Journal of Informetrics 8 (2014) 217–233

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Visual topical analysis of Chinese and American Library and Information Science research institutions

Lu An[a], Chuanming Yu[b,*], Gang Li[a]

[a] School of Information Management, Wuhan University, Luojia Shan, Wuhan, Hubei Province 430072, PR China
[b] School of Information and Safety Engineering, Zhongnan University of Economics and Law, 182# Nanhu Avenue, East Lake High-tech Development Zone, Wuhan 430073, PR China

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form 1 November 2013
Accepted 4 December 2013
Available online 25 December 2013

Keywords:
Self-Organizing Map
Compound Component Plane
Topical analysis
Research institution

ABSTRACT

Research institutions play an important role in scientific research and technical innovation. The topical analysis of research institutions in different countries can facilitate mutual learning and promote potential collaboration. In this study, we illustrate how an unsupervised artificial neural network technique Self-Organizing Map (SOM) can be used to visually analyze the research fields of research institutions. A novel SOM display named Compound Component Plane (CCP) was presented and applied to determine the institutions which made significant contributions to the salient research fields. Eighty-seven Chinese and American LIS institutions and the technical LIS fields were taken as examples. Potential international and domestic collaborators were identified based upon their research similarities. An approach of dividing research institutions into clusters was proposed based on their geometric distances in the SOM display, the *U*-matrix values and the most salient research topics they involved. The concepts of swarm institutions, pivots and landmarks were also defined and their instances were identified.

1. Introduction

Research institutions are major contributors of the scientific and technical innovation. The development of research institutions attracts the attention from the research funding agencies, the higher education system and the institutions themselves. Besides excellent academic performance, an ideal research institution should stand a leading position in hot research topics and fronts in the discipline that it involves, keep balances among several important research fields, and be able to keep its competitive advantages. To attain this goal, research institutions need to evaluate their current research fields and make proactive plans about their future research instead of developing as they will.

The research fields of a research institution can be tracked in many ways. The publications, such as journal articles, proceedings, monographs, research reports and patents, and the thematic indexes, such as keywords, classification codes and controlled terms, are typical sources for topical analysis. As we know, a research institution may produce a lot of publications in a certain time span, for example ten years, which may involve a great number of research fields. Thus, the relationship between research institutions and the research fields are multiple and high-dimensional. Such high-dimensional data can be visualized in a low-dimensional space by information visualization techniques and the complicated connections among objects can be intuitively revealed (Zhang, 2008). The Self-Organizing Map (SOM) is an unsupervised learning algorithm

* Corresponding author. Current address: School of Information and Safety Engineering, Zhongnan University of Economics and Law, 182# Nanhu Avenue, East Lake High-tech Development Zone, Wuhan 430073, PR China.
Tel.: +86 1897 1635 021.
E-mail addresses: anlu97@163.com (L. An), yuchuanming2003@126.com (C. Yu), imiswhu@aliyun.com (G. Li).

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.12.002

which transforms high-dimensional input data into a two-dimensional or three-dimensional configuration and the topology of the input data is preserved. It has the advantages of being applicable to general data distributions, requiring little a priori knowledge (Zhang & Li, 1993), and enabling users to understand the comprehensive data structure by analyzing the map (Rauber, Merkl, & Dittenbach, 2002). The adoption of an information visualization technique such as SOM provides a convenient way to topical analysis.

With the development of the information technology and the Internet, the research focus of Library and Information Science (LIS) has shifted from the traditional research fields such as library theory, document classification to some technical LIS fields, such as digital libraries, information systems and knowledge systems, Information Communication Technology and its usage (Mammo, 2011), E-governance, computer applications and literacy (Sethi & Panda, 2012), and information storage and retrieval. Rana (2011) stated that the open access system, repository system and digital libraries were emerging themes in LIS in India. Blessinger and Hrycaj (2010) found that the technology-related subjects, such as Internet, Automation, Indexes/Databases, Electronic publishing, and software dominated 22% of all subjects among the highly cited papers in the LIS field.

The prevalent global trend raises some unanswered questions: How do the technical LIS fields develop in the LIS research institutions in different countries? How similar or dissimilar are the LIS research institutions of different countries in terms of the technical LIS fields? Which LIS research institutions in the world have similar technical LIS fields? Which technical LIS fields are salient in the LIS research institutions worldwide? What are the research advantages, distinguished features and weaknesses of the LIS institutions in different countries? The answers will be revealed upon the analysis of the first-hand data about individual research institutions and their research fields.

The purpose of this study is to establish the methods of (1) revealing the similarities/dissimilarities among the research institutions in terms of their research fields; (2) examining the main research fields of the institution clusters; and (3) identifying the salient research fields and determining the institutions which made significant contributions to the salient research fields. To illustrate the process of the methods, the Chinese and American LIS research institutions and the technical research fields were taken as examples. The constructed methods can also be applied to other disciplines or other types of research entities, such as experts and research teams.

2. Related research

2.1. The analysis of topical research of institutions

Regarding the analysis of topical research of institutions, plenty of works have been done in the field of medicine (Shen et al., 2011a, 2011b), information science (Moed, Moya-Anegón, López-Illescas, & Visser, 2011), accounting (Daigle & Arnold, 2000), criminology (Reid & Chen, 2007), and material science (Huang, Chen, Chen, & Roco, 2004). Early researches mainly calculated the quantities of the research outputs of individual institutions in a certain research field and lacked of detailed analysis of the specific research topics, such as the work by Daigle and Arnold (2000) and Reid and Chen (2007).

Later, some researchers not only discerned the important research institutions in a specific field, but also identified leading research topics. However, they failed to scrutinize the relationship between research institutions and their research topics (Huang et al., 2004; Shen et al., 2011a, 2011b).

Only a few researchers studied the relationship between research institutions and their research topics. However, their main objectives were to discover whether the concentration of institution research was associated with better research performance (Moed et al., 2011).

From the above analysis, it is seen that deep analysis of the sub-field of individual institutions is seldom conducted in a specific field, such as the LIS field. Some bibliometric institutes regularly conducted researches in revealing the activities of research institutions on different disciplines, such as the ISI Web of Knowledge. However, these studies usually aggregated at the university level or a broad discipline. For example, in Web of Knowledge, the Library and Information Science, together with some other specific disciplines, was included in the discipline of *Social Sciences*, *General*. The users can only get to know the research activities of a university on a broad discipline, such as Social Science, General. The public has difficulties in identifying the research activities of a school or department on specific fields, such as the LIS field.

The analysis of topical research of institutions also leads to identifying potential collaborations between institutions and to measuring the research leadership at a university or a nation. Boyack (2009) proposed a method for identifying collaboration opportunities between two institutions based on determining authors whose publications fell into the same paper cluster. A collaboration potential index was designed to measure the potential overlap of all US universities with the work of Sandia National Laboratories. Klavans and Boyack (2010) developed a novel approach to measure research leadership at a university and three nations. The new method took into account the multi-disciplinary activities of researchers and thus was proved to more accurately measure research leadership at the national level compared with the traditional method.

The existing literature usually employed statistical, bibliometric, content analysis or some other methods to study the research performance of institutions or the main research topics in a specific field. A small amount of studies adopted some visualization techniques, such as subject charts (Blessinger & Hrycaj, 2010), content map analysis (Mammo, 2011; Sethi & Panda, 2012), block-modeling (Mammo, 2011), to identify the core authors and institutions, or to reveal the main subjects and domains in a specific field. However, the visual analysis of the similarities among research institutions in terms of their

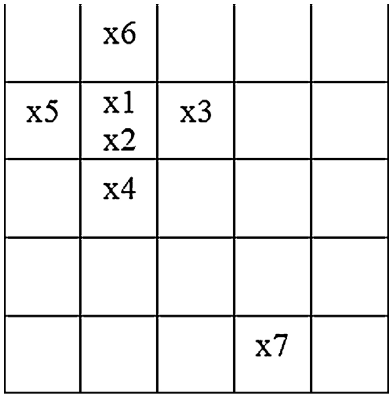

Fig. 1. An SOM display.

research topics was seldom conducted. In this study, we generate several comprehensive visual representations to show the collaboration potentials between individual institution pairs.

2.2. Information visualization techniques

As the scientific publications grow at an exponential rate, more and more information visualization techniques were developed, employed and evaluated by relevant researchers to generate useful knowledge from the huge amount of scientific publications. Examples include the iOPENER Workbench (Dunne, Shneiderman, Dorr, & Klavans, 2010), later renamed Action Science Explorer (ASE) (Dunne, Shneiderman, Gove, Klavans, & Dorr, 2011), GoPubMed (Transinsight, 2011), the citation tree by Web of Knowledge (Thomson Reuters, 2011), VxInsight (Boyack & Börner, 2003; Davidson, Hendrickson, Johnson, Meyers, & Wylie, 1998) and the CiteSpace-based visualization method by Chen (2004).

The above visualization researches usually aggregated at the author or paper level based on citation or co-citation networks. The abundant publications and the keywords, controlled terms or classification codes they involved provide potential clues to mine the research fields of institutions. However, few studies were conducted at the institution level based on the thematic contents they contain.

2.3. Self-Organizing Map

The Self-Organizing Map (SOM) technique is an unsupervised learning algorithm which visualizes the high-dimensional input data in a low-dimensional space as well as preserves the topology of the input data (Kohonen, 2001). The SOM display is usually composed of a two-dimensional regular grid of nodes. See Fig. 1.

Each SOM node is associated with a weight vector of n dimensions, where n equals the number of attributes of the input data. Through competitive machine learning, each of the input data (e.g. x1, x2, . . ., x7 in Fig. 1) is projected onto a SOM node. Since the SOM display can preserve the topology of the input data, the input data with similar attributes are projected onto the same SOM node (e.g. x1 and x2) or the immediate neighboring nodes (e.g. x1 and x3 (or x4, x5, x6)) while those with quite different attributes are projected onto the SOM nodes in far distance (e.g. x1 and x7).

The *U*-matrix map is a common SOM display presented by Ultsch and Siemon (1990). The *U*-matrix can effectively reveal the differences among the weight vectors associated with the SOM nodes and reveal the clustering structure of the input data. The definition of the *U*-matrix was later revised and improved. The updated *U*-matrix (Ultsch, 1992) has the same length and width as the SOM grids. The value in each unit of the *U*-matrix is equal to the sum of the distances of a SOM node to all its immediate neighboring nodes normalized by the largest occurring value in the SOM grid. A high *U*-matrix value means that the weight vector associated with the corresponding SOM node is quite different from the immediate neighboring SOM nodes and vice versa. As a result, the SOM nodes with high *U*-matrix values represent cluster boundaries, while those with low *U*-matrix values represent clusters themselves. The values in the *U*-matrix are then converted to specific colors and applied to the background color of the SOM display to illustrate the differences.

Therefore, the users may observe the similarity/dissimilarity degrees of the input data according to the positions of the input data in the SOM display and its background colors defined by the *U*-matrix.

Since its presence, the SOM technique has been widely used in many fields and disciplines, for example ontological analysis in the geographic domain (Mark, Skupin, & Smith, 2001), and geographic information visualization (Skupin, 2002). Recent examples of data analysis include clustering of abstracts (Skupin, 2009), journal subject analysis by An, Zhang, and Yu (2011), subject directory analysis by Zhang, An, Tang, and Hong (2009), and the analysis of electronic products and their attributes by An and Yu (2012).

In this study, we investigate the research fields of the American and Chinese LIS research institutions using the SOM technique, together with a novel SOM display named Compound Component Plane (CCP).

3. Data and methods

3.1. Selection of the Chinese and American LIS institutions

We selected the best 45 graduate schools in the category of library and information studies ranked by US News & World Report (Best Graduate School-Library and Information Studies, 2011), and the best 47 graduate schools in the category of library, information and archive management ranked by Research Center for China Science Evaluation (Qiu, Wang, & Wang, 2011). US News & World Report and Research Center for China Science Evaluation are the two recognized organization in education rankings in America and China, respectively. The top LIS institutions ranked by the two organizations in 2011 are considered as the ones of high academic performance so far. The investigated institutions are listed in the Appendix.

3.2. Selection of the data source

The data source in this study comes from the Engineering Village Compendex database. Compendex is a comprehensive scientific and technical engineering database, which incorporates millions of bibliographic citations and abstracts from thousands of technical journals and conference proceedings (Engineering Village, 2012). Since this study aimed to reveal the research activities of the LIS research institutions on the technical LIS field, the Compendex was selected as the data source in this study. To measure its coverage of the LIS field, we searched the databases of Compendex, Web of Science, Elsevier, Proquest, EBSCO, Wiley and Springer for the records with the author affiliation from the top five Chinese and five American institutions between 2001 and 2012 and found that Compendex produced the most records. Compendex also has the advantage of assigning several controlled terms, e.g. Information Retrieval, for each article.

3.3. Data collection

We used the function of author affiliation to retrieve the Compendex database for the articles published by the individual LIS institutions, respectively. Note that each institution may have many forms in Compendex. For example, “School of Information Management, Wuhan University”, “Information Management School, Wuhan University” and “Wuhan University, School of Information Management” refer to the same institution.

For each institution, the records from January 1, 2001 to December 31, 2012 were retrieved and the corresponding controlled terms were collected. The controlled terms were used because they can solve the problems of homographs, synonyms and polysemes. If the uncontrolled terms, such as author keywords or title words were used, similar concepts may be treated as different ones. For example, electronic library and aerospace digital library both belong to the research field of digital libraries. However, due to the different literal expressions, they will be treated as different research fields. Also, some important concepts may not be fully revealed by the paper title. Thus, the controlled terms were used in this study.

In total, the 87 LIS research institutions, including 43 American institutions and 44 Chinese ones, produced 10,036 articles in the past twelve years, which involved 878 controlled terms.

Since most Chinese LIS studies were conducted by the School of Management, School of Economics and Management, or the like, to remove the controlled terms not related to the LIS fields, three associate professors with PhD degrees in the LIS field were invited to evaluate the relatedness of each controlled term to the LIS field. The controlled terms which received negative judgment from more than one professor were considered not to be related to the LIS field and removed in the study.

As a result, two hundred and twelve (212) controlled terms were considered irrelevant to the LIS field and removed from the study, for example Agricultural Engineering, Carbon Dioxide, and Electric Power Supplies to Apparatus. The rest six hundred and sixty-six (666) controlled terms which were considered really related to the LIS field and the corresponding Chinese schools or departments were investigated in the study. Among these controlled terms, the minimum count is 1, e.g. analog to digital conversion, and the maximum count is 686, i.e. mathematical models.

3.4. Definition of the SOM input matrix

In this study, an SOM input matrix M1 with m rows and n columns was constructed. See Eq. (1), where the rows (m) of the matrix represent the objects (i.e. research institutions) to be visualized in the SOM display, and the columns (n) of the matrix define the attributes (i.e. the controlled terms provided by Compendex) of the objects. Rank all the investigated research institutions alphabetically and number them from 1 to m. Rank all the involved controlled terms alphabetically and number them from 1 to n. In Eq. (1) c_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) stands for a cell of the matrix. Element c_{ij} in the matrix

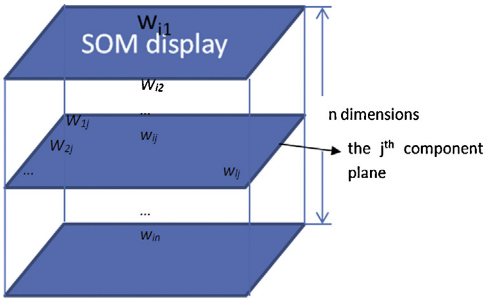

Fig. 2. The SOM display and the jth component planes.

is defined as the number of the articles with the jth controlled terms published by the ith research institutions. If the ith research institutions published no article with the jth controlled terms, then c_{ij} is equal to 0.

$$M1 \begin{pmatrix} c_{11} & c_{12\ldots} & c_{1n} \\ c_{21} & c_{22\ldots} & c_{2N} \\ & \ldots & \\ c_{m1} & c_{m2\ldots} & c_{mn} \end{pmatrix} \tag{1}$$

3.5. Definition of Compound Component Plane

As we know, each SOM node is associated with a weight vector. Each weight vector has n elements where n is the number of attributes of the input samples. The jth element of the weight vectors corresponds to the jth attribute of the input samples. The component plane, a common SOM display, can be generated for any attribute of the input samples. In total, n component planes can be generated because the input samples have n attributes. See Fig. 2 for the illustration of the SOM display and the jth component plane. w_i is the weight vector associated with the ith node and denoted by $(w_{i1}, w_{i2}, \ldots, w_{ij}, \ldots, w_{in})$, where n is the number of the attributes, namely dimensionality. The jth component plane can be denote by $(w_{1j}, w_{2j}, \ldots, w_{ij}, \ldots, w_{lj})$, where l is the number of SOM nodes.

For the jth attribute, the jth component plane is generated by allocating a color level proportional to the jth element values of weight vectors for each node (i.e. $w_{1j}, w_{2j}, \ldots, w_{ij}, \ldots, w_{mj}$) in the SOM grid (Zhang & An, 2010). That is to say, each component plane can only reflect the distribution of one attribute across the overall SOM display.

In this study, if a component plane for a specific controlled term is generated based on the input matrix $M1$, the component plane can only be used to reveal the contribution of each research institution to the controlled term in question. If more controlled terms are to be considered, one can only calculate the total occurrence of these controlled terms in the publications by each institution and observe the results, which is a very time-consuming and demanding task.

To solve the problem, we propose a novel SOM display named Compound Component Plane (CCP), which can reflect the contribution of more than one attribute to the overall SOM display. It can be used to identify the main contributing institutions to the salient research fields. The definition of CCP is described as follows.

Suppose a SOM display has l nodes. w_i is the weight vector associated with the ith node and denoted by $(w_{i1}, w_{i2}, \ldots, w_{in})$, where n is the number of the attributes, namely dimensionality. Construct a Compound Component vector V, denoted by $(v_1, v_2, \ldots, v_l)$. Suppose the CCP is generated for the k attributes denoted by $(s_{p1}, s_{p2}, \ldots, s_{pk})$, the value of V_i, the ith element of V is calculated as Eq. (2), where $i = 1, 2, \ldots, l$.

$$V_i = \sum_{j=p1}^{pk} w_{ij}. \tag{2}$$

It is seen in Eq. (2) that the value of V_i equals to the sum of the k elements (corresponding to the k attributes) of the weight vector associated with the corresponding node.

The values of V_i are converted to colors and applied to coloring the background of the SOM display. Then the CCP is generated. As explained before, a component plane demonstrates the distribution of the values of the specified component (attribute) (An & Yu, 2012). The CCP is actually the overlapping display of a certain number of component planes. Thus, the CCP reveals the distribution of the sum values of the specified components (a certain number of attributes). See Fig. 3 for the illustration of CCP.

In this study, the input matrix $M1$ was processed by the SOM algorithm and the CCP is generated for the salient controlled terms. Thus the CCP becomes the overall institution contribution visualization for the particular components chosen. The institutions which are projected onto the nodes with high V_i values are the ones which published a lot of articles regarding the salient controlled terms.

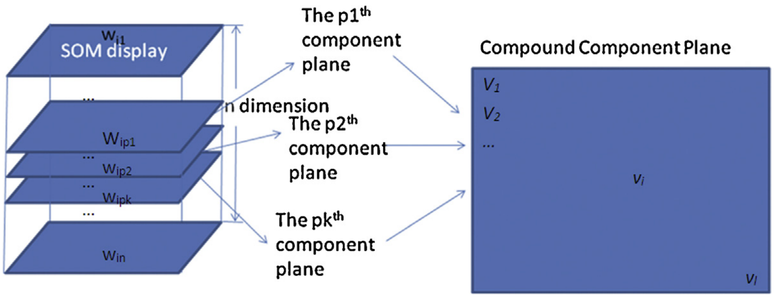

Fig. 3. The component plane and the Compound Component Plane.

3.6. Analysis method

The SOM technique has the advantage of preserving the topology of input data. The institutions projected onto the same node or neighboring nodes in the SOM display are considered to share similar controlled terms. As Chen (2004) designed a method to identify the landmark, hub and pivot articles from the perspectives of citation and cocitation, we are interested in three similar types of institutions from the perspective of the controlled terms. If a number of institutions are projected onto the same or neighboring SOM nodes with low *U*-matrix values, the institutions in question can be called *swarm institutions*. If the LIS institutions of the same country form a continuous area in the SOM display, the institutions which were projected onto the node within the group of the other country are called *pivot institutions*. The *landmark institutions* are those which made significant contributions to the salient research fields.

The *U*-matrix method provides supplementary information to institution similarity analysis. The SOM display can be colored based on its corresponding *U*-matrix to reveal the similarity degrees of the institutions. The SOM nodes with dark blue color represent low values in the *U*-matrix, while the nodes with red or orange color represent high values in the *U*-matrix. An SOM node with a high *U*-matrix value implies that there are significant differences between this node and its adjacent nodes, and vice versa. When two groups of institutions are projected onto two adjacent nodes in the SOM display and the color(s) of either or both nodes are red or orange, it means that these two groups of the institutions have very different controlled terms. If two groups of institutions are projected onto two adjacent nodes and the colors of these two nodes are blue or dark blue, then the two groups of the institutions have similar controlled terms. Thus, the research similarities among institutions can be analyzed based on the geometric distances between the nodes they were projected onto and the *U*-matrix values. Institutions with similar research fields are considered as potential collaborators. To reveal the main research fields of the investigated institutions, the most frequent controlled terms that each LIS institutions involved was explored. Combined with the vicinity of the SOM nodes and the *U*-matrix values, the LIS institutions were categorized into five clusters. The controlled terms for each institution cluster was summarized and discussed.

A thorough investigation of the salient and distinguished research fields was conducted for the Chinese and American institutions, respectively. The common research foci, as well as their differences, were compared.

Finally, two Compound Component Planes (CCPs) were generated for the top twenty salient controlled terms to identify the main contributing institutions to the salient research fields. One CCP was based on the ten Chinese salient controlled terms and another CCP was based on the ten American salient controlled terms. The institutions with high CCP values are good candidates for significant contribution to the salient research fields, which are called *landmark* institutions.

4. Results analysis and discussion

4.1. Analysis of the research similarities among the American and Chinese LIS institutions

The input matrix *M*1 was constructed, which had 87 rows and 666 columns. Notice that different controlled terms may have different occurring frequencies in the investigated LIS institutions. For example, the frequency of *electronic commerce* varied from 0 to 76 while the frequency of *abstracting* varied from 0 to 8. Thus, different attributes (columns) in the input matrix *M*1 may have different value ranges. If the input matrix *M*1 was directly processed by the SOM algorithm, the attributes with large value ranges would dominate the SOM display. To solve this problem, the SOM input matrix was first normalized with the 'Var' method (SOM_norm_variable, 2012), in which the variances of the attributes are linearly normalized to 1. In this case, the attributes with different value ranges were equally treated in the SOM training process. The toroid space was applied in the SOM display to avoid "border effect" (Kohonen, 2001).

The normalized input matrix was trained by linear initiation and batch learning since related studies (An et al., 2011) showed that this combination obtained the smallest final quantization error compared with other combinations of random/linear initiation and sequential/batch learning. The values of the *U*-matrix were calculated and applied to the background color for the SOM display (see Fig. 4 for the results). The color bar on the right indicated the *U*-matrix values for each color. The labels (namely the abbreviations of the LIS institutions) in Fig. 4 represent different institutions. See

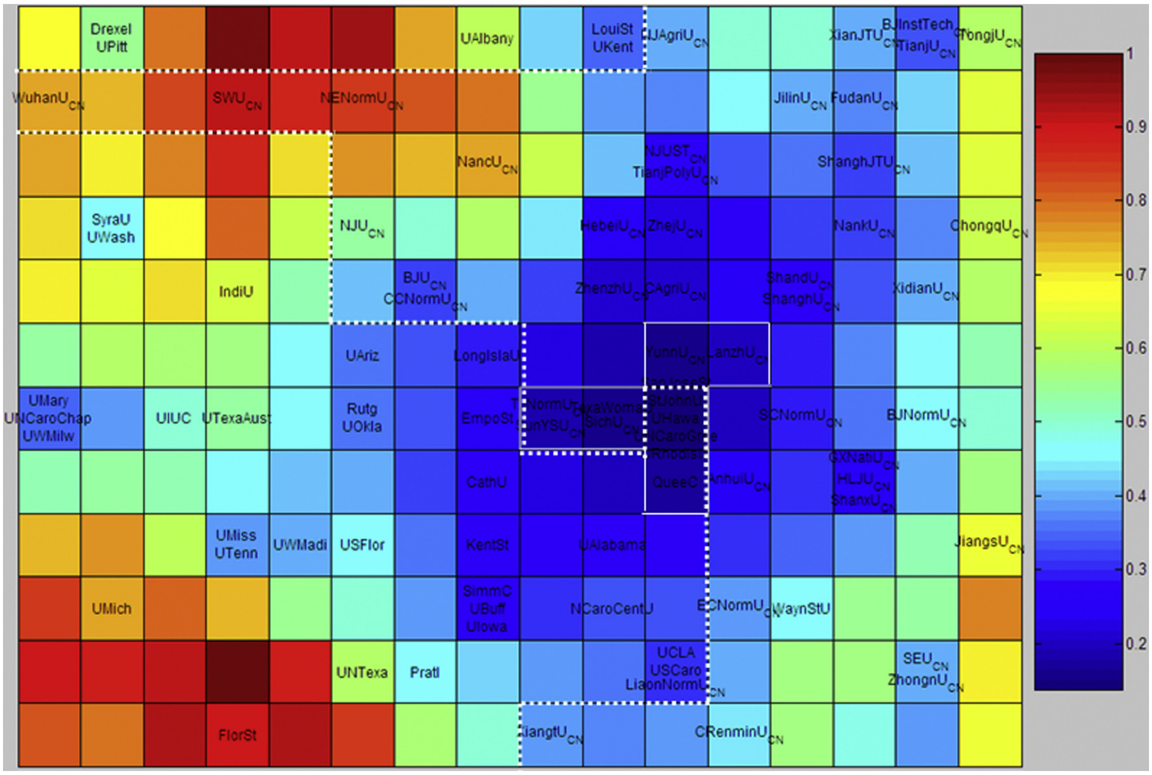

Fig. 4. The SOM display of the American and Chinese LIS institutions. *Note*: the white dash lines in the figure are the boundaries between the projected positions of Chinese and American LIS institutions in the SOM display. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

the Appendix for the relationships between the labels and corresponding institutions. In order to distinguish the Chinese institutions from the American institutions, the abbreviations for the Chinese institutions were appended with the suffix of "CN".

As explained before, the institutions which were projected onto the SOM nodes with low values of *U*-matrix tend to have similar controlled terms while the institutions which were projected onto the SOM nodes with high values of *U*-matrix tend to have quite different controlled terms. The distances between the SOM nodes onto which the input data are projected can also reveal the similarity degree of the input data. In this study, institutions which were projected onto the SOM nodes in vicinity had similar controlled terms while institutions which were projected onto the SOM nodes in far distance had dissimilar controlled terms.

Fig. 4 shows that an obvious dark area exists in the "central" part of the SOM display, as the white lines indicate. Seven American LIS institutions and five Chinese ones were projected onto the SOM nodes with low values of *U*-matrix, which means that these institutions had very similar research fields. See Table 1 for the list. They were considered as *swarm institutions*.

Table 1 shows that the five Chinese and seven American LIS institutions formed a large cluster. The research fields that they involved may be very important and worth exploring further.

On the other hand, SWU$_{CN}$, NENormU$_{CN}$ and FlorSt were projected onto the SOM nodes with high *U*-matrix values and they were separated from other institutions by some empty nodes. It means that the three institutions were highly different from other ones in terms of the controlled terms that it involved.

It is also observed that most of the Chinese and American institutions were projected onto separated area in the SOM display. To make it clear, we drew white dash lines in Fig. 4 and divided the SOM display into three parts. Please note that the white dash lines are not the boundaries between different clusters. They are the boundaries between the projected positions of Chinese and American LIS institutions in the SOM display. Most American institutions were projected onto the "upper"

Table 1
Swarm institutions.

Chinese institutions	American institutions
TJNormU$_{CN}$, SunYSU$_{CN}$, SichU$_{CN}$, YunnU$_{CN}$, LanzhU$_{CN}$	SanJoseSt, StJohnU, TexaWomaU, UHawa, UNCaroGree, URhodIsla, QueeC

Table 2
Pivot institutions.

Chinese pivot institution	American pivot institutions
LiaonNormU$_{CN}$	TexaWomaU, WaynStU

and "lower left" part while most Chinese institutions were projected onto the "right" part. It means that the Chinese and American institutions had very different research fields. The pivot institutions, namely the institutions which were projected onto the nodes within the group of the other country were summarized in Table 2.

Table 2 shows that one Chinese and two American LIS institutions had more similar research fields to their foreign counterparts than to their domestic ones. Please understand that if two or more LIS institutions are projected onto the same or adjacent SOM nodes, it means that the set of the controlled terms they involved as a whole are similar. Here we only listed several of the common controlled terms. For example, the frequent controlled terms of the Chinese LIS institution LiaonNormU$_{CN}$ were *Distributed Computer Networks*, *Mobile Computing*, *Sensor Networks*, *Wireless Telecommunication* Systems, etc., which were more similar to those of UCLA, the American LIS institution in the same SOM node than to other Chinese LIS institutions.

As for the American LIS institutions, the frequent controlled terms of TexaWomaU were Image Retrieval, Information Science, E-learning, etc., which were more similar to those of SichU$_{CN}$, the Chinese LIS institution in the same SOM node than to other American LIS institutions. The same was true for WaynStU. It involved more similar research fields, such as *Behavioral Research*, *Competition*, and *Data mining* to ECNormU$_{CN}$ than to other American LIS institutions.

To further analyze the similarities among the Chinese and American LIS institutions in detail, we identified the most similar Chinese/American institution(s) for each Chinese institution based on the geometric distance between two or more institutions and the *U*-matrix values that they involved. The method of determining the most similar institution(s) to the institution *i* is described as follows.

(1) Find the institution(s) in the shortest geometric distance to the institution *i* in the SOM display.
(2) If one or more institutions in the same node were found, the institution(s) were considered as the most similar institution to the institution *i*.
(3) If more than one institution in different SOM nodes were the same geometric distance to the institution *i*, compare the *U*-matrix values that they involved. The institution(s) with the lower *U*-matrix value(s) were considered as the most similar institution(s) to the institution *i*.

The most similar Chinese/American institution(s) for each Chinese institution and their geometric distances to the institution in question were summarized in Table 3.

It is seen in Table 3 that the average distance from a Chinese institution to its most similar Chinese counterpart was 1.09. However, the average distance from a Chinese institution to its most similar American counterpart was 2.48. It means that Chinese institutions tend to have more similar research fields to their domestic counterparts than to their American counterparts.

A thorough examination of Table 3 discovers twenty-eight Chinese institutions (approximately 63.6% of all Chinese institutions) to which the most similar institution(s) are also Chinese, i.e. their distances to the most similar Chinese institution(s) are shorter than those to the most similar American institution(s). For example, the distance from ShanghU$_{CN}$ to its most similar Chinese institution ShandU$_{CN}$ is 0 while the distance from ShanghU$_{CN}$ to its most similar American institutions SanJoseSt, StJohnU, UHawa, UNCaroGree, and URhodIsla is 4.

On the contrary, there are seven Chinese institutions (underlined in Table 3, approximately 15.9% of all Chinese institutions) to which the most similar institution(s) are American, i.e. their distances to the most similar American institution(s) are shorter than those to the most similar Chinese institution(s). For example, the distance from LiaonNormU$_{CN}$ to its most similar American institutions UCLA and USCaro is 0 while the distance from LiaonNormU$_{CN}$ to its most similar Chinese institutions ECNormU$_{CN}$ and NJAgriU$_{CN}$ is 2.

In addition, nine Chinese institutions (shown in bold font in Table 3, occupying 20.5% of all Chinese institutions) had equal distances to the most similar Chinese institution(s) and to the most similar American institution(s). For example, the distance from NJU$_{CN}$ to its most similar American institution UAriz and the distance from NJU$_{CN}$ to its most similar Chinese institutions BJU$_{CN}$ and CCNormU$_{CN}$ are both 2.

In a word, most of Chinese institutions had more similar technical LIS research fields to their domestic counterparts than to their foreign counterparts. More than three-fifths of Chinese institutions had more similar technical LIS research fields to their domestic counterparts than to their American counterparts. About one-fifth of Chinese institutions had the same similarities in technical LIS research fields with their American counterparts as with their domestic counterparts.

4.2. Analysis of potential domestic and international collaboration institutions

As Fig. 4 reveals the similarities among the institutions in terms of technical LIS research fields, the research similarities among the institutions build sound foundations for the potential collaboration between the involved institutions. To

Table 3
Chinese institutions and the most similar Chinese/American counterpart(s).

Institution	Most similar Chinese institution(s)	Distance	Most similar American institution(s)	Distance
NENormU$_{CN}$	NJU$_{CN}$	2	UNTexa	3
BJU$_{CN}$	CCNormU$_{CN}$	0	LongIslaU	2
CCNormU$_{CN}$	BJU$_{CN}$	0	LongIslaU	2
TJNormU$_{CN}$	SunYSU$_{CN}$	0	TexaWomaU	1
SunYSU$_{CN}$	TJNormU$_{CN}$	0	TexaWomaU	1
HebeiU$_{CN}$	ZhejU$_{CN}$, ZhenzhU$_{CN}$	1	TexaWomaU	3
ZhenzhU$_{CN}$	CAgriU$_{CN}$	1	TexaWomaU	2
NJUST$_{CN}$	TianjPolyU$_{CN}$	0	LouiSt, UKent	3
TianjPolyU$_{CN}$	NJUST$_{CN}$	0	LouiSt, UKent	3
ZhejU$_{CN}$	CAgriU$_{CN}$	1	SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	3
CAgriU$_{CN}$	YunnU$_{CN}$	1	SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	2
LanzhU$_{CN}$	YunnU$_{CN}$	1	SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	2
JilinU$_{CN}$	FudanU$_{CN}$	1	LouiSt, UKent	4
ShandU$_{CN}$	ShanghU$_{CN}$	0	SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	4
ShanghU$_{CN}$	ShandU$_{CN}$	0	SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	4
XianJTU$_{CN}$	BJInstTech$_{CN}$, TianjU$_{CN}$	1	Drexel, UPitt	4
FudanU$_{CN}$	ShanghJTU$_{CN}$	1	Drexel, UPitt	5
ShanghJTU$_{CN}$	NankU$_{CN}$	1	SyraU, UWash	5
NankU$_{CN}$	ShanghJTU$_{CN}$	1	SyraU, UWash	4
GXNatiU$_{CN}$	HLJU$_{CN}$, ShanxU$_{CN}$	0	QueeC	3
HLJU$_{CN}$	GXNatiU$_{CN}$, ShanxU$_{CN}$	0	QueeC	3
ShanxU$_{CN}$	HLJU$_{CN}$, GXNatiU$_{CN}$	0	QueeC	3
BJInstTech$_{CN}$	TianjU$_{CN}$	0	Drexel, UPitt	3
TianjU$_{CN}$	BJInstTech$_{CN}$	0	Drexel, UPitt	3
XidianU$_{CN}$	ShandU$_{CN}$, ShanghU$_{CN}$	2	SyraU, UWash	4
SEU$_{CN}$	ZhongnU$_{CN}$	0	WaynStU	3
ZhongnU$_{CN}$	SEU$_{CN}$	0	WaynStU	3
TongjU$_{CN}$	BJInstTech$_{CN}$, TianjU$_{CN}$	1	Drexel, UPitt	2
WuhanU$_{CN}$	**TongjU$_{CN}$**	**2**	**Drexel, UPitt**	**2**
SWU$_{CN}$	**NENormU$_{CN}$**	**2**	**FlorSt**	**2**
NJU$_{CN}$	**BJU$_{CN}$, CCNormU$_{CN}$**	**2**	**UAriz**	**2**
YunnU$_{CN}$	**LanzhU$_{CN}$**	**1**	**SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla**	**1**
CRenminU$_{CN}$	**LiaonNormU$_{CN}$**	**2**	**UCLA, USCaro**	**2**
SCNormU$_{CN}$	**LanzhU$_{CN}$**	**2**	**SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla**	**2**
BJNormU$_{CN}$	**HLJU$_{CN}$, GXNatiU$_{CN}$, ShanxU$_{CN}$**	**2**	**UMary, UNCaroChap, UWMilw**	**2**
ChongqU$_{CN}$	**NankU$_{CN}$**	**2**	**SyraU, UWash**	**2**
JiangsU$_{CN}$	**HLJU$_{CN}$, GXNatiU$_{CN}$, ShanxU$_{CN}$**	**3**	**UMary, UNCaroChap, UWMilw**	**3**
NancU$_{CN}$	HebeiU$_{CN}$, NJUST$_{CN}$, TianjPolyU$_{CN}$	3	UAlbany	2
XiangtU$_{CN}$	LiaonNormU$_{CN}$	3	LouiSt, UKent	2
SichU$_{CN}$	TJNormU$_{CN}$, SunYSU$_{CN}$	1	TexaWomaU	0
NJArgiU$_{CN}$	NJUST$_{CN}$, TianjPolyU$_{CN}$	2	LouiSt, UKent	1
LiaonNormU$_{CN}$	ECNormU$_{CN}$, NJAgriU$_{CN}$	2	UCLA, USCaro	0
AnhuiU$_{CN}$	LanzhU$_{CN}$	2	QueeC	1
ECNormU$_{CN}$	AnhuiU$_{CN}$	2	WaynStU	1
Average distance		1.09		2.48

Note: If two institutions were projected onto the same node, the distance is 0. If two institutions were projected onto immediate neighboring nodes, the distance is 1. If two institutions were projected onto the diagonal neighboring nodes or the two nodes separated by another node, the distance is 2, and so on.

promote international collaboration, we screened Table 3 for the Chinese–American institution groups in which the distances between the institutions of the two countries were 0 or 1, which means that the involved institutions had very high similarities. The similarities between the LIS institutions also depend on the background colors of the SOM display. If an institution is projected onto a blue SOM node and the distance between this institution and another institution is 1, the two institutions really have similar controlled terms. If an institution is projected onto a red SOM node and the distance between this institution and another institution is 1, the two institutions actually have dissimilar controlled terms. The potential international collaboration institutions, the distances between the group members, and their common research fields were summarized in Table 4.

Table 4 shows that eight groups of Chinese and American institutions have very similar technical LIS research fields and thus may conduct international collaboration. For example, SunYSU$_{CN}$ was recommended to collaborate with TexaWomaU in the field of Information science, Security of Data, and Social Networking (Online).

Since both Chinese and American LIS institutions had similar technical LIS research fields to the institutions of their home country, the distance threshold 0 for recommended domestic collaboration was selected. The institutions of the same country with the distance 0 were recommended to collaborate with each other. Tables 5 and 6 summarize the potential domestic collaboration institutions, the distances between the group members, and their common research fields for Chinese/American institutions, respectively.

Table 4
Highly recommended international collaboration institutions.

Chinese institution	Most similar American institution(s)	Distance	Common research fields
$TJNormU_{CN}$	TexaWomaU	1	Digital Libraries, Information Science, Surveys
$SunYSU_{CN}$	TexaWomaU	1	Information science, Security Of Data, Social Networking (Online)
$YunnU_{CN}$	SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	1	Electronic Commerce, Data Processing, Cognitive Systems
$SichU_{CN}$	TexaWomaU	0	Image Processing, Information analysis, E-learning
$NJArgiU_{CN}$[a]	LouiSt, UKent	1	Computer Simulation, Application, Information Management
$LiaonNormU_{CN}$	UCLA, USCaro	0	Computer Networks, Sensor Networks, Information Systems
$AnhuiU_{CN}$	QueeC	1	Computer systems, Economic Analysis, Problem Solving
$ECNormU_{CN}$	WaynStU[a]	1	Behavioral Research, Competition, Data mining

Note: If two or more institutions were projected onto the same node, the distance is 0. If two or more institutions were projected onto immediate neighboring nodes, the distance is 1.

[a] $NJArgiU_{CN}$ and WaynStU were projected onto a light blue and cyan SOM node, respectively, which means that the two institution groups actually had less similar technical controlled terms than other institution groups in the table.

Table 5
Highly recommended Chinese domestic collaboration institutions.

Most similar Chinese institutions	Distance	Common research fields
BJU_{CN}, $CCNormU_{CN}$	0	Management Science, World Wide Web, Information Services
$TJNormU_{CN}$, $SunYSU_{CN}$	0	Electronic Commerce, Information Technology, Information Management
$NJUST_{CN}$, $TianjPolyU_{CN}$	0	Management Science, Competition, Electronic Commerce
$ShandU_{CN}$, $ShanghU_{CN}$	0	Industry, Management Science, Electronic Commerce
$GXNatiU_{CN}$, $HLJU_{CN}$, $ShanxU_{CN}$	0	Electronic Commerce, Information Management, Analytic Hierarchy Process
$BJInstTech_{CN}$, $TianjU_{CN}$	0	Mathematical Models, Decision Making, Game Theory
SEU_{CN}, $ZhongnU_{CN}$	0	Decision Making, Industry, Mathematical Models

It is seen in Tables 5 and 6 that seven groups of Chinese LIS institutions and nine groups of American LIS institutions were recommended to build collaborative research relationships. For example, BJU_{CN} may collaborate with $CCNormU_{CN}$ in the fields of Management Science, World Wide Web, and Information Services. SimmC may collaborate with UBuff and UIowa in the fields of Information Retrieval, User Interfaces, and Digital Libraries.

4.3. Analysis of the main research fields of the institution clusters

To analyze the main research fields of the investigated LIS institutions, we labeled the SOM display with the three most frequent controlled terms that each LIS institution involved. See Fig. 5 for the labeled SOM display. To save space, the controlled terms were abbreviated by retaining the first four letters of the word.

According to their principles, the SOM display and the *U*-matrix can reveal the clustering structure of the input data. The institutions which were projected onto the neighboring SOM nodes with low *U*-matrix values are considered to form a cluster. The institutions which were projected onto the SOM nodes in far distance or with high *U*-matrix values are considered to form different clusters separately.

To summarize the main research fields of the LIS institutions, we categorized the 87 LIS institutions into different clusters. The method of determining institution clusters is described as follows.

(1) If two or more institutions were projected onto the same SOM nodes (such as $BJInstTech_{CN}$ and $TianjU_{CN}$) or neighboring nodes (such as $BJInstTech_{CN}$ and $XianJTU_{CN}$, also including diagonal neighboring, such as $BJInstTech_{CN}$ and $FudanU_{CN}$) and all their *U*-matrix values were not higher than 0.8 (the background color is blue, cyan, green, yellow or orange), the institutions form a cluster.

Table 6
Highly recommended American domestic collaboration institutions.

Most similar American institutions	Distance	Common research fields
UMary, UNCaroChap, UWMilw	0	Search Engines, Information Retrieval, Internet
Drexel, UPitt	0	Information Retrieval, World Wide Web, Semantics
SyraU, UWash	0	Societies and Institutions, Digital Libraries, Information Management
UMiss, UTenn	0	Search Engines, Information Retrieval, Students
Rutg, UOkla	0	Information Science, Professional Aspects, Information Retrieval
SimmC, UBuff, UIowa	0	Information Retrieval, User Interfaces, Digital Libraries
LouiSt, UKent	0	World Wide Web, Arctic Engineering, Communication
SanJoseSt, StJohnU, UHawa, UNCaroGree, UrohdIsa	0	Students, Data Reduction, Information Systems
UCLA, USCaro	0	Behavioral Research, Computer Architecture, Design

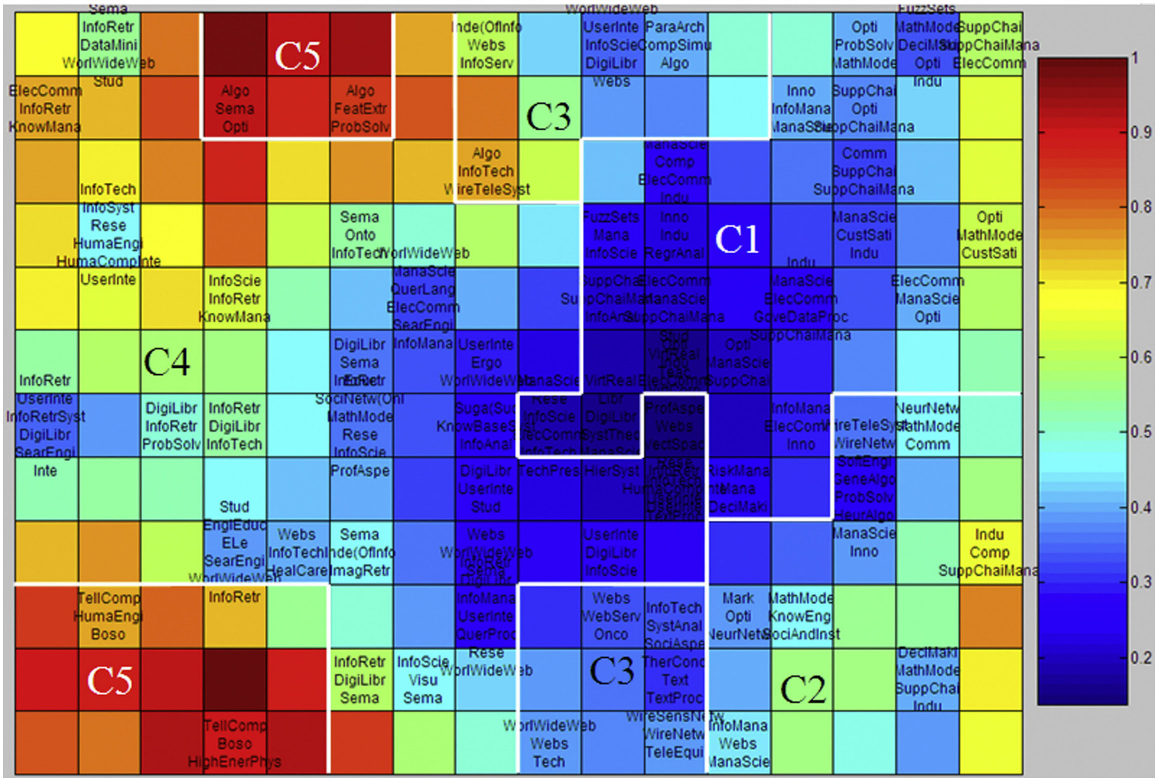

Fig. 5. The SOM display labeled with the most frequent controlled terms. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

(2) If two or more institutions were projected onto SOM nodes separated by empty nodes (such as $\mathrm{NancU_{CN}}$ and $\mathrm{HebeiU_{CN}}$) and one of their *U*-matrix values were higher than 0.8 (the background color is red), the institutions form different clusters, respectively.
(3) If the most frequent controlled terms of an institution were obviously different from those of another neighboring institution (such as EmpoSt and $\mathrm{TJNormU_{CN}}$), the institutions form different clusters, respectively. This rule can ensure clear thematic boundaries between different institution clusters.

Thus, considering the vicinity of the SOM nodes, the *U*-matrix values and the most frequent controlled terms that each LIS institutions involved, the LIS institutions can be divided into five clusters, as the white lines show in Fig. 5.

The five LIS institution clusters and their main research fields were summarized in Table 7.

Table 7
Research institution clusters and their main research fields.

Cluster no.	Research institutions	Main research fields	Number of institutions (%)
C1	$\mathrm{XianJTU_{CN}}$, $\mathrm{BJInstTech_{CN}}$, $\mathrm{TianjU_{CN}}$, $\mathrm{TongjU_{CN}}$, $\mathrm{JilinU_{CN}}$, $\mathrm{FudanU_{CN}}$, $\mathrm{NJUST_{CN}}$, $\mathrm{TianjPolyU_{CN}}$, $\mathrm{ShanghJTU_{CN}}$, $\mathrm{ChongqU_{CN}}$, $\mathrm{NankU_{CN}}$, $\mathrm{ZhejU_{CN}}$, $\mathrm{HebeiU_{CN}}$, $\mathrm{XidianU_{CN}}$, $\mathrm{ShandU_{CN}}$, $\mathrm{ShanghU_{CN}}$, $\mathrm{CAgriU_{CN}}$, $\mathrm{ZhenzhU_{CN}}$, $\mathrm{LanzhU_{CN}}$, $\mathrm{YunnU_{CN}}$, $\mathrm{TJNormU_{CN}}$, $\mathrm{SunYSU_{CN}}$, TexaWomaU, $\mathrm{SichU_{CN}}$, $\mathrm{SCNormU_{CN}}$, $\mathrm{AnhuiU_{CN}}$	Electronic Commerce, Management Science, Supply Chain	26 (29.9%)
C2	$\mathrm{BJNormU_{CN}}$, $\mathrm{GXNatiU_{CN}}$, $\mathrm{HLJU_{CN}}$, $\mathrm{ShanxU_{CN}}$, $\mathrm{JiangsU_{CN}}$, WaynStU, $\mathrm{ECNormU_{CN}}$, $\mathrm{SEU_{CN}}$, $\mathrm{ZhongnU_{CN}}$, $\mathrm{CRenminU_{CN}}$	Mathematical Model, Neural Network, Industry	10 (11.5%)
C3	$\mathrm{XiangtU_{CN}}$, UCLA, USCaro, $\mathrm{LiaonNormU_{CN}}$, NCaroCentU, UAlbany, LouiStU, UKent, $\mathrm{NJArgrU_{CN}}$, $\mathrm{NancU_{CN}}$	World Wide Web, Information Technology, Webs	10 (11.5%)
C4	Drexel, UPitt, $\mathrm{WuhanU_{CN}}$, SyraU, UWash, $\mathrm{NJU_{CN}}$, IndiU, $\mathrm{BJU_{CN}}$, $\mathrm{CCNormU_{CN}}$, UAriz, LongIsaU, UMary, UNCaroChap, UWMilw, UIUC, UTexaAust, Rutg, UOkla, EmpoSt, CathU, UMiss, UTenn, UWMadi, USFlor, KentSt, UAlabama, SimmC, UBuff, UIowa, UNTexa, PratI, QueeC, SanJoseSt, StJohnU, UHawa, UNCaroGree, URhodIsla	Information Retrieval, Semantics, Digital Libraries	37 (42.5%)
C5	UMich, FlorSt, $\mathrm{SWU_{CN}}$, $\mathrm{NENormU_{CN}}$	Miscellaneous	4 (4.6%)

Table 8
The top 10 salient/distinguished feature controlled terms.

Salient in China	Salient in America	Chinese distinguished features	American distinguished features
Electronic Commerce	Information Retrieval	Electronic Commerce	Information Retrieval
Algorithms	User Interfaces	Algorithms	User Interfaces
Competition	Digital Libraries	Competition	Human Engineering
Information Management	Human Engineering	Innovation	Human Computer Interaction
Innovation	Human Computer Interaction	Computer Simulation	Digital Libraries
Computer Simulation	World Wide Web	Government Data Processing	World Wide Web
Information Technology	Search Engines	Information Management	Search Engine
Government Data Processing	Semantics	Management	Metadata
Artificial Intelligence	Information Science	Neural Networks	Social_Networking_(Online)
Neural Networks	Information Systems	Artificial Intelligence	Websites

It is seen in Table 7 that Cluster C4, the largest cluster contains 33 American LIS institutions and 4 Chinese ones, occupying 42.5% of all the institutions. It means that a significant portion of LIS institutions, especially American ones focused on the core LIS research fields, such as of *Information Retrieval*, *Semantics*, and *Digital Libraries*.

Cluster C1, the second largest cluster contains 26 LIS institutions, occupying 29.9% of all the institutions. Twenty-five Chinese LIS institutions and one American institution were found to focus on the research fields of *Electronic Commerce*, *Management Science*, and *Supply Chain*. It is noteworthy that in China the discipline of Library and Information Science is set under the discipline of Management by Ministry of Education of China (MOE, 2013). Thus, the researches by many Chinese LIS institutions showed the subject characteristics of management science. Also, Sugimoto, Pratt, and Hauser (2008) found that the field of Management Information System (MIS) had a great and growing impact on the LIS field, especially on the journals with emphases on technology systems and digital information.

Cluster C2 contains 9 Chinese LIS institutions and one American institution, totally occupying 11.5% of all the institutions. It means that a significant portion of LIS institutions focused on the research fields of *Mathematical Model*, *Neural Network*, and *Industry*.

In addition, ten LIS institutions (11.5%) in Cluster C3 were found to focus on the research field of *World Wide Web*, *Information Technology*, and *Webs*, four LIS institution (4.6%) in Cluster C5 on miscellaneous research fields, such as *Human Engineering*, and *Feature Extraction*.

A comparison between the clustering results in Table 7 and the citation-based clustering results by Yan and Ding (2012) lends support to our findings and also supplements with different contents. According to Yan and Ding, IndiU was clustered with FlorSt in the bibliographic coupling network and with USFlor in the citation network. In Table 7, IndiU was clustered with USFlor in C4. However, FlorSt was clustered with other institutions in C5.

UOkla was clustered with UIUC in the bibliographic coupling network, with UMary in the citation network, and with Dexel in the coauthor network by Yan and Ding (2012). In Table 7, UOkla was clustered with the three institutions in C4.

As Yan and Ding (2012) stated that topical networks were the most similar to coword networks, followed by bibliographic coupling networks and cocitation networks, then citation networks, and finally coauthor networks, differences and similarities may exist among the clustering results in difference networks.

4.4. Identification of the salient/distinguished feature research fields of the American and Chinese LIS institutions

In the above section, the 87 LIS institutions were categorized into five clusters and each cluster was labeled with the main research fields. In fact, each institution usually involved a wide range of research fields, besides the listed controlled terms. Thus, it is interesting to identify the salient research fields of the investigated institutions. In addition, people may be curious about the differences between the salient research fields of the two countries. That is why we reveal the distinguished feature research fields of the institutions in the two countries.

In this study, the salient research fields were defined as the controlled terms which occurred highly frequently (not necessarily the most frequently) in the collection. The top ten salient controlled terms of the American/Chinese LIS institutions were summarized in Table 8.

To compare the distinguished features of the LIS institutions in the two countries, we calculated the average occurrences for each controlled term in America and China and located the controlled terms with the largest average occurrence differences between the two countries. The controlled terms whose average occurrences in Chinese LIS institutions were much more than those in American LIS institutions were considered the distinguished features of Chinese LIS institutions. On the contrary, the controlled terms whose average occurrences in American LIS institutions were much more than those in Chinese LIS institutions were considered the distinguished features of American LIS institutions. The distinguished features of Chinese/American LIS institutions were also summarized in Table 8.

Table 8 shows that in the top 10 salient controlled terms, American and Chinese LIS institutions shared two research fields in common, i.e. *Information Management/Information Science* and *Information Technology/Information Systems*. The other salient research fields of Chinese LIS institutions were Electronic Commerce, Information Technology (e.g. *Algorithms*, *Computer Simulation*, *Artificial Intelligence*, and *Neural Networks*), Competition/Innovation, and Information Management (e.g.

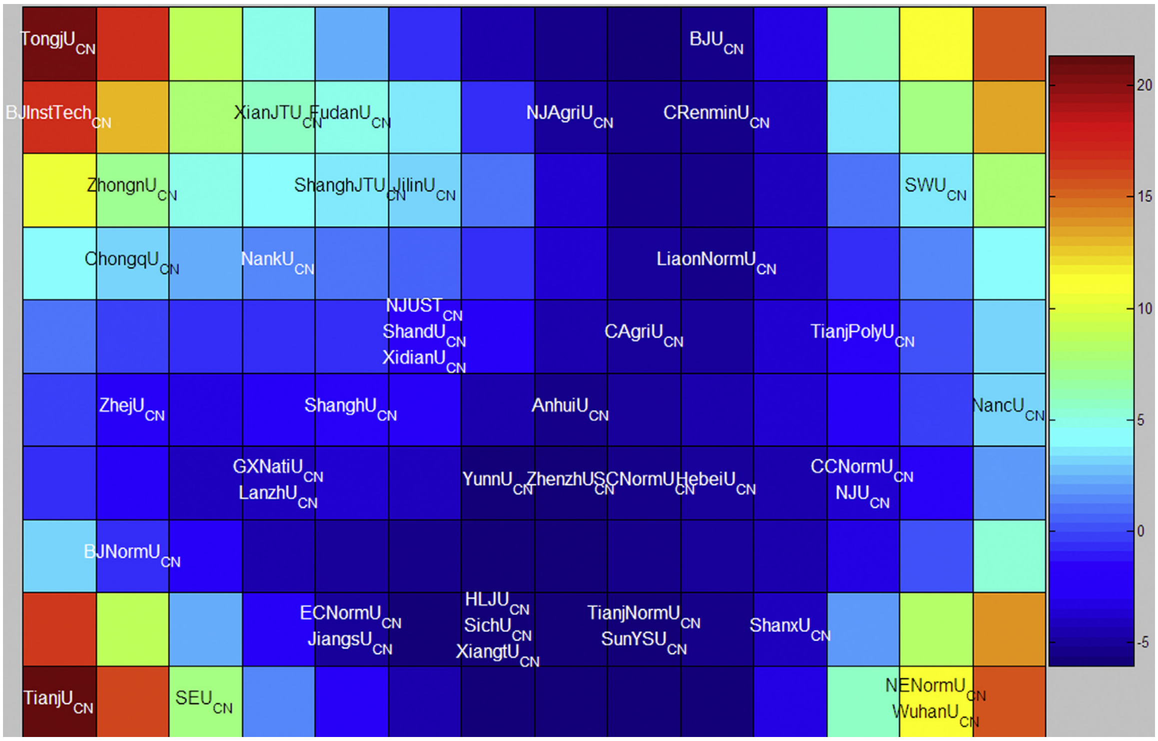

Fig. 6. The CCP for the ten salient controlled terms of the Chinese LIS institutions. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

Government Data Processing). Most of the salient controlled terms were also the distinguished feature research fields of the Chinese LIS institutions, which means that the American LIS institutions seldom did such researches.

The salient research fields of the Chinese LIS institutions identified in Table 8 partially coincided with the findings by Su (2007) and Yang, Du, Zhang, and Li (2008). Su found that the salient research topics in Chinese core LIS journals were Information Services, Information Resources, Competitive Intelligence, etc. Yang et al. discovered that the frequent occurring research topics in funded Chinese LIS projects were Information Resource Management, Knowledge Innovation, Electronic Commerce, etc.

Different from China, the other salient research fields of American LIS institutions were Information Retrieval (e.g. Search Engines), Human Engineering (e.g. User Interfaces, Human Computer Interaction), Digital Libraries, World Wide Web, Semantics, etc., many of which were the research fronts in the LIS field. Most of them, except for Semantics, Information Science and Information Systems, were also the distinguished feature research fields of the American LIS institutions, which means that the investigated Chinese LIS institutions seldom did researches on these topics. The distinguished feature research fields of the American LIS institutions also included Metadata, Social Networking (Online) and Websites, which were the weaknesses of the Chinese LIS institutions.

The salient research fields of the American LIS institutions identified in Table 8 partially coincided with the findings by An et al. (2011) and Yang et al. (2008). An et al. revealed the salient research subjects in American LIS journals were Digital Libraries, Information Technology, Websites, etc. Yang et al. (2008) found that the salient subjects of the *Journal of the American Society for Information Science and Technology* were Intelligent Information Retrieval, Web Users and Behaviors, Digital Libraries, etc., and the salient subjects of *Library and Information Science Abstracts* (LISA) were Web Information Resource Management, Digital Libraries, Web Information Science, etc.

4.5. Determination of main contributing institutions to salient research fields

As explained in Section 3.5, the Compound Component Plane (CCP) which is generated for a certain number of attributes can reveal the contribution of the attributes in question to the SOM display. To visualize the contribution of the American/Chinese LIS institutions to the salient controlled terms, we generated two CCPs separately. See Figs. 6 and 7. Fig. 6 was based on the ten Chinese salient controlled terms in Table 8. Fig. 7 was based on the ten American salient controlled terms.

The background colors of the CCPs are determined by the V_i (an element of the CCP) values as Eq. (2) in Section 3.5 shows. The color bar on the right indicated the V_i values for each color, which indicates that the red color represents high V_i values and the blue color represents low V_i values. The institutions projected onto the red SOM nodes are those who contributed a

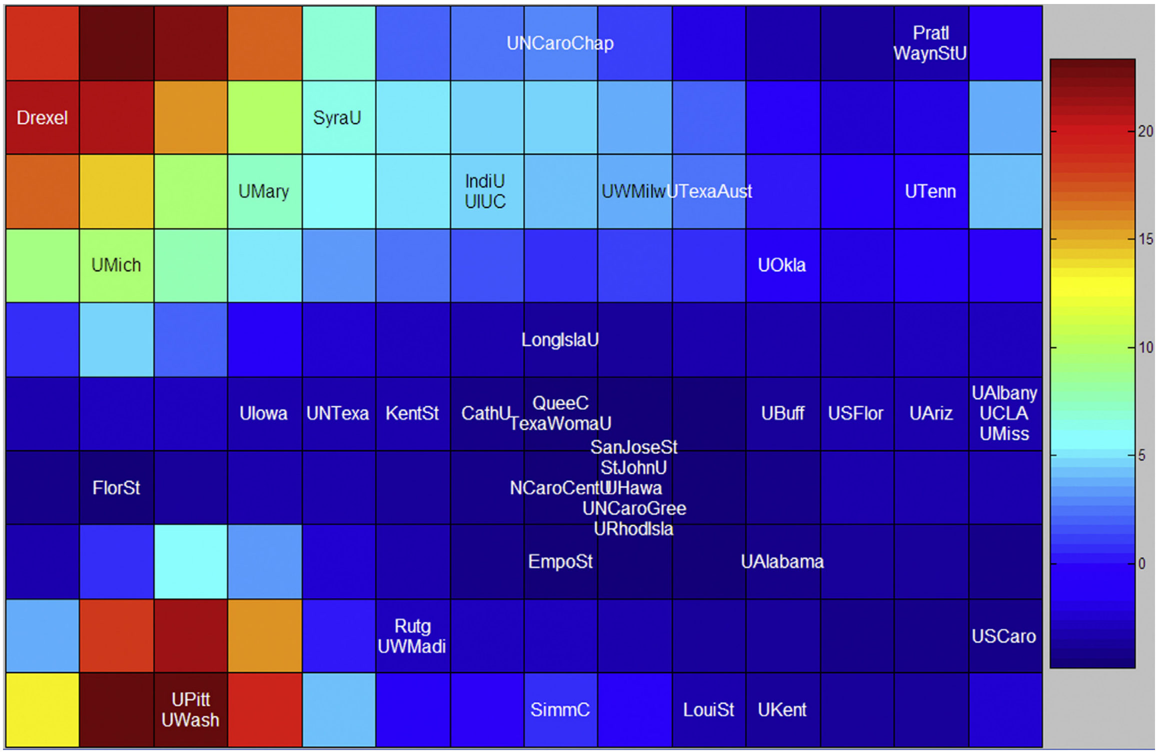

Fig. 7. The CCP for the ten salient controlled terms of the American LIS institutions. (For interpretation of the references to color in the text, the reader is referred to the web version of the article.)

Table 9
Landmark institutions.

CCP color (V_i value)	Chinese landmark institutions	American landmark institutions
Red (>15)	$\text{TongjU}_{\text{CN}}$, $\text{TianjU}_{\text{CN}}$, $\text{BJInstTechU}_{\text{CN}}$,	UPitt, UWash, Drexel
Orange or yellow (10–15)	$\text{WuhanU}_{\text{CN}}$, $\text{NENormU}_{\text{CN}}$	
Green or cyan (5–10)	$\text{ZhongnU}_{\text{CN}}$, $\text{XianJTU}_{\text{CN}}$, SEU_{CN}	UMich, UMary, SyraU

relatively large number of publications to the salient controlled terms in question, while the institutions projected onto the blue SOM nodes are those who contributed little to the salient controlled terms in question.

Figs. 6 and 7 show that most SOM nodes are colored blue, which means that most institutions contributed little to the salient controlled terms. $\text{TongjU}_{\text{CN}}$, $\text{TianjU}_{\text{CN}}$, UPitt and UWash were projected onto the nodes with high V_i values. It means that these institutions contributed a relatively large number of publications to the salient controlled terms in question. According to the corresponding V_i values, the Chinese and American institutions which made unique contributions to the salient controlled terms, namely the *landmark institutions* were summarized in Table 9.

It is seen in Table 9 that eight Chinese institutions and six American institutions made major contributions to the ten salient controlled terms for their own county. The fourteen institutions in Table 9 stood a leading position in salient research fields.

As explained before, the CCP was designed to visualize the contribution of the investigated LIS institutions to the salient controlled terms. To check whether this goal was attained, two Spearman correlation coefficient tests were conducted to verify the correlations between the V_i values (constituting the CCPs) and the total occurrences of the salient controlled terms in the publications of each LIS institution. The results show that the correlations between them are statistically significant. Thus, the CCP can be used to effectively identify the main contributing institutions to the salient controlled terms.

5. Conclusion

The visual topical analysis of research institutions can facilitate understanding the thematic characteristics of the research institutions in a specific discipline, reveal the topical similarities and differences between individual research institutions in different countries, help the institutions find their suitable benchmark counterparts, and potentially promote domestic and international collaboration.

In this study, we illustrate how an effective information visualization technique Self-Organizing Map (SOM) in combination with a self-defined Compound Component Plane (CCP) method can be used to analyze the similarities between the research topics of different research institutions, to find potential collaborators, and to identify the main contributing institutions to the salient research topics. Eighty-seven Chinese and American Library and Information Science (LIS) research institutions and the technical LIS research fields were taken as example. The data investigated came from the Compendex database of Engineering Village. The controlled terms were extracted from the articles published by the investigated institutions for research field analysis.

The investigated LIS institutions were projected onto the SOM nodes. The *U*-matrix was applied to the background colors. Five Chinese LIS institutions and seven American institutions were identified as swarm institutions since they formed a dark area in the SOM display, which means that their research fields were highly similar.

There exists a clear boundary between the projected positions of Chinese and American institutions in the SOM display, which means that most LIS institutions had more similar technical LIS research fields to their domestic counterparts than to the foreign ones. However, one Chinese LIS institution and two American institutions were identified as pivot institutions as they fell into the area of the other country.

A method of determining potential collaborators was presented based on research similarities. Upon it, eight international collaboration institution groups, seven Chinese domestic collaboration institution groups, and nine American domestic collaboration institution groups were found. In each group, every institution was recommended to collaborate with its group member(s) due to their high similarity in technical LIS research fields.

An approach of dividing research institutions into clusters was proposed based on their geometric distances in the SOM display, the *U*-matrix values and the most frequent controlled terms they involved, upon which the investigated LIS institutions were categorized into five clusters. The main research fields for each cluster were revealed.

The concept of distinguished feature was defined and the distinguished features, together with the salient controlled terms were compared between the Chinese and American LIS institutions. The self-defined Compound Component Plane (CCP) was employed to determine the main contributing institutions to the salient research fields, namely the landmarks. The findings can help research institutions examine their own current research fields and make decisions to bridge the gap between those leading institutions and themselves. The constructed methods in this study can also be applied to other disciplines or other types of research entities, such as experts and research teams.

Acknowledgements

We thank the National Social Science Foundation of China (11CTQ025) for financial support. We are also grateful to Qingling Pan and Li Dong for collecting the controlled terms of the articles published by the Chinese and American LIS institutions in this study.

Appendix A.

The list of investigated institutions in this study.

Institution name	Label
Catholic University of America, School of Library and Information Science	CathU
CUNY-Queens College, Graduate School of Library and Information Studies	QueeC
Drexel University, College of Information Science and Technology	Drexel
Emporia State University, School of Library and Information Management	EmpoSt
Florida State University, College of Information	FlorSt
Indiana University – Bloomington, School of Library and Information Science	IndiU
Kent State University, School of Library and Information Science	KentSt
Long Island University – Brookville (Palmer), Palmer School of Library and Information Science	LongIslaU
Louisiana State University – Baton Rouge, School of Library and Information Science	LouiSt
North Carolina Central University, School of Library and Information Sciences	NCaroCentU
Pratt Institute, School of Information and Library Science	PratI
Rutgers, the State University of New Jersey, School of Communication and Information	Rutg
San Jose State University, School of Library and Information Science	SanJoseSt
Simmons College, Graduate School of Library and Information Science	SimmC
St. John's University, Division of Library and Information Science	StJohnU
Syracuse University, School of Information Studies	SyraU
Texas Woman's University, School of Library and Information Studies	TexaWomaU
University at Albany – SUNY, Department of Information Studies	UAlbany
University at Buffalo – SUNY, Department of Library and Information Studies	UBuff
University of Alabama, School of Library and Information Studies	UAlabama
University of Arizona, School of Information Resources and Library Science	UAriz
University of California – Los Angeles, Department of Information Studies	UCLA
University of Hawaii—Manoa, Library and Information Science Program	UHawa
University of Illinois Urbana Champaign School of Library and Information Science	UIUC
University of Iowa, School of Library and Information Science	UIowa

Appendix A (*Continued*)

Institution name	Label
University of Kentucky, School of Library and Information Science	UKent
University of Maryland – College Park, College of Information Studies	UMary
University of Michigan – Ann Arbor, School of Information	UMich
University of Missouri, School of Information Science and Learning Technologies	UMiss
University of North Carolina Chapel Hill, School of Information and Library Science	UNCaroChap
University of North Carolina – Greensboro, Department of Library and Information Studies	UNCaroGree
University of North Texas, School of Library and Information Sciences	UNTexa
University of Oklahoma, School of Library and Information Studies	UOkla
University of Pittsburgh, School of Information Sciences	UPitt
University of Rhode Island, Graduate School of Library and Information Studies	URhodIsla
University of South Carolina, School of Library and Information Science	USCaro
University of South Florida, School of Library and Information Science	USFlor
University of Tennessee – Knoxville, School of Information Sciences	UTenn
University of Texas – Austin, School of Information	UTexaAust
University of Washington, The Information School	UWash
University of Wisconsin – Madison, School of Library and Information Studies	UWMadi
University of Wisconsin – Milwaukee, School of Information Studies	UWMilw
Wayne State University, School of Library and Information Science	WaynStU
Anhui University, School of Management	$\text{AnhuiU}_{\text{CN}}$
Peking University, Department of Information Management	BJU_{CN}
Beijing Institute of Technology, School of Management and Economics	$\text{BJInstTech}_{\text{CN}}$
Beijing Normal University, School of Management	$\text{BJNormU}_{\text{CN}}$
Northeast Normal University, School of Computer Science and Information Technology	$\text{NENormU}_{\text{CN}}$
Southeast University, School of Economics and Management	SEU_{CN}
Fudan University, School of Management	$\text{FudanU}_{\text{CN}}$
Guangxi University for Nationalities, School of Management	$\text{GXNatiU}_{\text{CN}}$
Hebei University, School of Management	$\text{HebeiU}_{\text{CN}}$
Heilongjiang University, School of Information Management	HLJU_{CN}
East China Normal University, School of Business	$\text{ECNormU}_{\text{CN}}$
South China Normal University, School of Economics and Management	$\text{SCNormU}_{\text{CN}}$
Central China Normal University, Department of Information Management	$\text{CCNormU}_{\text{CN}}$
Jilin University, School of Management	$\text{JilinU}_{\text{CN}}$
Jiangsu University, School of Management	$\text{JiangsU}_{\text{CN}}$
Lanzhou University, School of Management	$\text{LanzhU}_{\text{CN}}$
Liaoning Normal University, School of Management	$\text{LiaonNormU}_{\text{CN}}$
Nanchang University, School of Information Engineering	NancU_{CN}
Nanjing University, School of Information Management	NJU_{CN}
Nanjing University of Science and Technology, School of Economics and Management	NJUST_{CN}
Nanjing Agricultural University, School of Information Science and Technology	$\text{NJAgriU}_{\text{CN}}$
Nankai University, School of Business	NankU_{CN}
Shandong University, School of Management	$\text{ShandU}_{\text{CN}}$
Shanxi University, School of Management	$\text{ShanxU}_{\text{CN}}$
Shanghai University, School of Management	$\text{ShanghU}_{\text{CN}}$
Shanghai Jiaotong University, Antai School of Economics and Management	$\text{ShanghJTU}_{\text{CN}}$
Sichuan University, School of Public Administration	SichU_{CN}
Tianjin University, Division of Management and Economics	$\text{TianjU}_{\text{CN}}$
Tianjin Polytechnic University, School of Management	$\text{TianjPolyU}_{\text{CN}}$
Tianjin Normal University, School of Management	$\text{TJNormU}_{\text{CN}}$
Tongji University, School of Management and Economics	$\text{TongjU}_{\text{CN}}$
Wuhan University, School of Information Management	$\text{WuhanU}_{\text{CN}}$
XiDian University, School of Economics and Management	$\text{XidianU}_{\text{CN}}$
Xi'an Jiaotong University, School of Management	$\text{XianJTU}_{\text{CN}}$
Southwest University, School of Computer and Information Science	SWU_{CN}
Xiangtan University, School of Public Administration	$\text{XiangtU}_{\text{CN}}$
Yunnan University, School of Public Administration	YunnU_{CN}
Zhejian University, School of Public Administration	ZhejU_{CN}
Zhenzhou University, Department of Information Management	$\text{ZhenzhU}_{\text{CN}}$
China Agricultural University, School of Economics and Management	$\text{CAgriU}_{\text{CN}}$
China Renmin University, School of Information Resource Management	$\text{CRenminU}_{\text{CN}}$
Zhongnan University, School of Business	$\text{ZhongnU}_{\text{CN}}$
Sun Yat-Sen University, School of Information Management	$\text{SunYSU}_{\text{CN}}$
Chongqing University, School of Economics and Business Administration	$\text{ChongqU}_{\text{CN}}$

References

An, L., & Yu, C. (2012). Self-organizing maps for competitive technical intelligence analysis. *International Journal of Computer Information Systems and Industrial Management Applications*, *4*, 83–91.

An, L., Zhang, J., & Yu, C. (2011). The visual subject analysis of library and information science journals with self-organizing map. *Knowledge Organization*, *38*(4), 299–320.

(2011). *Best Graduate School-Library and Information Studies*. Retrieved from http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/search.result/program+top-library-information-science-programs/top-library-information-science-programs+y.
Blessinger, K., & Hrycaj, P. (2010). Highly cited articles in library and information science: An analysis of content and authorship trends. *Library & Information Science Research*, *32*(2), 156–162.
Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, *79*(1), 27–44.
Boyack, K. W., & Börner, K. (2003). Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *Journal of the American Society for Information Science and Technology*, *54*(5), 447–461.
Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(1), 5303–5310.
Daigle, R. J., & Arnold, V. (2000). An analysis of the research productivity of AIS faculty. *International Journal of Accounting Information Systems*, *1*(2), 106–122.
Davidson, G. S., Hendrickson, B., Johnson, D. K., Meyers, C. E., & Wylie, B. N. (1998). Knowledge mining with vxinsight: Discovery through interaction. *Journal of Intelligent Information System*, *11*(3), 259–285.
Dunne, C., Shneiderman, B., Dorr, B., & Klavans, J. (2010). iOpener Workbench: Tools for rapid understanding of scientific literature. In *Human–computer interaction lab 27th annual symposium* University of Maryland, College Park, MD,. Retrieved from ftp://ftp.umiacs.umd.edu/pub/bonnie/iOPENER-5-27-2010.pdf.
Dunne, C., Shneiderman, B., Gove, R., Klavans, J., & Dorr, B. (2011). Rapid understanding of scientific paper collections: Integrating statistics, text analysis, and visualization. In *Human-Computer Interaction Lab Tech Report*. University of Maryland.
(2012). *Engineering Village*. Retrieved from http://www.engineeringvillage.com/controller/servlet/Controller?CID=quickSearch&database=1.
Huang, Z., Chen, H., Chen, Z.-K., & Roco, M. C. (2004). International nanotechnology development in 2003: Country, institution, and technology field analysis based on USPTO patent database. *Journal of Nanoparticle Research*, *6*(4), 325–354.
Klavans, R., & Boyack, K. W. (2010). Toward an objective, reliable and accurate method for measuring research leadership. *Scientometrics*, *82*(3), 539–553.
Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Berlin: Springer.
Mammo, Y. (2011). Rebirth of library and information science education in Ethiopia: Retrospectives and prospectives. *The International Information & Library Review*, *43*(2), 110–120.
Mark, D. M., Skupin, A., & Smith, B. (2001). Features, objects, and other things: Ontological distinctions in the geographic domain. In D. R. Montello (Ed.), *Lecture notes in computer science* (Vol. 2205) *Spatial information theory: Foundations of geographic information science* (pp. 488–502). Berlin: Springer-Verlag.
Ministry of Education of China. (2013). *Discipline directory of degree granting and talent cultivation*. Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_834/201104/116439.html.
Moed, H. F., Moya-Anegón, F., López-Illescas, C., & Visser, M. (2011). Is concentration of university research associated with better research performance? *Journal of Informetrics*, *5*(4), 649–658.
Rana, R. (2011). Research trends in library and information science in India with a focus on Panjab University Chandigarh. *The International Information & Library Review*, *43*(1), 23–42.
Rauber, A., Merkl, D., & Dittenbach, M. (2002). The growing hierarchical self-organizing map: Exploratory analysis of high-dimensional data. *IEEE Transactions on Neural Networks*, *13*(6), 1331–1341.
Reid, E. F., & Chen, H. (2007). Mapping the contemporary terrorism research domain. *International Journal of Human–Computer Studies*, *65*(1), 42–56.
Research Center for China Science Evaluation. (2011). In J. Qiu, X. Wang, & B. Wang (Eds.), *Evaluation report on Chinese graduate education and disciplines*. Beijing: Kexue Press.
Sethi, B. B., & Panda, K. C. (2012). Growth and nature of international LIS research: An analysis of two journals. *The International Information & Library Review*, *44*(2), 86–99.
Shen, J., Yao, L., Li, Y., Clarke, M., Gan, Q., Fan, Y., et al. (2011a). Visualization studies on evidence-based medicine domain knowledge (series 1): Mapping of evidence-based medicine research subjects. *Journal of Evidence-Based Medicine*, *4*(2), 73–84.
Shen, J., Yao, L., Li, Y., Clarke, M., Gan, Q., Fan, Y., et al. (2011b). Visualization studies on evidence-based medicine domain knowledge (series 2): Structural diagrams of author networks. *Journal of Evidence-Based Medicine*, *4*(2), 85–95.
Skupin, A. (2002). On geometry and transformation in map-like information visualization. In K. Börner, & C. Chen (Eds.), *Lecture notes in computer science* (Vol. 2539) *Visual interfaces to digital libraries* (pp. 161–170). Berlin, Germany: Springer-Verlag.
Skupin, A. (2009). Discrete and continuous conceptualizations of science: Implications for knowledge domain visualization. *Journal of Informetrics*, *3*(3), 233–245.
(2012). *SOM_norm_variable*. Retrieved from http://www.cis.hut.fi/somtoolbox/package/docs2/som_norm_variable.html.
Su, X. (2007). Hot spots and trends of library, information and documentation science (2000–2004). *Journal of the China Society for Scientific and Technical Information*, *26*(3), 373–383.
Sugimoto, C. R., Pratt, J. A., & Hauser, K. (2008). Using field cocitation analysis to assess reciprocal and shared impact of LIS/MIS fields. *Journal of the American Society for Information Science and Technology*, *59*(9), 1441–1453.
Thomson Reuters. (2011). *ISI web of knowledge*. Retrieved from http://www.isiwebofknowledge.com.
Transinsight. (2011). *GoPubMed*. Retrieved from http://www.gopubmed.org.
Ultsch, A. (1992). Self-organizing neural networks for visualization and classification. In *Proceedings of conference of society for information and classification* Dortmund, Germany.
Ultsch, A., & Siemon, H. P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In *Proceedings of international neural network conference* (pp. 305–308). Dordrecht, Netherlands: Kluwer Press.
Yang, W., Zhang, L., Li, L., & Du, X. (2008). Research frontiers of information science: A document-based survey. *Library and Information Service*, *52*(3), 11–14, 61.
Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology*, *63*(7), 1313–1326.
Zhang, J. (2008). *Visualization for information retrieval*. Berlin, Germany: Springer.
Zhang, J., & An, L. (2010). Visual component plane analysis for the medical subjects based on a transaction log. *Canadian Journal of Information and Library Science*, *34*(1), 83–111.
Zhang, J., An, L., Tang, T., & Hong, Y. (2009). Visual health subject directory analysis based on users' traversal activities. *Journal of the American Society for Information Science and Technology*, *60*(10), 1977–1994.
Zhang, X., & Li, Y. (1993). Self-organizing map as a new method for clustering and data analysis. *Proceedings of International Joint Conference on Neural Networks*, *3*, 2448–2451.