Visualization Analysis on Knowledge Organization based on Web of Science

Jie Liu [1, a], Shimin Guo [2, b], Zhijuan Yang [1, c]

[1]School of Economics and Management, Xidian University, Xi'an 710071, China

[2]Institute of Technology, Huazhong Agricultural University, Wuhan 430070, China

[a]jiecando@163.com, [b]884975713@qq.com, [c]763014107@qq.com.

Abstract

In order to meet the development demands of the field of knowledge organization. Through literature review on knowledge organization study made readers know the latest research progress more comprehensively, intuitively and clearly in the field of knowledge organization. In this paper, 586 literature records regarding to Knowledge organization are regarded as the research object from the Web of Science between 2001 and 2015, then the time, the author, the state and the co-citation about literatures were analyzed and the knowledge mappings were drawn by CiteSpace software. The results show that: in the field of knowledge organization, the overall cooperation of the researchers is higher, and the United States and Canada have much stronger international influence. The research hotspots of this field mainly include concept study, foundation et al. The result of retrieving shows that citing journals mainly concentrate on the areas of psychology and education, while cited journals are distributed over in the areas of sociology and computer et al.

Keywords

Knowledge Organization; Visualization; Knowledge Mapping; CiteSpace.

1. Introduction

Knowledge organization has always been a hot topic in academia. Knowledge organization is the word which is first put forward by the famous American library scientist, classification expert Bliss HE in his book "knowledge organization and the science system" in 1929, it refers to the knowledge object for finishing, processing, reveal, control and a series of organizational processes and methods [1]. And through the International Organization of Knowledge Organizations (ISKO) to carry out a series of activities, and constantly improve the content of knowledge organizations. In this way, the knowledge granularity of knowledge organization was condensed to the level of knowledge element [2]. In the method, Yongfu Jiang and Jingzheng Li summed up the representation, reorganization, clustering, checking and editing of knowledge organization and so on[3].

The above studies provide some reference for the readers to understand the knowledge organization, but few scholars have studied it in the way of knowledge map. Therefore, this paper introduces CiteSpace software into the research of knowledge organization, and displays the temporal and spatial distribution, representative personages and representative works in the domain of knowledge organization by knowledge map.

2. Data sources and research tools

2.1 Data sources

This paper takes the core collection of Web Of Science as the data source, and takes "knowledge organization" or "organization of knowledge" as the retrieval style, and selects the time span from 2001 to 2015. There are 586 items of data, include: author (AU), title (TI) and abstract (AB) and so on. In the following analysis process, the time scaling interval is modified to 1, which is divided into 15 segments by year for segment processing.

2.2 Research tool

This paper uses the software developed by Dr. Chen Chaomei based on the Java platform visualization software, CiteSpace3.9.R5. CiteSpace is able to identify and show new trends and developments in scientific literature in scientific literature, it is capable to perform complex, time-varying, and dynamic network analysis. By drawing the map of scientific knowledge, it can show the development trend of knowledge domain in a certain period of time and form the development course of several research frontier fields[4].

3. Time distribution and author′s analysis

3.1 Time distribution

The number of academic papers often as one of the important indicators of the development trend in the field, the total amount of published literature statistics, the status of the field can be measured and development trends. Figure 1: 2001-2015 Web of Science database in the field of knowledge organization published the results, which can be seen from the figure, the relevant literature started in 2000, 2001-2015 the overall trend of rising waves, 2001-2004 issued a relatively small amount, basically single digits, in 2005 it started increased to ten digits, in 2013 reached a peak of 105 articles.
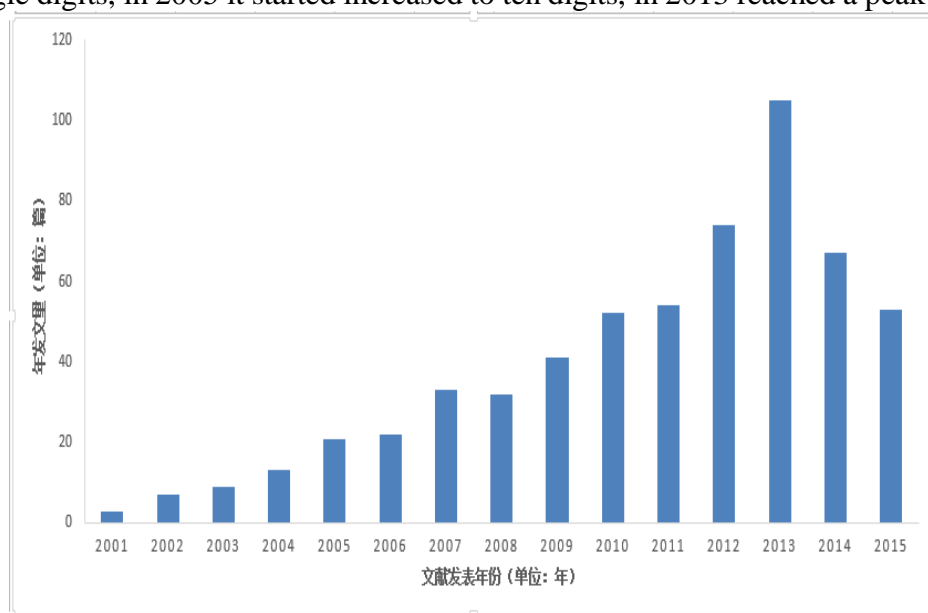

Fig.1 Time distribution

3.2 Country and Institutional Distribution

The study of the distribution of countries and institutions in the field of knowledge organization will provide a more intuitive understanding of the current situation of the exchanges and cooperation between countries and research institutions in this field. In the CiteSpace III interface, the network node selects "Organization" and "Country". The top 50 names are selected, the thresholds are (1,2,10) (1,2,12) and (1,2,12) respectively. After clustering, 402 nodes and 406 edges map of countries and institutions is obtained, as shown in Figure 2.

Figure 2, the circular node refers to the country, attached to a large round node mainly refers to the affiliated research institutions in the country, the greater node , the greater total number of papers issued, the stronger the impact; The thickness of the line between the nodes symbolizes the degree of inter-agency cooperation. At the same time, the concept of intermediate centrality is introduced, referring to the ratio of the shortest path connecting the other two points to the total number of the shortest path lines between these two points. Reference to the results of expert research, the center is greater than or equal to 0.1 points known as the key point.

In Figure 2, the main areas of knowledge organization papers are form the United States, Canada and Spain. Combined with the field of knowledge organization research strength of the table, shown in

Table 1. From the point of view, the contribution rate of the United States is the largest, and the number of papers published is 165, accounting for 28.24% of the total literatures, more higher than other countries, followed by Canada, 63 articles, Germany and China. At the same time, the United States and Canada, the middle of the highest ranking, indicating that most countries in the network are directly or indirectly, and the two have a cooperative relationship. Cooperation with the United States are Spain, Canada and Denmark, with Canada, the main countries of France, China and Italy. The Royal College of Library and Information Sciences in Denmark, the University of Washington in the United States and the University of Alberta in Canada are important institutions in the field of knowledge organization.
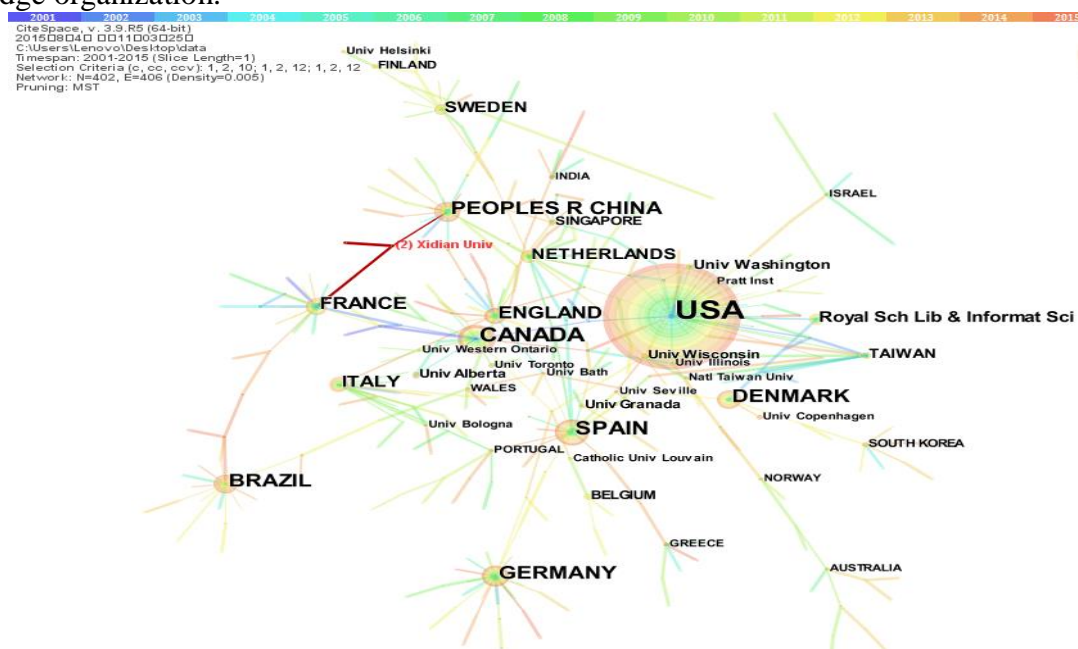

Fig.2 Countries and Institutions map

Tab.1 Table of knowledge organization field research force statistics

country (area)	frequency	intermediate Centralization	institution	frequency	intermediate Centralization
United States	165	0.58	Royal Library and Information Science	13	0.07
Canada	63	0.30	Washington University	8	0.00
Spain	49	0.42	University of Alberta	7	0.00
Germany	34	0.19	University of Granada	7	0.06
China	33	0.26	University ofWisconsin	7	0.23
Brazil	21	0.12	University of Toronto	5	0.01
Denmark	21	0.10	University of Copenhagen	5	0.00
France	19	0.20	University of Illinois	5	0.13
United Kingdom	18	0.17	University of Western Ontario, Canada	4	0.02
Italy	17	0.15	University of Seville	4	0.11

3.3 Authorship analysis

In recent years, the influx of new knowledge and methods has posed new challenges to the field of knowledge organization. In order to meet the new challenges, the author can break through the bottleneck of the subject to become a new choice. (1,2,20) (2,3,20) and (2,3,20) respectively, and select the network simplification algorithm and static clustering to get a network of 267 nodes and 145 sides of the composition of the network, as shown in Figure 3.
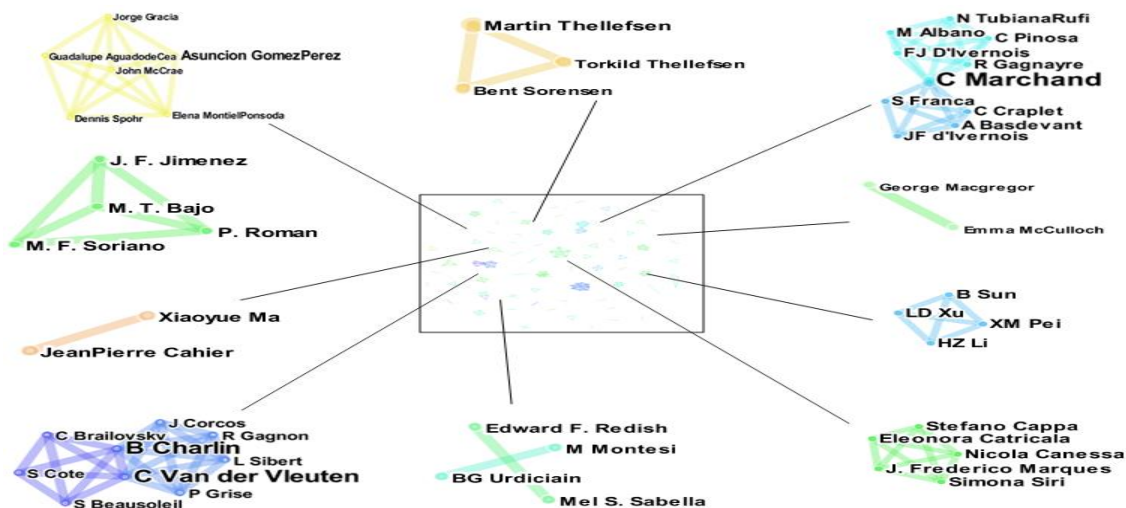
Fig.3 Authors' cooperative network diagram

From the network map and the statistical data, it can be concluded that among the 586 papers, 477 are authors whose number is 2 or more, that is, the overall cooperation rate is 79.79%. The overall cooperation of the researchers is higher and the small groups are more, but the connection between the sub-network are less. According to the network structure in the knowledge map and the related literature[5] , this paper divides the cooperative subnetworks of knowledge organization into point to point mode, media mode and interworking mode.

(1) Point to point mode

It is composed of two authors form a co-operation form. The authors involved in this model are basically from the same institution or university. For example, both Molberg K and Shiri A are from the University of Alberta, Canada, and they have collaborations in 2005, 2007, and 2008 on digital libraries, knowledge organization systems and services and search interface design and research on how to implement them Organic combination [6] . For example, Prof. Cahier JP and Dr. Ma Xiaoyue, both from the CICO Laboratory at the Technical University of Troyes in France, jointly published an article entitled "Graphical graphics for knowledge tagging" in the Journal of Information Science in 2014. The result shows that the new label can improve the label structure and the connotation of the label[7].

(2) Media model

It is a form of a subnet in which a small number of nodes (typically 1 or 2) are connected to another community (the number of authors ≥ 2). In Figure 3, the most obvious media model is the two small groups that Professor Marchand C connects. The authors are Gagnayre R, D'lvernois FJ, Pinosa C, Albanno M, Tubiana Rufi N, who are researchers at the University of Paris, France, and Franca S, Craplet C, Basdevant A, D'lvernois JF from the University of Paris VI , They joint together to study the knowledge organization, concept maps and other technology to explore the impact of preoperative mental state factors, according to the relevant factors to take appropriate psychological intervention, it can effectively improve the quality of care of patients [8].

(3) Interworking mode

It is in the formation of the sub-network, any two authors have a cooperative relationship. For example, from Spain, Bajo MT, Jimenez JF, Soriano MF and Roman P four medical experts in 2009 jointly published a paper entitled Inhibitory processes in memory are impaired in schizophrenia: Evidence from retrieval induced forgetting, the knowledge of cognitive analogy appears barren, the main reason for schizophrenic patients' reactive depression and selective depressive disorder is the clogging of memory retrieval memory.

3.4 Co-citation analysis

Citation frequency is positively correlated with academic ability. Through citation analysis, the highly cited authors and their team play an important role in the research of knowledge organization. We choose (3, 2, 20), (2, 2, 20) and (3, 2, 20) respectively, and then objects were selected by the CiteSpace to the co-citation analysis of the knowledge organization domain. For the top 50, we get 583 nodes and 1954 edges, it formed the network, as shown in Figure 4. According to the frequency of citation and the center of the middle, the authors will be cited as the core author and key authors, in Table 2.
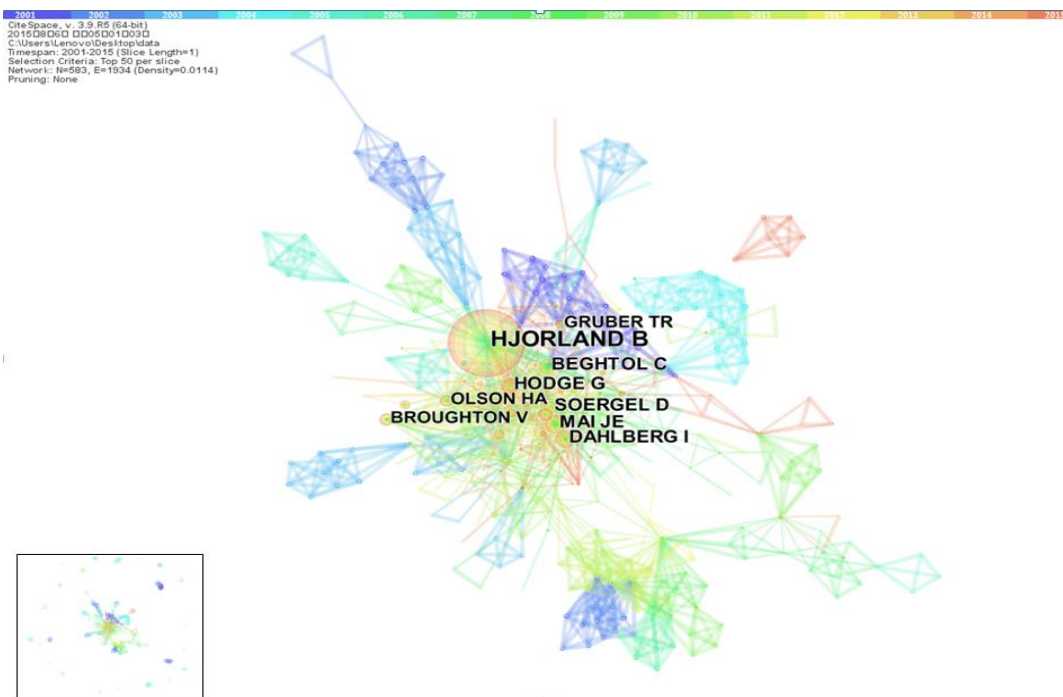

Fig. 4 Co-citation author network

Tab. 2 Co-authorship of core authors and list of key authors

Core author (frequency ≥12)			Key authors (median ≥0.1)		
frequency	Citation time	Author	frequency	Citation time	Author
90	1997	Hjorland B	0.45	1994	Svenonius E
31	1992	Dahlberg I	0.33	1996	Martin A
29	2000	Mai JE	0.23	1997	Hjorland B
27	2002	Bechtol C	0.22	2000	Hodge G
27	2000	Hodge G	0.21	1953	Wittgenstein L
26	1974	SoergeL D	0.21	2003	Spelke ES
23	2002	Olson HA	0.18	2007	Negri GAL
22	1993	Gruber TR	0.17	1997	Medin DL
22	2005	Broughton V	0.17	1988	Zingeser LB
21	2005	Gnoli C	0.17	2006	Saccuman MC
20	1995	Nonaka I	0.17	1999	Coley JD
19	2001	Berners-Lee T	0.16	2002	Devlin JT

Among the core authors, the most frequently cited were Prof. Hjorland B who is come from the Royal Danish Library of Library and Information Sciences, it was cited 90 times in the field of information

science, library and information science, knowledge organization and philosophy. Followed by Prof. Dahlberg I from France, and Prof. Mai JE from Copenhagen University in the third place.

Among the key authors, the center author is Professor Svenonius E who is come from the University of California, USA. His main research interests are in the area of library and information science; the second most important author is Martin A, a senior researcher at the National Institutes of Health, Neuropsychology, radiology and nuclear medicine; the third is Prof. Hjorland B.We can see that knowledge organization is a multidisciplinary integration of the field.

4. Knowledge foundation

According to the frequency of citation and the median sorting, six important documents were selected. Three core documents and three key documents were selected and analyzed by the search engine to retrieve the relevant data. In this paper, the CiteSpace III tools were used to map the co-citation network, Important documents, visual display in the field of knowledge base. Core documents and key documents are important components of the domain knowledge base, and the knowledge base has an inestimable role in specifying the frontier of research in this field and maintaining the steady development of the field itself.We choose (2, 2, 20) (2, 2, 20) and (3, 3, 20), the top 50 is selected, and the final result are 369 nodes and 1395 edges of the authors were co-citation network, shown in Figure 5.
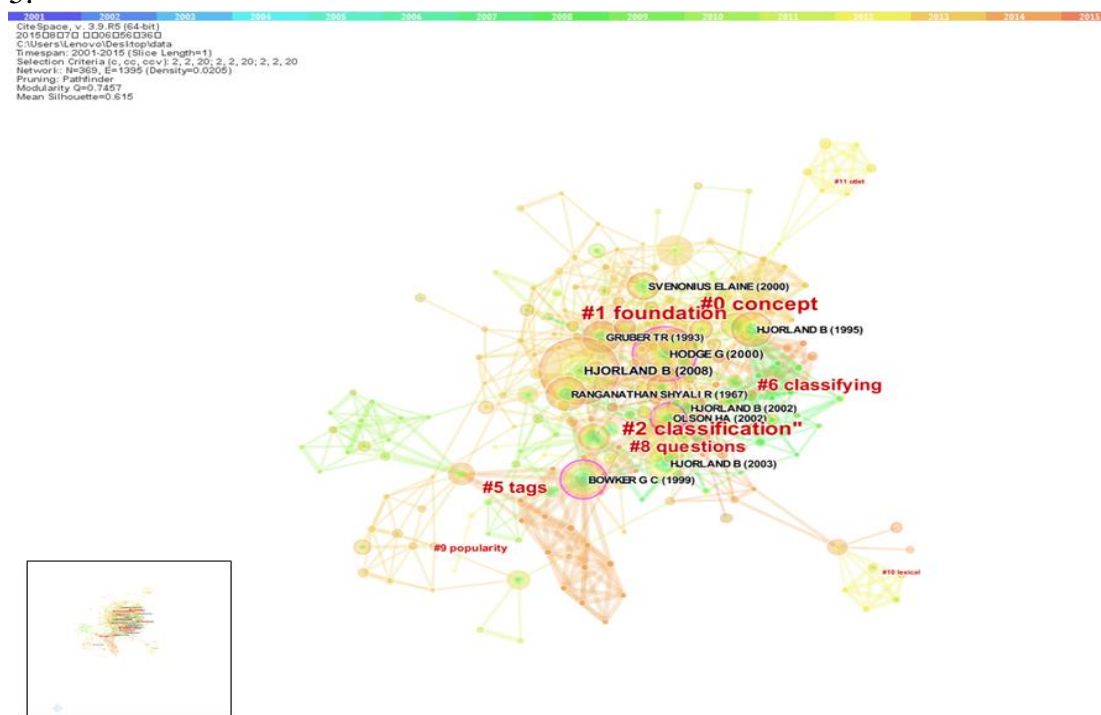

Fig.5 Clustering of co-citation literatures

4.1 Core Document Analysis

Core literature refers to the literature which is closely related to the development of the field, which is often repeatedly cited, often contains the theoretical knowledge of the field of study [9].

The first article is from Professor Hjorland B who is come from the Royal Danish Library of Library and Information Sciences (1997). He published the most informative work entitled Information seeking and subject representation in 1997. He was cited 371 times in Google Scholar's search (in 2016 5 Month 5). This paper analyzes the relationship between Information Science (IS) and Information Retrieval (IR) from the perspective of social cognition, and gives an overview of the differences between traditional cognitive and social cognitive views.[10]

The second article is the Knowledge organization and terminology: Philosophical and linguistic bases paper, published by Professor Dahlberg Ingetraut, France in 1992. In this paper, the value of

philosophical contribution to knowledge organization and terminology discipline leadership is discussed and exemplified. With particular emphasis on the use of logic, science, ontology and metaphysics and other philosophical theory. [11]

The third article is from the University of Copenhagen, Denmark, Professor Mai JE published in 2000. The subject indexing process: an investigation of problems in knowledge description papers, proposed a knowledge based on the model of the subject of the problem index method. In the same year, the organization of information was published on the Journal of the American society for information science. In the comments, UCE cataloging and classification were introduced to guide students to use the basic model for cataloging and classification. At the same time, the introduction of a broader framework Concept, more emphasis on access and control, which will focus on knowledge from sublimation to knowledge search up. Professor Mai JE is in favor of Taylor's view that knowledge organization is made up of controlled and structured vocabularies, and important information is still in recession and remains to be explored and discovered. [12]

4.2 Key Literature Analysis

Key documents are the key nodes connecting different networks, often referred to as transitional literature, which can present research hotspots in the field over a period of time[13] .

The first key document is from the United States University of California Professor Svenonius E in 1994 entitled Access to nonbook materials: The limits of subject indexing for visual and aural languages. The purpose of this paper is to find out the subject index factors from the existing different media in the absence of translate ability, and to find out the topic indexing factors in the visual and auditory language. Emphasis on this issue needs to explore the nature of subject index, linguistics, auditory and linguistic symbolism. Through the empirical analysis of attributes, the transition from image retrieval to multimedia retrieval is realized, which makes the knowledge organization get a qualitative leap [14].

The second article, Prof. Martin A, a senior researcher at the National Institutes of Health, published a article entitled Neural correlates of category-specific knowledge in 1996, which was cited 1009 times in the Web of Science search (May 5, 2016). Found an interesting and puzzling phenomenon, the human brain will selectively amnesia specific objects, the knowledge organization into the brain and language function of medicine up. Linguistic functions and brain-region mapping have been updated much more than in the classic Broca / Wernicke area. It can be said that the description of knowledge organization is mainly embodied in the processing and processing of knowledge semantics, and the perfect knowledge organization can promote the development of medical cause [15] .

4.3 Cluster analysis

In this paper, the clustering of network nodes is clustered into two classes by CiteSpace III, and 12 clusters are automatically identified by LLR (Log-likelihood ratio), as shown in Fig.5 and Tab. 3.

Tab. 3 Citation table

Cluster ID	Size	Silhouette	mean(year)	Lable(LLR)	Representative Authors
0	44	0.777	1998	concept	Friedman, Alon
1	43	0.658	2000	foundation	LopezHuertas, Maria
2	40	0.631	1990	classification	Jacob, Elin K
3	39	0.905	1998	case	Pillon, Agnesa
4	36	0.907	2000	case	Pillon, Agnesa
5	31	0.94	2003	tags	Ma, Xiaoyue
6	31	0.814	2000	classifying	Szostak, Rick
7	27	0.989	1996	thematic	Sachs, Olga
8	27	0.837	1989	questions	Pinho, FabioAssis

9	15	0.932	1995	popularity	Sha, Xiqing
10	8	0.989	2003	Lexical	Chen, Shujun
11	7	0.977	1983	otlet	Buckland, Michael K

From the statistical analysis of Figure 5 and Table 3, it is found that # 0 concept has the largest number of authors, with 44 authors. The most representative authors are Friedman and Alon. The two clusters with the largest clustering intensity are # 7 thematic and # 10 Lexical, all reached 0.989, indicating that their clusters are more closely connected, # 7 representative of Sachs and Olga; # 10 representative of Chen, Shujun. The authors of # 0, # 1, # 2, and # 4 are basically highly cited authors in the field of knowledge organization. Concepts # 0, # 10, # 7 and # 8 play an indispensable role in the development of knowledge organizations. Basic research (# 1), classification (# 2, # 6), cases (# 3, # 4) and labels (# 5) are all important applications of knowledge organization. In addition, the rapid development of knowledge organizations should also consider the cultural background of users in different regions, but also produce different fame (# 9).
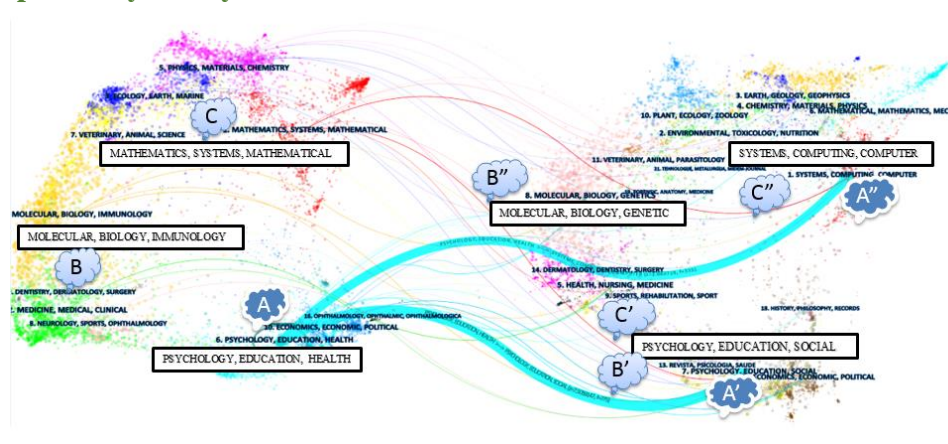
4.4 Dual-map overlays analysis

Fig. 6 Dual-map overlays

Dual-map overlays are divided into left and right sub-graphs. The left-side graphs are called citation literature periodicals. The right-side graphs are named as the periodicals in the cited literatures. In figure 6, the field of knowledge organization cited journals are mainly distributed in the three regions: A blue mark to identify the psychological / pedagogy / health-based curve B Molecular / biology / immunology identified in yellow and curve C are supplemented by mathematical / systematic identification of the red color. The reference curve originating from the blue-yellow-red region points to the right-side reference graph: where curve A mainly points to (Curve A ') Psychology / Pedagogy / Sociology, (curve A ") system / computer; (Curve B ') Psychology / Pedagogy / Sociology, (Curve B') Molecule / Biology / Genetics Two Regions; Curve A Main Point (Curve C ') Psychology / Pedagogy / Sociology C ") system / computer. This fully shows that the development and application of knowledge organization involves a variety of different disciplines, such as computer, psychology and pedagogy, etc. The data proved that knowledge organization is an emerging interdisciplinary research field.

5. Conclusion

(1) In the era of information explosion, the research field of knowledge organization is a increase trend. The United States and Canada have strong influence in this field. The Royal Library and Information Science College in Denmark and the University of Washington in the United States are important organizations. Thus, China should learn more from the latest research institutions, combined with domestic resources out of a knowledge with Chinese characteristics Organization of research roads.

(2) From the author's point of view: the overall level of cooperation of the higher, but little exchange between groups. Through co-citation analysis, Hjorland B, Dahlberg I, Mai JE, Svenonius E and Martin A are the leaders in the field of knowledge organization. Their academic achievements are of great reference value.

(3) Through the co-citation analysis and clustering analysis, the research area of knowledge organization is composed of a group of highly cited and high-center literatures, which include Professor Hjorland B, entitled Information seeking and subject Representation and Professor Svenonius E. By CiteSpace spectral clustering, a number of research hotspots, such as conceptual, taxonomy and basic research categories.

(4) Citation journals are mainly distributed in psychology, education and other fields. Citation journals are mainly distributed in Sociology, systems, computers and other fields. It is proved that knowledge organization is an interdisciplinary research field, strengthening the communication and cooperation among disciplines is one of the directions of the development of knowledge organization in the future.

References

[1] Bliss H E. The Organization of Knowledge and the System of the Sciences [M]. Holt and Company, 1929:92-94.
[2] Youkui Wen. Knowledge Organization and Retrieval Based on "Knowledge Element"[J]. Computer Engineering and Applications, 2005, 01:55-57, 91.
[3] Yongfu Jiang, Jingzheng Li. Knowledge Organization Method [J]. Chinese Journal of Hospital Administration, 2001, 01:3-7.
[4] Yue Chen,Chaomei Chen. The Methodological Function of Knowledge [J]. Science of Science research, 2015, 33(2): 242-253.
[5] Argote, Linda. Organizational Learning: From Experience to Knowledge [J]. Organization Science, 2011, 22(5):1123-1137.
[6] Shiri A, Molberg K. Interfaces to knowledge organization systems in Canadian digital library collections [J]. Online Information Review, 2005, 29(6): 604-620.
[7] Ma,X.Y, Cahier,J.P. Graphically structured icons for knowledge tagging[J]. Journal of Information Science, 2014, 40(6):779-795.
[8] Marchand C, Poitou C, Dehaye B. Cognitive structures of obese patients undergoing bariatric surgery: a concept mapping analysis [J]. Obesity Surgery, 2007, 17(10):1350-1356.
[9] Junping Qiu, Ke Dong. The author reveals the comparative research on the capabilityof revealing the network structure of scientific research [J]. Chinese Journal of Hospital Administration, 2014, 40(1): 15-24.
[10] Hjorland B. Information seeking and subject representation [M]. ABC-CLIO/ Greenwood, 1997.
[11] Dahlberg I. Knowledge organization and terminology: Philosophical and linguistic bases [J]. International classification, 1992, 19(2): 65-71.
[12] Dahlberg I. Knowledge organization and terminology: Philosophical and linguistic bases [J]. International classification, 1992, 19(2): 65-71.
[13] Qiang Xu. Bibliometrical Analysis of Social Computing [J]. Journal of Information, 2015, 34(2): 148-153.
[14] Svenonius E. Access to nonbook materials: The limits of subject indexing   for visual   and aural languages [J]. JASIS, 1994, 45(8): 600-606.
[15] Martin A, Wiggs C L, Ungerleider LG, et al. Neural correlates of category specific knowledge [J]. Nature, 1996, 379(6566): 649-652.