topic in recently.

Besides, as a leader of LIS research in China, Wuhan University has also invested its research energy on LIS education and gained corresponding advantage on this subject. However, we also find in the past decade, Wuhan University seems to have given up their longstanding research advantages on traditional library science.

2. The National Science Library of China

NSLC is also known as the Documentation and Information Centre of the Chinese Academy of Sciences, which is the highest academic institution of science in China. With its dual role as a service-focused library as well as a research-focused institution, NSLC has contributed far more research outputs than other libraries in China, confirming its leading role in LIS research. The functions of NSLC are described on its homepage (National Science Library 2013) as:

> NSLC functions as the national reserve library for information resources in natural sciences, inter-disciplinary fields and high-tech fields. It also provides services in information analysis, research information management, digital library development, scientific publishing, and promotion of sciences. NSLC is actively participating and leading national efforts to build a powerful National Scientific Information Infrastructure.

The institution-specific keywords of NSLC in Table 2 show that, it has unique advantages in scientific research oriented service and open access, which correspond to its research functions. As "the national server library for information resources", its research emphasizes topics about scientific information management; as an institution which is "leading national efforts to build a powerful National Scientific Information Infrastructure", its research emphasizes topics about open access in the last decade. Beside of these two unique research advantages, NSLC also has unique research advantage in Classification. Actually, NSLC has compiled the "Classification for Chinese Academy of Sciences Library", which has been widely used in scientific institutions and university libraries in China. Other research advantages of NSLC include the knowledge organization technology, and the competition intelligence. Its research on knowledge organization technology is more related to scientific research serving comparing with ISTIC, such as

Mashups, linked data, and metadata. Its research on competition intelligence focuses more on strategic intelligence than Nanjing University and ISTIC.

3.   School of Information Management, Nanjing University

Nanjing University is well known in China as the founder and manager of the Chinese Social Science Citation Index (CSSCI) Database, which is the most widely used database for academic evaluation of Humanity and Social Science in China (Su et al. 2014). As we can see from the institution-specific keywords of Nanjing University in Table 2, it has paid high attention to the research evaluation of Humanity and Social Science based on the CSSCI, such as academic evaluation, journal evaluation and evaluation system. Another unique research advantage of Nanjing University is the social informatization, in which researchers there have conducted a series of studies including the social informatization measurement system and the investigation of digital divide in China. Besides, competition intelligence is an advantage subject of Nanjing University. Comparing with the application-oriented research of NSLC and ISTIC, its research in competition intelligence focus more on basic theory, in which the counterintelligence research is a highlight. Another research advantage of Nanjing University is the informetrics, in which it focuses more on the relation mining (such as co-citation analysis and link analysis) comparing with Wuhan University.

4.   Department of Information Management, Peking University

As one of the most famous university in China, Peking University is renowned for its humanistic spirit, which is also reflected in its LIS research. Researchers in this institution have paid high attention to the public library, library policy, and the information industry, showing their keen interest in social utilities. Other research advantages of Peking University include the basic theory of LIS, information retrieval, and LIS education; all of them are traditional subjects of LIS. The research focus of Peking University shows that it still insists on the traditional LIS researches and pays high attention to the social re-sponsibility of Library, which is rare because too much emphasis has been placed on technologies of LIS in China nowadays.

5.   Department of Information Management, Jilin University

Jilin University is very special in the top LIS institutions in China. The Department of Information Management in Jilin University is set up in the School of Management, which is originated from the old engineering economics specialty, while the LIS schools or de-partments in other 5 universities are all independent. Therefore, it seems obvious that the research focus of Jilin University in LIS would be influenced by its unusual research background.

As we can see from Table 2, the institution-specific keywords of Jilin University show very different research preferences with other institutions. It has unique advantages in enterprise information management and knowledge management, making it the only one focusing on the LIS research from the perspective of enterprise in the top LIS institutions in China. Obviously, this could be contributed to its special research background. Other research advantages of it include the knowledge organization technology and the infor-mation ecology. Comparing with NSLC and ISTIC, its research in knowledge organization technology focus more on the technologies used in digital library, such as formal concept analysis and semantic grid. Its research in information ecology is not as mature as Central China Normal University.

6. School of Information Management, Sun Yat-sen University

Sun Yat-sen University has unique advantages in web analysis, in which the online public opinion is a merging topic recently in China. Other research focuses of it include the public library, user and service, and the LIS education. Its research on public library is very similar with Peking University. Its research on user and service focuses more on the service usability and user satisfaction comparing with Wuhan University.

7. School of Information Management, Central China Normal University

Central China Normal University has a unique research advantages in the topic of knowledge community, despite of its small output scale. Other research advantages of it include the information economics and information ecology, which overlap with Wuhan University and Jilin University. Its research in information ecology has a certain scale and has more obvious advantages than Jilin University. Notice that, there are many scattered research focus of CCNU, showing their development potentials.

8. Institute of Scientific and Technical Information of China

The ISTIC is a national research and service institute subordinated to the Ministry of Science and Technology of China. This background endows it with a special research orientation. The functions of ISTIC are described on its homepage (ISTIC 2013) as:

> ISTIC is designed to provide decision making support to the government agencies that take care of S&T activities in the country, in addition to its mandate of providing comprehensive information services to industry, universities, research institutes, and research personnel. It functions as a major pillar in the national science and technology innovation system, providing guidance to S&T activities and staging demonstrations for the same purpose.

According to its special responsibilities, it has many unique advantages despite of its small output scale. As we can see in Table 2, it has unique advantage in the evaluation of S&T in China, which is directly related to its function of "providing decision making support to the government agencies that take care of S&T activities in the country". It has established the Chinese Science and Technology Paper Citation Database (CSTPCD), which is the most authority database for S&T evaluation in China. Besides, from its research advantages in language process, competitive intelligence, and knowledge organization technology, we can see its effort on S&T information analysis for providing "comprehensive information services". Its unique research advantage in information retrieval, especially in machine translation is a natural result of tracing foreign S&T development and providing related information services nationwide. Its competitive intelligence focuses more on the industry comparing with Nanjing University and NSLC. Another unique research advantage of ISTIC is the thesaurus. Actually, the Chinese Library Classification and the Chinese thesaurus are both constructed by ISTIC. So there is no doubt that ISTIC has put high attention on the word analyst and thesaurus construction.

Conclusion and discussions

In this paper, we use the Keyword Activity Index (learnt from the Activity Index) to identify which topics are emphasized by a given institution in the context of LIS research in China. We chose the top 8 most prolific LIS research institutions in China for analysis.

The corresponding keywords with the highest KAI values are selected from the top 200 high-frequency keywords of each institution as their institution-specific keywords. With the aim of investigating the research focus of these institutions, we cluster the institution-specific keywords of each institution as a representation of its research advantages, because the institution is more productive on those research topics as compared to its overall productions.

Our research, based on a large-scale dataset with 65,653 publications, finds that: (1) the KAI indeed helps to better position the research advantages of each LIS research institute in China; and (2) the KAI reveals that the top 8 LIS research institutions in China have their own research advantages, some of them are unique advantages which have been emphasized by only one institution, and others are shared by several institutions, but some differences in the same research subject of different institutions can be distinguished according to the institution-specific keywords. In the analysis part, we have discovered strongly relation between the research advantage of institutions with their research function and research background.

Keywords of publications have been widely used as proxy of research topics in many bibliometric researches. Some researchers prefer to use high-frequency keywords in a research group's publications as its topical focus. However, some high-frequency keywords are general concepts shared by many other researchers. Thus, it may be inefficient to uncover the topical specialization of a research group with its high-frequency keywords, especially when we want to identify the diversity of research topics between different research groups. In this paper, we proposed a method to select institution-specific keywords from a more holistic perspective, that is, to consider the frequency of keywords both in the institution and in other similar institutions. A background corpus representing the whole LIS field in China is constructed to help us identifying whether a keyword is specific in the few institution or commonly shared by many institutions. As we can see, the institution-specific keywords in Table 2 are diverse between different institutions. Thus, we can use less keyword to reveal the research advantages of institutions, and the result fits well with the research function and research background of those institutions.

Our method can also be applied to other researches in many ways. Firstly, keywords assigned by indexers or extracted automatically can also be measured with the KAI using the same formulas, so as to get better representation of research topics of institutions. Secondly, our method can also be able to generalize to discover the research focus of other research entities (authors, journals, countries, etc.), research domains, time spans and so on. The research focus of them can be identified by extracting their specific keywords which are more concentrated in their publications than in all publications.

There are some limitations of our method. The first problem is the source of keywords. In this paper we use the author keywords because the LIS journals and database in China do not provide options for the authors to select keywords from controlled thesaurus. As a matter of fact, the author-assigned keywords are relatively subjective and some idiosyncratic keywords may be used; the effects are more noticeable when identifying the institution-specific keywords. It cost us a lot of time and energy to clean the keywords and match keywords with the established thesaurus. Therefore, we would like to suggest the follow researchers using controlled terms as the analysis objects if possible. Another problem is how to set the criteria to select keywords for later analysis, which is also a common problem in keyword bibliometrics. In this paper the criteria is set manually taking account to the real data. We believe that there is still room for improvement and we are looking for a more quantitative method in the following study.

References

Abrahamson, E. (1996). Management fashion. *Academy of Management Review, 21*(1), 254–285.

Abramo, G., Cicero, T., & D'Angelo, C. A. (2013). National peer-review research assessment exercises for the hard sciences can be a complete waste of money: The Italian case. *Scientometrics, 95*(1), 311–324.

Abramo, G., & D'Angelo, C. A. (2014). Assessing national strengths and weaknesses in research fields. *Journal of Informetrics, 8*(3), 766–775.

Aguillo, I. F., Bar-Ilan, J., Levene, M., & Ortega, J. L. (2010). Comparing university rankings. *Scientometrics, 85*(1), 243–256.

An, L., Yu, C., & Li, G. (2014). Visual topical analysis of Chinese and American Library and Information Science research institutions. *Journal of Informetrics, 8*(1), 217–233.

Assefa, S. G., & Rorissa, A. (2013). A bibliometric mapping of the structure of STEM education using co-word analysis. *Journal of the American Society for Information Science and Technology, 64*(12), 2513–2536.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage 1. *The Manchester School, 33*(2), 99–123.

Belter, C., & Sen, A. (2014). Bibliometric performance measures for the evaluation of NOAA R&D.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*(1), 197–245.

Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2012). The new excellence indicator in the world report of the SCImago Institutions Rankings 2011. *Journal of Informetrics, 6*(2), 333–335.

Bornmann, L., Stefaner, M., de Moya Anegón, F., & Mutz, R. (2014). Ranking and mapping of universities and research-focused institutions worldwide based on highly-cited papers: A visualisation of results from multi-level models. *Online Information Review, 38*(1), 43–58.

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

Choi, J., Yi, S., & Lee, K. C. (2011). Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information and Management, 48*(8), 371–381.

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.

Frame, J. D. (1977). Mainstream research in Latin America and the Caribbean. *Interciencia, 2*(3), 143–148.

Glänzel, W., Thijs, B., Schubert, A., & Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics, 78*(1), 165–188.

Harzing, A. W., & Giroud, A. (2014). The competitive advantage of nations: An application to academia. *Journal of Informetrics, 8*(1), 29–42.

Hazelkorn, E. (2014). Reflections on a decade of global rankings: What we've learned and outstanding issues. *European Journal of Education, 49*(1), 12–28.

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

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy, 41*(2), 251–261.

Holten, D., Cornelissen, B., & Van Wijk, J. J. (2007). Trace visualization using hierarchical edge bundles and massive sequence views. In *Proceedings of the 4th IEEE international workshop on visualizing software for understanding and analysis, 2007. VISSOFT 2007* (pp. 47–54). IEEE.

Hu, C. P., Hu, J. M., Deng, S. L., & Liu, Y. (2013). A co-word analysis of Library and Information Science in China. *Scientometrics, 97*(2), 369–382.

Hu, C. P., Hu, J. M., Gao, Y., & Zhang, Y. K. (2011). A journal co-citation analysis of Library and Information Science in China. *Scientometrics, 86*(3), 657–670.

Huang, M. H., Chang, H. W., & Chen, D. Z. (2006). Research evaluation of research-oriented universities in Taiwan from 1993 to 2003. *Scientometrics, 67*(3), 419–435.

ISTIC. (2013). About ISTIC. Retrieved August 21, 2013. http://www.istic.ac.cn/English/

Law, J., & Whittaker, J. (1992). Mapping acidification research: A test of the co-word method. *Scientometrics, 23*(3), 417–461.

Leta, J., Glänzel, W., & Thijs, B. (2006). Science in Brazil. Part 2: Sectoral and institutional research profiles. *Scientometrics, 67*(1), 87–105.

Liu, X., & Ma, F. (2013). Transfer and distribution of knowledge creation activities of bio-scientists in knowledge space. *Scientometrics, 95*(1), 299–310.

López-Illescas, C., de Moya-Anegón, F., & Moed, H. F. (2011). A ranking of universities should account for differences in their disciplinary specialization. *Scientometrics, 88*(2), 563–574.

McCloskey, D. N. (1998). *The rhetoric of economics*. Wisconsin: University of Wisconsin Press.

Miguel, S., Moya-Anegón, F., & Herrero-Solana, V. (2008). A new approach to institutional domain analysis: multilevel research fronts structure. *Scientometrics, 74*(3), 331–344.

Milojević, S., Sugimoto, C. R., Yan, E., & 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.

Moed, H. F., de Moya-Anegón, F., López-Illescas, C., & Visser, M. (2011). Is concentration of university research associated with better research performance? *Journal of Informetrics, 5*(4), 649–658.

National Science Library. (2013). About the Libraries Home. Retrieved August 21, 2013. http://english.las.cas.cn/au/

Niu, B., Hong, S., Yuan, J., Peng, S., Wang, Z., & Zhang, X. (2014). Global trends in sediment-related research in earth science during 1992–2011: A bibliometric analysis. *Scientometrics, 98*(1), 511–529.

Pouris, A., & Ho, Y. S. (2014). Research emphasis and collaboration in Africa. *Scientometrics, 98*(3), 2169–2184.

Quoniam, L., Balme, F., Rostaing, H., Giraud, E., & Dou, J. M. (1998). Bibliometric law used for information retrieval. *Scientometrics, 41*(1), 83–91.

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

STolZ, I., Hendel, D. D., & Horn, A. S. (2010). Ranking of rankings: Benchmarking twenty-five higher education ranking systems in Europe. *Higher Education, 60*(5), 507–528.

Su, X., Deng, S., & Shen, S. (2014). The design and application value of the Chinese Social Science Citation Index. *Scientometrics, 98*(3), 1567–1582.

Thijs, B., & Glänzel, W. (2008). A structural analysis of publication profiles for the classification of European research institutes. *Scientometrics, 74*(2), 223–236.

Thijs, B., & Glänzel, W. (2010). A structural analysis of collaboration between European research institutes. *Research Evaluation, 19*(1), 55–65.

Van Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology, 60*(8), 1635–1651.

Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*(1), 33–143.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E., Tijssen, R. J., Eck, N. J., et al. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419–2432.

Yan, E., Ding, Y., & Zhu, Q. (2010). Mapping Library and Information Science in China: A coauthorship network analysis. *Scientometrics, 83*(1), 115–131.

Yan, E., & Sugimoto, C. R. (2011). Institutional interactions: Exploring the social, cognitive, and geographic relationships between institutions as demonstrated through citation networks. *Journal of the American Society for Information Science and Technology, 62*(8), 1498–1514.

Yi, S., & Choi, J. (2012). The organization of scientific knowledge: the structural characteristics of keyword networks. *Scientometrics, 90*(3), 1015–1026.

Zhao, R., & Wang, J. (2011). Visualizing the research on pervasive and ubiquitous computing. *Scientometrics, 86*(3), 593–612.

Zhu, J., Hassan, S. U., Mirza, H. T., & Xie, Q. (2014). Measuring recent research performance for Chinese universities using bibliometric methods. *Scientometrics*, *101*(1), 429–443.

Zong, Q. J., Shen, H. Z., Yuan, Q. J., Hu, X. W., Hou, Z. P., & Deng, S. G. (2013). Doctoral dissertations of Library and Information Science in China: A co-word analysis. *Scientometrics, 94*(2), 781–799.