

# Exploring the topic hierarchy of digital library research in China using keyword networks: a K-core decomposition approach

Lu Xiao<sup>1</sup> · Guo Chen<sup>2</sup> · Jianjun Sun<sup>1</sup> · Shuguang Han<sup>3</sup> · Chengzhi Zhang<sup>2</sup>

Received: 5 August 2015 / Published online: 2 July 2016  
© Akadémiai Kiadó, Budapest, Hungary 2016

**Abstract** Exploring the topic hierarchy of a research field can help us better recognize its intellectual structure. This paper proposes a new method to automatically discover the topic hierarchy, in which the keyword network is constructed to represent topics and their relations, and then decomposed hierarchically into shells using the K-core decomposition method. Adjacent shells with similar morphology are merged into layers according to their density and clustering coefficient. In the keyword network of the digital library field in China, we discover four different layers. The basic layer contains 17 tightly-interconnected core concepts which form the knowledge base of the field. The middle layer contains 13 mediator concepts which are directly connected to technology concepts in the basic layer, showing the knowledge evolution of the field. The detail layer contains 65 concrete concepts which can be grouped into 13 clusters, indicating the research specializations of the field. The marginal layer contains peripheral or isolated concepts.

**Keywords** Intellectual structure · Topic hierarchy · Keyword network · K-core decomposition · Digital library in China

## Introduction

In China, digital library (DL) has attracted much attention from academia in the past decades. Research topics range from the theoretical perspective of DL to practical applications for DL techniques (Zhou 2005; Qiu and Wang 2010; Shen et al. 2008). Nowadays,

---

✉ Guo Chen  
chenguo@njust.edu.cn

<sup>1</sup> School of Information Management, Nanjing University, Nanjing 210046, China

<sup>2</sup> Department of Information Management, Nanjing University of Science and Technology, Nanjing 210094, China

<sup>3</sup> School of Information Sciences, University of Pittsburgh, Pittsburgh, PA 15213, USA

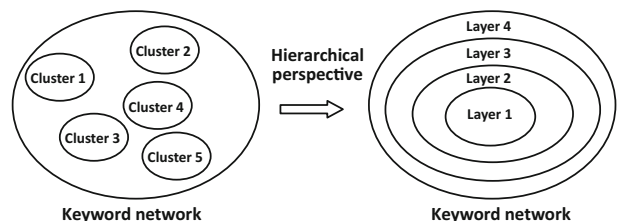
it has become one of the most important subfields of Library and Information Science (LIS) in China, with the highest number of LIS publications (Su and Xia 2011; Liu et al. 2012). Zhao and Zhang (2011) found that DL studies in China are more diversified and decentralized compared to international studies. To provide a fine-grained analysis of DL research topics, scholars have attempted to depict its internal structure (Dong 2009; Zhang and Lv (2010); Su and Xia 2011; Xu and Yang 2011; Zhao and Zhang 2011; Liu et al. 2012). In these studies, keyword co-occurrence relationships are often utilized to describe the association between concepts, and researchers have mainly focused on mapping knowledge structure of the field based on cluster analysis of keywords.

Yet, clustering is considered to be a crude approach for organizing and structuring knowledge concepts (Ma and Du 2007). In one research domain, there are three main kinds of relationships between knowledge concepts: equivalence relationships, hierarchical relationships, and associative relationships (Green 2001). Clustering only aims to find the associative relations and the hierarchical relations are ignored. Previous studies identified that human intellectual knowledge is often structuralized as a set of knowledge concepts and their hierarchical relations (Corbett and Anderson 1994). Nguyen and Chowdhury (2013) believed that the hierarchical structure of a domain map is more comprehensive, systematic and can show the knowledge evolution of a domain. Therefore, it is necessary to take hierarchical relationships into account when mapping the knowledge structure of domain concepts.

To identify the hierarchical relations from keyword networks, we can learn the past successes from the studies of complex networks. Recent studies have provided evidence that real-world networks (keyword network as one example) are often hierarchically organized (Ravasz et al. 2002; Barabási et al. 2003; Clauset et al. 2007; Choi et al. 2011). Many methods have already been proposed to identify the hierarchical structure of complex networks (Alvarez-Hamelin et al. 2005; Carmi et al. 2007; Sales-Pardo et al. 2007; Clauset et al. 2008). However, little attention has been given to the hierarchical structure of keyword networks.

In this paper, we propose a quantitative approach to explore the topic hierarchy of the DL field in China based on its keyword network. The keyword network is constructed and abstracted based on keyword co-occurrence relationships, and then decomposed hierarchically into shells using the K-core decomposition method. Finally, adjacent shells with similar morphology are merged into layers according to their density and clustering coefficient. Our research offers a different perspective of uncovering the intellectual structure of the DL field in China. The differences between our method and the traditional method are illustrated in Fig. 1, in which we focus on dividing the keyword network into hierarchical layers (right) instead of simply clustering them into small groups (left).

**Fig. 1** A new perspective of uncovering the intellectual structure of a research field



## Related work

### Intellectual structure of the DL in China

At present, many researchers have analyzed the DL field in China using both qualitative and quantitative methods. Specifically, there are many quantitative studies based on keyword analysis, a method that has been widely utilized to reveal the intellectual structure of a research field. Keywords are believed to be the most basic fundamental carrier of knowledge (Lee et al. 2010). Keyword co-occurrence within an article suggests the keywords are relevant to the topics they refer to (Cambrosio et al. 1993).

Dong (2009) utilized the co-word method to cluster high-frequency keywords of DL research papers in China from 1999 to 2008. He clusters the research topics into four parts, including *resource organization technologies*, *resource building*, *information service*, and *copyright*. Zhang and Lv (2010) reviewed the keywords of DL research papers in China from 2004 to 2008. They indicate three main directions including *resource*, *technology* and *copyright*. Su and Xia (2011) analyzed the high-frequency keywords of DL research papers in China from 2000 to 2009 and summarize the field with six clusters, including: *information resource building and sharing*, *information service*, *information storing and description*, *copyright*, *digital library construction*, and *key techniques*. They also conclude that resource building and sharing is always a hotspot in the field. Xu and Yang (2011) extracted the author-keyword network of the field based on papers from 1998 to 2010, and identify four topic clusters including *technology*, *resource*, *service*, and *copyright*. Zhao and Zhang (2011) constructed the co-word network of China's DL field based on research papers from 1994 to 2010. After analyzing four topic clusters, including *services*, *copyright*, *basic theories*, and *content & technologies*, they point out that DL research in China is much more decentralized, and focuses more on right issues and basic theories, than international research. Liu et al. (2012) identify seven topic clusters based on the co-word analysis of DL research papers in China from 2002 to 2011; they conclude that the field is based on *resource problems*, supported by *technologies* and centered on *service*.

### Exploring the hierarchical structure of complex networks

Real-world networks often have an inherently hierarchical organization (Ravasz et al. 2002; Barabási et al. 2003; Clauset et al. 2007; Sales-Pardo et al. 2007), in which nodes are clustered into many small, highly-connected groups, which are then gathered into groups at higher levels. Some important topological properties of networks, such as the scale-free property, high clustering coefficient and small short path, are consequences of hierarchical organization (Clauset et al. 2008). Yi and Choi (2012) found that the correlation between the degree of each keyword and its clustering coefficient exhibit a scaling behavior; such a correlation indicates a network's inherently hierarchical organization (Barabási et al. 2003).

Many studies have focused on identifying the hierarchical structure of real-world complex networks. Sales-Pardo et al. (2007) adopted the modularity measure and proposed a box-clustering method to identify modules at hierarchical levels. Clauset et al. (2008) produced a dendrogram with a set of probabilities as a graphical representation and summary of a network's hierarchical structure. Alvarez-Hamelin et al. (2005) proposed a decomposition method in which nodes are partitioned into hierarchical shells based on their K-core value. Carmi et al. (2007) decomposed the Internet into 41 shells by using the

K-core decomposition method, and then divided them into three components based on percolation theory.

However, these studies mainly focus on revealing the macroscopic properties of hierarchical structures in complex networks, such as interpersonal networks, the World Wide Web, the Internet, power grids and so on. There is a lack of research that focuses on the hierarchical details of keyword networks, and this remains a challenging research problem.

## Methodology

To explore the topic hierarchy of the DL field in China, we firstly extract the author-assigned keywords of papers in the field as proxies of its research topics. Then a keyword network is constructed based on the co-occurrences of keywords, and a subnet is abstracted, which is quite different from the traditional high-frequency keyword networks. Afterwards, we adopt the K-core decomposition method and improve it so as to divide the keyword network into hierarchical layers.

### Data collection and preprocessing

The Chinese Journal Full-Text Database (CJFTD) is used as our data source because it contains almost all of the important journal papers in China. Target papers are collected by retrieving the term digital library (in Chinese) within the title or author keywords of the papers, setting the time span from 2003 to 2012. “Core journal” is selected as the data source category. After eliminating those informal papers, such as conference notices and book reviews, we obtained 2107 papers as a proxy of the DL research field in China. Then, the author keywords of these papers are extracted and their co-occurrence relationships are accumulated.

Before constructing the keyword network, we have to manually remove keywords that are too general (He 1999). We firstly eliminate meaningless terms such as *research*, *counter measure*, *problem* and so on. The term *digital library* is also removed because it is presumably related to all other keywords. Since different authors may use various keywords when describing the same concept, we map all keywords with the same meaning into a standard form. After the preprocessing, 2136 keywords with a total frequency of 5488 are retained.

### Constructing and abstracting the keyword network

The overall keyword network of the DL field in China is constructed based on the co-occurrence relationships of keywords. Since there are so many nodes and edges in the network, abstraction is necessary for analysis and visualization.

In previous studies, researchers have mainly focused on identifying research topics (for example, research theme clustering and network community discovering), such that high-frequency keywords are usually considered to be important and keyword networks are often abstracted by eliminating keywords below a certain frequency (Choi et al. 2011). However, keywords may be used frequently because they are generalizations or represent popular themes; these words may be useful in showing a rough overview of a research field, but are less successful at displaying its detailed themes (Chen and Xiao 2016). This is because mid-frequency keywords are more likely to represent specific concepts which can

help people better recognize a field (Luhn 1958; Salton 1975; Rokaya et al. 2008), and low-frequency keywords can express emerging new concepts (Quoniam et al. 1998). Thus, the high-frequency network is partial in exploring the topical hierarchy of a research field. A more comprehensive keyword network containing keywords at different levels (for example, basic concepts, intermediate concepts, and detailed concepts) should be abstracted from the overall keyword network of China’s DL.

As Zhao et al. (2014) conclude, the most traditional network metrics concentrating on node measures (especially the ones coming from social network analysis) are designed for unweighted networks. They proposed a method called the *h*-subnet to naturally simplify a weighted complex network into a small and concise sub-network that retains the most important links within its core structure. The key function of their method is to extract the network based on the link strength of nodes. The subnet is defined as *h*-subnet, which includes all nodes connected by links with strength larger than or equal to *h*. This has been proven to cover a large segment of important nodes and is more efficient in presenting the network’s major structure.

The keyword network in our case is weighted, so we will abstract it using the *h*-subnet. The distribution of link strength in the overall keyword network is shown in Table 1, based on which we can set the link strength threshold as 2. The 2-subnet is accordingly abstracted by eliminating keyword relations with strength below 2.

Compared with the high-frequency keyword network, this new subnet is more reliable because it retains all strong co-word relations, which have been omitted in the high-frequency keyword network. It can also eliminate the interference effect from low-strength relations. Besides, important keywords with low frequency are included, resulting in a more comprehensive knowledge map of the research field.

### Decomposing the keyword network with K-core

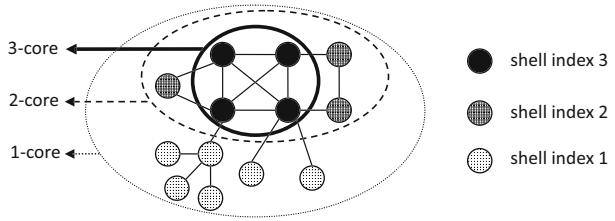
The K-core of a network is the largest sub-network in which each node has at least *k* interconnections (Dorogovtsev et al. 2006). The K-core value (also called the shell index) of a node is defined as *k* if it belongs to the K-core but not to the (*k* + 1)-core. Thus, the K-shell of a network is accordingly composed of all nodes with a shell index of *k* (Fig. 2). Nodes in higher shells are considered to be more central (Carmi et al. 2007). Moreover, the K-core value is a quite robust measure (Kitsak et al. 2010) compared with node degree, because its range extends far more slowly when the network grows rapidly, and the size of higher shells stays more stable (Zhang et al. 2008). Thus, K-core decomposition has been widely used to decompose networks hierarchically (Tong et al. 2002; Alvarez-Hamelin et al. 2005; Carmi et al. 2007; Zhang et al. 2010; Kitsak et al. 2010).

A potential problem with K-core decomposition is that it may result in some similar shells. In the study by Carmi et al. (2007), the Internet was decomposed into 41 shells. They believed that some adjacent shells are similar and should be merged for analysis, so

**Table 1** The distribution of link strength in the overall keyword network of the DL field in China

Link strength	Number of links
1	4152
2	231
3	81
≥4	79

**Fig. 2** The K-core and K-shell of a network (Alvarez-Hamelin et al. 2005)

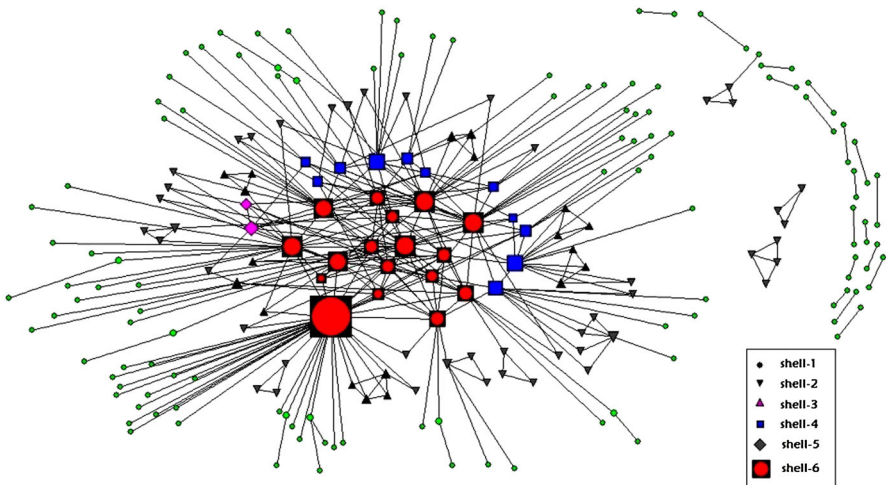


the 41 shells were divided into three components based on the percolation theory. We will use the same idea and merge the K-shells of our keyword network so as to make its structure simpler and more logical. In doing this, selecting a parameter to identify the similarity of different shells is the key. Since the hierarchical structure of networks can be indicated by the inverse relationship between the clustering coefficient and degree of nodes (Barabási et al. 2003), we will use the clustering coefficient as the parameter for shell recombination. The clustering coefficient is a well-defined and widely-used characteristic index of networks. It measures the amount of cliquishness of the network, that is, the fraction of neighboring nodes that are also connected to one another (Collins and Chow 1998). We view the shells as sub-networks so that their clustering coefficients can be calculated as the average of all nodes local clustering coefficients within the shell (Watts and Strogatz 1998).

### Result and discussion

#### The keyword network of the DL field in China and its hierarchical layers

Based on the method described above, we abstract the subnet of the keyword network of DL research in China. The keyword network can be divided into 6 shells after the K-core



**Fig. 3** The K-shells of the keyword network of the DL field in China

decomposition, as shown in Fig. 3. The clustering coefficients of each shell are calculated and listed in Table 2; their density and size are also listed for an additional description of their morphology.

In Table 2, we can see that as the shell index increases, the clustering coefficients of the K-shells fluctuate sharply. The clustering coefficients of shell-2 and shell-3 are extremely high despite their low density, while the clustering coefficients of shell-4 and shell-5 decrease sharply, and the clustering coefficient of shell-6 reaches a high value. With reference to Carmi et al. (2007), we can divide these shells into four layers by recombining adjacent shells with similar morphology based on the presented data (see Fig. 4), as follows:

- The basic layer: contains shell-6, with a high clustering coefficient and a high density;
- The middle layer: contains shell-5 and shell-4, with an extremely low clustering coefficient and a medium level of density;
- The detail layer: contains shell-3 and shell-2, with an extremely high clustering coefficient and a low density;
- The marginal layer: contains shell-1, with a zero-value of clustering coefficient and an extremely low density.

The clustering coefficient, density and size of each layer are calculated and shown in Table 3. As seen in Fig. 3, keywords in shell-1 are either isolated pairs or peripheral words that reveal less about intellectual structure, so we will focus on the other three layers, which can be visualized in NetDraw as Fig. 5.

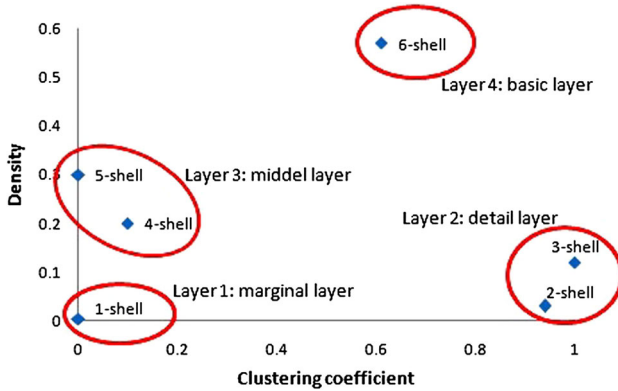
### The basic layer

As seen Fig. 5, the basic layer is the core of the keyword network. Keywords here have the highest K-core value, which indicates that they are the most influential DL concepts in China. There are 17 keywords in the basic layer, all of which are general concepts with broad semantics. This reflects reality very well, as most Chinese research papers in the DL field contain at least one of these basic concepts. The basic concepts are interconnected tightly with a high density of 0.57, which is far above the density of the whole network (only 0.05). The high redundancy of links between basic concepts is beneficial for the exploration of new research directions. Thus, the link strength between these basic concepts expands more rapidly than others with the development of the research field, forming a solid knowledge base. Other concepts with more concrete semantics can be considered to be the expansion of the knowledge base.

The basic layer can be viewed as the main research subfields of the DL field in China. We visualize the basic layer independently in NetDraw using the Gower Metric Scaling

**Table 2** Properties of each K-shell

Shell index	Clustering coefficient	Density	Size
6	0.61	0.57	17
5	0	0.3	2
4	0.1	0.2	11
3	1	0.12	19
2	0.94	0.03	46
1	0	0.004	119



**Fig. 4** Merging adjacent shells into layers according to their density and clustering coefficient

**Table 3** Properties of each layer

Layers	Clustering coefficient	Density	Size
Basis layer	0.61	0.57	17
Middle layer	0.08	0.15	13
Detail layer	0.92	0.02	65
Marginal layer	0	0.004	119

Layout, in which keywords with intense relations (either directly or through other keywords) are plotted closely together (Verspagen and Werker 2004). As shown in Fig. 6, four main clusters including technology, resource, service and copyright can be marked off, which can be viewed as the main subfields of DL research in China. The basic concepts in each subfield are listed as follows:

- The technology subfield: *information retrieval, grid, metadata, information organization, information technology;*
- The resource subfield: *digital information resource, network, information resource sharing, database, information resource, digitalization, information resource construction;*
- The service subfield: *information service, service, user research, personalize service;*
- The copyright subfield: *copyright.*

This result is in accordance with the conclusions of many previous studies (Dong 2009; Zhang and Lv (2010); Xu and Yang 2011; Liu et al. 2012). The resource subfield has the most basic concepts, indicating that the resource problem is the basic concern of the DL field in China. Meanwhile, there is only one basic concept in the copyright subfield, showing that research about copyright is less influential in the field.

To reveal the relationship between four subfields, we calculated the link strength between different clusters (as shown in Table 4). We find that in China’s DL field:

- The subfields of resource and technology are closely associated with each other, indicating that these technologies are mainly resolving resource problems;
- The relation strength between technology and service, service and resource, are at the middle level;



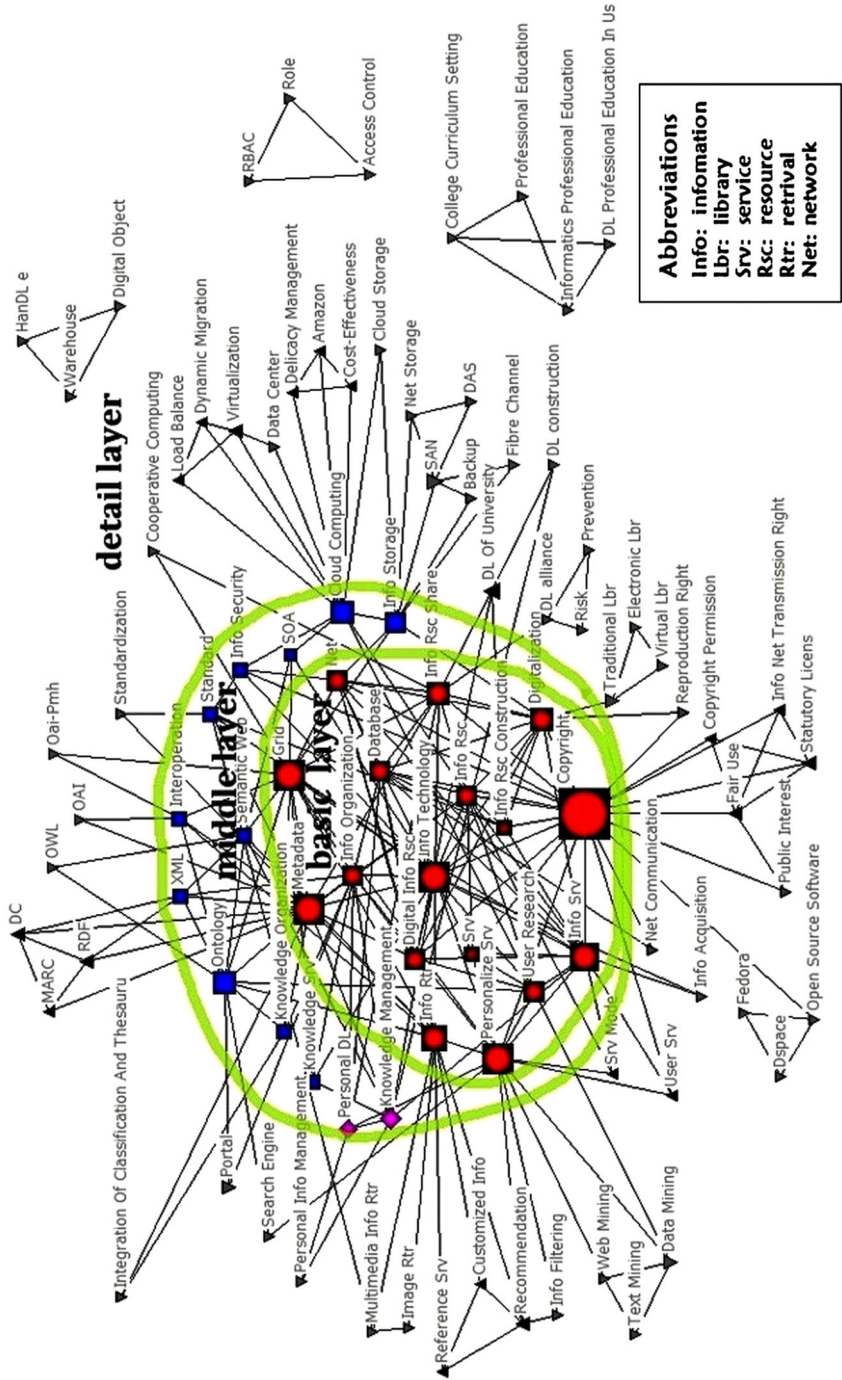
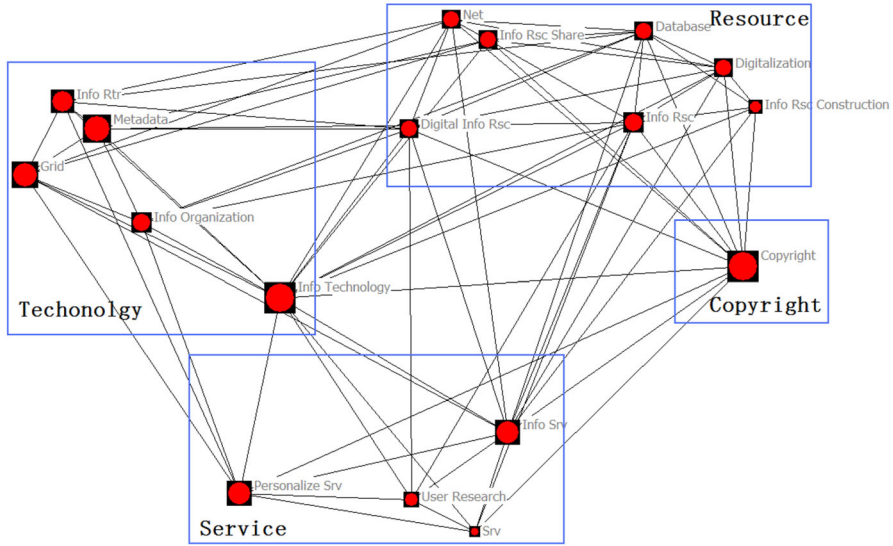


Fig. 5 Three main layers in the keyword network of the DL field in China



**Fig. 6** Clusters in the basic layer, using the gower metric scaling layout method

**Table 4** Relation strength between basic concepts in 4 subfields

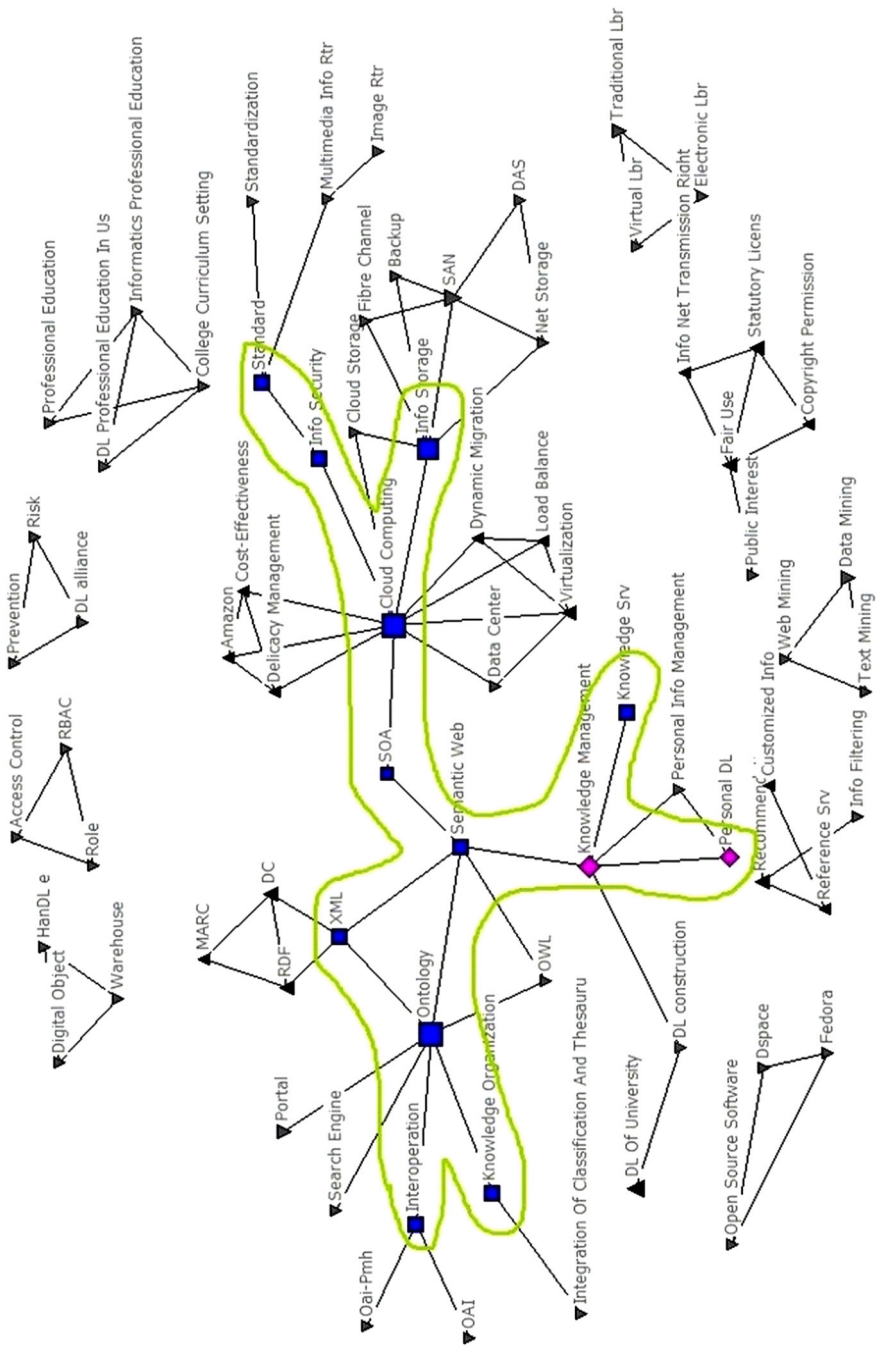
Subfields	Technology	Resource	Service	Copyright
Technology	9	17	8	1
Resource	17	13	9	7
Service	8	9	6	3
Copyright	1	7	3	0

- The relation strength between copyright and technology, and copyright and service are both very weak, indicating the special-purpose of copyright studies for resolving resource problems in the field.

### The middle layer

The middle layer contains 13 keywords including: *cloud computing, information security, standard, information storage, SOA, semantic web, XML, ontology, knowledge management, interoperation, knowledge management, and personal digital library*. The semantic meanings of these keywords are narrower than the general keywords in the basic layer, so they are less influential in the network. Although some of their degrees are higher than basic keywords (for example, *cloud computing, ontology, and information storage*), their K-core values are lower because they have not yet connected to enough basic concepts. Besides, the clustering coefficient of the middle layer is extremely low (as shown in Table 3), indicating that these keywords do not tend to group together.

The middle layer is an intermediate region between the basic layer and detail layer. Keywords in the middle layer can help us recognize how those general concepts evolve into detailed concepts in China’s DL field. Figure 7 shows the relationship between keywords in the middle layer and the detail layer, from which we can see that keywords in the



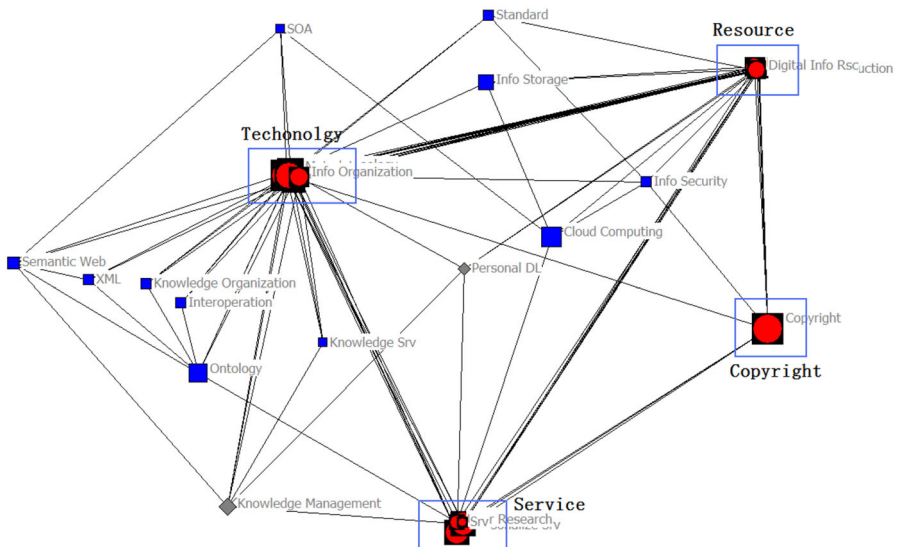
**Fig. 7** Keywords in the middle layer and the detail layer

middle layer are more like the representative of different keyword clusters. Thus, they can be viewed as mediator concepts in the field. Figure 8 shows the relationships between keywords in the middle layer and four subfields in the basic layer. We can see that most middle-layer keywords have a direct relationship with basic concepts in the technology subfield, indicating that the development of research in the field is mainly driven by the improvement and merging of technologies.

From Figs. 7 and 8, we can identify some important concepts in the middle layer. For example, *ontology*, *semantic web*, *cloud computing*, and *knowledge management* are typical “hot spots” in the field. They connect to lots of other keywords in the middle layer, showing that they are important intermediate topics and are more likely to become basic concepts in the future. *Personal digital library* and *information security* are two typical multi-domain DL topics in China. They have connected with basic concepts covering three subfields. *Personal digital library* connects with basic concepts including *personal service*, *information resource share*, and *information organization*, covering the subfield of technology, resource and service. *Information security* connects with basic concepts including *information technology*, *network*, and *copyright*, covering the subfield of technology, resource and copyright. Since they have not yet differentiated into detailed clusters, they have great potential for further enhancement in future studies.

### The detail layer

The detail layer is composed of low-frequency keywords whose semantics are more explicit than keywords in the basic or middle layers. These detailed concepts are evolved from general concepts. They can be used to express emerging topics of the field and therefore are very important in revealing its intellectual structure. However, many of them were ignored in previous studies because of their low frequency.



**Fig. 8** Relations between keywords in the middle layer and the subfields in the basic layer

Specifically, the detail layer in our network contains 65 keywords. Most have formed into independent clusters, which usually contain three to six similar concepts. Keywords are tightly interconnected in the same cluster but rarely connected in different clusters, leading to an extremely high clustering coefficient of the detail layer despite its low density.

There are 13 detailed DL keyword clusters in China, as listed in Table 5. The parent nodes of each cluster, defined as higher-layer keywords that connect directly to the cluster, are also listed. The cluster names are labeled according to their parent nodes. It is believed that most complex networks are accompanied by hierarchical modularity (Ravasz et al. 2002). In our case, these clusters can be viewed as basic modules of the DL research in China, from which we can identify detailed topical specializations of the field. Besides these clusters, there are 12 keywords in the detail layer which have not yet clustered together. Compared with keywords in the marginal layer, these are more influential because they have connected directly with more basic concepts or mediator concepts in the field, and therefore are more likely to be grouped into clusters in the future.

**Table 5** Topical clusters in the detail layer

Cluster name	Keywords	Parent nodes
Cloud computing technologies	Load Balance, dynamic migration, virtualization, data center	Cloud computing
Cloud computing management	Delicacy management, cost-effectiveness, Amazon	Cloud computing
Information storage technologies	Backup, SAN, DAS, network storage, fiber channel	Information storage
Metadata in XML format	DC, RDF, MARC	Metadata, XML
Copyright system	Statutory license, fair use, copyright permission, Information Network Transmission Right, public Interest	Copyright
Copyright of open source software	Open source software, Fedora, Dspace	Copyright
Information retrieval in personalized service	Customized information, recommendation, reference service, information filtering	Personalize service, information retrieval
User research for personalized service	Data mining, web mining, text mining	Personalize service, user research
Professional education	College curriculum setting, professional education, informatics professional education, DL professional education in US	–
Access permission	RBAC, role, access control	–
DL architecture	Digital object, warehouse, handle	–
Digitalization of library	Traditional Library, virtual Library, electronic library	Digitalization
DL alliance for resource sharing	DL alliance, risk, prevention	Information resource sharing

## Discussion and conclusions

Based on the K-core decomposition method, we find that there are four hierarchical layers in the backbone of the DL keyword network in China. The basic layer contains 17 basic concepts which are tightly interconnected, forming a steady knowledge base of the research field. The middle layer contains 13 keywords which can be viewed as mediator concepts between the knowledge base and the detailed clusters. The detail layer contains 13 topical clusters and some immature concepts; these detailed clusters represent the field's research specialization. The research topic hierarchy of the DL field in China can be simplified in Fig. 9.

The analysis of topic hierarchy can help us investigate the DL field in China from a new perspective. Based on the micro-morphology of the knowledge base, mediator concepts and detail clusters, we can draw the following conclusions:

1. The knowledge base of the field covers four main subfields including resource, technology, service and copyright. The resource and technology subfields are emphasized and are highly associated with each other. The copyright subfield is less influential in the knowledge base; it has very weak relation with other subfields except for resource.
2. The research field is expanded through the merging and development of new technologies. *Ontology*, *semantic web*, *cloud computing* and *knowledge management* are important intermediate concepts which connect many other localized "hot spots" in the field; they are more likely to join the knowledge base in the future. *Personal digital library* and *information security* are typical multi-domain concepts in the field, which cover most of the main subfield. They have great potential for further study.
3. The detailed DL research in China is mainly focused on cloud computing (including the technology and management aspects), information storage technologies, metadata in XML format, copyright (including the copyright system and open source software), personal service (including the information retrieval aspect and the user research aspect), professional education, access permission technologies, DL architecture components, digitalization of libraries, and information sharing in DL alliance. There are also some topics which have not yet been well-studied.

Although hierarchical structure has been widely used to represent knowledge, it has received little attention when mapping the intellectual structure of research fields through keywords networks. Nguyen and Chowdhury (2013) defined DL keywords in both broader and narrower terms, and then mapped the DL concepts at three levels (core topics, clusters of subtopics, and subtopics) based on the relationships between these broader terms and narrower terms. Our method provides a more quantitative approach of exploring the hierarchical levels of DL concepts.

Hierarchically decomposing the keyword network is a new method of exploring the intellectual structure of a research field, and is quite different from traditional co-word clustering. We can determine the advantage of our method by comparing topic hierarchy with the topic clustering results of traditional co-word methods (Su and Xia 2011; Liu et al. 2012). On the one hand, our result retains both the main subfields of the DL research in China and the most important keywords in these subfields. On the other hand, our result provides more detailed information about the subfields, and the topical hierarchy reveals the different role of concepts in each subfield, such as the knowledge base, evolutionary mediator, detailed topical clusters, and marginal topics. Concepts in a domain are very

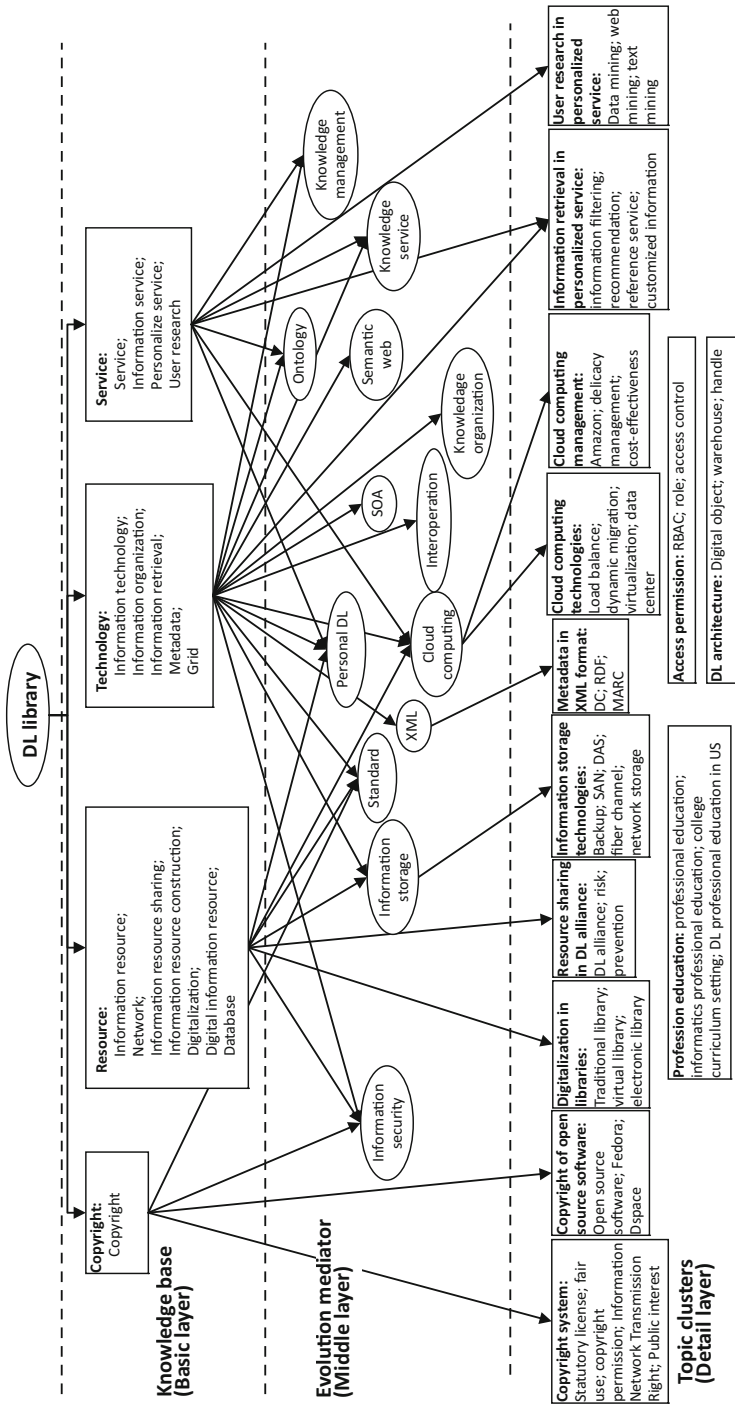


Fig. 9 Research topic hierarchy of the DL field in China

different in a semantic scope. Hierarchical decomposition of the keyword network can help us distinguish individual differences among concepts and achieve a more in-depth understanding of the research field.

**Acknowledgments** This study was supported by the Major Project of the National Social Science Foundation of China (12&ZD221), the Project of National Natural Science Foundation of China (71273125), the Fundamental Research Funds for the Central Universities (No. 30916013101). The authors are grateful to anonymous referees and editors for their invaluable and insightful comments.

## References

- Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). K-core decomposition: A tool for the visualization of large scale networks. arXiv preprint cs/0504107.
- Barabási, A. L., Dezső, Z., Ravasz, E., Yook, S. H., & Oltvai, Z. (2003). Scale-free and hierarchical structures in complex networks. *Sitges Proceedings on Complex Networks*, 661(1), 1–16.
- Cambrosio, A., Limoges, C., Courtial, J. P., & Lavoie, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics*, 27(2), 119–143.
- Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., & Shir, E. (2007). A model of Internet topology using K-shell decomposition. *Proceedings of the National Academy of Sciences*, 104(27), 11150–11154.
- Chen, G., & Xiao, L. (2016). Selecting publication keywords for domain analysis in bibliometrics: a comparison of three methods. *Journal of Informetrics*, 10(1), 212–223.
- 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.
- Clauset, A., Moore, C., & Newman, M. E. (2007). Structural inference of hierarchies in networks. In E. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, & A. X. Zheng (Eds.), *Statistical network analysis: Models, issues, and new directions* (pp. 1–13). Berlin: Springer.
- Clauset, A., Moore, C., & Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191), 98–101.
- Collins, J. J., & Chow, C. C. (1998). It's a small world. *Nature*, 393(6684), 409–410.
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278.
- Dong, W. (2009). Analysis on hotspot of digital library in home during 10 years based on co-word analysis. *Document Information and Knowledge*, 5, 58–63.
- Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2006). K-core organization of complex networks. *Physical Review Letters*, 96(4), 040601.
- Green, R. (2001). Relationships in the organization of knowledge: An overview. In A. Bean & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 3–18). Berlin: Springer.
- He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends*, 48(1), 133–159.
- Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, 6(11), 888–893.
- Lee, P. C., Su, H. N., & Chan, T. Y. (2010). Assessment of ontology-based knowledge network formation by vector-space model. *Scientometrics*, 85(3), 689–703.
- Liu, G. Y., Hu, J. M., & Wang, H. L. (2012). A co-word analysis of digital library field in China. *Scientometrics*, 91(1), 203–217.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159–165.
- Ma, W. F., & Du, X. Y. (2007). Some theoretical issues relating to knowledge organization system. *Journal of Library Science in China*, 33(2), 13–17. (in China).
- Nguyen, S. H., & Chowdhury, G. (2013). Interpreting the knowledge map of digital library research (1990–2010). *Journal of the American Society for Information Science and Technology*, 64(6), 1235–1258.
- Qiu, J. P., & Wang, M. Z. (2010). The analysis of the digital library research paper in China from the years of 1999 to 2008. *Journal of Intelligence*, 29(2), 1–5. (in China).
- Quoniam, L., Balme, F., Rostaing, H., Giraud, E., & Dou, J. M. (1998). Bibliometric law used for information retrieval. *Scientometrics*, 41(1), 83–91.
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586), 1551–1555.



- Rokaya, M., Atlam, E., Fuketa, M., Dorji, T. C., & Aoe, J. I. (2008). Ranking of field association terms using co-word analysis. *Information Processing and Management*, 44(2), 738–755.
- Sales-Pardo, M., Guimera, R., Moreira, A. A., & Amaral, L. A. N. (2007). Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences*, 104(39), 15224–15229.
- Salton, G. (1975). *Theory of indexing*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Shen, X., Zheng, Z., Han, S., & Shen, C. (2008). A review of the major projects constituting the China Academic Digital Library. *The Electronic Library*, 26(1), 39–54.
- Su, X. N., & Xia, L. X. (2011). Topic analysis of digital library research from 2000 to 2009 in China: Based on the statistical data of key words released by CSSCI. *Journal of Library Science in China*, 37(7), 60–69. (in China).
- Tong, A. H. Y., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., et al. (2002). A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*, 295(5553), 321–324.
- Verspagen, B., & Werker, C. (2004). Keith Pavitt and the invisible college of the economics of technology and innovation. *Research Policy*, 33(9), 1419–1431.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684), 440–442.
- Xu, J., & Yang, S. L. (2011). Research status and frontier about digital library based on mapping knowledge domain. *Library*, 6, 012. (in China).
- Yi, S., & Choi, J. (2012). The organization of scientific knowledge: The structural characteristics of keyword networks. *Scientometrics*, 90(3), 1015–1026.
- Zhang, X., & Lv, Y. J. (2010). Research overview on development of digital library in China in the past five years. *Researches in Library Science*, 2, 18–22. (in China).
- Zhang, G. Q., Yang, Q. F., Cheng, S. Q., & Zhou, T. (2008). Evolution of the Internet and its cores. *New Journal of Physics*, 10(12), 123027.
- Zhang, H., Zhao, H., Cai, W., Liu, J., & Zhou, W. (2010). Using the k-core decomposition to analyze the static structure of large-scale software systems. *Journal of Supercomputing*, 53(2), 352–369.
- Zhao, L., & Zhang, Q. (2011). Mapping knowledge domains of Chinese digital library research output, 1994–2010. *Scientometrics*, 89(1), 51–87.
- Zhao, S. X., Zhang, P. L., Li, J., Tan, A. M., & Ye, F. Y. (2014). Abstracting the core subnet of weighted networks based on link strengths. *Journal of the Association for Information Science and Technology*, 65(5), 984–994.
- Zhou, Q. (2005). The development of digital libraries in China and the shaping of digital librarians. *Electronic Library*, The, 23(4), 433–441.