

Knowledge Mapping of Citation Analysis Domains

Yongxia Liang¹ Zeyuan Liu¹ Zhongkai Yang¹ Xianwen Wang¹

30 May 2008

Abstract

With the development of information technology, mapping of science is an emerging research field and become hot topic in scientometrics. The authors select “citation analysis” as search word and 747 records of articles and 23487 citation references in 1974 – 2006 are retrieved from SCI-Expanded and SSCI in WoS. Then the authors study the dataset with Citation Analysis, Cluster Analysis, Multidimensional Scaling, Social network analysis and get knowledge maps in “citation analysis” field. Some softwares for example SPSS, NETDRAW, CITESPACE are used. These maps could reflect the development trend, find out the pivotal people and work in citation analysis field and exhibit the relation between knowledge structure and evolution process.

1 Introduction

With the development of information technology, mapping of science is an emerging research field and become hot topic in scientometrics. Mapping of science is a kind of graph which display the development and structure of science, it is the scientometrics’ production which from maths expression to figure, the result that knowledge from geography distributing map to visualizing knowledge structure and evolution disciplinary(Chen &Liu2005) 。 The forerunner of visualizing knowledge domains that based on citation data is the history map for DNA research, which is handmade in 40 years ago (Garfield, Sher, & Torpie, 1964) 。 From then on, Derek Price uses

the same data in his famous article, The Science Citation Network, which described visualizing knowledge domains of the relation between research fronts by space expression (Price 1961;Price 1963; Price 1965) 。 Garfield also presented the idea of longitudinal mapping in 1994. In longitudinal maps, a series of maps that ordered by years can be used to look over the evolution of science. Analysts and domain experts could predict the emerging trend of discipline by longitudinal mapping. Due to the visualizing knowledge domains could describe the representative work in a field. Researchers could be familiar with a field easily by reading the landmark articles, books and know the members of invisible college in this field, it’s useful especially for those newcomers to a new field. To celebrate Garfield’s 75 birthdays, Cronin & Atkins (Cronin & Atkins, 2000)published a book The Web of Knowledge, particularly introduced the history, theory and application of Citation Index and analysis. Katy Börner, Chaomei Chen and Kevin W. Boyack(2003) reviewed the course of Visualizing Knowledge, they used several emerging visualizing tools and methods to describe the main domains of scientometrics. One of their graphs indicated that there are four embranchments in citation analysis field from 1997 to 2001, which is ordered by factor analysis: mapping of science and visualizing technology, social study of science, bibliometrics and quantitative analysis and evaluation of science, knowledge diffusion and cocitation analysis. Although a lot of scholars have studied many fields based on visualizing tools and citation data, the study for “citation analysis” itself is rare. In this paper, the authors try to use several visualizing tools to show the citation analysis field

¹ WISE Lab ,Dalian University of Technology,Dalian, 116085, P.R.China

clearly. There are four parts in this paper. The first part is the introduction, following is the data source and method, the third part is the analysis result, the last is the discussion and conclusion.

2 Data source and method

The authors select “citation analysis” as search word and 747 records of articles and 23487 citation references in 1974 – 2006 are retrieved from SCI-Expanded and SSCI in WOS. In order to nicety, the data have been cleaned up and standardization.

Using Bibexcel, we have got the relationships between reference authors, selected the top 53 authors which time cited above 35, making co-citation matrix; using SPSS, we carry through Multidimensional Scaling and Cluster Analysis and get the knowledge map of representative scientists and knowledge clusters of embranchments of citation analysis; using Netdarw, we prosecute Social network analysis and protract the network of cocitation authors which indicate the status of authors; Using CiteSpace, we select 3 in “Time Scaling, that is to say, every three years is a time slice, 23 years which between 1974 – 2006 were divided into 8 slices. This action have two advantages: one is propitious to the speed and accurate of software, the other is easy to discriminate the burst point of discipline and tense mode of research front. With CiteSpace, We can get the graph of evolvement of citation analysis.

3 Results

3.1 Graph of Multidimensional Scaling

Scientometrics. Multidimensional Scaling graph is the most widely used visualizing technology information science. It is especially used for literature visualization(Chalmers, 1992), author cocitation analysis (White &McCain, 1998), document analysis(Hetzler et al., 1998), mapping of science (Small, 1999b), visualizing memorizer(McQuaid, Ong, Chen, & Nunamaker,

1999), and some performance evaluation of scientists (Noyons, Moed, & Van Raan, 1999).

We have got the representative scientists map (Fig 1) of citation analysis by Multidimensional Scaling and Cluster analysis on 53 authors which time cited above 35. It is show clearly in fig 1 that authors were formed four clusters by assemble and colony force. In Cluster 1 (red), Garfield and I SCI INF which are the base and origin of citation analysis. There are the most authors in cluster 2 (green) which research is widely, they study on the theory (pay more attention to maths method and model) and application (pay more attention to evaluation of discipline, institution, journal et al, and making policy) of citation analysis. Cluster 3 (yellow) is included researchers of visualizing of citation analysis. The scholars in cluster 4 (blue) who engaged in webometrics.

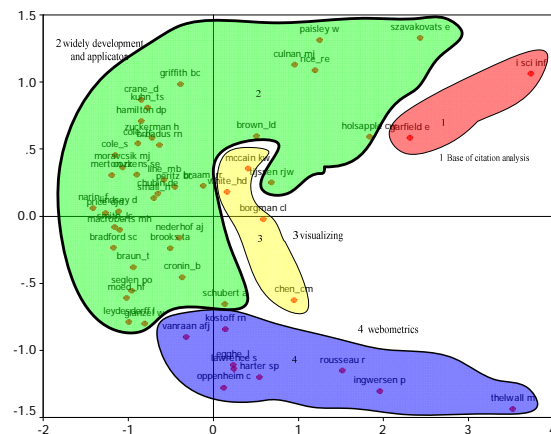


Figure 1: Author co-citation MDS map of citation analysis

From the situation of the clusters we can see that cluster 1 is the origin, cluster 3&4 is the two wings, and cluster 2 is the body. If cluster 1 not created the Science Citation Index, then the following will not be development. On one hand, Cluster 2 derive from cluster 3&4, it is supply abundant materials and base for visualizing and web analysis; on the other, cluster 3 &4 are the advanced tools which accelerate the development of cluster 2. Cluster 2 is the development period of citation analysis, in this period, it is developed the basic methods and theories of citation analysis which expanded the

path for the whole field. Cluster 3 is the visualizing of citation analysis. Due to cite and cited are complex process which could be most explained clearly by graph, so researcher apply themselves to knowledge visualizing. Cluster 4 is network analysis, relations between references are natural network, the development of network also hastens the time efficient of citation and challenges the traditional citation.

According to principle component analysis, we have 4 components, the result are similar to cluster analysis. From the fig 1, we can see that these four components interaction each other. Although some scholars are the representative scientists in one subfield, they also study on the other subfield and act as bridge between these fields. The scholars in cluster 2 are the most great influence groups which have great contribution to theory, method and application of citation analysis. Mapping of science and visualizing technology is the emergent field which integrated the theory and method of scientometrics and information science. In later 1990, webometrics has come into being. Fig 1 shows that webometrics are closed to mapping of science and visualizing technology.

Table 1: Representative scientists and time cited of embranchments of citation analysis

2 broad development and application of citation analysis					
AUTHOR	TIME CITED	AUTHOR	TIME CITED	AUTHOR	TIME CITED
SMALL H	261	NARIN F	110	HARTER SP	70
PRICE DJD	206	SEGLEN PO	93	LEYDESDORFF L	69
CRONIN B	181	MERTON RK	80	GLANZEL W	53
MACROBERTS MH	131	EGGHE L	77	BRAUN T	43
MOED HF	112	CULNAN MJ	74	SCHUBERT A	38
4 webometrics		3 mapping of science and visualizing		1 base of citation analysis	
AUTHOR	TIME CITED	AUTHOR	TIME CITED	AUTHOR	TIME CITED
THELWALL M	166	WHITE HD	228	GARFIELD E	933
ROUSSEAU R	76	MCCAIN KW	150	I SCI INF	110
INGWERSEN P	52	CHEN CM	57		
VANRAAN AFJ	75	BORGMAN CL	57		

According to Citation Analysis, we find in 1974-2006, there are 9312 authors who had been

cited in citation analysis field and the total cited times are 23471 and average cited times are 2.5 per author. These show that there is a large citation analysis researcher's group which has close to 10000 people in the world. In this group, the total cited times of 50 highest authors are 4612 which occupy the 19.6 % in the whole. The average cited times of per high author are 92.24. The higher cited times had proved that these authors had made huge and far-reaching influence on citation analysis's development. The research group which they are represented had improved the different sub-fields of citation analysis came into being and development.

3.2 Graph of social network analysis

Social network analysis is a sociology method which measure and exhibit the relationship between peoples or organizations. In network, Nodes represent people or organization , links express their relationship. Social network analysis have supply visualizing and mathematics analysis for relationship between members(nodes) in networks. This method has been widely used in academia with the development and free use of software (eg. UCINET, Pajek) of social network.

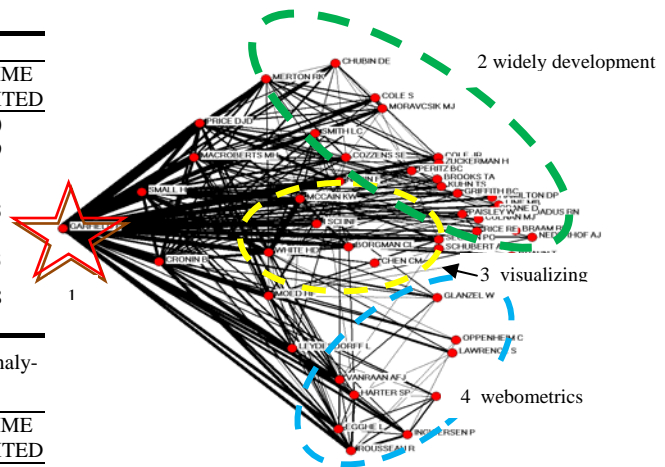


Figure 2: Network of co-cited authors above ten of citation analysis

In 2004, Kretschmer H according to social network analysis, took the science collaboration network of members of COLLNET as example, they compared science collaboration net-

work of based literature with web and discussed the visibility of collaboration on the Web (Kretschmer & Aguillo, 2004).Liu and Yin analyzed the coword network of high frequency keyword in 6 core journal of international scientometrics by social network analysis and visualizing tools[16]. Fig2 is the author co-citation network of citation analysis by Netdraw.

Fig 2 shows the relationship of authors who were co-cited. The thicker of links the more co-cited times. We can see that Garfield are co-occurrence with all other authors and have most status in network. Form the affinity of links, we can say that the similar with the works of authors. It is interesting that the author's situation in network is similar with the clusters in Fig1. I SCI INF situated in the network and have relation with all other members which indicated that it is the base which supply data for citation analysis.

3.3 Graph of CiteSpaceII

From Fig 1, Fig 2 and Tab1, we can see that Chaomei Chen is the important representative scholar in visualizing citation analysis and has made great contribution for information visualizing . He has published a lot of papers about knowledge visualization in some authority international journal such as Journal of the American Society for Information Science and Technology from 1998. He has developed sofeware CiteSpace based on Java which can analyze the science literature and visualizing knowledge. According to analyze and transact records(reference and keyword) of literature, it is can exhibit the evolvment and research front of science development. Using CiteSpace, he investigated the research front and trend of “terrorism” in1990-2003 and studied the development and evolution of “mass extinction” in1981-2003 by coword analysis and co-citation analysis. In this section, we use CiteSpaceII to protract the knowledge map of research front and evolution of citation analysis.

Using CiteSpaceII , import the record data(including title, keyword, reference et al), select “cluster” , at the same time , set threshold as (3, 2, 15) 、(4, 3, 20) 、(4, 3, 20), the nodes are selected as reference, the source is title, abstract, descriptor and identifiers,

term is None, not select pruning, and get the merged network of co-citation cluster. There are 84 nodes and 685 links in Fig3.

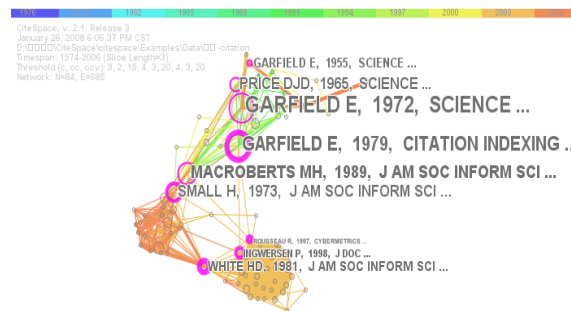


Figure 3: Cluster of “ citation analysis”

Fig3 show that there are 3 clusters in citation analysis; the three clusters are also the three develop periods. In the first cluster, we can see the pivotal points are Garfield E and Price DJD, they two are the founder of citation analysis. In second cluster, Small H develop the citation analysis and put forward the famous co-citation theory; Macroberts MH give a review of citation analysis, he discussed the citation problems by system. The third cluster is the fast and widely development period. In 1980's based on the former, the visualization of analysis had great development and White HD, McCain KW, Chen CM are the representative scholars. In late 1990, with the great development of internet, the web citation analysis has become the hot topic in citation analysis. The famous researchers are Ingwersen P, Thelwall M and Rousseau R. Of course ,due to the threshold of setting , the graph can only reflect foremost scholars and works.

In order to indicate the form and development vividly and pick up the skeleton of citation analysis clearly , we still using CiteSpaceII, import the record data(including title, keyword, reference et al), select “cluster” , at the same time , set threshold as (3, 2, 15) 、(4, 3, 20) 、(4, 3, 20), the nodes are selected as reference, the source is title, abstract, descriptor and identifiers, term is None, minimum spanning tree, pruning sliced networks, pruning merged the network are selected, click time-zone, get the evolution network of citation analysis. There are 202 nodes and 2033links in Fig4.

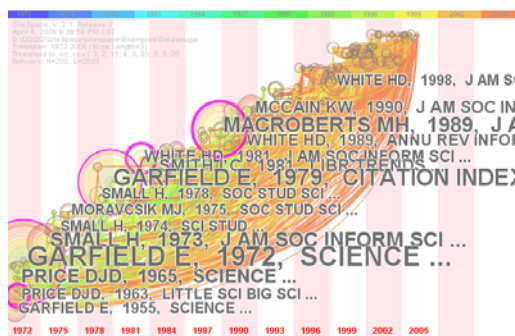


Figure 4: Evolution network of citation analysis

From Fig 4 we can see the evolution network of citation analysis along with time and more representative scholars. Fig4 are coincident with Fig 3. The pink circles are the pivotal articles in citation analysis field. They are the research fronts of the day and the intellectual base of the latter. We can get the role of development of citation analysis, the more early published the more cited times. If the new published articles are the pivotal points and get more cited times, we can deem that it is the hot topics in the filed. The character of CiteSpace is that the turning point is the specifically author and work. It is can predicate the trend and hot topic of discipline.

4 Discussion

In this paper, we use Citation Analysis, Cluster Analysis, Multidimensional Scaling, Social network analysis to analyze the dataset, and get knowledge maps in "citation analysis" field. These maps could reflect the development trend, find out the pivotal people and work in citation analysis field and exhibit the relation between knowledge structure and evolution process.

We can conclude that the knowledge maps of citation analysis which draw by three methods are coincident with each other. They all express some phenomena in citation analysis and indicate the representative scholars and their contribution to citation analysis. The common of these maps are helped us to find information which we wanted from a great deal of information quickly and directly. It is easy for us to understand the citation analysis field.

The character of CiteSpace map is that bring forth some pivotal points and develop trends which indicate the evolution process and emer-

gent field of citation analysis; the MDS map could express the similarity of authors through the space location; the SNA map show the relationship by the links in network. Every map has its advantage and character. We should select the adaptive method based on our requirement.

Acknowledgement

This research was supported by the National Natural Science Foundation of China under Grant 70773015, Project of DUT under Grant DUTHS1002, Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20070141059.

References

- Börner, K., C. Chen, and K. Boyack (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology* 37 (1), 179–255.
- Chalmers M, Chitson P (1992). *Bead: explorations in information visualization*. ACM Press New York, NY, USA.
- Chen C (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*.
- 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 (PNAS)*, 101 (Supp. 1), 5303 -5310.
- Chen Yue, Liu Ze-yuan(2005). The rise of mapping knowledge domain. *Studies in Science of Science*23(2),149-154.
- Garfield E, Sher IH and Torpie RJ (1964). The use of citation data in writing the history of science. *Philadelphia: Institute for Scientific Information*.
- Hetzler B, et al (1998). Multi-faceted Insight Through Interoperable Visual Information Analysis Paradigms. *Proceedings of IEEE Symposium on Information Visualization, InfoVis' 98*, p. 137-144.

- Hetzler B, et al (1998). Visualizing the Full Spectrum of Document Relationships. Proceedings of the Fifth International Society for Knowledge Organization (ISKO) Conference, 1998.
- Hou Jian-hua ,Chen Yue (2007). Research on visualization of the evolution of strategic management front . *Studies in Science of Science* 25(S1),15-21.
- Kretschmer H, Aguillo IF (2004). Visibility of collaboration on the Web. *Scientometrics* 61(3),405-426.
- Liu Zeyuan, Yin Lichun (2006). Visualization of International Science of Science Co-word Network . *Journal of the China Society for Scientific and Technical Information* 5(25), 634-640.
- McQuaid MJ, et al (1999). Multidimensional scaling for group memory visualization. *Decision Support Systems*, 27(1-2),163-176.
- Noyons, ECM, Moed HF, Van Raan AFJ (1999). Integrating research performance analysis and science mapping. *Scientometrics* , 46(3): 591-604.
- Price DJD (1961). Science since Babylon. New Haven: Yale University Press.
- Price DJD (1963). Little Science, Big Science. New York: Columbia University Press,1963.
- Price DJD (1965). Networks of Scientific Papers. *Science*, 149(3683):510-515.
- Small H (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science* 50(9): 799-813.
- White HD , McCain KW (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science* 49(4), 327-355.