Scientometrics (2015) 104:737–762
DOI 10.1007/s11192-015-1629-z

Detecting the intellectual structure of library and information science based on formal concept analysis

Ping Liu[1] · **Qiong Wu**[1] · **Xiangming Mu**[2] · **Kaipeng Yu**[1] · **Yiting Guo**[1]

Received: 15 December 2014 / Published online: 25 June 2015

Abstract Detecting intellectual structure of a knowledge domain is valuable to track the dynamics of scientific research. Formal concept analysis (FCA) provides a new perspective for knowledge discovery and data mining. In this paper we introduce a FCA-based approach to detect intellectual structure of library and information science (LIS). Our approach relies on the mathematical theory which formulates the understanding of "concept" as a unit of extension (scholars) and intension (keywords) as a way of modelling the intellectual structure of a domain. By analyzing the papers published in sixteen prominent journals of LIS domain from 2001 to 2013, the intellectual structure of LIS in the new century has been identified and visualized. Nine major research themes of LIS were detected together with the core keywords and authors to describe each theme. The significant advantage of our approach is that the mathematical formulae produce a conceptual structure which automatically provides generalization and specialization relationships among the concepts. This provides additional information not available from other methods, especially when shared interests of authors from different granularities are also visualized in concept lattice.

Keywords Formal concept analysis · Intellectual structure · Library and information science

Electronic supplementary material The online version of this article (doi:10.1007/s11192-015-1629-z) contains supplementary material, which is available to authorized users.

✉ Ping Liu
p.liu@whu.edu.cn

[1] School of Information Management, Wuhan University, Wuhan 430072, People's Republic of China

[2] School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee, WI 53102, USA

Introduction

Library and information science (LIS) is a discipline combining the study of library science and information science (Bates and Maack 2010). Since the expansion of the Internet and the World Wide Web in the 1990s, there have been radical shifts in issues relating to information science. LIS has transformed from a specialized field focusing on librarianship to an academic field focusing on information and use (Larivière et al. 2012). Besides, LIS is coupled with various disciplines such as general science, business and management, computer science, education and sociology (Chang and Huang 2012). Therefore, it is of great significance to identify the intellectual structure of LIS under new circumstance.

Intellectual structure usually refers to hierarchical knowledge system consisting of knowledge elements of a discipline and their inter-relationships. It represents the basic knowledge and the relationship among knowledge content of a domain in a systematic way. Nicolai and Torben (2003) emphasized that knowledge element includes both explicit knowledge element (basic concepts to sum up topics in a domain) and implicit knowledge element (people as carriers of knowledge). Scholars should be considered as important as research topics in intellectual structure. Actually, Kedrov (1980) pointed out that the general intellectual structure of science consists of general science, branches of science, and individual scientists in specific domains. The disciplinary intellectual structure consisting of topics, subtopics, and individual scholars, is formed through the integration and interaction of knowledge elements with the development of disciplines. Therefore, the disciplinary intellectual structure contains not only subject concepts and their hierarchical relationships, but also scholars and their research topics.

There have been lots of investigations on the intellectual structure, which frequently employ bibliometric techniques as analyzing tools. A significant portion of bibliometric analysis has been focused on citation analysis, which traces through citations and reveals the intellectual structure from a macro perspective. The representatives are direct citation (Garfield 2004), journal co-citation analysis (Hu et al. 2011; Mustafee et al. 2014), document co-citation analysis (Li et al. 2010; Gao et al. 2012), document bibliographic coupling (Ahlgren and Jarneving 2008), author co-citation analysis (ACA) (White and Griffith 1981; Chen and Lien 2011), and author bibliographic coupling (Zhao and Strotmann 2008a; Ma 2012). Another important branch of bibliometric analysis is content analysis, which relies on natural language processing to analyze large corpora of literature text and depicts the intellectual structure from a micro perspective. The representatives are co-word analysis (Hu et al. 2013; Assefa and Rorissa 2013) and author-keyword coupling analysis (Liu and Zhang 2010). Scholars are the major focus of citation analysis and topics are the major focus of content analysis. To the best of our knowledge, there are few researches using the method of simultaneously combining scholars and topics to detect intellectual structure. Herein, we explore new ways to integrate scholar and topic analysis in detecting intellectual structure.

In this paper, a formal concept analysis (FCA) based method is proposed to identify the intellectual structure of LIS. FCA is a method attempting to formalize logic and being used to identify, sort, and display the concept structure of data sets. Unlike traditional statistical analysis and knowledge representation methods, FCA emphasizes human cognition (Wille 2005), which provides a unique angle in knowledge discovery. The present study addresses the following research questions.

1. What is the intellectual structure of the LIS field during the period 2001–2013?

2. Is the analysis of research activity based on FCA an effective approach to the study of intellectual structures of LIS?

We hope to contribute a new perspective of detecting intellectual structure of the LIS field, using FCA as a means to generate a conceptual hierarchy of the domain and reveal useful insights about the nature of the intellectual structure in terms of themes and scholars.

The details of our presentation are laid out as follows: first the FCA based definition and representation model of intellectual structure are given; then the "Methodology" section presents a detailed description of how to construct formal context of LIS field based on selected journal papers, and how to visualize the formal concepts and their complex relationships through concept lattice; the "Results" section describes the detected nine major research themes, associated authors, and the complex intrinsic relationships among them; the "Evaluation and discussion" section compares our result with ACA based method and identifies the additional value of our approach. The "Conclusion" section presents conclusions and outline future work.

Related definitions of the FCA based intellectual structure

FCA has been originally developed as a subfield of Applied Mathematics, the aim and meaning of FCA as mathematical theory of concepts and concept hierarchies is to support the rational communication of humans by mathematically developing appropriate conceptual structures which can be logically activated (Wille 2005). It thereby activates mathematical thinking for conceptual data analysis and knowledge processing (Ganter and Wille 2012). According to the main philosophical tradition, a concept is constituted by its extension, comprising all objects which belong to the concept, and its intension, including all attributes (properties, meanings) which apply to all objects of extension (Wille 1982). A formal context is defined as a triple (G, M, I) consisting a set of formal objects G, a set of formal attributes M, and a relation $I \subseteq G \times M$ between G and M. Given a formal context, FCA then derives all concepts from this context and orders them according to a sub-concept super-concept relation. This results in a concept lattice.

The intellectual structure of a discipline is a hierarchical knowledge system consisting of scholars, research topics, and their relationships. To employ FCA in detecting the intellectual structure, authors are taken as objects, and keywords as attributes. The formal concept corresponds to an intellectual structure unit which is composed of a set of authors (i.e., extent) and a set of keywords (i.e., intent). The formal context of the intellectual structure is created based on the relation of authors and keywords. The concept lattice can be generated with hierarchical structure based on the formal context. It reflects the clustering and associations among concepts, and visually displays the intellectual structures of a discipline. The basic definitions are introduced and explained by an example as follows.

Definition 1 The formal context of disciplinary intellectual structure is defined as a triple $KS = (A, K, R)$, where A is a set of authors (objects), K is a set of keywords (attributes), and R is a binary relation between A and K (i.e., $R \subseteq A \times K$, briefly ARK). aRk means that a relationship exists between a and k, where $a \in A$ and $k \in K$, indicating that "author a" (object) has "keyword k" (attribute). Table 1 shows a formal context that consists of eight authors and nine keywords possessed by these authors.

Definition 2 Suppose P is a subset of the author set A, then define $f(P) := \{k \in K | \forall a \in P, \text{aRk}\}$, and $f(P)$ means the set of keywords that are shared by all the authors in P.

Table 1 An example of formal context based on eight authors and nine keywords

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
a_1	×	×					×		
a_2	×	×					×	×	
a_3	×	×	×				×	×	
a_4	×		×				×	×	×
a_5	×	×		×		×			
a_6	×	×	×	×		×			
a_7	×		×	×	×				
a_8	×		×	×		×			

Correspondingly, suppose T is a subset of the keyword set K, then define $g(T) := \{a \in P | \forall k \in T, \text{aRk}\}$, and $g(T)$ means the set of authors who have all the keywords in T. Take Table 1 as an example, suppose $P_1 = \{a_1, a_2\}$, then $f(P_1) = \{k_1, k_2, k_7\}$; suppose $T_1 = \{k_1, k_2, k_7\}$, then $g(T_1) = \{a_1, a_2, a_3\}$.

Definition 3 A formal concept of the formal context (A, K, R) is defined as a pair (P, T), where $P \subseteq A$, $T \subseteq K$, and $f(P) = T$, $g(T) = P$. P is called the extent and T is called the intent of formal context (P, T). In the example in definition 2, $f(P_1) = T_1$, but $g(T_1) \neq P_1$, so (P_1, T_1) is not a formal concept. But suppose $P_2 = \{a_1, a_2, a_3\}$, $T_2 = \{k_1, k_2, k_7\}$, according to the formal context in Table 1 $f(P_2) = T_2$, whilst $g(T_2) = P_2$, so (P_2, T_2) is a formal concept.

Definition 4 Suppose (P_1, T_1) and (P_2, T_2) are two formal concepts of formal context (A, K, R), if $P_1 \subseteq P_2$ and $T_2 \subseteq T_1$, then (P_1, T_1) is the sub-concept of (P_2, T_2), (P_2, T_2) is the super-concept of (P_1, T_1), which is written as $(P_1, T_1) \leq (P_2, T_2)$. The $\leq$ relation is called the partial order among formal concepts, and the subconcept–superconcept relations refer to the partial ordering relations against the set of all concepts of a particular context. As in Table 1, take formal concepts $C_1 = (\{a_2, a_3\}, \{k_1, k_2, k_7, k_8\})$ and $C_2 = (\{a_1, a_2, a_3\}, \{k_1, k_2, k_7\})$ as examples, since the extent $\{a_2, a_3\}$ of $C_1 \subseteq$ the extent $\{a_1, a_2, a_3\}$ of C_2, whilst the intent $\{k_1, k_2, k_7\}$ of $C_2 \subseteq$ the intent $\{k_1, k_2, k_7, k_8\}$ of C_1, the concept C_1 is a sub-concept of concept C_2, and concept C_2 is a super-concept of C_1.

Definition 5 Let $\beta\ (A, K, R)$ be the set of all formal concepts of (A, K, R) and contain the subconcept–superconcept relations among the concepts. $\beta\ (A, K, R)$ is called the concept lattice corresponding to the formal context (A, K, R). Since all the concepts are arranged by partial order, the concept lattice can be displayed visually by a line diagram. Figure 1 shows the complete concept lattice generated from the formal context in Table 1.

Each node in Fig. 1 represents a formal concept composed of a set of authors (extent) and a set of keywords (intent); each edge connects a pair of adjacent concept nodes and represents the superconcept–subconcept relation among the nodes. It is not necessary to write the full extent and intent for each formal concept. A much simpler, reduced labelling is achieved if each object and each attribute (keyword) is entered only once in the diagram. If a node is marked by a blue filled upper semi-circle, then there is an attribute attached to this concept, which is the largest concept whose intent contains the attribute. If a node is marked by a black filled lower semi-circle then there is an object (author) attached to this concept, which is the smallest concept whose extent contains the object. A node with whole circle means there are at least two attributes or at least two objects attached to this

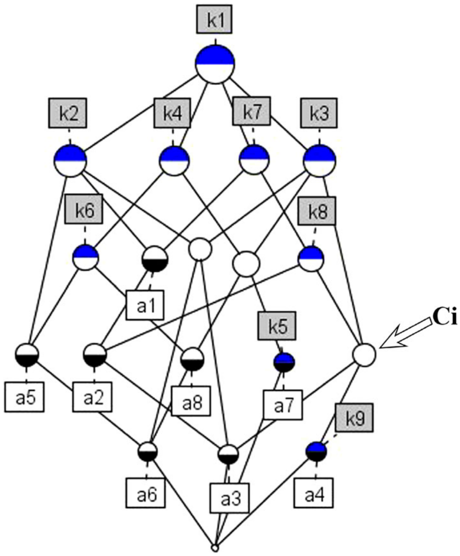

Fig. 1 The concept lattice corresponding to the formal context in Table 1

concept, and therefore the attributes and objects are not shown in the lattice by the space limitation but can be interpreted by two simple rules. The extent of the node is the union of the extents of its sub-concept nodes on a path going downward. Conversely, the intent of the node is intersection of the intents of its super-concept nodes.

It is easy to know that a more general concept is represented by a larger set of authors who share a smaller set of keywords. For instance, the top concept in Fig. 1 is represented by eight authors who share one keyword k_1. Conversely, a more specific concept has a smaller set of authors and a larger set of keywords. For example, a concept C_i in Fig. 1 is represented by authors a_3 and a_4 who share keywords k_1, k_3, k_7, k_8.

Methodology

Scientific publications are the key element in practice of science, the terminologies provide shared conceptual system of scientific communities. By extracting papers from LIS domain from 2001 to 2013, and taking authors as objects and keywords as attributes, the intellectual structure of LIS can be detected. Firstly, core authors were identified from the collected papers based on their productivity and impact, then the high-frequency keywords annotated by these authors are extracted and selected. A formal context of the intellectual structure was then constructed based on relations between core authors and high-frequency keywords. FCA was applied to produce the concept lattice. By zooming in the concept lattice, the detailed intellectual structure of LIS was revealed and interpreted.

Data collection

Papers published in sixteen world's leading journals of LIS (Table 2) were considered as the data sources for the present study. These papers reflect the vital research findings and latest progress of LIS. The bibliographic records from 2001 to 2013 were downloaded from

Table 2 The list of LIS journals and selected journal articles for analysis in the present study

Journals (abbreviation)	Journal articles (2001–2013)	%[a]
Annual Review of Information Science And Technology (ARIST)	133	1.2
Information Processing and Management (IPM)	935	8.8
Scientometrics (SciMetr)	1995	18.7
Journal of Information Science (JIS)	614	5.9
Journal of Documentation (JDoc)	482	4.5
Journal of The American Society For Information Science And Technology (JASIST)	2007	18.8
Online Information Review (OIR)	590	5.5
Journal of Informetrics (J INFORMETR)	389	3.7
Library Resources and Technical Services (LRTS)	232	2.2
Program-Electronic Library And Information Systems (ELECTRON LIB)	310	2.9
Library and Information Science Research (LIBR INF SCI RES)	354	3.3
Journal of Academic Librarianship (JAL)	741	7.0
College and Research Libraries (CRL)	402	3.8
Electronic Library (EL)	676	6.3
Library Trends (LT)	545	5.1
Library Quarterly (LQ)	243	2.3
Total	10,648	100

[a] Percentages of selected journal articles for analysis

the Science Citation Index (SCI) and Social Science Citation Index (SSCI). Only reviews, journal articles, and proceeding papers were selected for further investigations, totaling 10,648 documents. Journal titles and the total number of selected papers are listed in Table 2. We developed software to parse these downloaded records, and to store the resulting data fields, such as author names, publishing sources, affiliation, title, abstract, keywords, email address, in a data structure that was convenient for later data analysis, including the determination of core authors and associated keywords. We then sorted records in ascending alphabetical order based on the last name of the first listed author.

Core author selection

In the processing of core author selection, author name ambiguity is a critical issue in managing information at the individual level. Some authors publish their papers under several variations of their own name, (for example, “Xie, HI”, “Xie, I” and “Xie, H”). On the other hand, the same name form may, depending on the context, refer to several different authors. In our study, context information (including author full name, affiliation, and e-mail address) is used to distinguish one author from others. We have designed a simple program to automatically disambiguate authors. 6058 individual authors were identified. Core authors are selected based on two types of measures, the productivity (quantitate measures) and impact (quality measures). In this paper, an author’s productivity was indicated by the total number of publications written by this author as first author in the selected 16 journals during 2001–2013. An author’s impact was indicated by the citation frequency of this author’s publications. Authors were ranked according to the number of publications. Price (1965) proposed a law $N = 0.749 \times (\eta_{\max})^{1/2}$ as the threshold of determining the number of documents of prolific authors, where $\eta_{\max}$ represents the maximum number of documents of the most prolific author within the given sample. According to this law, authors who published more than threshold were selected as prolific authors. On the other hand, the impact-based author selection was carried out by calculating the number of citations of the authors in the selected dataset and set a threshold of citation counts. In the present study, core authors were identified as prolific authors with high citation frequencies.

Keywords selection

Followed the identification of core authors, papers and keywords annotated by authors were extracted. Since some papers do not include keywords in the original paper, we have developed software to extract keywords from titles and abstracts. First, each keyword was transformed into lower case, then, general terms were filtered (e.g., United Kingdom, Italy) and similar terms were normalized, such as merging synonyms, unifying singular and plural terms, unifying full names with abbreviations, combining terms with and without hyphens. The following examples will illustrate the process.

(1) Different spellings and variants of the same terms are unified: digital library + digital libraries = digital libraries; social network analysis + SNA = social network analysis; h-index + h index + hirsch index + hirsch-index = h-index; information seeking behaviour + information-seeking behavior = information seeking behavior; and so on.
(2) Synonyms: bibliometric analysis + bibliometric study + bibliometrics + bibliometric methodology + bibliometric measures = bibliometrics; digital libraries + electronic libraries = digital libraries; and so on.
(3) Broad term/narrow term: scientometric + spatial scientometrics = scientometrics; information systems + geographic information systems = information systems; and so on.

After standardization process, the remaining keywords were then ranked according to the frequency of occurrence. The high-frequency keywords were selected since these words may properly reflect the research topics.

Construction of formal context

The formal context is a matrix describing the relationship between core authors (objects) and keywords (attributes). Here an author is related to a keyword if the author has annotated this keyword in his/her papers (either in title, abstract, or keyword section). Suppose m core authors and n keywords have been identified in the previous section, the formal context is constructed as shown in Table 3. This cross table was used as a basis to generate the concept lattice.

Table 3 Author-keyword incidence matrix (formal context)

	k_1	k_2	…	k_n
a_1	×		…	×
a_2		×	…	×
…	…	…		…
a_m	×	×	…	

Construction of concept lattice

Concept lattice is a fundamental structure in FCA. Construction of concept lattice consists of generating of all formal concepts and building concept lattice from the given formal context. Many algorithms have been developed for constructing concept lattices (Godin et al. 1995; Kuznetsov and Obiedkov 2002). These algorithms can be divided into batch algorithm and incremental construction algorithm. Despite different algorithms being used, the generated concept lattice is always same, that is, the concept lattice is unique. Some tools are available (such as ConExp[1]) to help users automatically calculating set of all concepts and construction of concept lattices from a given context.

Results

According to Price's law (Price 1965), the number of core authors N can be calculated with the formula $N = 0.749 \times (\eta_{\max})^{1/2}$, where $\eta_{\max}$ is the number of papers published by the author with most papers. In the present study, author "Egghe, L." was found publishing the most number of papers (97 papers). Therefore, $\eta_{\max}$ was 97 and the number of core authors N was calculated as 7.4. Therefore, authors with at least 8 papers were selected, which identified a total number of 117 authors. Subsequently, taking the cited frequencies into consideration, 60 authors with a citation of more than 85 times were chosen from the 117 authors as the core authors in LIS (Table 4).

The 60 authors have published a total number of 1224 papers. A total of 1426 unique keywords (4147 keywords including duplicated words) were collected from the chosen 1224 papers. These keywords were standardized using the methods described in the "Methodology" section. After processing, 1379 keywords were obtained. About a quarter of these keywords were with singe appearance. Since keywords with high frequency are more appropriate for describing LIS topics than those with low frequency, 99 keywords with the frequency of more than 4 were selected as core keywords. The accumulative frequency of these keywords was amounted to 2559 (about 54 % of the total). These 99 keywords were chosen as the representative words of LIS (Table 5).

Taking the 60 core authors identified as objects and the 99 high-frequency keywords acquired as attributes, it was constructed the formal context of LIS based on the association relationship between objects and attributes. Table 6 shows the partial formal context related to citation analysis as an example, where the rows represent the authors related to citation analysis, the columns represent the keywords that these authors have used, and "×" in the table illustrate the keyword k_j is associated with author a_i.

Based on the constructed formal context, all the formal concepts were calculated and the corresponding concept lattice was built using the open source lattice visualization tool ConExp. The subconcept–superconcept relationships between formal concepts was reflected in the concept lattice (as shown in Fig. 2), which can be seen as a semantic net providing hierarchical conceptual clustering of the authors and keywords.

Each node of the concept lattice is a pair, composed of a subset of the authors as extent and a subset of the keywords as intent; in each pair, the subset of keywords contains just the keywords shared by the subset of authors, and, similarly, the subset of authors contains just the authors sharing the subset of keywords. In Fig. 2, an author label is attached below the node of the smallest concept whose extent contains the author, while a keyword label is

[1] http://conexp.sourceforge.net/.

Table 4 The list of 60 core authors in LIS

Author	Pub	Cit	Author	Pub	Cit	Author	Pub	Cit
Egghe, L	97	1238	Zhang, J	17	118	Campanario, JM	12	105
Jacso, P	69	527	Huang, MH	17	102	Shachaf, P	12	87
Thelwall, M	64	1346	Vinkler, P	16	247	Costas, R	11	214
Leydesdorff, L	62	1705	Kostoff, RN	16	191	Guan, JC	11	148
Bornmann, L	57	1235	Frandsen, TF	15	123	Lewison, G	11	137
Glanzel, W	39	1453	Ortega, JL	15	114	Kretschmer, H	11	129
Abramo, G	39	244	Yu, G	15	91	Julien, H	11	125
Bar-Ilan, J	37	823	van Raan, AFJ	14	716	Walters, WH	11	100
Nicholas, D	31	389	Chen, CM	14	410	Yang, CC	11	88
Burrell, QL	26	368	Ding, Y	14	275	Moed, HF	10	307
Spink, A	24	857	Lariviere, V	14	228	Davis, PM	10	306
Pinto, M	24	100	Liang, LM	14	154	Vakkari, P	10	297
Rousseau, R	23	258	Schubert, A	13	248	Zitt, M	10	274
Jansen, BJ	22	909	Xie, HI	13	115	Meyer, M	10	265
Savolainen, R	22	278	Jaeger, PT	13	91	Small, H	10	198
Vaughan, L	21	535	White, HD	12	400	Zhao, DZ	10	167
Hjorland, B	21	464	Braun, T	12	312	Franceschet, M	10	139
Waltman, L	20	257	Kousha, K	12	258	Bilal, D	9	280
Cronin, B	19	632	Ford, N	12	209	Stvilia, B	9	154
Schreiber, M	19	187	Hernon, P	12	112	Kim, KS	9	153

Pub publications, *Cit* citation frequency

attached above the node of the largest concept whose intent contains the keywords. Therefore keywords are shown in the upper half in Fig. 2 while authors are shown in the lower half in Fig. 2. All concept nodes above a node labelled by an author have the author in their extent. All concept nodes below a node labelled by a keyword have the keyword in their intent. The bottom concept is defined by the set of all keywords and contains no author; the top concept contains all authors and is defined by their common keywords (none).

The concept lattice as shown in Fig. 2 displays the LIS knowledge structure in a reasonable manner. The concepts hierarchy is represented by the sub-concept/super-concept relation in the concept lattice, following the principle of partial order. Since the intent (keywords) of a concept is inherited by relevant sub-concepts, the intents of the top-level concepts can be recognized as the main research themes. Lower-level nodes inherit more keywords as their intents along the hierarchical relation. In this way, each research theme splits into various research topics and subtopics. The keywords of each research theme are identified by intents of super-concepts. Correspondingly, authors associated with each topic can be identified by the extent of the concept. The upper-level node represents broader topics with more authors, while the lower-level node represents more specific topics with fewer authors. The details and subtopics of each theme can be visualized through zooming in on the concept lattice. For example, from the attribute window it can be seen that the concept generated by the term "citation analysis" is a super-concept generated by the term "impact factor", and "impact factor" is more broad term than terms "h-index" and "g-index". From the author window it can be seen that the labelled authors

Table 5 The list of 99 representative keywords of LIS with frequencies

Keyword (frequency)	Keyword (frequency)	Keyword (frequency)
Information retrieval (111)	Research performance (22)	Performance evaluation (10)
h-Index (109)	Scientometrics (22)	Reference services (10)
Bibliometrics (95)	Citation distributions (21)	Source normalization (9)
Search engines (84)	Electronic journals (21)	Field normalization (9)
impact factor (82)	Link analysis (20)	Information organization (9)
Databases (67)	Information literacy (20)	Precision (9)
Worldwide web (62)	Information seeking behavior (18)	Data mining (9)
Information behavior (55)	Indexing (18)	Collaboration patterns (9)
Co-authorship (46)	Abstracting (18)	Citation behavior (8)
Science mapping (45)	Scholarly communication (17)	Retrieval effectiveness (8)
Web sites (45)	Open access (17)	Evaluation criteria (8)
Citation analysis (43)	Content analysis (17)	Ontology (8)
Information seeking (43)	Scopus (16)	Information sources (8)
Libraries (43)	Co-citation analysis (16)	Co-link analysis (8)
Peer review (43)	Similarity measure (16)	Knowledge organization (8)
Journal impact (42)	Digital libraries (16)	Scientific collaboration (8)
Web searching (41)	Power laws (15)	Electronic publishing (7)
Research evaluation (37)	Information systems (14)	Modeling (7)
Informetrics (37)	Cluster analysis (14)	Web links (7)
Citation impact (34)	r-Index (14)	Web citations (7)
User studies (32)	Information retrieval system (13)	Catalogues (6)
Information searches (30)	Google scholar (13)	Crown indicator (6)
Webometrics (30)	skills (13)	Co-authorship networks (6)
Query (28)	Scientometric indicators (12)	Journal rankings (5)
Bibliometric indicators (28)	Patent citations (12)	Search strategies (5)
Information visualization (28)	Research collaboration (12)	Knowledge management (5)
Patent (28)	Log analysis (12)	University rankings (5)
g-Index (28)	Citation network (11)	Text retrieval (5)
Social network analysis (28)	Citation window (11)	Network analysis (5)
Relevance (27)	Metadata (11)	Quantitative research (5)
Information services (24)	Librarians (11)	Quality indicators (5)
Self-citations (24)	Recall (10)	User interfaces (5)
Web of science (23)	Citation pattern (10)	Interface design (5)

are all associated with theme "citation analysis", normally, the lower the author label, the narrower the concept. For example, "Zitt, M" is attached to the narrow concept whose intent contains "citation analysis", "impact factor", "citation distribution", "field normalization", and "source normalization" (as shown later in Table 8).

By examining the concept lattice in Fig. 2, nine topics were identified as main themes in LIS. These topics were "bibliometrics, scientometrics, and informetrics", "citation analysis", "information retrieval", "information behavior", "libraries", "user studies", "social network analysis", "information visualization", and "webometrics". Detailed analysis of the nine themes is as follows.

Table 6 Partial formal context (citation analysis) based on the author-keyword association relationship

	CA	US	IB	IV	Wm	Lb	Pt	Rl
Hjorland, B	×	×	×			×		×
Bar-Ilan, J	×	×	×		×			
Leydesdorff, L	×			×			×	×
Thelwall, M	×	×			×			
Jacso, P	×			×		×		
Bornmann, L	×			×				×
Glanzel, W	×						×	×
Vaughan, L	×				×			
Burrell, QL	×							
Cronin, B	×							
Egghe, L	×							
Moed, HF	×							
van Raan, AFJ	×							

CA citation analysis, *US* user studies, *IB* information behavior, *IV* information visualization, *Wm* webometrics, *Lb* libraries, *Pt* patent, *Rl* relevance

Theme analysis: bibliometrics, scientometrics, and informetrics

Figure 3 presents the concepts and their hierarchical relationships in the theme "bibliometrics, scientometrics, and informetrics", which is illustrated as concept lattices of keywords and associated authors (Panel a) and the distribution of authors on each concept/subtopics (Panel b). These three metric terms are integrated into one big theme because they are closely related and some authors use "bibliometrics" synonymously for all three metrics (Hood and Wilson 2001). In Fig. 3, it can be seen that these three metric terms share same keywords (e.g., "co-authorship", "scientific collaboration", "power laws") as their sub-concepts. There are 34 authors in this theme. Bibliometrics is the most popular topic, 31 of the 34 authors are related to bibliometrics topic. 10 authors (about 29 % of the 34 authors) engage in scientometrics, and 5 authors (about 14 %) work on informetrics. The main keywords in the "metrics" theme include "co-authorship" (16 authors/47 %), "scientific collaboration" (10 authors/29 %), "patent" (6 authors/18 %), "power laws" (5 authors/15 %), "scholarly communication" (5 authors/15 %), "performance evaluation" (4 authors/12 %), "modeling" (4 authors/12 %), and "quantitative research" (3 authors/9 %).

The concept lattice reveals the hierarchical relation of the formal concepts of the intellectual structure in "bibliometrics, scientometrics, and informetrics" theme. Table 7 shows the intent and extent of partial concepts in this theme. For example, the extent of concept C1 contains a group of 10 authors (e.g., "Abramo, G"), while the intent of concept C1 includes "bibliometrics" and "scientific collaboration". Therefore, these authors are interested in "bibliometrics" and "scientific collaboration". To detect research details of the 10 authors, we examined concept C2, the sub-concepts of concept C1, it ca be found that 8 of them (e.g., "Franceschet, M") concentrate on "co-authorship". Through examining concept C3, another sub-concept of concept C1, it can be found that 4 of them (e.g., "Glanzel, W") focus on "patent". The concept C4 is the sub-concept of concepts C2 and C3, therefore the intent of C4 include the intent of C2 and C3, the extent of C4 is smaller than that of concepts C2 and C3, only 3 authors ("Leydesdorff, L", "Glanzel, W", "Huang, MH") were identified working on both the "co-authorship" and "patent".

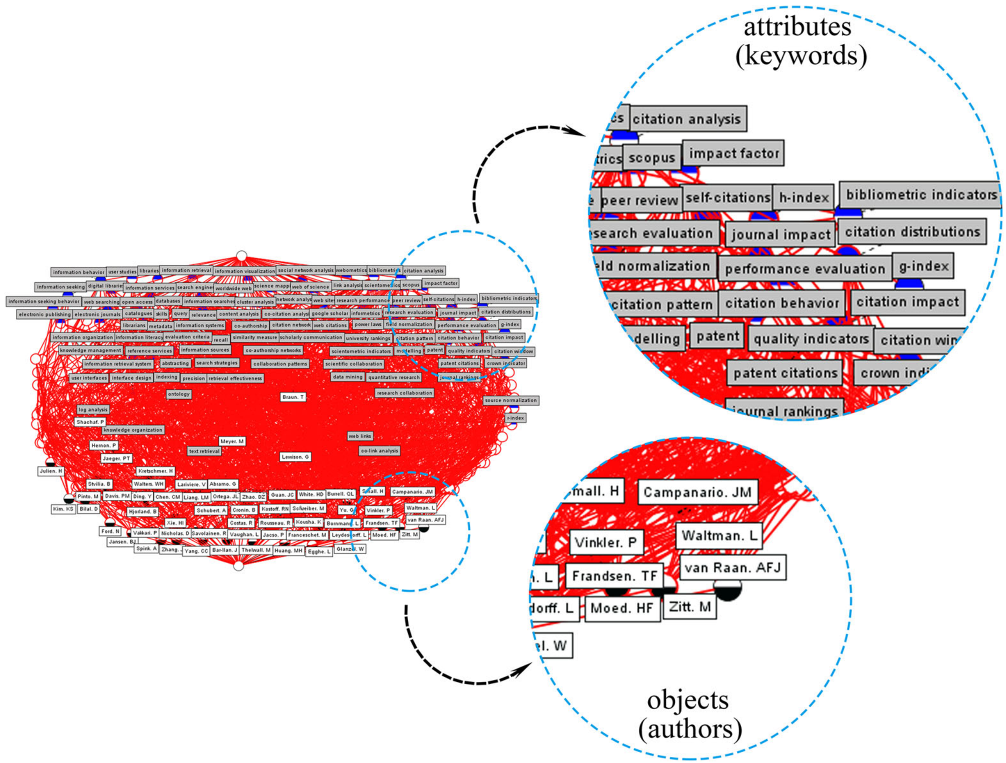

Fig. 2 Concept lattice corresponding to the LIS formal context. Each node of the concept lattice represents a formal concept (authors, keywords). Keywords are noted above the nodes in *grey-shaded boxes* as intent, while authors are noted below the nodes in *blank boxes* as extent

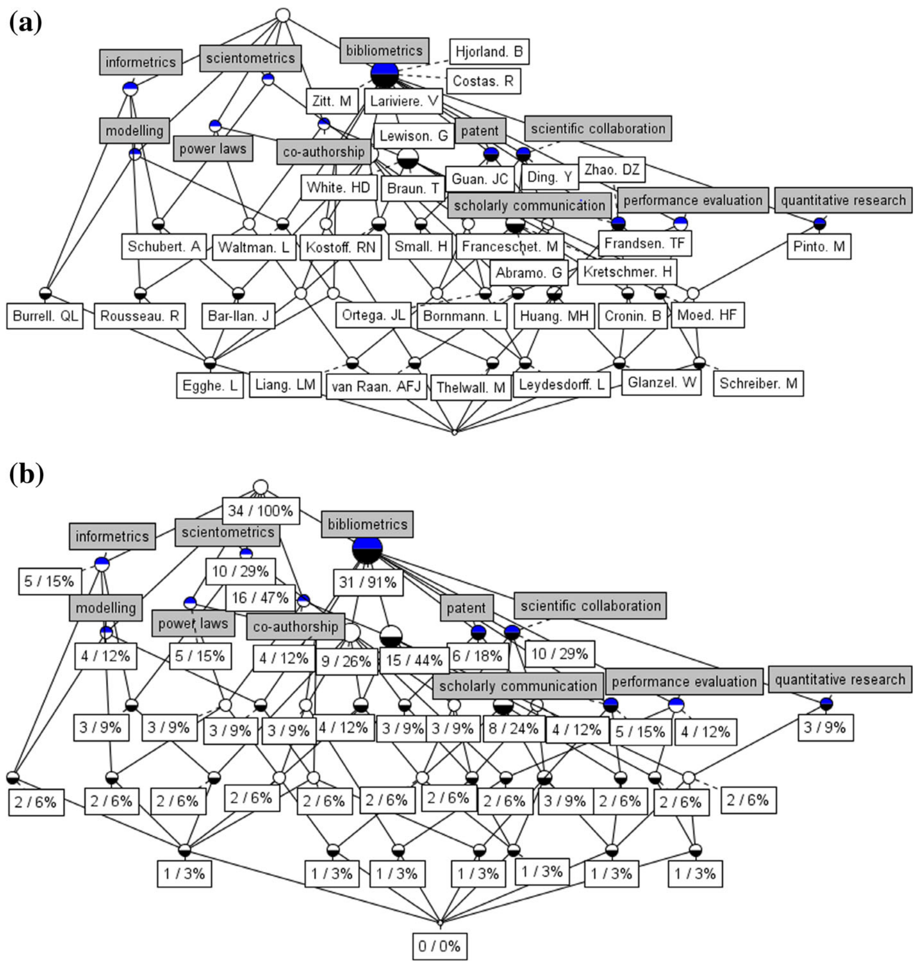

Fig. 3 Concept lattices for theme "bibliometrics, scientometrics, and informetrics" with keywords and associated authors (**a**) and the distribution of authors on each concept/subtopics (**b**)

Theme analysis: citation analysis

Fig. S1 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "citation analysis". The core keywords associated with this theme include "impact factor" (21 authors/70 %), "h-index" (18 authors/60 %), "journal impact" (16/53 %), "self-citation" (13 authors/43 %), "citation distribution" (12 authors/40 %), "citation window" (6 authors/20 %), and "patent citations" (4 authors/13 %).

Table 8 displays some formal concepts of the theme "citation analysis" with intent and extent. By examining the intent and extent of concepts C1 and C2, it revealed that studies on "h-index" as well as "impact factor" attract attention of many authors. As concept C3 is sub-concept of concept C2, we detected that 6 authors (e.g., "Burrell, QL"), who are included in the extent of concept C3, share "g-index" in its intent besides the keywords in the intent of concept C2, indicating that they devote themselves not only to "h-index" but

Table 7 Partial formal concepts of the theme "bibliometrics, scientometrics, and informetrics"

Concepts	Intent—keywords	Extent—authors
C1	Bibliometrics, scientific collaboration	Leydesdorff, L; Glanzel, W; Cronin, B; Ding, Y; Abramo, G; Liang, LM; Franceschet, M; Kretschmer, H; Ortega, JL; Huang, MH
C2	Bibliometrics, scientific collaboration, co-authorship	Leydesdorff, L; Glanzel, W; Cronin, B; Abramo, G; Liang, LM; Franceschet, M; Kretschmer, H; Huang, MH
C3	Bibliometrics, scientific collaboration, patent	Leydesdorff, L; Glanzel, W; Ortega, JL; Huang, MH
C4	Bibliometrics, scientific collaboration, co-authorship, patent	Leydesdorff, L; Glanzel, W; Huang, MH
C5	Bibliometrics, performance evaluation	Bornmann, L; van Raan, AFJ; Moed, HF; Schreiber, M
C6	Bibliometrics, quantitative research	Glanzel, W; Schreiber, M; Pinto, M.
C7	Informetrics, modeling	Burrell, QL; Egghe, L
C8	Informetrics, power laws, co-authorship	Egghe, L; Rousseau, R

also to its variant "g-index". By analyzing concept C4 (the sub-concept of C3), 3 of the 6 authors (i.e., "Egghe, L; Rousseau, R; Schreiber, M") were found to share more interests in "r-index" apart from studies on "h-index" and "g-index". The concept C7 shows that there are 12 authors (e.g., "van Raan, AFJ") also made efforts on the analysis of "citation distributions" and the study of "impact factor". The Concept C8 is a sub-concept of C7, indicates 7 of the 12 authors (e.g., "Bornmann, L") have special interest on "field normalization".

Theme analysis: information retrieval

Fig. S2 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "information retrieval". The concept lattice shows that 18 authors are included (about 30 % of the total authors) and the associated keywords are "search engines" (14 authors/78 %), "information searches" (11 authors/61 %), query (10/56 %), "web searching" (9 authors/50 %), "relevance" (10 authors/56 %), "information retrieval system" (6 authors/33 %), "information seeking behavior" (8 authors/44 %), and "retrieval effectiveness" (7 authors/39 %). Researchers in this subfield have been exploring to meet high-level requirements of users when searching for information on the Web.

Table 9 shows some concepts of the theme "information retrieval" with intent and extent. For example, the extent of concept C1 contains a group of 14 authors (e.g., "Jansen, BJ"), while the intent of concept C1 includes information retrieval and search engines. Table 9 also illustrates concepts C2, C3, C4, and C5 are the sub-concepts of concept C1, and concept C6 is the sub-concept of C5. The extent and intent of concept C2 indicate that 3 authors (i.e., "Kim, KS", "Xie, HI", "Vaughan, L") have special interest on "web search" and "evaluation criteria". The concept C3 indicates 8 authors (e.g., "Yang, CC") center on users' "information seeking behavior"; the concept C4 indicates 7 authors (e.g., "Zhang, J") concentrate on comparisons of "retrieval effectiveness" of search engines. "Relevance" is a very important topic and concept C5 illustrates that there are 7 authors

Table 8 Partial formal concepts of the theme "citation analysis"

Concepts	Intent—keywords	Extent—authors
C1	Citation analysis, impact factor	Leydesdorff, L; Glanzel, W; Thelwall, M; Egghe, L; Bornmann, L; Bar-Ilan, J; van Raan, AFJ; Vaughan, L; Jacso, P; Burrell, QL; Moed, HF; Zitt, M; Rousseau, R; Waltman, L; Schubert, A; Vinkler, P; Franceschet, M; Frandsen, TF; Campanario, JM; Huang, MH; Yu, G
C2	Citation analysis, h-index	Leydesdorff, L; Glanzel, W; Egghe, L; Bornmann, L; Bar-Ilan, J; van Raan, AFJ; Cronin, B; Jacso, P; Burrell, QL; Rousseau, R; Waltman, L; Schubert, A; Vinkler, P; Costas, R; Schreiber, M; Liang, LM; Franceschet, M; Huang, MH
C3	Citation analysis, h-index, g-index	Burrell, QL; Glanzel, W; Egghe, L; Rousseau, R; Costas, R; Schreiber, M
C4	Citation analysis, h-index, g-index, r-index	Egghe, L; Rousseau, R; Schreiber, M
C5	Citation analysis, impact factor, h-index	Leydesdorff, L; Glanzel, W; Egghe, L; Bornmann, L; Bar-Ilan, J; van Raan, AFJ; Jacso, P; Burrell, QL; Rousseau, R; Waltman, L; Schubert, A; Vinkler, P; Franceschet, M; Huang, MH
C6	Citation analysis, impact factor, h-index, quality indicators	Bornmann, L; Jacso, P; Rousseau, R
C7	Citation analysis, impact factor, citation distributions	Leydesdorff, L; Glanzel, W; Egghe, L; Bornmann, L; van Raan, AFJ; Burrell, QL; Moed, HF; Zitt, M; Waltman, L; Vinkler, P; Franceschet, M; Yu, G
C8	Citation analysis, impact factor, citation distributions, field normalization	van Raan, AFJ; Leydesdorff, L; Bornmann, L; Zitt, M; Waltman, L; Vinkler, P; Jacso, P
C9	Citation analysis, impact factor, citation distributions, field normalization, source normalization	Leydesdorff, L; Zitt, M; Waltman, L
C10	Citation analysis, impact factor, citation window	Leydesdorff, L; Glanzel, W; Bornmann, L; Zitt, M; Frandsen, TF; Campanario, JM
C11	Citation analysis, impact factor, self-citations	Leydesdorff, L; Glanzel, W; Egghe, L; van Raan, AFJ; Schubert, A; Vinkler, P; Frandsen, TF; Campanario, JM; Huang, MH; Yu, G

(e.g., "Jansen, BJ") concentrating on "relevance evaluation of search engine". Among these 7 authors, 3 authors ("Ford, N", "Xie, HI", "Yang, CC") have additional research interest on "search strategy". By examining Concept C7 it can be found that "query" is also considered as important concept, which attracts the interests of 9 authors (e.g., "Vakkari, P", "Jansen, BJ"). C8 indicates "Egghe, L", "Bar-Ilan, J", and "Zhang, J" focus on "query" and "similarity measure".

Theme analysis: information behavior

Fig. S3 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "behavior". The concept lattice shows that 18 authors (about 30 %

Table 9 Partial Formal concepts of the theme "information retrieval"

Concepts	Intent—keywords	Extent—authors
C1	Information retrieval, search engines	Jansen, BJ; Spink, A; Bar-Ilan, J; Vaughan, L; Jacso, P; Nicholas, D; Vakkari, P; Bilal, D; Ford, N; Kostoff, RN; Kim, KS; Zhang, J; Xie, HI; Yang, CC
C2	Information retrieval, search engines, web searching, evaluation criteria	Kim, KS; Xie, HI; Vaughan, L
C3	Information retrieval, search engines, information seeking behavior	Spink, A; Nicholas, D; Bilal, D; Savolainen, R; Ford, N; Kim, KS; Zhang, J; Yang, CC.
C4	Information retrieval, search engines, retrieval effectiveness	Zhang, J; Jansen, BJ; Spink, A; Bar-Ilan, J; Vakkari, P; Ford, N; Yang, CC
C5	Information retrieval, search engines, relevance	Jansen, BJ; Spink, A; Nicholas, D; Vakkari, P; Ford, N; Xie, HI; Yang, CC
C6	Information retrieval, search engines, relevance, search strategies	Ford, N; Xie, HI; Yang, CC
C7	Information retrieval, search engines, query	Jansen, BJ; Spink, A; Bar-Ilan, J; Vaughan, L; Jacso, P; Vakkari, P; Kostoff, RN; Zhang, J; Yang, CC
C8	Information retrieval, query, similarity measure	Egghe, L; Bar-Ilan, J; Zhang, J

of the total authors) are related to this subtopic. The associated keywords include "information seeking" (11 authors/61 %), "web searching" (9 authors/50 %), "information seeking behavior" (9 authors/50 %), "information searches" (8 authors/44 %), "metadata" (4 authors/22 %), "log analysis" (3 authors/17 %), "electronic journals" (3 authors/17 %), and "electronic publishing" (3 authors/17 %). There is a wide range of elements for behavior analysis covering from users, environment, resources, to channels of information.

Table 10 shows some concepts of theme "information behavior" with intent and extent. Concept C1, C2, C3 and C4 are four layer hierarchical concepts. Concept C2 indicates 11 authors (e.g., "Jansen, BJ") share interest of "information seeking" in "information behavior"; among these 11 authors 9 of them are interested in "information seeking behavior". Concept C4 indicates 2 authors ("Nicholas, D" and "Davis, PM") have special interests in "electronic journals" and "log analysis". C5 indicates that "web searching" is a shared topic by 9 authors (e.g., "Spink, A"). By examining concept C6 (the sub-concept of concept C5), it is found that authors "Davis, PM" and "Bar-Ilan, J" have special interest in "electronic journals". Besides, C7 indicates that researches by authors "Zhang, J", "Nicholas, D" and "Bar-Ilan, J" are relevant to "information searches" and "electronic publishing".

Theme analysis: libraries

Fig. S4 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "libraries". The concept lattice shows that there are 14 authors (about 23 % of the total authors) related to this subtopic. The associated keywords are "libraries" (13 authors/93 %), "librarians" (6 authors/43 %), "information seeking" (6 authors/43 %), "digital libraries" (5 authors/36 %), "information services" (5 authors/36 %), "reference services" (4 authors/29 %), "electronic journals" (4 authors/29 %), and "information organization" (3 authors/21 %).

Table 10 Partial formal concepts of the theme "information behavior"

Concepts	Intent—keywords	Extent—authors
C1	Information behavior	Zhang, J; Yang, CC; Vakkari, P; Thelwall, M; Stvilia, B; Spink, A; Savolainen, R; Pinto, M; Nicholas, D; Kim, KS; Julien, H; Jansen, BJ; Jaeger, PT; Hjorland, B; Ford, N; Davis, PM; Bilal, D; Bar-Ilan, J
C2	Information behavior, information seeking	Jansen, BJ; Hjorland, B; Zhang, J; Kim, KS; Yang, CC; Spink, A; Bilal, D; Savolainen, R; Nicholas, D; Ford, N; Davis, PM
C3	Information behavior, information seeking, information seeking behavior	Zhang, J; Kim, KS; Yang, CC; Spink, A; Bilal, D; Savolainen, R; Nicholas, D; Ford, N; Davis, PM
C4	Information behavior, information seeking, information seeking behavior, electronic journals, log analysis	Nicholas, D; Davis, PM
C5	Information behavior, web searching	Spink, A; Savolainen, R; Kim, KS; Jansen, BJ; Ford, N; Davis, PM; Bilal, D; Bar-Ilan, J; Thelwall, M
C6	Information behavior, web searching, electronic journals	Davis, PM; Bar-Ilan, J
C7	Information behavior, information searches, electronic publishing	Zhang, J; Nicholas, D; Bar-Ilan, J

Table 11 shows part of the formal concepts of libraries with intent and extent. By examining the intent and extent of concepts C1 and C2, it was found that 5 authors (e.g., "Jacso, P", "Xie, HI") share interests in "information seeking" and 4 authors (e.g., "Jacso, P", "Walters, WH") focus on "electronic journals". As concept C3 is sub-concept of concepts C1 and C2, a group of 3 authors (i.e., "Jacso, P", "Nicholas, D", "Davis, PM") work on "electronic journals" and "information seeking". Concept C4 (the sub-concept of concept C3) indicates two authors (i.e., "Jacso, P", "Nicholas, D") have additional interests in "information services" and "digital libraries".

Likewise, as indicated by concept C6 and its sub-concepts C7 and C8, it reveals that 3 authors (i.e., "Pinto, M", "Hjorland, B", "Stvilia, B") work on "information organization", of whom "Hjorland, B" and "Stvilia, B" also work on "knowledge organization", while "Pinto, M" and "Stvilia, B" focus on "metadata". By examining concept C9 and its sub-concepts C10, C11, C12, and C13 it was detected that 6 authors (e.g., "Julien, H") work on "librarians", of whom "Hernon, P", "Vakkari, P" and "Shachaf, P" also show interests in "information services" or "reference services", while "Julien, H" focuses on "information literacy", "Walters, WH" and "Jaeger, PT" focus on "open access".

Theme analysis: user studies

Fig. S5 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "user studies". The concept lattice shows that 12 authors (about 20 % of the total authors) are related to this topic. The associated keywords include "user studies" (12 authors/100 %), "search engines" (8 authors/67 %), "web searching" (7 authors/58 %), "libraries" (6 authors/50 %), "information retrieval system" (5 authors/42 %), "information seeking behavior" (4 authors/33 %), and "user interfaces" (3 authors/

Table 11 Partial Formal concepts of the theme "libraries"

Concepts	Intent—keywords	Extent—authors
C1	Libraries, information seeking	Jacso, P; Nicholas, D; Davis, PM; Hjorland, B; Xie, HI
C2	Libraries, electronic journals	Jacso, P; Nicholas, D; Davis, PM; Walters, WH
C3	Libraries, information seeking, electronic journals	Jacso, P; Nicholas, D; Davis, PM
C4	Libraries, digital libraries, information seeking, electronic journals, information services	Jacso, P; Nicholas, D
C5	Libraries, digital libraries, information seeking, evaluation criteria	Xie, HI
C6	Libraries, information organization	Hjorland, B; Stvilia, B; Pinto, M
C7	Libraries, information organization, knowledge organization	Hjorland, B; Stvilia, B
C8	Libraries, information organization, metadata	Stvilia, B; Pinto, M
C9	Libraries, librarians	Julien, H; Vakkari, P; Hernon, P; Walters, WH; Jaeger, PT; Shachaf, P
C10	Libraries, librarians, information services	Hernon, P; Vakkari, P
C11	Libraries, librarians, reference services	Vakkari, P; Shachaf, P
C12	Libraries, librarians, information literacy	Julien, H
C13	Libraries, librarians, open access	Walters, WH; Jaeger, PT

25 %). The theme "user studies" has close relationship with themes "information retrieval" and "libraries", since information retrieval is an important activity to obtain information for most users in digital environments. The studies of user need and behavior have also benefited the study of libraries and retrieval service. Therefore, "user studies" has attracted much attention from LIS community.

Table 12 shows some concepts of the theme "user studies" with intent and extent. For example, the concept C1 illustrates that 7 authors (e.g., "Bar-Ilan, J") share interest in

Table 12 Partial formal concepts of the theme "user studies"

Concepts	Intent—keywords	Extent—authors
C1	User studies, search engines, information searches	Bar-Ilan, J; Ford, N; Jansen, BJ; Nicholas, D; Spink, A; Vakkari, P; Xie, HI
C2	User studies, search engines, information searches, electronic journals	Bar-Ilan, J; Nicholas, D
C3	User studies, search engines, information searches, web searching, user interfaces	Bar-Ilan, J; Jansen, BJ
C4	User studies, search engines, information searches, relevance	Vakkari, P; Spink, A; Nicholas, D; Jansen, BJ; Ford, N; Xie, HI
C5	User studies, search engines, information searches, relevance, information seeking behavior	Ford, N; Nicholas, D; Spink, A
C6	User studies, search engines, information searches, relevance, web searching, information retrieval system	Spink, A; Jansen, BJ; Ford, N; Xie, HI
C7	User studies, search engines, information searches, relevance, libraries	Nicholas, D; Vakkari, P; Xie, HI

"information search using searching engine". By examining concepts C2, C3, C4, which are the sub-concepts of C1 concept, it was found that authors "Bar-Ilan, J" and "Nicholas, D" focus on "electronic journals", while "Bar-Ilan, J" and "Jansen, BJ" work on "comparisons of various interfaces in the assessments of search engines". The concept C4 indicates "relevance" is an important topic in "user studies", 6 authors (e.g., "Vakkari, P") share interests in "relevance judgment". By examining concepts C5, C6, and C7 (the sub-concepts of concept C4), it was detected 3 authors of them (e.g., "Ford, N") focus on "information seeking behavior"; 4 authors (e.g., "Spink, A") concentrate on "web searching" and "information retrieval systems"; 3 authors (e.g., "Vakkari, P") work on "libraries".

Theme analysis: social network analysis

Fig. S6 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "social network analysis". The concept lattice shows that 12 authors (about 20 % of the total authors) are related to this topic. The associated keywords include "scientific collaboration" (6 authors/50 %), "cluster analysis" (6 authors/50 %), citation network (5 authors/42 %), "collaboration patterns" (4 authors/33 %), "network analysis" (3 authors/25 %), "co-authorship networks" (3 authors/25 %), "bibliometric indicator" (3 authors/25 %), and "co-citation analysis" (2 authors/17 %).

Table 13 shows some concepts of the theme "social network analysis" with intent and extent. For example, concept C1 indicates 6 authors (e.g., "Leydesdorff, L") share interest of "social network analysis" in "scientific collaboration". Concepts C2, C3, C4 and C5 (all are sub-concepts of concept C1) indicate that 2 authors ("Glanzel, W" and "Huang, MH") have special interest in "patent citations network"; 2 authors ("Glanzel, W" and "Kretschmer, H") explore to detect collaboration patterns through analysis of topological characteristics and structure of collaboration networks; while 3 authors ("Leydesdorff, L", "Ding, Y", "Kretschmer, H") focus on "co-authorship network", of whom "Leydesdorff, L" and "Ding, Y" have same interests in "citation network", "co-citation network", and "cluster analysis". Besides, concept C6 indicates that 6 authors (e.g., "Waltman, L") apply

Table 13 Partial formal concepts of the theme "social network analysis"

Concepts	Intent—keywords	Extent—authors
C1	Social network analysis, scientific collaboration	Leydesdorff, L; Glanzel, W; Ding, Y; Franceschet, M; Kretschmer, H; Huang, MH
C2	Social network analysis, scientific collaboration, patent citations	Glanzel, W; Huang, MH
C3	Social network analysis, scientific collaboration, collaboration patterns	Glanzel, W; Kretschmer, H
C4	Social network analysis, scientific collaboration, co-authorship networks	Leydesdorff, L; Ding, Y; Kretschmer, H
C5	Social network analysis, scientific collaboration, co-authorship networks, co-citation analysis, citation network, cluster analysis	Leydesdorff, L; Ding, Y
C6	Social network analysis, cluster analysis	Waltman, L; Ding, Y; Leydesdorff, L; Glanzel, W; Franceschet, M; Ortega, JL
C7	Social network analysis, bibliometric indicators	Rousseau, R; Kretschmer, H; Guan, JC

"cluster analysis" into social network analysis (SNA), and concept C7 indicates that "Rousseau, R", "Guan, JC" and "Kretschmer, H" share interests in "bibliometric indicators" and "social network analysis". SNA has been providing potential as a powerful tool for analyzing complicated relationship and researchers of LIS make new explorations on traditional relationship like author co-citation, co-authorship from a new perspective of SNA. New methods are proposed for identification of thematic clusters, community detection, detection of collaboration patterns, and impact analysis.

Theme analysis: information visualization

Fig. S7 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "information visualization". The concept lattice shows that 12 authors (about 20 % of the total authors) are related in this topic. The associated keywords include "information visualization" (12 authors/100 %), "science mapping" (10 authors/83 %), "co-citation analysis" (6 authors/50 %), "co-authorship" (5 authors/42 %), "cluster analysis" (5 authors/42 %), and "information searches" (3 authors/25 %). Visualization techniques are mainly used in science mapping and information retrieval.

Table 14 shows some concepts of the theme "information visualization" with intent and extent. As indicated by intent and extent of concept C1, authors "Zhang, J" and "Jacso, P" apply "visualization techniques into information searches". One of the challenges of information searches is the expansion of search engines and index pages, which makes it difficult for users to effectively find information. Since it is intuitive for people to identify visual patterns, the research on "information visualization" aims to transform searching data into a visual format. The concept C2 revealed that great efforts have been made by 10 authors (e.g., "Bornmann, L") on "application of visualization in science mapping". The concept C3 reveals 5 authors (e.g., "Jacso, P") use "cluster analysis" in information visualization. C4 is the subconcept of C2 and C3, indicates 4 authors (e.g., "Waltman, L") work on "science mapping" and "cluster analysis". The concepts C5 shows that 5 authors (e.g., "Chen, CM") work on "science mapping" based on "co-citation analysis". By

Table 14 Partial formal concepts of the theme "information visualization"

Concept	Intent—keywords	Extent—authors
C1	Information visualization, information searches	Zhang, J; Jacso, P
C2	Information visualization, science mapping	Bornmann, L; Chen, CM; Leydesdorff, L; Ortega, JL; Small, H; Waltman, L; White, HD; Zhao, DZ; Zitt, M; Glanzel, W
C3	Information visualization, cluster analysis	Jacso, P; Leydesdorff, L; Small, H; Waltman, L; Glanzel, W
C4	Information visualization, science mapping, cluster analysis	Waltman, L., Leydesdorff, L; Small, H; Glanzel, W
C5	Information visualization, science mapping, co-citation analysis	Chen, CM; Zhao, DZ; Leydesdorff, L; Small, H; White, HD
C6	Information visualization, co-authorship	White, HD; Bornmann, L; Leydesdorff, L; Glanzel, W
C7	Information visualization, co-authorship, research performance (2)	Bornmann, L; Glanzel, W

examining the concept C7 (the sub-concept of concept C6), authors "Bornmann, L" and "Glanzel, W" are identified to share interests in "research performance" besides "co-authorship analysis". Visualization explorations include tracking knowledge diffusion based on visualization of co-citation network, mapping research fronts for a specialty, or mapping institutions, regions or universities of excellence in the world. The application of visualization explorations makes research results easy to understand and enables researchers to detect implicit relationship and distributing characteristics.

Theme analysis: webometrics

Fig. S8 (see Supplementary Materials) presents the concepts and their hierarchical relationships in the theme "webometrics". The concept lattice shows that there are 5 authors (about 8 % of the total authors) related to this topic. The associated keywords include "web citations" (5 authors/100 %), "search engines" (4 authors/80 %), "scholarly communication" (3 authors/60 %), "research collaboration" (3 authors/60 %), "link analysis" (3 authors/60 %), "web sites" (3 authors/60 %), and "co-link analysis" (2 authors/40 %). Researches of the theme "webometrics" are based on "web links", "co-link analysis", "search engines", and "web sites". Webometrics inherits bibliometrics with flexibility in web environment. "Web links" of "webmometrics" are similar to "citation relations" of "bibliometrics". "Co-link" of "webmometrics" is similar to "co-citation" of "bibliometrics" (Yang and Sun 2013). "Search engines" of "webmometrics" are widely used to find quantitative data such as the number of pages matching a query and the international spread of pages (Thelwall 2008). "Web sites" of "webmometrics" are regarded as the unit for analysis.

Table 15 shows some concepts of the theme "webometrics" with intent and extent. For example, concept C2 indicates that 3 authors (e.g., "Thelwall, M") focus on "scientific collaboration under web environment". Researches in webometrics are distinguished from traditional bibliometrics by indicators such as "web links" and "web visibility", which are introduced as collaborative studies instead of that based on co-authorship. The concept C3 (the sub-concept of concept C2) indicates that "Kretschmer, H" and "Ortega, JL" of the 3 authors apply "social network analysis" into "webometrics". By analyzing concept C4, 3 authors (e.g., "Kousha, K") are detected to work on "explorations of online informal scholarly communication". By examining concept C5 (the sub-concept of concept C4),

Table 15 Partial formal concepts of the theme "webometrics"

No.	Intent—keywords	Extent—authors
C1	Webometrics	Thelwall, M; Vaughan, L; Kousha, K; Kretschmer, H; Ortega, JL
C2	Webometrics, research collaboration	Thelwall, M; Kretschmer, H; Ortega, JL
C3	Webometrics, research collaboration, social network analysis	Kretschmer, H; Ortega, JL
C4	Webometrics, search engines, web citations, scholarly communication	Kousha, K; Vaughan, L; Thelwall, M
C5	Webometrics, search engines, web citations, scholarly communication, performance evaluation	Kousha, K; Vaughan, L
C6	Webometrics, search engines, web citations, web sites, web links	Thelwall, M; Vaughan, L; Ortega, JL

two of the 3 authors (i.e., “Kousha, K” and “Vaughan, L”) were found to also work on “performance evaluation”. The concept C6 indicates 3 authors (i.e., “Thelwall, M”, “Vaughan, L”, “Ortega, JL”) work on “web citations” and “links to web sites”.

Evaluation and discussion

Over the past 20 years, many researchers (Milojević et al. 2011; Åström 2007; Moya-Anégon et al. 2006; Janssens et al. 2006; White and McCain 1998; Zhao and Strotmann 2008b) have examined the intellectual structure of LIS. However the results attained by the researchers are different from each other, and the number of the main themes range from 3 to 16. It is probably due to several factors, e.g., the data collected for analysis covered different core LIS journals, different time period, or the methods frequently used have relatively strong subjective judgments. Given that there is no ground truth or gold standard available for the analysis of our result, the results detected using ACA in Zhao and Strotmann’s study (Zhao and Strotmann 2008b) was chosen as comparison since ACA is the most widely used approach to detect intellectual structure. Zhao and Strotmann detected the intellectual structure of information science based on 12 core journals from 1996 to 2005 including four library-oriented journals, the detected intellectual structure can be comparable to ours. The main themes detected by two approaches are listed in Table 16.

In general, nine research themes are identified by FCA-based approach in the present study, while twelve themes are detected by ACA-based method. Three research themes (citation analysis, user studies, webometrics) were revealed by both methods, and another five research themes detected by FCA are similar or close to some themes by ACA.

The research theme “bibliometrics and informetrics and scientometrics” detected by FCA corresponds to “bibliometric models and distributions” by ACA, but demonstrates a broader scope. The research theme “information retrieval” detected by FCA-based method encompasses three themes (experimental retrieval, users’ judgments of relevance, Information seeking and context) by ACA. Concept lattice illustrated that the sub-keywords of information retrieval include information seeking, relevance and information retrieval

Table 16 Comparison between the intellectual structures of LIS detected by FCA and ACA

FCA	ACA
Bibliometrics and informetrics and scientometrics	Bibliometric models and distributions
Citation analysis	Citation analysis
Information retrieval	Experimental retrieval users’ judgments of relevance (situational relevance) information seeking and context
Information behavior	Children’s information searching behavior
Libraries	Metadata and digital resources
User studies	User studies
Social network analysis	–
Information visualization	Visualization of knowledge domains
Webometrics	Webometrics
–	Science communication
–	Structured abstracts (academic writing)

systems. The research theme “information behavior” identified by FCA corresponds to the specialty “children’s information searching behavior” by ACA, while the former is much broader. The research theme “libraries” detected by FCA corresponds to the specialty “metadata and digital resources” by ACA. Metadata was initially used in the library catalogue as part of Library Management System (Bagley 1968; Solntseff and Yezerski 1974). Digital resource is an essential part in the transformation from traditional library to digital library, which focuses on the organization and process of digital resources. The research theme “information visualization” identified by FCA corresponds to the specialty “visualization of knowledge domains” by ACA. The intellectual structures recognized by FCA and by ACA are consistent to a large degree.

Two specialties, “science communication” and “structured abstracts (academic writing)”, were solely detected by ACA-based approach. In FCA-based approach, a similar phrase “scholarly communication” is found to be the intent of sub-concept of “bibliometrics and informetrics and scientometrics” and “webometrics”, and “abstracting” is the associated keyword in the theme “information retrieval” and “libraries”. The ACA-based method detected only one author (i.e., Hartley) in the specialty “structured abstracts”, which was not recognized as a research theme in FCA-based method.

A relatively new research theme “social network analysis (SNA)” was identified by FCA-based method. The emerging SNA was bound up with the rise and development of social network sites (SNSs), and studies on SNA were initially seen in sociology, then introduced to LIS and resulted in the growing of academic research, including data mining, network modeling and sampling, structure analysis through user attributes and behavior, community-maintained resource support and recommender systems development. SNA is supposed to be an emerging specialty of LIS (Yin and Ma 2009; Zhu and Li 2008). Since publications relevant to SNA were rare before 2005, this theme was not detected in the Zhao and Strotmann’s ACA study.

The previous researches on detecting intellectual structure seldom revealed the hierarchical structure of the discipline. The associations between themes and authors often depend on subjective assessments. Instead, FCA-based exploration discovers main themes, the keywords and active authors associated with each theme. By analysis of the extent and intent of sub-concepts, authors related to one theme are further clustered, revealing shared interests of authors from different granularities. FCA-based approach enables a better organization of authors and keywords, and provides an aid to understanding the hidden and complex interactions between authors and keywords. The hierarchical conceptual clustering is better displayed through semantic net in concept lattice than traditional tree diagram. By clustering authors and keywords simultaneously, our approach is also applicable to other applications like detection of academic community, recommendation of search queries, or construction of discipline knowledge map.

Conclusion

It is always a hot topic in the study of LIS for visualizing and mapping the intellectual structure of specific knowledge domains, especially when analytical methods are widely adopted such as co-citation analysis and co-word analysis. The present study proposes a FCA-based method to identify the intellectual structure. By taking the authors (set) as the extent and the keywords as the intent in FCA analysis, the intellectual structure is detected and displayed with the concept lattice. The empirical analysis verified that the FCA-based intellectual analysis clearly revealed the complicated relationships between research themes, between authors, as well as between research themes and authors at different levels.

Comparing with traditional bibliographic methods such as ACA, author bibliographic coupling analysis, and author keyword coupling analysis, there are at least three advantages of FCA. Firstly, FCA decomposes the authors and their keywords into different modules by constructing formal context of authors and keywords at the same time. As a result, intellectual structure can be identified at a fine-grained level of granularity. Secondly, by analyzing the similarity of different portions in the concept lattice, similarities and differences between authors in multi-levels are easily recognized. Thirdly, it allows for describing the research topics of a group of authors with the corresponding keywords in the intent directly, which reduces manual interference and allows the presentation of an objective and actual intellectual structure of a discipline. It also avoids the limitation of time delay in co-citation analysis, which is crucial in detecting emerging research topics.

However, there are still remaining issues. One of them is that the large number of concepts makes it too complicated to visualize. A possible solution is to combine approaches such as rough sets and concept stability for reducing the complexity of the concept lattice structure. Another future work may focus on how to formulate a multiple-valued formal context to better represent authors' interests.

Acknowledgments This project is supported by National Natural Science Foundation of China (71203164) and the Fundamental Research Funds for the Central Universities (2012GSP058, Wuhan University).

References

Ahlgren, P., & Jarneving, B. (2008). Bibliographic coupling, common abstract stems and clustering: A comparison of two document–document similarity approaches in the context of science mapping. *Scientometrics, 76*(2), 273–290. doi:10.1007/s11192-007-1935-1.

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. doi:10.1002/asi.22917.

Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology, 58*(7), 947–957. doi:10.1002/asi.20567.

Bagley, P. R. (1968). Extent of programming language concepts. University City Science Center technical report, November 1968.

Bates, M. J., & Maack, M. N. (2010). *Encyclopedia of library and information sciences*. Boca Raton: CRC Press.

Chang, Y. W., & Huang, M. H. (2012). A study of the evolution of interdisciplinarity in library and information science: Using three bibliometric methods. *Journal of American Society for Information Science and Technology, 63*(1), 22–33.

Chen, L. C., & Lien, Y. H. (2011). Using author co-citation analysis to examine the intellectual structure of e-learning: A MIS perspective. *Scientometrics, 89*(3), 867–886. doi:10.1007/s11192-011-0458-y.

Ganter, B., & Wille, R. (2012). *Formal concept analysis: Mathematical foundations[M]* (p. 2012). Berlin: Springer.

Gao, J. P., Ding, K., Teng, L., & Pang, J. (2012). Hybrid documents co-citation analysis: Making sense of the interaction between science and technology in technology diffusion. *Scientometrics, 93*(2), 459–471. doi:10.1007/s11192-012-0691-z.

Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science, 30*(2), 119–145. doi:10.1177/0165551504042802.

Godin, R., Missaoui, R., & Alaoui, H. (1995). Incremental concept formation algorithms based on Galois (concept) lattices. *Computational Intelligence, 11*(2), 246–267. doi:10.1111/j.1467-8640.1995.tb00031.x.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics, 52*(2), 291–314. doi:10.1023/A:1017919924342.

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. doi:10.1007/s11192-013-1076-7.

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. doi:10.1007/s11192-010-0313-6.

Janssens, F., Leta, J., Glanzel, W., & Moor, B. D. (2006). Towards mapping library and information science. *Information Processing and Management, 42*(6), 1614–1642. doi:10.1016/j.ipm.2006.03.025.

Kedrov, B. M. (1980). On modern classification of sciences. *Philosophic Issues, 10*, 85–103.

Kuznetsov, S. O., & Obiedkov, S. A. (2002). Comparing performance of algorithms for generating concept lattices. *Journal of Experimental and Theoretical Artificial Intelligence, 14*(2–3), 189–216. doi:10.1080/09528130210164170.

Larivière, V., Sugimoto, C. R., & Cronin, B. (2012). A bibliometric chronicling of library and information science's first hundred years. *Journal of the American Society for Information Science and Technology, 63*(5), 997–1016. doi:10.1002/asi.22645.

Li, H. Y., Cui, L., Cui, M., & Tong, Y. Y. (2010). Active research fields of acupuncture research: A document co-citation clustering analysis of acupuncture literature. *Alternative Therapies in Health and Medicine, 16*(6), 38–45.

Liu, Z. H., & Zhang, Z. Q. (2010). Author keyword coupling analysis: An empirical research. *Journal of the China Society for Scientific and Technical Information, 29*(2), 268–275. **(in Chinese)**.

Ma, R. M. (2012). Author bibliographic coupling analysis: A test based on a Chinese academic database. *Journal of Informetrics, 6*(4), 532–542. doi:10.1016/j.joi.2012.04.006.

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. doi:10.1002/asi.21602.

Moya-Anégon, F., Herrero-Solana, V., & Jimenez-Contreras, E. (2006). A connectionist and multivariate approach to science maps: The SOM, clustering and MDS applied to library and information science research. *Journal of Information Science, 32*(1), 63–77. doi:10.1177/0165551506059226.

Mustafee, N., Katsaliaki, K., & Fishwick, P. (2014). Exploring the modelling and simulation knowledge base through journal co-citation analysis. *Scientometrics, 98*(3), 2145–2159. doi:10.1007/s11192-013-1136-z.

Nicolai, J. F., & Torben, P. (2003). The MNC as knowledge structure: The roles of knowledge sources and organizational instruments in MNC knowledge management. *Danish Research Unit for Industrial Dynamics, 2003*(5), 1–33.

Price, D. J. S. (1965). The scientific foundations of science policy. *Nature, 206*(4), 233–238. doi:10.1038/206233a0.

Solntseff, N., & Yezerski, A. (1974). A survey of extensible programming languages. *Annual review in automatic programming, 7*, 267–307. doi:10.1016/0066-4138(74)90001-9.

Thelwall, M. (2008). Quantitative comparisons of search engine results. *Journal of the American Society for Information Science and Technology, 59*(11), 1702–1710. doi:10.1002/asi.20834.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for information Science, 32*(3), 163–171. doi:10.1002/asi.4630320302.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American society for information science, 49*(4), 327–355. doi:10.1002/(SICI)1097-4571(19980401)49:4<327:AID-ASI4>3.0.CO;2-W.

Wille, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered sets* (pp. 445–470). Dordrecht (Vol. 83 of NATO Advanced Studies Institute), Boston: Reidel.

Wille, R. (2005). Formal concept analysis as mathematical theory of concepts and concept hierarchies. In B. Ganter, G. Stumme, & R. Wille (Eds.), *Formal concept analysis: Foundations and applications, state-of-the-art survey* (pp. 1–33). Berlin: Springer.

Yang, B., & Sun, Y. (2013). An exploration of link-based knowledge map in academic web space. *Scientometrics, 96*(1), 239–253. doi:10.1007/s11192-012-0919-y.

Yin, Z. M., & Ma, R. M. (2009). Review of the application research of SNA on the library and information science in China. *Document, Information and Knowledge, 2009*(6), 64–69. **(in Chinese)**.

Zhao, D., & Strotmann, A. (2008a). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology, 59*(13), 2070–2086. doi:10.1002/asi.20910.

Zhao, D., & Strotmann, A. (2008b). Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology, 59*(6), 916–937. doi:10.1002/asi.20799.

Zhu, Q. H., & Li, L. (2008). Social network analysis and the application achievement of SNA in information science. *Information Studies: Theory and Application, 31*(2), 179–183. **(in Chinese)**.