DSI DATA SCIENCE AND INFORMETRICS
VOL.2, NO.3, AUG. 2022

Identification of potential interdisciplinary cooperative topics based on co-word network: Taking LIS and computer science as an example

Changling Li*, Guoyang Rong, Kun Luan, Xinxin Wang, Lixu Shen, Qingqing Fan

Institute of Information Management, Shandong University of Technology, Zibo, China

ABSTRACT

[Purpose/significance] Interdisciplinary knowledge fusion plays a key role in promoting the development of interdisciplinary integration and providing new ideas for interdisciplinary cooperative research. This study sets out to identify potential interdisciplinary cooperative topics between Library and Information Science (LIS) and Computer Science. [Method / Process] We built an interdisciplinary co-word network to identify potential interdisciplinary cooperative topics by closed and opened irrelevant knowledge discovery methods. We also constructed the topic interdisciplinary cooperation potential index (TICPI) to calculate the interdisciplinary cooperation potential of the topic and found the best contact path of the cooperation topic by constructing the practicable value (PV) of the contact patch. [Result / Conclusion] The experimental data suggested that both methods can identify the same potential interdisciplinary cooperative topics, such as knowledge service & matrix decomposition, online comments & social media processing, academic text & generative adversarial network, network public opinion & smart home. Exploiting the cooperation potential of these topics can help the knowledge fusion between disciplines.

KEYWORDS

Interdisciplinary cooperative topics; Interdisciplinary research; Irrelevant knowledge discovery; Library and Information Science; Computer Science

Introduction

Interdisciplinary technologies and methods have solved numerous scientific research problems and injected new vitality into the disciplines. The prerequisite of the interdisciplinary cooperation is to find the appropriate topics. The principle is shown in Figure 1:

The common problems in interdisciplinary cooperation are:

1. Which concepts, technologies, and methods of which disciplines can cooperate with the knowledge we have to solve the relevant problems?

2. How to identify them scientifically?

The discovery of interdisciplinary relevant knowledge is the basis of interdisciplinary cooperative research. Therefore, taking Library and Information Science (LIS) and Computer Sci-

* Corresponding Author
E-mail: lichl69@163.com，ORCID: 0000-0001-6266-4820

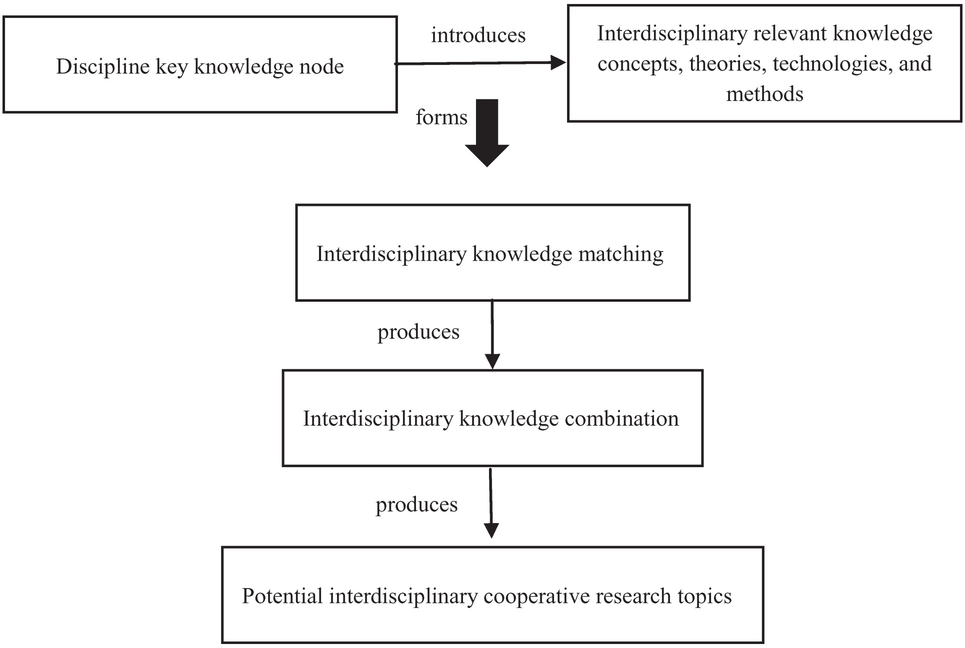


Figure 1 Interpretation diagram of interdisciplinary relevant knowledge

ence as examples, we applied closed and opened irrelevant knowledge discovery methods to identify potential interdisciplinary cooperative topics and found the difference between the two methods.

Literature review

With the development of science and society, scientific innovation has been promoted through different circulation methods and channels, and the participation of various scientific disciplines under different disciplinary backgrounds. Moreover, interdisciplinary cooperative research is of great significance to promote the output of scientific achievements and explore the integration and development of interdisciplinary knowledge. The interdisciplinary cooperative research at home and abroad mainly involves the following aspects:

(1) Exploring the mechanism and influencing factors of interdisciplinary cooperation from the theoretical perspective.

From the analysis of the driving force of interdisciplinary research, the researcher expounded the obstacles, importance and reality of interdisciplinary research, and emphasized interdisciplinary research has become an inevitable requirement of current scientific development (Yang, 2011). Interface rules are the basis for interdisciplinary cooperation. Researchers constructed a theoretical analysis framework for interdisciplinary cooperation network interface rules and emphasized interdisciplinary team coordination and cross-team relationship governance (Huang & Wang, 2012). On the basis of Porter and Lawler's comprehensive motivation theory, researchers pointed out that the individual characteristic factors and intrinsic motivation factors of university graduate students will affect their interdisciplinary cooperative research behavior (Xiong & Yu, 2011). Koenig (2013) found that labor division in an interdisciplinary cooperation depends on their disciplinary background. Analyzing the disciplinary integration mechanism of interdisciplinary research, Huang and Yang (2016) summarized the key elements of disciplinary integration and proposed three interdisciplinary cooperative development models: the emerging discipline derived network model, the leading discipline derived network model and the multidisciplinary joint network model. By constructing a theoretical model of the influencing factors of researchers' potential interdisciplinary cooperative behavior, researchers found the correlation of perceived behavior control, subjective norms, and behavior attitudes to potential interdisciplinary cooperative behavior (Feng & Zhang, 2020).

(2) Studying the phenomenon of knowledge flow in the process of interdisciplinary cooperation.

The basic characteristics of interdisciplinary research are reflected in the mutual exchange, transfer, flow and diffusion of concepts and methods between disciplines (Porter et al., 2007). Eto (2003) evaluated the output behavior of nanotechnology projects from an interdisciplinary perspective, and analyzed how interdisciplinary information plays a role in the project as input and output. Topic correlation analysis (TCA) methods could extract common and unique topic latent features from multiple disciplines to facilitate the efficient transfer of knowledge (Li et al., 2012). Yan (2014) combined knowledge transaction theory with citation network to analyze the knowledge transaction relationship between discipline categories and relevant characteristics, and analyzed the impact of discipline knowledge on trade by calculating the trade surplus amount. By analyzing the evolution of keywords, researchers found three stages of interdisciplinary research: incubation, budding and maturity. And each research area would play one of the four roles of knowledge source, knowledge receiver, knowledge responder and interdisciplinary participant at different times (Jian et al., 2018).

(3) Measuring the interdisciplinary degree by the correlation between citations of scientific literature and co-occurrence network or the citation relationship between media data.

The interdisciplinary degree could be measured by factor analysis and similarity of the interdisciplinary citation matrix (Rafols et al., 2010). Researchers conducted interdisciplinary degree measures from the diversity of reference disciplines, target literature disciplines, and cooperative institution disciplines (Huang et al., 2019). Wu and Zhang (2019) constructed a kinship tree from users' friends, comments and recommendations. They found that Computer Science and Management Science are the most closely relevant disciplines for users in the LIS on social media.

(4) Identifying intersecting research topics by using text clustering, citation network, topic correlation and overlapping community methods.

The overlapping community analysis tool CFinder could be used to identify the intersection of Information Science and Computer Science (Li et al., 2013). Piepenbrink and Nurmammadov (2015) used the LDA topic model to identify intersecting research topics in Transition Economies and Emerging Markets. Researchers identified the intersecting research topics in the fields of Information Science and Computer Interdisciplinary Application from two aspects: interdisciplinary field basis and interdisciplinary association basis (Yue et al., 2016). Researchers used the improved discipline correlation analysis method to extract the common topics and independent topics of Agriculture Reproductive Biology and Veterinary Science. In addition, they quantified the correlation between the independent topics (Wu et al., 2017).

(5) Identifying potential interdisciplinary cooperation by clustering community, citation relationship and contribution network methods.

The clustering community detection method could discover the potential cooperative relationship of researchers (Santos et al., 2009). From the perspective of citation, a potential cooperation possibility measurement model for researchers in two different research fields was constructed to demonstrate the feasibility of using direct citation to tap potential cooperators (Ma, 2015). By analyzing five different interdisciplinary citation relationships, a potential interdisciplinary cooperation intensity model of authors was constructed and the identified potential interdisciplinary cooperators were divided into three types: knowledge absorption type, knowledge reciprocity type and knowledge radiation type (Li et al., 2018). Researchers identified potential cooperative research topics in Information Science and Computer Science by constructing an interdisciplinary keyword co-occurrence network and using irrelevant knowledge discovery methods (Li et al., 2018; Liu et al., 2017). Feng and Li (2018) identified potential interdisciplinary cooperators by analyzing the authors' interdisciplinary citation relationships. The method of screening core keywords was used to construct interdisciplinary authors-core keyword 2-model network and authors' interdisciplinary cooperation potential index model to identify the best combination of interdisciplinary cooperation (Liu et al., 2018).

In summary, the current research on interdisciplinary cooperation mainly includes:

1. Theoretical analysis explored the mechanisms factors of interdisciplinary cooperation.
2. The phenomenon of interdisciplinary cooperation knowledge flow was analyzed through disciplinary relationships.
3. Interdisciplinary degree were measured by correlations between scientific literature citation and co-occurrence network or citation relationships between media data.
4. not aligned with others. Text clustering, citation network, topic correlation, and overlapping community methods identified interdisciplinary research topics.
5. Potential interdisciplinary cooperation identification.

At present, there are more and more researches on interdisciplinary cooperation theory and interdisciplinary evaluation indicators, and many scientific research problems need to be solved in interdisciplinary. Therefore, we used closed and opened irrelevant knowledge discovery methods to identify potential interdisciplinary cooperative research topics between LIS and Computer Science, in order to solve the scientific research problems of LIS with the method of Computer Science, and provide suggestions on the direction of interdisciplinary cooperation for the authors of two disciplines.

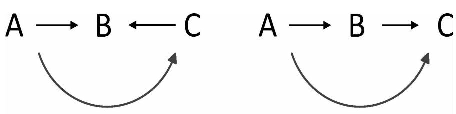
Methodology and Results

Strategy

"Strength of weak ties" originated in sociological research, which was defined as a brief social contact between two actors (Granovetter, 1973). Weak ties promote the flow of information between different groups and spread information that people are unlikely to see. "Strength of weak ties" has the characteristics of extensive sources, and its role is mainly reflected in information transmission and interpersonal communication. "Strength of weak ties" is measured by the amount of time, emotional intensity, intimacy, and reciprocal services. After "strength of weak ties" was proposed, some problems also appeared: (1) How to divide the strength and weak ties? (2) How to quantitatively calculate the strength and weak ties?

Weak ties play an important role in social network relationships. Strength ties make the internal connections of the organization close and stable, while weak ties provide an important way for the exchange of information between different groups and organizations. So that the isolated sub-groups began to establish connections, and with the continuous strengthening of weak ties, the scope of different information exchanges has been further expanded, accelerating the dissemination, integration, development, and innovation of information. And other researchers also discussed this point in their research (Zhao et al., 2010; Aalabaf-Sabaghi, 2012; Bakshy et al., 2012; Onnela et al., 2007). Compared with strength ties, weak ties can transmit more potential, diversified, and non-redundant knowledge resources (Genuis, 2005). Weak ties made it easier to establish more extensive relationships for individuals in other subnets of the knowledge and individuals network, which has a positive effect on promoting scientific research cooperation (Yang et al., 2009; Abbasi et al., 2011; Bettoni et al., 2015).

Irrelevant knowledge discovery method, also named Swanson theory (Swanson, 2015). The basic idea of irrelevant knowledge discovery is to establish the association between two groups of irrelevant literature A and C through an intermediate word or intermediate literature B, and to mine the undiscovered knowledge in the literature through the relationship among the three, which is also a form of Weak ties. The method includes two types: closed and opened.



(1)Closed (2)Opened

Figure 2 Irrelevant knowledge discovery research method

Now, A, B, and C represent three different concepts. We assumed that there is a co-occurrence relationship between A and B, B and C in the published literature, and there is no co-occurrence relationship between A and C. Then, the closed method can be expressed in Figure 2 (1). If B can be retrieved from A, and B can be retrieved from C, there is an internal connection between A and C; The open method can be expressed in Figure 2 (2). If B can be retrieved from A, and C can be retrieved from B, there should also be an internal connection between A and C.

In recent years, Huang Shuiqing and others have successively adopted closed (Liu et al., 2017) and opened (Li et al., 2018) irrelevant knowledge discovery method to verify that Swanson's knowledge discovery theory and method are also applicable to Chinese literature; Zhang (2009) proposed that retrieval theory, Bibliometrics and Logic are the theoretical basis of irrelevant knowledge discovery (Zhang & Leng, 2009); Liu et al. (2016) proposed a directional recognition algorithm between semantic relationships of irrelevant documents based on data mining; Liu (2019) designed a knowledge association retrieval system for irrelevant documents to reveal the knowledge association between documents from a finer granularity level.

Therefore, at present, researchers mainly explored the method of knowledge discovery in theory or technology, and do not use this method in interdisciplinary research field. We applied the idea of irrelevant knowledge discovery method to potential interdisciplinary cooperative research, tried to explore the irrelevant research topics indirectly related to different disciplines, and revealed the connection bridge between the two irrelevant topics, so as to help researchers understand the development and extension of the discipline and carry out targeted innovative research.

Data

We selected three journals in each of LIS and Computer Science as samples. The samples include *Documentation, Information & Knowledge, Library and Information Service, Journal of the China Society for Scientific and Technical Information, Acta Automatica Sinica, Journal of Software, and Chinese Journal of Computers.* We took the keywords marked by the author as the research topics of the documents. And topics of the above six journals from 2016 to 2020 were selected as the sample data for empirical research.

Collecting the bibliography of six journals from CNKI, we got 6,533 valid documents, including 3,392 documents in LIS and 3,141 documents in Computer Science, a total of 29,747 non-repeating topics were obtained. The retrieval time is November 1, 2021.

We saved the above 29,747 topics as an Excel file, and processed them into a format that can be recognized by Bibexcel through the VBA program. To distinguish the discipline source of each topic, we added a suffix to each topic through the VBA program. Topics were called out suffix "(Q)" if they only appeared in LIS, called out suffix "(J)" if they only appeared in Computer Science.

Closed irrelevant knowledge discovery method and results

(1) Cooperative network construction and cooperative topic identification

We got topic frequency and co-word matrix without distinguishing disciplines by Bibexcel, which was shown in Table 1.

Table 1 Co-word extracted by Bibexcel

Co-occurrence times	Topic	Topic
33	Deep learning	Convolutional neural network
12	Internet of things	Information physical fusion system (J)
11	Blockchain	Smart contract
11	Library (Q)	Reading promotion (Q)
10	Artificial intelligence	Library (Q)
9	University Library (Q)	Maker space (Q)
9	Deep learning	Computer vision (J)
9	Reading promotion (Q)	University Library (Q)
9	Government data (Q)	Open data
9	Smart library (Q)	Smart service (Q)
…	…	…

The indirect irrelevant co-words in Table 1 include topics in the same discipline and different disciplines. We excluded the co-words of the same discipline and both intermediate word pairs. There were 1903 pairs of interdisciplinary co-words, including 871 pairs of Computer Science and intermediate words, and 1032 pairs of LIS and intermediate words.

(2) Potential interdisciplinary cooperative topics identification

The distance of co-words can reflect the correlation between topics. Distance 1, means they have appeared in the same article. Distance 2, means that the two topics are second-order irrelevant topics, which is an indirect connection through a intermediate word.

To get the distance 2 co-word pairs, we constructed an interdisciplinary co-word set under the same intermediate word for Table 1. Each LIS topic co-occurrence with the Computer Science topic through intermediate word, as shown in Table 2.

Table 2 Interdisciplinary co-word set

Co-occurrence times	Topic set of LIS	Intermediate word	Topic set of Computer Science	Co-occurrence times
2	Knowledge service (Q)	Recommendation system	Matrix decomposition (J)	6
1	E-commerce (Q)		Comment text (J)	3
1	Information service (Q)		Location privacy (J)	1
1	Smart Library (Q)		Load balancing (J)	1
			Context aware (J)	1
			Frequent mode (J)	1
			Factorization machine (J)	1

We permuted and combined each co-word set, and finally got 5034 pairs of interdisciplinary topics. But there existed repeated interdisciplinary co-word pairs due to different intermediate paths.

We removed the repeated interdisciplinary topic pairs and obtained interdisciplinary co-word pairs with a distance of 2, which are the most likely research topics for interdisciplinary cooperation in the fields of LIS and computer science in the future. They are respectively the discipline word referred to above a_i and c_i.

Taking "Knowledge service (Q)"as an example, the second-order irrelevant topics corresponding to the intermediate word "Recommendation system" are obtained through the above procedures: matrix decomposition (J), comment text (J), location privacy (J), load balancing (J), context-awareness (J), frequent mode (J), factorization machine (J), etc. These topics and "Knowledge service (Q)"have the potential for interdisciplinary cooperation, which is the identification goal of this paper.

After processing, a total of 4693 pairs of interdisciplinary topics with indirect connections were obtained. So, how likely are these 4693 pairs of potential interdisciplinary cooperation topics? What is the potential of interdisciplinary cooperation between the two?

(3) Construction and calculation of interdisciplinary cooperation potential index

The feasibility intensity and the discipline interdisciplinary cooperation potential index defined in preliminary research (Liu et al., 2017) are applied to evaluate the cooperation potential of co-word pairs with potential interdisciplinary cooperation.

It is assumed that there are n intermediate words between topics a_i and c_i, that is, there are n contact paths between them, as shown in Figure 3.

In Figure 3, b_j=(j=1, ⋯, n) is the j^{th} intermediate word of a_i and c_i , L_{j1} is the co-occurrence times of topic a_i and intermediate word b_j , and L_{j2} is the co-occurrence times of topic c_i and intermediate word b_j .

The practical value (PV) of the contact path b_j is calculated as follows:

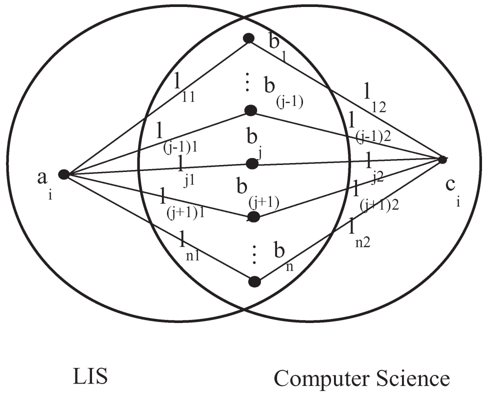


Figure 3 Labeling diagram of each quantity between target topics

$$\mathrm{PV}_j = \frac{(l_{j1} \times l_{j2})^2}{|l_{j1} - l_{j2}| + \beta} \quad (1)$$

We introduced thematic interdisciplinary cooperation potential index (TICPI) (Liu et al., 2017) to measure the connection strength of a_i and c_i.

$$\mathrm{TICPI} = 0.5 \times n + 0.5 \times \sum_{j=1}^{n} \frac{(l_{j1} \times l_{j2})^2}{|l_{j1} - l_{j2}| + \beta} \quad (2)$$

TICPI comprehensively calculates the strength of multiple paths, which can measure the strength of the total path more reasonably. In addition, TICPI can consider the cooperation breadth and intensity of potential knowledge combinations by assigning weights to the number and intensity of paths.

We calculated the TICPI of 4693 pairs of topics, which are arranged from high to low. Table 3 shows the top ten pairs of topics and its relevant data.

Table 3 LIS and Computer Science interdisciplinary cooperative topic identification data

Initial topic a_i	**Target topic c_i**	**TICPI index**	**Intermediate word b_i**	**PV_i**
Information engineering (Q)	MapReduce (J)	56.75	Big data	112.50
Academic text (Q)	Generative adversarial network (J)	25.50	Deep learning	50.00
Knowledge service (Q)	Matrix decomposition (J)	21.00	Recommendation system	36.00
Network public opinion (Q)	Smart home (J)	19.50	Knowledge graph	36.00
Emergency decision (Q)	Parallel computing (J)	18.50	Big data	36.00
Online comments (Q)	Social media processing (J)	18.50	Emotional analysis	36.00
Scientometrics (Q)	Graph clustering (J)	8.50	Complex network	16.00
Digital Library (Q)	Access control (J)	7.85	Blockchain	7.20
Digital humanities (Q)	Computer vision (J)	7.05	Deep learning	11.11
Knowledge sharing (Q)	Social media (J)	6.60	Social networks	7.20

* The data in the table are rounded estimates.

Thence, the potential interdisciplinary cooperation topics are identified but still did not identify how they cooperated. To be precise, how can each topic in the word set A to be connected with the topic in word set C? Now we look for the bridge: intermediate words.

(4) Intermediate path query of potential interdisciplinary cooperative topics

We obtained all the contact paths and co-occurrence times of 4693 pairs of co-words through the interdisciplinary co-word network by the VBA program. For a pair of interdisciplinary cooperation topics, there could be multiple intermediate words. Using the practical value (PV) of contact path b_j we found the intermediate words that are most likely to promote the interdisciplinary cooperation of this pair of topics.

Figure 4 shows three different contact paths by taking "Digital Library (Q)" and "Access control (J)" as examples.

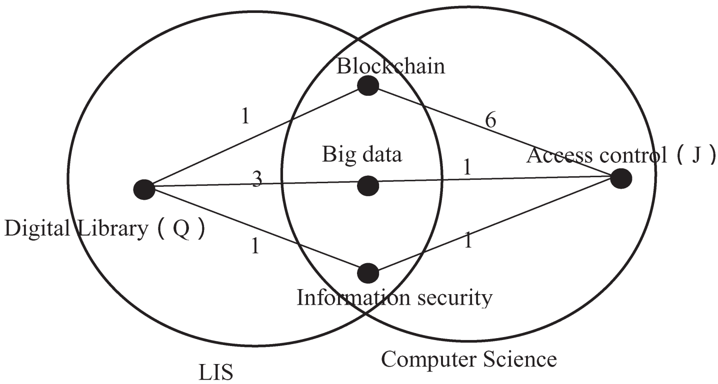


Figure 4 Schematic diagram of irrelevant topic path

In Figure 4, "Digital Library (Q)" and "Access control (J)" have three contact paths, "Blockchain", "Big data" and "Information security". According to formula (1), the feasibility strengths of the three paths PV_j are 7.2, 4.5 and 1. Therefore, it is more likely that "Digital Library" and "Access control" are associated through "Blockchain". So the most effective intermediate word is "Blockchain".

We found 4693 pairs of intermediate words of interdisciplinary cooperation topics by this method, and calculated the feasibility intensity of each contact path of each pair of cooperation topics. Then, we found the best contact path with the largest feasibility intensity PV_j . The contact path with the largest feasibility intensity of interdisciplinary cooperation topics in Table 3 and the corresponding feasibility intensity are shown in columns 4 and 5 of Table 3.

This paper shows the second-order irrelevant knowledge discovery process in the two disciplines of LIS and Computer Science. But in practice, many seemingly irrelevant discipline topics in the two different disciplines may have indirect connections through more order topics. Using the above methods, we can query the multi-order indirect contact paths of any two irrelevant topics, Such as third-order irrelevant topics, fourth-order irrelevant topics, fifth-order irrelevant topics, etc. However, the greater order of crossing between the two irrelevant topics, the lower correlation of the two topics, and the smaller the practical research significance. Because the speed of interdisciplinary cooperation is limited, it is less likely that the research topics of various disciplines will achieve too much integration in the next few decades (Wang & Zhao, 2015).

(5) Application and analysis of results

The second-order irrelevant topics and their intermediate words in the two disciplines are found, which can provide new ideas for solving the problems in a certain field. And we made the knowledge or methods of various disciplines play a greater value through interdisciplinary cooperation. Based on the data in Table 3, we analyzed partial conclusions and explored computer technology solutions to frontier issues in LIS.

Knowledge service (Q) → Recommendation system ← Matrix decomposition (J): Matrix decomposition method can help realize accurate recommendation of knowledge service.

We found that the word "Knowledge service" in LIS is indirectly related to "Matrix decomposition" in Computer Science through "Recommendation system". Knowledge service refers to the information service process of extracting knowledge and information content from various explicit and tacit knowledge resources according to people's needs, building a knowledge network, and providing knowledge content or solutions for users' problems. It is a service centered on user needs and oriented to knowledge content and solutions. Matrix decomposition is to disassemble a matrix into the product of several matrices. It can simplify the calculation and deepen the theory. The most significant advantage of matrix decomposition is that it can maintain good recommendation performance in sparse data sets. The recommendation system can recommend the information and products that users are interested in to users according to users' information needs and interests. Applying matrix decomposition to user demand analysis can improve the efficiency and quality of knowledge service.

Digital Library (Q) →Blockchain ← Access control (J): Protect the legitimate use of digital resources through secure access technology.

We found that there is an indirect relationship between "Digital Library" in LIS and "Access control" in Computer Science. The digital library is a library that processes and stores all kinds of illustrated documents with digital technology. In essence, it is a distributed information system made of multimedia, which makes people's access to information consumption free from space and time constraints to a great extent. Access control is a technology that restricts users' access to certain information items or the use of certain control functions according to the user's identity and a defined group to which they belong. The application of blockchain technology to the construction of digital library system and secure access control mechanism can enable the digital library to effectively control its digital resources, so as to ensure the legitimate rights and interests of information publishers. It can also ensure that legitimate users make rapid and efficient use of network information resources, effectively prevent unauthorized access and intrusion of illegal users, and protect the intellectual property rights of electronic resources.

Digital humanities (Q) →Deep learning ← Computer vision (J): Applied research on ancient books combined with digital humanities and computer vision technology.

"Computer vision" is the indirect topic of "Digital humanities" in Computer Science discovered by irrelevant knowledge discovery method in this paper. Computer vision is a science that studies how to make machines "see". Further, it refers to machine vision such as using cameras and computers to identify, track, and measure targets instead of human eyes. Then, graphics processing get images more suitable for human eyes to observe or transmit to instruments for detection. Computer vision technology has been applied in digital humanities data collection, labeling, analysis, automatic identification of ancient book entities, and ancient book corpus construction combined with deep learning models.

Network public opinion (Q) → Knowledge graph ← Smart home (J): The network public opinion monitoring system developed by informatics can apply to user feedback identification of smart home.

Network public opinion has played a pivotal role in modern society. Companies have also begun to pay attention to the monitoring of network public opinion. We found that the network public opinion monitoring system has great potential for cooperation with smart home related companies. The system can detect users' reviews on products and automatically identify them. It can also combine with the knowledge graph and develop into a user comment evolution graph of smart home. In this way, informatics may cooperate with smart home companies to develop user feedback monitoring systems in the future.

Opened irrelevant knowledge discovery method and results

(1) Two-dimensional vector space model of topics

Co-word network can intuitively show the co-occurrence relationship between two nodes with a direct connection, but the second-order relevant topics without a direct connection are difficult to find. And the interdisciplinary second-order relevant topics are dazzling. We constructed a two-dimensional vector space model of network nodes to accurately express the attributes of nodes in the interdisciplinary co-word network and facilitate the acquisition of potential interdisciplinary cooperation topics of each topic through operation.

An undirected network W= (N, L) is composed of node set N and edge set L. the point adjacent to the node N_i is called the "adjacent point" of the point. Co-word network is an undirected network, and the size of the node in the network represents the frequency of the topic. The thickness of the edge represents the weight of the edge between the two nodes, and the weight represents the co-occurrence times of the corresponding two nodes.

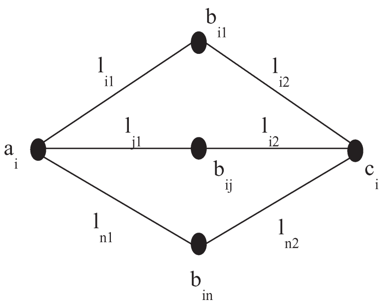


figure 5 Co-word network

In figure 5, a_i represents the initial topic node,b_{ij} (j=1, ... ,n) is the j[th] approach point of a_i and c_i is approach point of b_{ij} (excluding a_i). l_{j1} is the co-occurrence times of the initial topic a_i and the intermediate word b_{ij} , and l_{j2} is the co-occurrence times of the co-word b_{ij} and the target topic c_i .

The second-order vector space model of topics refers to a representation method to describe nodes in a co-word network through three attributes: topic frequency, edge weight, and adjacent points. For example, node a_i in Figure 4 shall be expressed as:

$$a_i = \begin{bmatrix} T_i & l_{11} & b_{i1} \\ T_i & l_{21} & b_{i2} \\ \vdots & \vdots & \vdots \\ T_i & l_{n1} & b_{in} \end{bmatrix} \tag{3}$$

In the two-dimensional vector space model of nodes, a_i is an n × 3 matrix. Matrix first column T_i is the frequency of a_i The second column I_{j1} (j=1, … , n) is the weight corresponding to the edge between the node and each adjacent point. The third column b_{ij} is all adjacent points of node a_i.

We edited the .cit files and. coc files of sample data through Excel, and represent each topic node a_i in the interdisciplinary co-word network with the above vector model. The relationship between each topic a_i and its adjacent point b_{ij} is clear at a glance. At the same time, we found the potential interdisciplinary cooperation topic of each topic through the opened irrelevant knowledge discovery method and the adjacent point in each topic matrix is the intermediate word candidate corresponding to the topic. However, not every candidate word is suitable to be an intermediate word because some neighbors have little connection with a_i. Therefore, adjacent points need to be filtered.

(2) Construct vector model of effective topics

In co-word networks, not all topic pairs are effective. We eliminated the invalid topic pairs, only retained the effective topic pairs, and finally constructed the vector model of effective topics.

For example, the relevant data of the topic "Informatics (Q)" in the vector model of equation (3) are sorted as shown in Table 4.

Table 4 "Informatics (Q)" adjacent points related data

a_i	Frequency T_i of a_i	Co-occurrence times I_{ij}	Adjacent point b_{ij}
Informatics (Q)	97	33	Library and information service (Q)
		22	Library(Q)
		13	Library science(Q)
		6	Big data
		…	…

In table 4, it is not difficult to find that " Library and information service (Q)", " Library(Q)", and" Library science (Q)" are disciplinary co-words. Thus, we cannot directly find the interdisciplinary co-word of "Informatics (Q)" through these words with the help of the substitution method. Therefore, we eliminated them from the co-word network. Topic pairs without suffixes also cannot help us find interdisciplinary co-word. They need to be removed from the co-word network. Finally, only one topic with suffix (Q) and one topic without suffix are retained. Computer Science do the same, retaining only one topic with suffix (J) and one topic without suffix. The screened effective topic pairs are used to construct effective co-word networks of Informatics (Q) and Computer Science.

The vector model of effective topics refers to the effective co-word network according to the informatics obtained in the previous step. It is used to represent the topic matrix collinear with all nodes a_i , which is recorded as a_i'. In order to highlight the relationship between topic a_i and intermediate word b_{ij} , the vector model matrix only retains the relevant information of intermediate word b_{ij} . For example, the effective topic vector model of "Informatics (Q)" is:

$$\text{Informatics(Q)}' = \begin{bmatrix} 8 & \text{Big data} \\ 5 & \text{Knowledge graph} \\ 2 & \text{Deep learning} \\ & \cdots \end{bmatrix} \tag{4}$$

It can be seen that in the data samples, "Big data", "Knowledge graph" and "Deep learning" are the intermediate words of screened "Informatics (Q)", which have appeared 8 times, 5 times, and 2 times respectively.

(3) Identification of potential interdisciplinary cooperative research topics by substitution method

In the above process, we got the topic pairs of A and B and the topic pairs of B and C. To identify the corresponding A and C, we proposed an operation the method suitable for topic vector model - substitution method. The second-order correlation matrix which add topic a_i is written as a_i*. Take "Informatics (Q)" as an example to show the calculation process.

According to the above process and sample data, the relevant core vector model is calculated and constructed as follows:

$$\text{Informatics(Q)}' = \begin{bmatrix} 8 & \text{Big data} \\ 5 & \text{Knowledge graph} \\ 2 & \text{Deep learning} \\ & \cdots \end{bmatrix}$$

and

$$\text{Big data}' = \begin{bmatrix} 3 & \text{Parallel computing (J)} \\ 2 & \text{MapReduce(J)} \\ \cdots & \cdots \end{bmatrix}$$

$$\text{Artificial intelligence}' = \begin{bmatrix} 3 & \text{Deep reinforcement learning (J)} \\ 1 & \text{Generative adversarial network(J)} \\ \cdots & \cdots \end{bmatrix}$$

$$\text{Deep learning(Q)}' = \begin{bmatrix} 12 & \text{Generative adversarial network(J)} \\ 10 & \text{Computer vision(J)} \\ \cdots & \cdots \end{bmatrix}$$

Core vector model was substituted to obtain Informatics (Q)*:

$$\text{Informatics(Q)}^{*} = \begin{bmatrix} 8 & \text{Big data} & 3 & \text{Parallel computing(J)} \\ 8 & \text{Big data} & 2 & \text{MapReduce(J)} \\ 8 & \text{Big data} & \cdots & \cdots \\ 2 & \text{Knowledge mapping} & 2 & \text{Smart home(J)} \\ 2 & \text{Deep learning} & 12 & \text{Generative adversarial network(J)} \\ 2 & \text{Deep learning} & 10 & \text{Computer vision(J)} \\ 2 & \text{Deep learning} & \cdots & \cdots \end{bmatrix} \quad (5)$$

In a_i*, They are the co-occurrence times l_{j1} of a_i and the intermediate words in the first column. They are the intermediate word b_{ij} in the second column. They are the co-occurrence times l_{j2} of intermediate words and second-order relevant topics in the third column. They are the second-order relevant topic c_i corresponding to a_i in the fourth column. Therefore, the substitution method is a process of substituting the effective topic vector model of each Computer Science discipline into the effective topic vector model of Informatics discipline. This process can be realized by Excel and VBA program.

Because of the different co-occurrence times, the connection strength of each topic pair is also different. Therefore, we need to calculate the connection strength of each topic pair selected. Finally, the potential interdisciplinary cooperation topic pairs identified are the topic pairs with greater cooperation possibility.

(4) Cooperation potential calculation for potential interdisciplinary cooperative topics.

In the above process, a two-dimensional vector space model of undirected network was established. We found all nodes a_i in the interdisciplinary co-word network and express them with a matrix. The intermediate word set B of a_i is screened by certain rules, and the effective topic vector model a_i ' of a_i is established. The substitution operation was defined. Through the substitution method, the potential cooperation topic c_i of Computer Science corresponding to a_i in LIS from 2016 to 2020 is identified. Due to the different co-occurrence times, the identified potential interdisciplinary cooperation topic pairs will have different cooperation possibilities.

According to formula (2), we calculated the cooperation potential of potential interdisciplinary cooperation topics identified by opened irrelevant knowledge discovery method. The calculation results are sorted from large to small, and the top 10 pairs of potential interdisciplinary cooperation topics are shown in Table 5.

Table 5 Potential interdisciplinary cooperative topics of LIS and Computer Science

a_i	b_{ij}	c_i	cooperation potential
Academic text(Q)	Deep learning	Generative adversarial network(J)	72.5
Information literacy(Q)	Big data	Parallel computing(J)	41
Smart city(Q)	Big data	MapReduce(J)	41
Informatics(Q)	Deep learning	Deep reinforcement learning(J)	38
knowledge service(Q)	Recommendation system	Matrix decomposition(J)	21
Smart library(Q)	Artificial intelligence	Deep reinforcement learning(J)	18.5
Online reviews(Q)	Emotional analysis	Social media processing(J)	18.5
Internet public opinion(Q)	Knowledge Mapping	Smart home(J)	18.5
Intelligence analysis(Q)	Big data	Functional Dependency(J)	18.5
Patent analysis(Q)	Topic model	Matrix decomposition(J)	8.5
...	...	...	...

(5) Application and analysis of identification results

Table 5 shows the potential interdisciplinary cooperative research topics in LIS and Computer Science identified by the opened irrelevant knowledge discovery method. It also analyzes several pairs of potential cooperation topics with great potential for interdisciplinary cooperation. We explored the technical scheme of using computer theory to solve the problems of LIS. Three pairs of topics are selected for application analysis.

Smart city (Q) → Big data → MapReduce (J): With the help of MapReduce technology and methods, it helps to realize the fine and dynamic management of the city.

We found that "MapReduce" in the field of Computer Science can cooperate with the "Smart city" of LIS through the opened irrelevant knowledge discovery method. The concept of "Smart city" originally came from the media field. It refers to the use of various information technologies or innovative concepts to connect and integrate urban systems and services, so as to improve the efficiency of resource utilization, optimize urban management and services, and help realize the fine and dynamic management of the city.

MapReduce is a computing model, framework, and platform for big data and parallel processing in the computer field. It can help us summarize all kinds of disorderly data generated in the process of urban management according to some characteristics, and then process them to help us solve the problems in the process of urban management.

Smart Library (Q) → Artificial intelligence → Deep reinforcement learning (J): Apply deep reinforcement learning technology to help build a more digital, networked and intelligent library.

"Deep reinforcement learning" is a potential cooperative research topic of "Smart Library" in Computer Science. Smart library refers to an intelligent building formed by applying intelligent technology to library construction. It is the organic combination and innovation of intelligent building and highly automated digital library. Deep reinforcement learning combines the perception ability of deep learning with the decision-making ability of reinforcement learning, which can be controlled directly according to the input information. It is an artificial intelligence method closer to human thinking mode. With the help of deep reinforcement learning technology, we can realize more intelligent service and management, and make the library have the characteristics of knowledge sharing, service efficiency, and use convenience.

Patent analysis (Q) → Topic model → Matrix decomposition (J)：The matrix decomposition algorithm is used to simplify the collection of a large amount of patent information from patent documents such as patent specifications and patent bulletins.

"Matrix decomposition" in Computer Science can carry out interdisciplinary cooperation with "patent analysis" in LIS. Patent analysis requires us to collect a large amount of disordered patent information from patent specifications, patent bulletins, and other patent documents. Then we need to process, sort, combine and analyze the patent information. One of the main ideas of matrix decomposition algorithms in Computer Science is cooperative filtering and optimization of objective function. Then we can use the matrix decomposition to help simplify various information in patent documents. Finally, the statistical method is used to transform this information into a set of activities that can predict competitive intelligence and provide reference for decision-making in various activities of enterprises.

Comparative analysis

The identification results of closed and opened irrelevant knowledge discovery methods in Table 3 and Table 5 can be found:

(1) The recognition results of the two methods are consistent. Among the ten pairs of recognition results displayed by the two irrelevant knowledge discovery methods, there are four pairs of the same interdisciplinary topics, which are academic text & generative adversarial network, knowledge service & matrix decomposition, online comments & social media processing, network public opinion & smart home. Although their interdisciplinary cooperation potential indexes are different, the intermediate words that play the role of intermediate bridge are also consistent. Without considering the influence of different choices of intermediate words with the same strength on the final co-occurrence results, the coincidence recognition results of the two recognition methods have been nearly one-half in the ten pairs of interdisciplinary topics displayed; If further requirements are made for the selection of intermediate words with the same intensity in the same interdisciplinary topic pair, so that the selection of intermediate words follows certain principles on the basis of the same intensity, the probability of co-occurrence of the same interdisciplinary topic pair through the same intermediate word in the two recognition methods will further increase. It can be concluded that the closed and opened irrelevant knowledge discovery methods have high consistency in the identification of potential interdisciplinary cooperation topics, and the two methods are fundamentally interlinked.

(2) There are also some differences in the recognition results of the two methods. These differences are intuitively reflected in the difference of the final co-word pairs, and also in the different cooperation potential intensity of the interdisciplinary topic pairs identified by the two methods. Through the previous research, we know that the interdisciplinary cooperation potential index is not only related to the practical value (PV) of contact path b_j , but also related to the number of contact paths. Different recognition methods will lead to differences in identified intermediate words and contact paths.

Therefore, we identified topics academic texts (Q), knowledge services (Q), online comments (Q), and network public opinion (Q) with great interdisciplinary cooperation potential, and the cooperation target topics identified in Computer Science are also the same through the closed and opened irrelevant knowledge discovery methods. It can be seen that: ① The research results of the two methods have the same part, which shows that although the research methods of closed and opened irrelevant knowledge discovery methods are completely different, they are fundamentally interlinked; ② The recognition results of the two methods are different, which shows that they have their own emphasis in the research process and can learn from each other.

At the same time, we can find that the topics of Computer Science with great interdisciplinary cooperation potential with LIS are mostly technical method topics, such as parallel computing, matrix decomposition, generative adversarial network, etc. The corresponding topics of LIS are mostly expressed in research objects and research fields, which shows that Computer Science provides a certain degree of technical support for the development of LIS. We can speculate that more computer technologies will be applied to LIS in the future to promote the development of LIS, and which technologies of Computer Science can be applied to the field of LIS need to be further explored.

Conclusion

We used the closed and opened irrelevant knowledge discovery methods to identify the topics with interdisciplinary cooperation potential. At the same time, we introduced the topic interdisciplinary cooperation potential index (TICPI) to identify potential cooperation topics for LIS and Computer Science. The index can calculate the feasibility strength of contact path, find the most effective topic contact path, and ultimately achieve interdisciplinary irrelevant cooperative topic recognition and irrelevant topic path query. We analyzed and prospected the application value of this method, and analyzed the identified cooperation topics, which is found that the recognition results of the closed and opened irrelevant knowledge discovery methods proposed in this paper do have certain interdisciplinary cooperation significance. Both methods are feasible in theory.

However, this method uses co-occurrence times screening to determine the contact paths between knowledge, which may make the measurement less than the actual paths. This will cause a certain error in the calculation results of TICPI. The closed and opened irrelevant knowledge discovery methods are not only applicable to the discovery of second-order irrelevant topics, but also can be extended to the discovery of irrelevant topics of any specified order and query the association paths between them. Although the identification of potential interdisciplinary cooperative research topics can explore interdisciplinary solutions to some problems, it is only the first step in the direction. For the development of interdisciplinary cooperation, it is also necessary to determine the appropriate interdisciplinary collaborators. Therefore, we should focus on the identification of interdisciplinary collaborators, so as to make the identification of potential interdisciplinary cooperation topics more practical in the future.

Acknowledgments

The researches described in this paper is supported by the National Social Science Foundation of China, Research on Identification of Interdisciplinary Potential Knowledge Growth Point and Innovation Trend Forecast (No. 19ATQ006). The authors are grateful to Prof. Qiu Junping from Hangzhou Dianzi University. Thanks are also due to the editors and reviewers for their constructive suggestions.

References:

Aalabaf-Sabaghi, M. (2012). Networks, crowds and markets: Reasoning about a highly connected world. *Journal of the Royal Statistical Society Series A, 175* (4), 1073.

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics, 5* (4), 594–607.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. *ACM*, 519.

Bettoni, M., Schiller, G., & Bernhard, W. (2015). *Weak ties cooperation in the CoRe knowledge network.* https://www.researchgate.net/publication/264840767_Weak_Ties_Cooperation_in_the_CoRe_Knowledge_Network

Eto, H. (2003). Interdisciplinary information input and output of a nano-technology project. *Scientometrics, 58*, 5–33. https://doi.org/10.1023/A:1025423406643

Feng, Z., Li, C., Liu, X., & Fu, X. (2018). Interdisciplinary analysis of library and information science based on citing and cited reference information. *Information Science, 36* (3), 105–111. http://doi.org/10.13833/j.issn.1007-7634.2018.03.018

Feng, Z., & Zhang, Z. (2020). Analysis on the influencing factors of potential interdisciplinary cooperation behavior. *Information studies: Theory & Application, 43* (2), 114–120. http://doi.org/10.16353/j.cnki.1000-7490.2020.02.018

Genuis, S. K. (2005). Published literature and diffusion of medical innovation: Exploring innovation generation. *Canadian Journal of Information & Library Science, 29* (1), 27–54.

Granovetter, M. (1973). The strength of weak ties. *The American Journal of sociology, 78* (6), 1360–1380.

Huang, C., & Wang, Y. (2012). Interface rules and governance mechanisms of university interdisciplinary partnerships. *Higher Education Exploration, 2012* (3), 11–16.

Huang, C., & Yang, Y. (2016). The disciplinary integration mechanism and mode selection of university interdisciplinary cooperation. *Higher Education Exploration, 2016* (12), 5–12.

Huang, Y., Zhang, L., Sun, P., Wang, Z., & Zhu, D. (2019). Interdisciplinarity measurement: External knowledge integration, internal information convergence and research activity pattern. *Studies in Science of Science, 37* (1), 25–35. http://doi.org/10.16192/j.cnki.1003-2053.2019.01.005

Jian, X., Yi, B., Ying, D., Sinan, Y., Hongli, Z., Chen, Y., & Lin, S. (2018). Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics, 117,* 973–995.

Li, C., Feng, Z., Liu, Y., & Liu, X. (2018). Identification of potential interdisciplinary collaborators based on citation network: Taking library and information science as an example. *Information and Documentation Services, 2018* (3), 93–98.

Li, C., Liu, F., & Guo, F. (2013). Analysis on interdisciplinary research topics with cinder of overlapping communities visualization software: Taking the information science and computer science for example. *Library and Information Service, 57* (7), 75–80.

Li, C., Liu, X., Liu, Y., & Feng, Z. (2018). Identifying potential disciplinary collaboration research topics by open literature–based discovery: Taking information science and computer science as examples. *Information Studies: Theory & Application, 41* (2), 100–104. http://doi.org/10.16353/j.cnki.1000–7490.2018.02.018

Li, L., Jin, X., & Long, M. (2012). *Topic correlation analysis for cross–domain text classification. AAAI Press*.

Liu, A., & An, T. (2019). System design and implementation of knowledge correlation and retrieval for non–related documents. *Journal of Modern Information, 39* (8), 52–58.

Liu, X., Fu, H., & Jiang, C. (2016). A directional recognition algorithm of semantic relation for literature–based discovery. *Computer Science Bibliography, 2016* (2), 281–288.

Liu, X., Li, C., Cui, B., & Liu, T. (2017). Research topics identification of potential interdisciplinary collaboration based on closed and irrelevant knowledge discovery. *Information studies: Theory & Application, 40* (9), 71–76. http://doi.org/10.16353/j.cnki.1000–7490.2017.09.014

Liu, X., Li, C., Liu, Y., & Fu, X. (2018). Identification of the potential interdisciplinary cooperation combinations based on 2–mode net of author and keywords: Taking library and information science and computer science for example. *Information Studies: Theory & Application, 41* (2), 105–110. http://doi.org/10.16353/j.cnki.1000–7490.2018.02.019

Ma, R. (2015). A new exploratory study on discovery of latent collaborators based on author citation network. *Journal of the China Society for Scientific and Technical Information, 34* (2), 182–191.

Onnela, J. P., Saramakai, J., & Hyvonen, J. (2007). Structure and tie strengths in mobile communication networks. *Proceedings Of The National Academy Of Sciences Of The United States Of America.*

Piepenbrink, A., & Nurmammadov, E. (2015). Topics in the literature of transition economies and emerging markets. *Scientometrics, 102* (3), 2107–2130.

Porter, A. L., Cohen, A. S., Roessner, J. D., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics, 72* (1), 117–147.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the Association for Information Science & Technology, 61* (9), 1871–1887.

Santos, C., Evsukoff, A. G., Lima, B., & Ebecken, N. (2009). Potential collaboration discovery using document clustering and community structure detection. Paper presented at *the Proceeding of the ACM First International Workshop on Complex Networks Meet Information & Knowledge Management*, CIKM–CNIKM 2009, Hong Kong, China, November 6, 2009.

Swanson, D. R. (2015). Fish oil, Raynaud´s syndrome, and undiscovered public knowledge. *Perspectives in Biology & Medicine, 30* (1), 7.

Wu, L., Tian, R., & Zhang, X. (2017). Interdisciplinary topic detection method and empirical research based on topic correlation analysis: A case study of animal resource and breeding. *Library and Information Service, 61* (1), 72–79. http://doi.org/10.13266/j.issn.0252–3116.2017.01.009

Wu, X., & Zhang, C. (2019). Interdisciplinary analysis of library and information science based on social media. *Library And Information Service, 63* (13), 66–74. http://doi.org/10.13266/j.issn.0252–3116.2019.13.007

Xiong, Y., & Yu, Y. (2011). An empirical study on the influencing factors of postgraduates´ interdisciplinary knowledge sharing behavior–Taking master´s students in three universities in region a as examples. *Modern University Education, 2011* (4), 80–86.

Yan, E. (2014). Finding knowledge paths among scientific disciplines. *Journal of the Association for Information Science and Technology, 65* (11), 2331–2347.

Yang, L. (2011). Theoretical basis of interdisciplinary science. *Library and Information Service, 55* (16), 29–32.

Yang, L., Morris, S. A., & Barden, E. M. (2009). Mapping institutions and their weak ties in a specialty: A case study of cystic fibrosis body composition research. *Scientometrics, 79* (2), 421–434.

Yue, Z., Xu, H., Guo, T., & Fang, S. (2016). Comparative research on the interdisciplinary of "LIS" and " Computer Science & Interdisciplinary Application". *Information and Documentation Services, 2016* (2), 7.

Zhang, Y., & Leng, F. (2009). Theoretical basis of knowledge discovery in unrelated documents. *Journal of Library Science in China, 35* (4), 25–30.

Zhao, J., Wu, J., & Ke, X. (2010). Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E, 82* (1 Pt 2), 16105.