Timeline and Landscape: A Case Study of Visualizing the Evolution of Science Communication Research Front

Ting Zhang[1] ZeYuan Liu[2] TaiYang Zhao[3]

19 June 2008

Abstract

Research Fronts are clusters of documents that tend to cite a fixed, time invariant set of base documents. In a specialty, research fronts are considered as major topics that most researchers are interested in. This paper plots the research fronts in science communication as timelines and landscapes for visualization. Timelines and landscapes, as new techniques of mapping knowledge domains, take advantage of a set of data processing and querying technologies based on Matlab and database prior to those based on text and memory. A set of data, collected from the reputable journal named Science Communication from 1994 to 2007 are used to manifest the dynamics and evolution of science communication research fronts over time. The mapping techniques are based on finding 'research fronts' consisting of groups of papers clustered using bibliographic coupling. Timelines of research fronts, two-dimension maps, show the specialty dynamics over time. Landscapes of research fronts are three-dimension plots with many peaks rising one after another over time. Landscapes show the panorama of research fronts in science communication, and comparison of various research fronts over time comes to be easier in fact. Such information will provide some references for technology forecasting and further knowledge development.

1 Introduction

Science communication, as one of social activities aims at spreading, diffusing and popularizing knowledge, method and spirit of science and technology. Science communication plays an important role in modern activities of science and technology. Meanwhile, the study on science communication attracts more and more attention in academia. And science communication is becoming a newly arising cross discipline. Science communication, namely scientific communication or scholar communication include science and technology communication, science and engineering communication (Liu, 2002).

In 1940s, American scholars started their communication science study on promoting agricultural technology pattern. After over 50-year long exploration, Science Communication has made a great progress in specialty research. With the advancement of modern science and technology, especially for the introduction of network and information techniques, a series of changes emerge in the domains, contents and measures of science communication. Thus, topics in science communication most of researchers are interested in are changing. Confirming research fronts over time in science communication and manifesting the rising and declining process of science communication research fronts will not

[1] WISE Lab., Dalian University of Technology, Dalian, People's Republic of China
[2] Institute of Science Studies and S&T Management in Dalian University of Technology
[3] Electronic and Information Engineering Institute of Dalian University of Technology

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

only provide the academic value for the disciplinary development and theory innovation in science communication, but also be useful for science communication to play an important role in promoting scientific quality of the public.

On account of the social responsibility and significance of science communication, scientists from home and abroad are actively exploring a set of mode and structure for science communication to fit for and support the rapid development of science and technology. Among them, some researchers use the Scientometrics method to study on science communication. Christin L. Bergman and Jonathan Furner (2003) from University of California in the USA have ever found the core journals in science communication with the Bibliometrics techniques and drawn the knowledge network of science communication.

Huang Dan(2005), a professor in Fudan University, a famous and key university in China plotted the knowledge map of communication in 2005. He tried to manifest the basic structure of communication science and structure the boarder and inner patterns through the knowledge map.

This paper explores the development trend of science communication by using the newest information visualization method. And it finds out which research fields play important roles in the development of science communication and discusses the "borrow" process of knowledge in science communication.

2 Data Resources and Techniques

The data in this paper are searched from the database of Social Sciences Citation Index owned by Institute for Scientific Information. The data were updated lately on Juan, 1st in 2007. 350 papers were collected from the journal of Science Communication from 1994 to 2007 with 8570 references. Many information can be read from the data including author's name, title, abstract, publication date, literature style, address and reference, etc.. SAGE publications make a specific introduction of Science Communication on its own website. The former name of the journal was Knowedge: Creation, Diffusion ,Utilization and changed to be Science Communication in 1994 (Zhai, 2002). Science Communication is defined as an interdisciplinary social science magazine, focusing on the academic and practical topics including the most active political and social problems. The journal discusses the hot topics in some specialties with no limit in different countries, culture and economy. The Scientometrics analysis results based on Science Communication can basically reflect the research situation in the field of the international science communication.

The techniques used in the paper are timeline and landscape. Timeline, defined time as X-axis, is a kind of visualization map manifesting the research topics' development trends. A function of research fronts with the variable of time is founded. The function was reflected on the plot through which the developing current of research fronts is easy to see. Small and Greenlee (1989) manifested the relations of research fronts using timelines on the case study of AIDS. Topics' variation in the research field of AIDS was shown on the timeline map through the co-citation analysis. Braam, Moed and Van Raan (1991) described the citation frequency of key literatures. This paper uses the techniques of timelines similar with those in the paper on anthrax research made by Steven Morris in 2003. Bibliographic coupling is used in the clustering analysis. This type of analysis produces groups of papers that tend to use the same base knowledge, as represented by the common references that they cite (Hou, Kretschmer & Liu, 2008). The data flow of the visualization process can be shown on Figure 1 below.

Using Matlab, a kind of technical computing software, we create landscape maps to visualize the research fronts of science communication. The three-dimension knowledge maps can be used in further comparing with the different peaks of research fronts. These peaks rising one

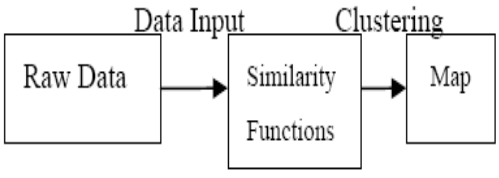

Figure 1. Visualization flow chart.

higher than another over time can manifest the research fronts' current very clearly. Three-dimension landscape, as one of new visualization techniques, ever was used by some foreign researchers. In 2001, Sandia National Laboratories in the National Department of Energy of the U.S.A took advantage of VxInsight Software to draw the three-dimension landscape on which science and technology in management field was visualized. Katy Borner, an Associate Professor of Information Science at the School of Library and Information Science, at Indiana University in the USA also used three-dimension map to plot the knowledge domains in 2003 (Borner&Chen, 2003). Visual Analysis Group in Pacific Northwest National Laboratory in 2004 ever found the hot topics of research fields through the landscape.

It is very easy for us to find the mutual relations of research fronts and the changing research topics over time through timeline and landscape. On these visualization knowledge maps, some messages can be shown as bellows for researchers:

- ·time when one research front appear and disappear
- ·basic references of research fronts
- ·core institutes and core authors in research fronts
- ·hierarchical structure of research fronts
- ·information flow in various specialties
- ·research fronts with outstanding peaks

An initial study on anthrax research was used by Morris, Yen, Wu and Asnake (2003) to show the use of bibliographic coupling to form research fronts of papers. A research front we defined as differs slightly from that of previous authors. Price (1965) considered a research front as a "growing tip, or epidermal layer" of current papers. Garfield (1994) defines a research front as "co-citation clusters and the documents that cite them". For Persson (1994), research fronts were defined as clusters of articles grouped by the co-citation clusters they cite. Chaomei Chen (2006) generalized and compared the definitions of research fronts in Scientometrics and he defined a research front as a group of abrupt dynamic conceptions and potential research problems. Actually, research fronts' exploration in specialties lies on the exploring techniques in Scientometrics to some extent. The meaning of research fronts is different for using various techniques. Bibliographic coupling is used in this paper to cluster papers. Bibliographic coupling (Jarneving, 2005), a dual concept with co-citation links documents that cite the same references with the similar meanings. The bibliographic coupling documents that cite the same groups of references continuously have the similar meanings. It means that the bibliographic coupling documents represent the research fronts and hot topics that most researchers pay attention to. As a result, research fronts in specialties can be shown by clustering in using bibliographic coupling.

Database Information Visualization and Analysis system or DIVA (Morris, 2002) is a computer program that helps to perform bibliometric analysis of collections of scientific literature and patents for technology forecasting. DIVA is run in the cases of Matlab and it is a kind of visualization software based on the database. Comparing to the previous visualization software based on documents and EMS memory, DIVA has superiorities in realizing the complex analysis process taking advantage of the existing database resources. The processed data volume of DIVA amounts to 1Giga byte. Comparing to other software in visualization, DIVA also quickens its processing speed. Because it is not a kind of commercial software designed open to the public, DIVA has the strong pertinence when processing data. To achieve the special research goal using DIVA, we developed the DIVA software for the second time on base of the original DIVA and some special functions are added to DIVA such as drawing three-dimension map. The developing techniques and tools include MS Acess, VBA,

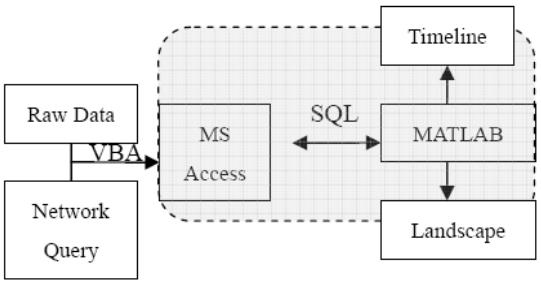

Figure 2. Software system chart.

SQL and Matlab. On Figure 2, the parts labeled with the dashed line is the core of the visualization processing software, of which MS Acess has the powerful function of database, Matlab has the functions of computing and three-dimension drawing. DIVA, collecting the advantages of various techniques and tools provides the powerful technological support for visualization analysis. The paper shows the visualization results in two kinds of knowledge visualization plots: (1) timeline, two-dimension map clustering using bibliographic coupling draws research fronts over time; (2) landscape, three-dimension map visualizes research fronts as range upon range of hills clustering using bibliographic coupling.

DIVA with updated functions is processing and analyzing method based on the database that can deal with literatures not only in English but also in Chinese. DIVA has superiorities in processing Chinese papers comparing to other information visualization software.

Timeline and landscape visualization techniques have some advantages summarized as below: (1) bibliographic coupling in research fronts of specialties; (2) Visualization analysis techniques based on the database deal with 1G-bye data; (3) using Matlab in functions of computing and drawing three-dimension map; (4) special research goal according to individual research need; (5) Chinese data processing becomes possible.

3 Timeline and Landscape Visualization Map of Research Fronts in Science Communication

The journal of Science Communication says scientific exploring is conducted in the fields of communication in scientific community, science communication to the public, science and technology communication policy. The topics for discussion involve health, education, international development, environment and risk management and so on. It is stressed that the content related with the science is almost always important including social science, education, international development, environment and risk management. We use knowledge domain mapping techniques of timeline and landscape to draw the research fronts of science communication and explore its development process as below.

The acquisition of papers published on Science Communication was conducted on June 1st, 2007, using ISI's Web of Science (WOS) product and 350 papers were acquired, limiting the search to papers available from WOS from 1994 to June of 2007. A paper to reference matrix consists of these 350 papers having links to 8570 references. Using bibliographic coupling, paper pairs with less than 3 coupling times were excluded. 105 papers are clustered into 10 research fronts. Figure 3 and 4 shows a research front timeline and landscape respectively for the science communication collection.

On Figure 3, X-axis presents time and Y-axis presents research fronts. Looking at the left of timeline, the identifying numbers for each research front are printed in a column to the right of the clustering dendrogram. The dendrogram shows the hiberarchy of research fronts. The papers in each research front are plotted by time in horizontal tracks, with the research front labels on the right side of the plot. Research front labels were found by manually searching the papers in each research front for themes.

The circles on the plot correspond to papers and the size of each circle is proportional to the number of times that the paper was cited. Papers were cited more times, the circles were larger with great influence in the research front. Each circle is shaded in proportion to the number of times its corresponding paper has been cited in the last year of the collection (June 2006 to June 2007). The shaded circle is the latest cited paper in the research front. On Figure 3, two shaded circles are situated in research front 3 and 6. Research front 3 is labeled as public understanding science and bioengineering communication. The largest shaded circle in research front 3 is the paper on public communication of science and technology in the 21st century written by Borchelt. Research front 6 has the largest shaded circle presenting the paper named biotechnology and the American media written by Nisbet. The process of information flow can be seen on the plot and the development discipline of research fronts is ladder-like (Morris & Yen, 2004). With the time

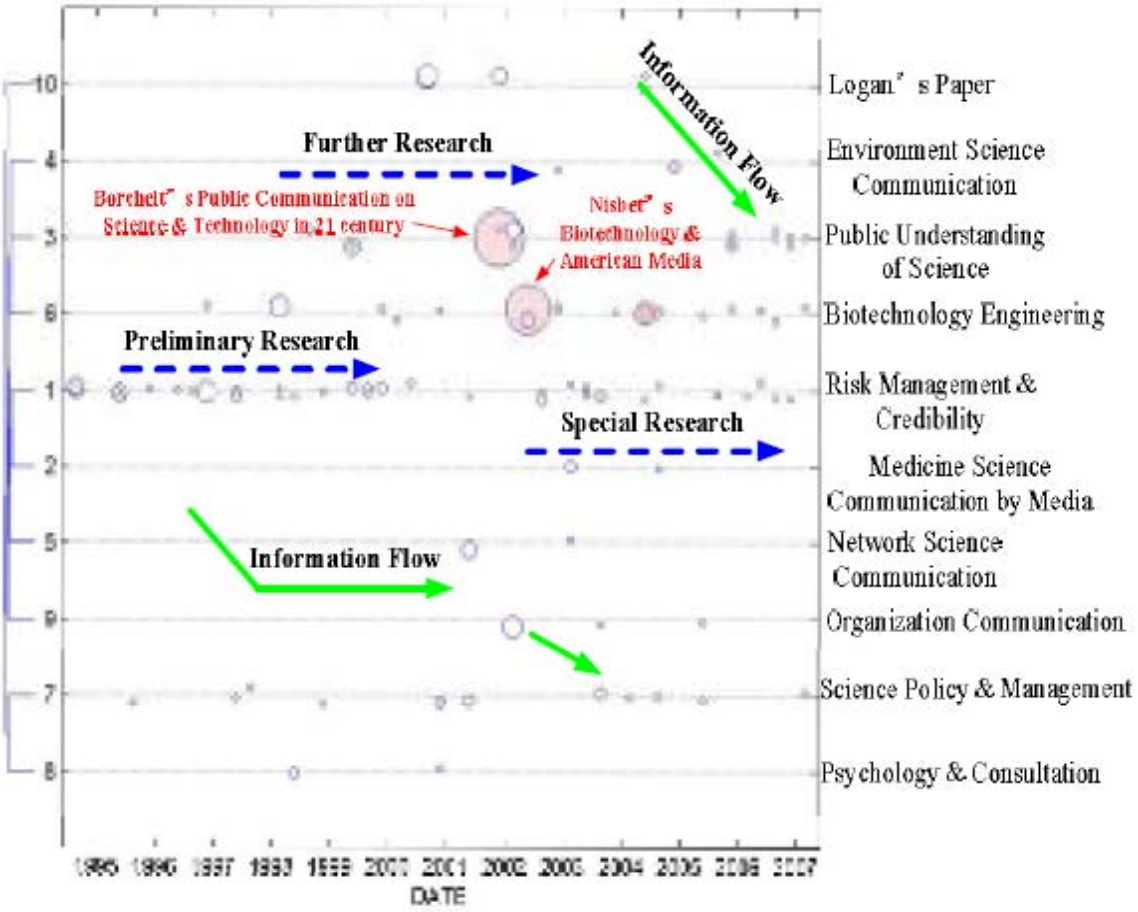

Figure 3. Timeline of Research fronts in Science Communication

goes by, the scientific information flow turns from the preliminary domain No.1 to the other domains. The information flow can give an impetus to the development of other scientific research fields. As a result, more hot topics have come into being.

It is easy to see three steps of the development of the science communication research on the timeline plot including the preliminary research, further research and special research. Most of the papers' topics focused on the research front No. 1 before the year of 2000. Research front No. 1 involves the topics of the credibility of science and risk communication. Papers in other research fronts haven't appeared in the preliminary research. After the year of 2000, more topics were discussed in the fields of the scientific policy and management, public understanding science, bioengineering communication and so on. The year of 2000 marks a new phase of science communication research. Some researchers began to set foot in various research fronts especially the public understanding science and bioengineering communication. With the rising-up of the research on scientific policy and management and network science communication, science communication entered into the second phase of the further research. In the preliminary research, the research hot topics were science credibility. And the hot topics changed during the second phase of the further research. In the third phase of the special research, various research fronts in multiple views appeared. During the period of the recent three years, science communication research enters into a special research phase, information flows in the professional domains. And the research on science communication is much

Table 1. Varying Trend of Research Fronts in Science Communication

Year	Research Front	Hiberarchy of Research Font
1994-2000	Science Credibility & Risk Communication	Preliminary Study
2000-2004	Public Understanding Science, Special Knowledge Communication	Further Study
2004-2007	Environmental Communication, Public Understanding Science, Special Knowledge Communication, Science Policy & Management	Special Study

more professionalized with increasing topics and views.

On Figure 3, until 2000, special scientific knowledge domains such as bioengineering science have overlapped with the research front No. 1. It means that since 2000, science communication in special knowledge domains have began to grow gradually. The research front of the public understanding science rose later with a lot of papers published by the end of 2001. Thus, the research fronts' dynamic changes in science communication is summarized as Chart one below. The dominant research fronts in science communication changed from those of science credibility and risk communication to public understanding science and special knowledge communication. Following that, research fronts in different views focusing on various topics in specialties have emerged.

Table 2. Ranking of Research Fronts in Science Communication by Paper Numbers

Number	Research Front	Paper	Year
1	Risk Communication & Credibility	37	1994-2006
3	Public Understanding Science & Effects	22	1998-2007
6	Bioengineering Communication	18	1996-2007
7	Science Policy & Management	11	1995-2007
10	Logan's Paper	5	2000-2004
4	Environmental Science Communication	3	2002、2004、2005
9	Organization Communication	3	2002-2005
2	Medical Communication by Mass Media	2	2003、2004
5	Network Science Communication	2	2001、2003
8	Psychology & Consultant	2	1998-2000

To better show the panorama of research fronts in science communication and further analyze the relations of different research fronts, we draw the landscape of research fronts in science communication using Matlab (Elzabeth Hetzler & Alan Turner, 2004). On Figure 4, Landscape visualization map is a three-dimension plot of research fronts in specialties. X, Y, Z-axis indicate time, research fronts and paper density. Mountains in the landscape represent papers in ten research fronts in science communication. These mountains in each research front are plotted by time in tracks parallel to X-axis with the same labels as timeline. Ten rows of these mountains are ten research fronts in science communication. On landscape, research fronts are labeled as numbers

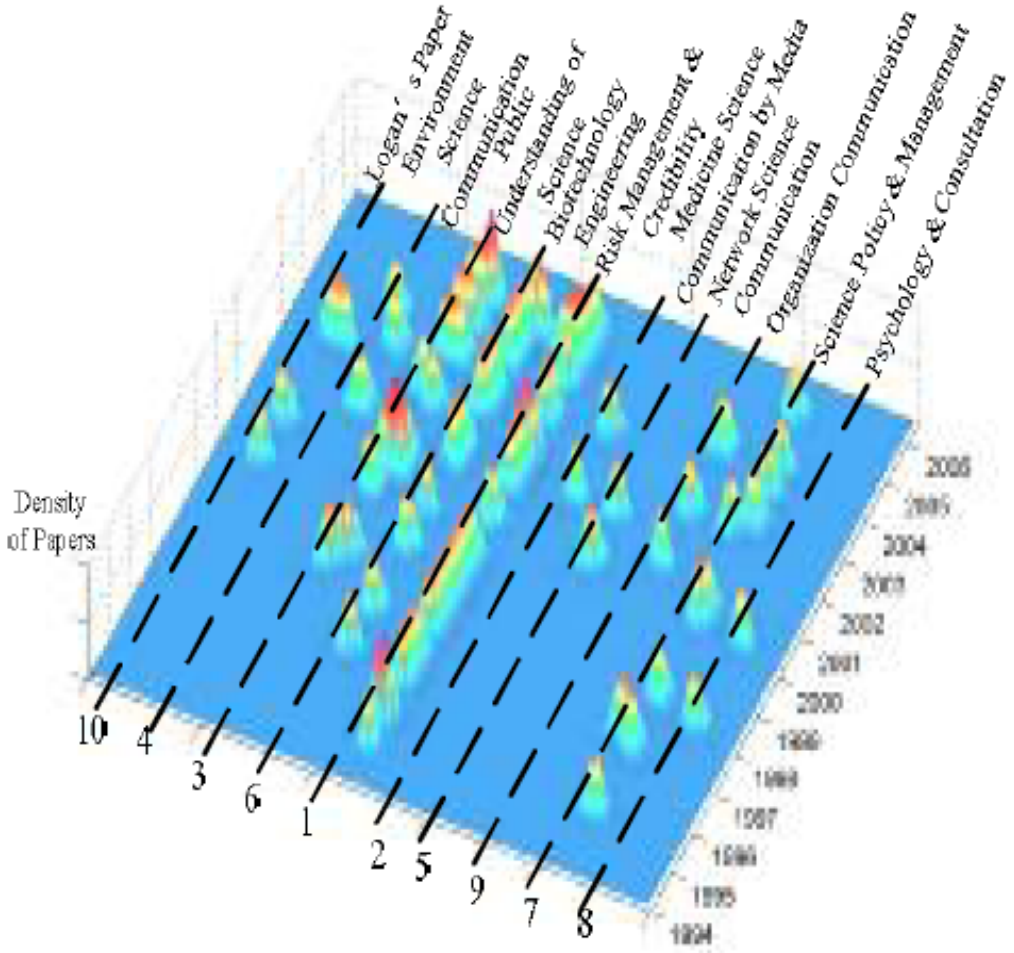

Figure 4. Visualization Map--Science Communication Research Front Landscape

that are corresponding to those in timeline. Ten research fronts are listed in the descending order according to the numbers of papers on Table 2. Research fronts having a lot of papers are easy to catch the researchers' eyes. Looking at Table 3, there are four research fronts with over ten papers, i.e., risk communication and credibility, public understanding science and effect, bioengineering communication and science policy and management. We defined these four research fronts as four main research fronts of science communication.

Among ten research fronts in science communication, different numbers of mountains are clustered in each research front. And each mountain indicates one research topic. On the landscape, the height of mountains indicates the density of papers. It means that the higher of mountain, the hotter topics of research fronts. Peaks of mountains rising one higher than another that indicate different research fronts consist of landscape visualization map.

And some topics that mountains represent are labeled on Figure 5. On the map, the mountains with the paper number of no less than 3 are defined as "hot topics" in science communication. Table 4 lists the ranking table of mountains according to the paper number. There are five "hot topics" in science communication, focusing on research front No.1 and No. 3, including those of credibility and social responsibility, science hierarchy and risk communication, the credibility of information source, scientific public and culture and public understanding science. These five topics have caught the researcher's eyes in science communication.

To manifest the main research fronts in science communication very clearly, we extract the first four research fronts with many papers among ten original research fronts. That is, the first four rows of papers in Table 2 are extracted to draw another landscape map of Figure 6 in science communication. On landscape of the main research fronts, four outstanding peaks are labeled as A, B, C and D respectively, of which A and B are located in the research front of public understanding and effects, C and D, in the research front of risk communication and credibility. As a result, among ten research fronts, the top two highest mountains of research fronts have maximal effects on the development of the specialty and attract most of researchers in science communication. Peaks of A, B, C and D are ranked as D, B, C and A by time on Table 5.

It is easy to find on the landscape that Peak D appeared in earliest time of 1995. Its topic focused on how to look on the credibility of the scientific information communication. The other hot topic the peak B represents appeared during the year of 2001 and 2002. Researchers are interested in how to conduct the science communication facing to the public and promoting the public understanding of science. And the study on science popularization was rated as the agenda of science communication agenda.

4 Conclusion

Timeline and Landscape are new techniques of visualizing the knowledge domains. The mapping knowledge domains' techniques take advantage of the powerful data processing and query function based on Database. On the base of original two-dimension visualization map, we develop for the second time to improve the visualization techniques. And three-dimension of landscape can manifest the dynamics of research fronts from multiple views. On case study of Science Communication, timeline and landscape indicate the development discipline and trend. It

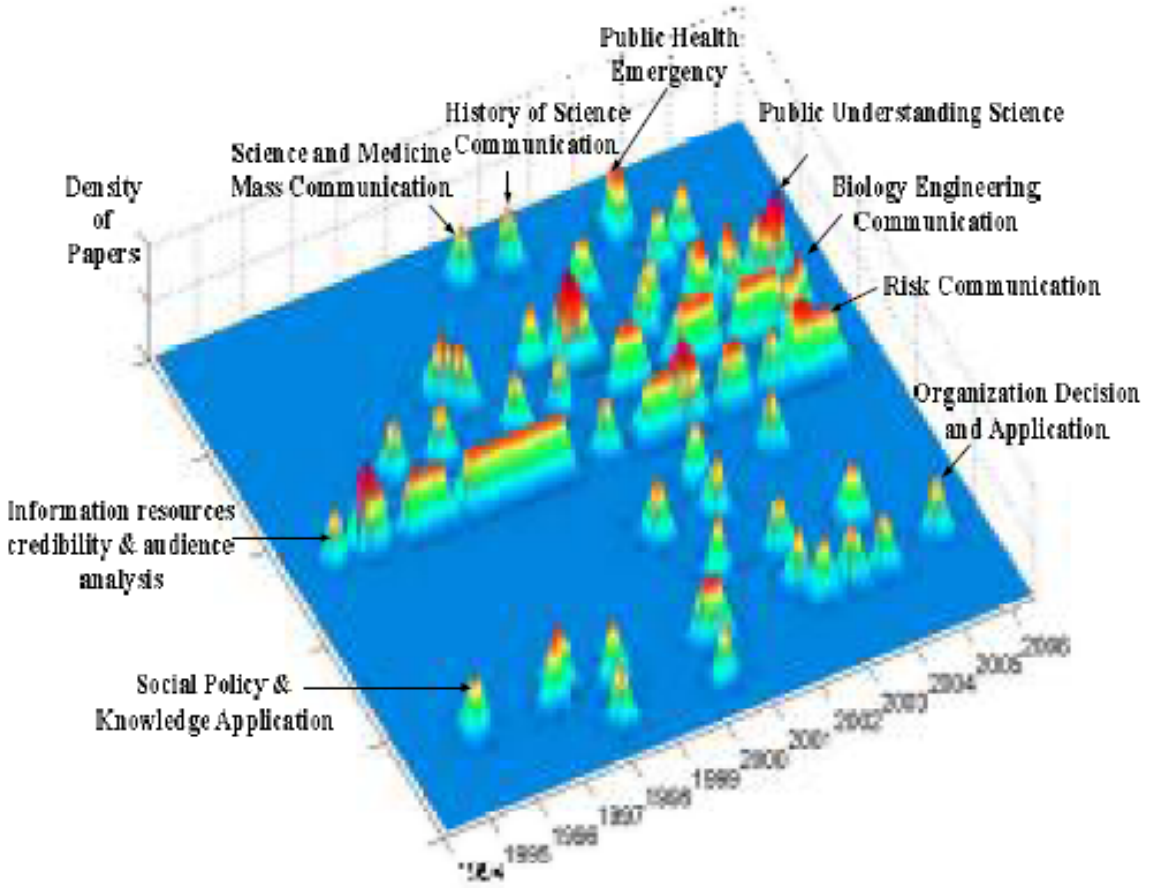

Figure 5. Landscape of Research Front in Science Communication

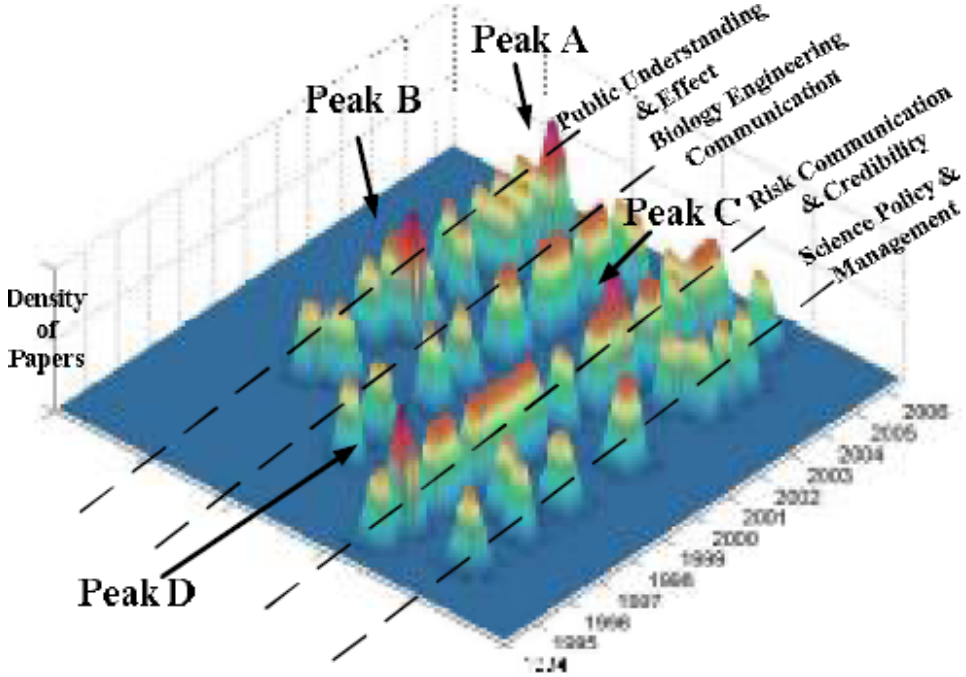

Figure 4. Landscape of Main Research Fronts in Science Communication

is easy to see which are the hot topics and helpful for further scientific forecasting. Timeline and landscape can realize different research aims according to the scholars' individual needs. The processing data volume increases and in near future, timeline and landscape can deal with data in different languages including Chinese because they are based on database not text or memory.

References

Berberich, K., S. Bedathur, M. Vazirgiannis, and G. Weikum (2006). Buzzrank … and the trend is your friend. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, New York, NY, USA, pp. 937–938. ACM Press.

Börner, K., C. Chen, and K. Boyack (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology* 37 (1), 179–255.

Chakrabarti, S. (2003). Mining the Web: discovering knowledge from hypertext data. Morgan Kaufmann.

Egghe, L. and R. Rousseau (2003). BRS-compactness in networks: Theoretical considerations related to cohesion in citation graphs, collaboration networks and the internet. *Mathematical and Computer Modelling* 37 (7–8), 879–899.

Folly, G., B. Hajtman, J. Nagy, and I. Ruff (1981). Some methodological problems in ranking scientists by citation analysis. *Scientometrics* 3 (2), 135–147.

Janssens, F., W. Glänzel, and B. De Moor

Table 3. Four main research fronts of Science Communication

Peak	Hot Topics	Year
D	Information Sources' Credibility & Audience Analysis	1995
B	Science Public and Culture	2001-2002
C	Credibility and Science Responsibility	2003-2004
A	Public Understanding Science	2007

(2007). A Hybrid Mapping of Information Science. In D. Torres-Salinas and H. F. Moed (Eds.), *Proceedings of ISSI 2007*, Volume 1, Madrid, pp. 408–420.

Katz, J. (2005). Scale-Independent Bibliometric Indicators. *Measurement: Interdisciplinary Research and Perspectives* 3 (1), 24–28.

Leskovec, J., J. Kleinberg, and C. Faloutsos (2006). Laws of Graph Evolution: Densification and Shrinking Diameters. *Arxiv preprint physics/0603229*.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16 (12), 317–323.

Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46 (5), 323–351.

Rao, I. (1980). The distribution of scientific productivity and social change. *Journal of the American Society for Information Science* 32, 111–122.

Vlachỳ, J. (1978). Frequency distributions of scientific performance: a bibliography of Lotka's law and related phenomena. *Scientometrics* 1 (1), 107–130.

Table 4. Hot topics of Science Communication

Research Front	Hot Topics	Numbers of Papers
Research Front No.1(Risk Communication & Credibility)	Credibility & Science Responsibility	5
	Science Hiberarchy & Risk Communication	4
Research Front No. 3 (Public Understanding Science & Effects)	Scientific Public and Culture	6
	Public Understanding Science	5

Table 5. Hot Topics of Outstanding Research Fronts in Science Cmmunication

Peak	Hot Topics	Year
D	Information Sources' Credibility & Audience Analysis	1995
B	Science Public and Culture	2001-2002
C	Credibility and Science Responsibility	2003-2004
A	Public Understanding Science	2007