BOSLA
Bombay Science Librarians' Association

सी डैक
CDAC

BEYOND LIBRARIANSHIP

Creativity, Innovation and Discovery

(BOSLA National Conference Proceedings)

16-17 September, 2011, Mumbai, India

Editors

Dr. B. S. Kademani

Dr. A. N. Bandi

Dr. Sangayya Sirurmath

Dr. Mallikarjun Angadi

Mr. I. C. Bandi

Ms. Truptee Shah

Ms. Savita Rao

Organised by

C-DAC, Mumbai

and

Bombay Science Librarians' Association (BOSLA), Mumbai

ISBN: 935050007-8

Co-citation Analysis: An Overview

Ganesh Surwase, Anil Sagar, B. S. Kademani and K. Bhanumurthy
Scientific Information Resource Division, Bhabha Atomic Research Centre, Trombay, Mumbai (India) - 400085. surwaseg@barc.gov.in

This article gives an overview of co-citation analysis and its applications in tracking the linkages among the intellectual works and mapping the evolutionary structure of scientific disciplines. It also focuses on the features, interface, terminology used, merits and demerits of co-citation based online database applications.

Keywords: Co-citation analysis, intellectual linkages, structure of science, SciVal Spotlight

1. INTRODUCTION

Today's advanced scientific knowledge is the cumulative effect of earlier research works. Each research paper reference/cite to the earlier research paper. In another way, each current paper is an extension of earlier research topic. Thus the intellectual linkages are established between the previous and current research topics. Knowledge is multidimensional, multifaceted, multidirectional and interdisciplinary in nature growing at a faster pace. It is essential to classify the knowledge to understand the growth and evolution in a better way. There are various methods of knowledge classification developed over a period of time. Co-citation analysis is one of the most effective and efficient methods of knowledge mapping. Many attempts have been made in the past manually to understand the structure of knowledge in a limited way. The advances in computer and information technology and developments of well-structured databases has paved way for co-citation applications.

Classification of research article based on the journal is too general or vague, because each scientific journal is classified into a major field despite the fact that journals are progressively covering a wider array of disciplines and paradigms. Co-citation analysis, involves classifying the scientific literature into small groups of papers that correspond to specific problems. Henry Small, known as father of co-citation analysis applied the technique to identify the strengths and weaknesses of an institution.

2. BIBLIOGRAPHIC COUPLING VS CO-CITATION COUPLING

The term 'bibliographic coupling' was first introduced by M.M. Kessler in 1963. Bibliographic coupling occurs when two works reference a common third work in their bibliographies. The bibliographic coupling strength increases with the increase in the number of references they have in common. Bibliographic coupling is used to extrapolate how similar the subject matter of the two works is. Bibliographic coupling is not subjectively valuable in all fields of research since it helps the researcher to find related research done in the past.

In 1973 Henry Small introduced the term concept of co-citation analysis, as a better indicator of subject similarity. Co-citation coupling is a method used to establish a subject similarity between two documents. The two documents are said to be co-cited when they both appear in the reference list of a third document. If papers A and B are both cited by paper C, they may be said to be related to one another, even though they don't directly cite each other. If papers A and B are both cited by many other papers, they have a stronger relationship. The more papers they are cited by, the stronger their relationship is. The co-citation frequency is defined as the frequency with which two documents are cited together. Figure-1 illustrates the concept of bibliographic coupling and co-citation coupling.

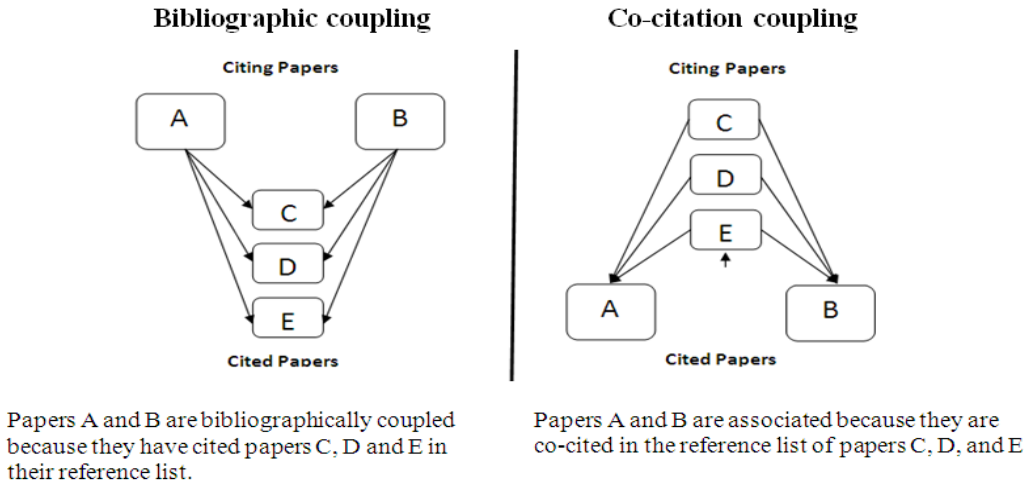


Figure-1: Bibliographic coupling and co-citation coupling

3. CO-CITATION ANALYSIS

In bibliometrics, there have been extensive studies on the assessment of document similarities. For instance, Kessler proposed the approach of bibliographic coupling, and Small proposed the co-citation approach. The purpose for the assessment of document similarities is to classify documents. The classified levels include the document itself, its author, and the journal that contains the document.

Documents co-citation is used to conduct searches on similar documents. Journal co-citation is of interest to the collection manager concerned with developing core journal lists, selecting journals and evaluating collections that serve particular research-oriented constituencies. Author co-citation analysis has been used in analyzing the intellectual structure of scientific disciplines.

Co-citation analysis is a unique method for studying the cognitive structure of science. Co-citation analysis involves tracking pairs of papers that are cited together in the source articles. When the same pairs of papers are co-cited by many authors, clusters of research begin to form. The co-cited papers in these clusters tend to share some common theme. Combined with single-link clustering and multidimensional scaling techniques, co-citation analysis can literally map the structure of specialized research areas as well as science as a whole.

Co-citation analysis creates paradigms/ clusters to reveal interdisciplinary research trends within institutions. The ideas of co-citation have been applied to commercial application such as *SciVal Spotlight* etc. It is one of the most effective methods of monitoring the development and structure of science.

4. APPLICATION OF CO-CITATION ANALYSIS - *SCIVAL SPOTLIGHT*

SciVal Spotlight is a web-based application brought out by Elsevier Publishers. It is based on the technique of co-citation analysis. In *SciVal Spotlight* the articles cited for a year are clustered into two groups:

1. Highly cited reference articles for the map year are selected and clustered based on co-citation counts.
2. Current articles are assigned to the clusters based on their references. These current articles can be fractionally assigned to the clusters based on the distribution of their references.

Co-citation clusters are thus comprised of reference articles and the current articles that cite them. Each cluster is assigned to a discipline or subject area. It uses ccommercial online bibliographic database (*Scopus*) to retrieve the co-citation frequency matrix.

Steps involved in the generating the results and maps by *SciVal Spotlight* are as follows:

1: **Selecting the publications:** *SciVal Spotlight* uses one year of publications in order to pick up current trends.

2: **Selecting the references:** *SciVal Spotlight's* threshold for highly cited references is 5 to 100 citations, and threshold for papers those published 4 years ago is age of paper + 1 citation.

3: **Calculating the relatedness between these references:** generate cocitation clusters by using a modified cosine index based on co-citation counts for similarity, and run the resulting matrix of cosine values through a visualization program (DrL) which would assign each reference paper an (x, y) position on a 2-D plane. Visualization program 'DrL' is a force directed placement algorithm with edge cutting, it acts as a filter that reduces the inherent noise in the system

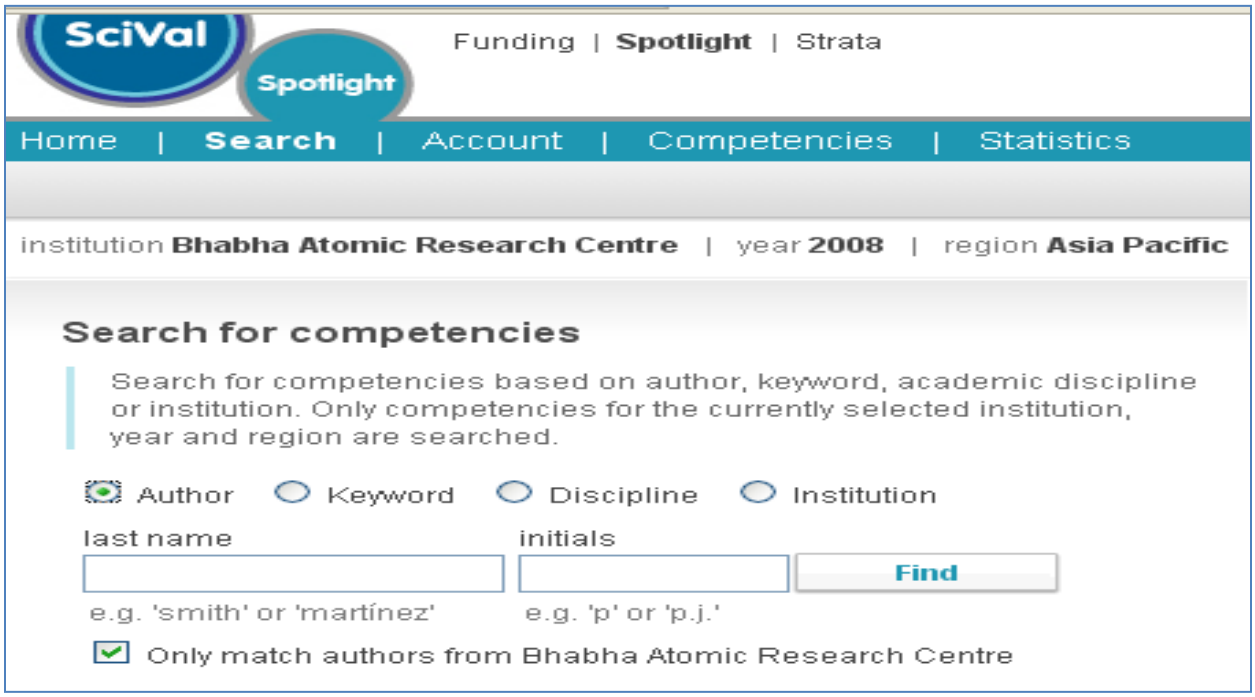
4: **Clustering the references:** SciVal Spotlight's average-link clustering algorithm is used to cluster the (x, y) positional data output of the DrL layout algorithm.

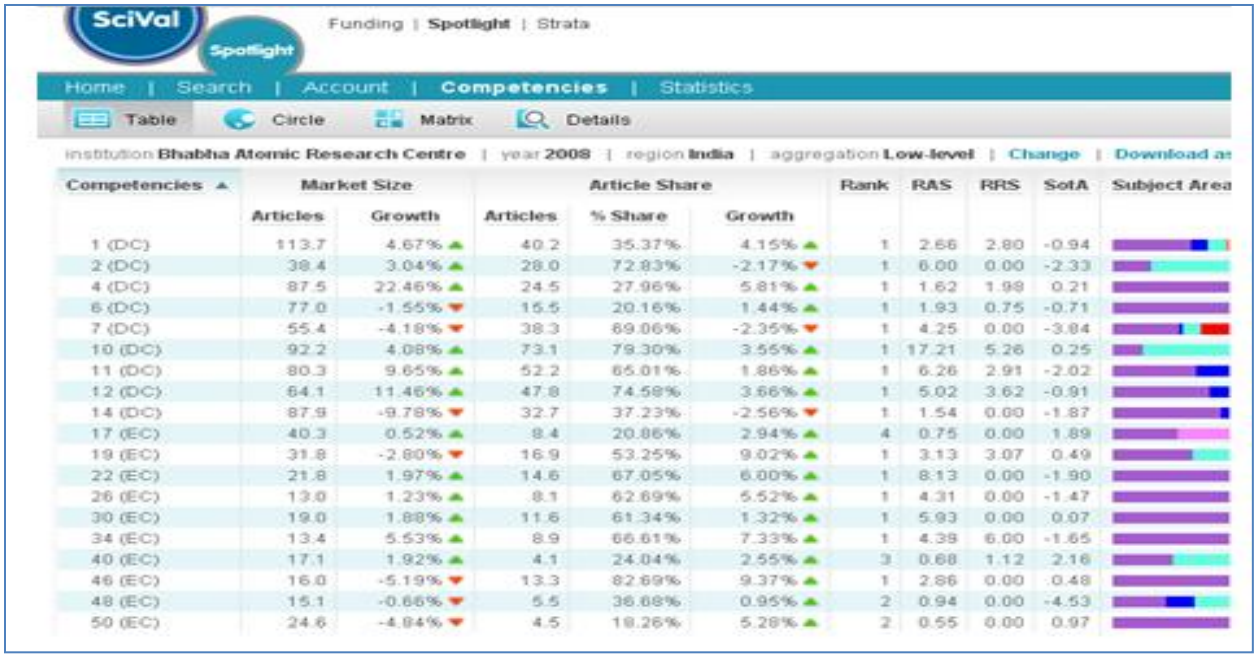
5: **Assigning current papers to these clusters**: Each current paper is assigned fractionally to clusters based on the references in the paper. *SciVal Spotlight* also assigns the four previous years of current papers to the reference structure identified above. This allows one to identify publications trends associated with a specific set of highly cited references.

In *SciVal Spotlight* the proprietary algorithm is run on the data produced by co-citation counts and the result, graphs and maps are displayed.

5. INTERFACE OF *SCIVAL SPOTLIGHT*

The screenshots of various functions of *SciVal Spotlight* are given below:





Competencies	Market Size		Article Share			Rank	RAS	RRS	SotA	Subject Area
	Articles	Growth	Articles	% Share	Growth					
1 (DC)	113.7	4.67% ▲	40.2	35.37%	4.15% ▲	1	2.66	2.80	-0.94	
2 (DC)	38.4	3.04% ▲	28.0	72.83%	-2.17% ▼	1	6.00	0.00	-2.33	
4 (DC)	87.5	22.46% ▲	24.5	27.96%	5.81% ▲	1	1.62	1.98	0.21	
6 (DC)	77.0	-1.55% ▼	15.5	20.16%	1.44% ▲	1	1.93	0.75	-0.71	
7 (DC)	55.4	-4.18% ▼	38.3	69.06%	-2.35% ▼	1	4.25	0.00	-3.84	
10 (DC)	92.2	4.08% ▲	73.1	79.30%	3.55% ▲	1	17.21	5.26	0.25	
11 (DC)	80.3	9.65% ▲	52.2	65.01%	1.86% ▲	1	6.26	2.91	-2.02	
12 (DC)	64.1	11.46% ▲	47.8	74.58%	3.66% ▲	1	5.02	3.62	-0.91	
14 (DC)	87.9	-9.78% ▼	32.7	37.23%	-2.56% ▼	1	1.54	0.00	-1.87	
17 (EC)	40.3	0.52% ▲	8.4	20.86%	2.94% ▲	4	0.75	0.00	1.89	
19 (EC)	31.8	-2.80% ▼	16.9	53.25%	9.02% ▲	1	3.13	3.07	0.49	
22 (EC)	21.8	1.97% ▲	14.6	67.05%	6.00% ▲	1	8.13	0.00	-1.90	
26 (EC)	13.0	1.23% ▲	8.1	62.69%	5.52% ▲	1	4.31	0.00	-1.47	
30 (EC)	19.0	1.88% ▲	11.6	61.34%	1.32% ▲	1	5.93	0.00	0.07	
34 (EC)	13.4	5.53% ▲	8.9	66.61%	7.33% ▲	1	4.39	6.00	-1.65	
40 (EC)	17.1	1.92% ▲	4.1	24.04%	2.55% ▲	3	0.68	1.12	2.16	
46 (EC)	16.0	-5.19% ▼	13.3	82.69%	9.37% ▲	1	2.86	0.00	0.48	
48 (EC)	15.1	-0.66% ▼	5.5	36.68%	0.95% ▲	2	0.94	0.00	-4.53	
50 (EC)	24.6	-4.84% ▼	4.5	18.26%	5.28% ▲	2	0.55	0.00	0.97	

6. TERMINOLOGY USED IN *SCIVAL SPOTLIGHT*

Competencies: A competency is a network of article clusters that are grouped together. If an institute is #1 or #2 in either RAS, RRS or SotA, it forms competency.

- *Distinctive Competency (DC)*: It is the subject area where an institute has a large relative market share compared to peers active in that same subject area. In order for a competency to become a distinctive competency market size shall be ≥ 200articles and the institute must be leader in either publication, reference or state of the art.

- *Emerging Competency (EC)*: It is the competency where your institute is not number 1 but it hold the potential to be #1.

Highly Cited References: Threshold for highly cited references is5 to 100 citations for the articles older than 5 years and article's age + a citation for the articles not older than 5 years. Maximum 100 citations are considered to avoid the aggregation effect.

Market Size: The total number of articles in each competency, not just from an Institute but from all institutes in that region which contribute to this competency. The values are fractionalized, meaning that only the fraction of each article that falls within the competency is counted.

***Growth Rate*:** It is nothing but the average yearly change in the number of articles within each competency from all institutions in the given region. The average growth is calculated from pairs of subsequent years. The two year part is formed by adding two subsequent years, then slide to the next year and calculate the growth.

***Relative Article Share (RAS*):** RAS means how many articles within each competency are from Institute relative to the institution ranked #1 in that competency. If Institute is ranked #1, the value is compared to the institution ranked #2.

***Relative Reference Share (RRS*):** Relative Reference Share (RRS): The number of highly cited reference articles that are authored Institute divided by that of the institute ranked #1 in that competency. If Institute is ranked #1, the value is compared to the institution ranked #2.

State of the Art (SotA): The recency of the work cited by the articles from an institute within each competency, relative to the average recency for that competency. Positive values indicate the work being cited is more recent than average. Example: If SotA value for Institute is 5, it means references cited by Institute are 5 years recent than the average recency of the competency.

***Subject Area*:** It is a graphical representation of the primary subject areas of each article cluster within the competency. There are 13 primary subject areas comprised of 566 minor subject categories. High-level competencies are formed by merging low level competencies for each of the primary subject areas.

***Rank*:** The rank of an Institute within each competency, calculated over the entire 5-year period. Relative Article Share (RAS) determines the rank of an institution in the competency.

7. MERITS OF *SCIVAL SPOTLIGHT*

This co-citation based web application '*SciVal Spotlight'* map the knowledge at interdisciplinary levels. It helps to accommodate scarce resources and reduce organisational in-efficiencies. Also helps to identify specific areas of research excellence, emerging competencies. Identify potential collaborating authors/ institutions. It enables to find out the leading researchers/experts/peers in various fields. One can compare the results with other competing institutions/authors. It helps to review competitive standings of an institute on a global scale.

8. DEMERITS/ LIMITATIONS OF *SCIVAL SPOTLIGHT*

SciVal Spotlight is database dependent and it carries forward the shortcomings and limitations of source (*Scopus*) database like literature coverage, input time lag, variations in name of author, institute etc. The maps generated are not clearly understandable (less expressive). In this proprietary application, data is hidden and just results are shown. Categorisation and numbering of competencies is not in a specific order. Average recency (RAS) in terms of number of years, is relative i.e. less interpretative. With the help of *SciVal Spotlight* the publication list for an author/ institution can not be generated.

9. CONCLUSION

Co-citation analysis based applications encourage researchers to explore the intellectual structure of science and guide them about the future developments. Intellectual linkages are established through the process of referencing and citation. It is a research tool for examining the intellectual development and structure of the scientific discipline.

REFERENCES

1. *Co-Citation Analysis: The Methodology of SciVal Spotlight* http://www.americalatina.elsevier.com/sul/pt-br/scival/pdf/Co-Citation_Analysis_SciVal_Spotlight.pdf
2. Culnan, Mary J. 1987. Mapping the intellectual structure of MIS, 1980-1985: A co-citation analysis. *MIS Quarterly*, 11(3), pp. 341-353.
3. Eom, Sean B. 1996. Mapping the intellectual structure of research in decision support systems through author co-citation analysis (1971-1993). *Decision Support Systems*, 16(4), pp. 315-338.
4. Kessler, M. M. 1963. Bibliographic coupling between scientific papers. *American Documentation*, 14(1), pp. 10-25.
5. McCain, K. W. 1986. Co-cited author mapping as a valid representation of intellectual structure. *Journal of the American Society for Information Science*, 37(3), pp. 111-122.
6. McCain, K. W. 1990. Mapping authors in intellectual space: a technical overview. *Journal of the American Society for Information Science*, 41(6), pp. 433-443.
7. Small, H. 1973. Cocitation in scientific literature: New measure of relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.
8. Small, H. 1999.Visualizing science by citation mapping. *Journal of the American Society for Information Science, 50*(9), 799-813.
9. Small, H. 2003. Paradigms, citations and maps of science: A personal history. *Journal of the American Society for Information Science and Technology, 54*(5), 394-399.
10. White, H. D. and Griffith, B. C. 1981. Author co-citation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, pp. 163-171.