Scientometrics (2015) 102:929–955
DOI 10.1007/s11192-014-1402-8

Mapping the intellectual structure of scientometrics: a co-word analysis of the journal Scientometrics (2005–2010)

S. Ravikumar · Ashutosh Agrahari · S. N. Singh

Received: 2 April 2014 / Published online: 4 October 2014

Abstract 959 full text articles has been studied to explore the intellectual structure of scientometrics in the period 2005–2010 using text mining and co-word analysis. The trends and patterns of scientometrics in the journal Scientometrics were revealed by measuring the association strength of selected keywords which represent the produced concept and idea in the field of scientometrics. All articles were collected from the journal Scientometrics through Springerlink (full text database) and keywords were added non-parametrically from the LISA database and the articles themselves (keywords provided by author). Other important keywords are extracted from the title and abstract of the article manually. These keywords are standardized using a vocabulary tool. With the objective of delineating dynamic changes of the field of scientometrics, the period 2005–2010 was studied and further divided into two consecutive periods: 2005–2007 and 2008–2010. The results show that publication has some well-established topics which are changing gradually to adopt new themes.

Keywords Co-word analysis · Scientometrics · Research trend

Introduction

Identifying most highlighting and emerging themes of any journal is indispensable for intellectual and policy reasons in science and social sciences. In the era of computer technology and automation, co-word analysis has become easier to extract knowledge from the text: research papers, conference papers, articles of the newspaper, and chapters of the book etc. Co-word analysis is one of the methods that use to identify themes and

S. Ravikumar (✉)
North Eastern Hill University, Shillong, India
e-mail: ravikumar@nehu.ac.in

A. Agrahari · S. N. Singh
Mizoram University, Aizawl, Mizoram, India

relationship among these. It is related to co-citation analysis (Small 1973; Small and Griffith 1974). It deals directly with set of terms shared by documents instead of shared citations. It counts and analyzes the co-occurrences of key-words in the publication on given subject. In addition, it has potential to describe interactions, which exist between different phases of innovation process and to show whether basic research or applied research is the moving force (Callon et al. 1991).

Co-word analysis is a technique that uses pattern of co-occurrence of words and phrases in a corpus. It establishes relation between idea and concept within the subject area, presented in corpus. Occurring of two keywords within the same paper indicates a relationship between the topics to which they refer (Cambrosio et al. 1993). The availability of many co-occurrences with a keywords or phrase indicates to central point which has many connections with other words in a corpus that may similar to a research theme. It identifies the co-occurrence strength of terms and creates a set of lexical graph that effectively illustrates the strongest association between various terms.

In this study an attempt has been made to trace the intellectual structure which developed under the umbrella of Scientometrics over the year 2005–2010. Scientometrics is an international journal, launched in 1978. The journal covers all aspects of scientometrics and published 46.31 % of scientometrics research paper of the world (Mooghali et al. 2011). The whole period (year 2005–2010) has been divided into two phases i.e. 2005–2007 and 2008–2010. Nine hundred and fifteen articles have been analyzed using co-word analysis, falling under the framed period. MDS has been used to plot and identify proximity between keywords. Salton index has been used to draw network and understand the relationship between keywords. Different characteristics of the same have been calculated using inner-link, inner-link key, outer-link, outer link-key, centrality and density and using these characteristics, developed and emerging themes been identified.

Literature review

Co-word analysis has been used by many researchers to explore conceptual network in different discipline such as Management information systems (Culnan 1986), Software engineering (Coulter et al. 1998), Polymer chemistry (Callon et al. 1991), Environmental acidification research (Law and Whittaker 1992; Law et al. 1988), Scientometrics (Courtial 1994), Neural network research (Van Raan and Tijssen 1993; Noyons and van Raan 1998), biological safety (Cambrosio et al. 1993), Patents (Courtial et al. 1993), Optomechatronics (Noyons and van Raan 1994), Bioelectronics (Hinze 1994), Medicine (Rikken et al. 1995), Biology (Rip and Courtial 1984; Looze and Lemarie 1997), Condensed matter physics (Bhattacharya and Basu 1998).

Currently, mature visualization skills of co-word analysis have been applied in many subjects and disciplines, such as Information retrieval (Ding et al. 2001), Nanotechnology (Kostoff et al. 2006), International scientific studies (Hou et al. 2006), Human genome (Musgrove et al. 2003) and Medical informatics (Wagner and Leydesdorff 2005), Management science (Yue 2012), Iranian nanoscience and technology (Mohammadi 2012) and Knowledge management (Ponzi 2002; Hou et al. 2006; Sedighi and Jalalimanesh 2014).

Methodology

This study has been made on the journal Scientometrics which fall under the framed period. A total of 959 papers were retrieved from the journal Scientometrics through ‘Springer Link’

database from the same period. A total of 44 papers are excluded as they don't come into category of articles and belong to reviews, editorials, letters, short communication etc. Finally, 915 articles are selected for co-word analysis. In this study, 'R' software (R Core Team 2012) has been used. It is open source statistics analysis software, freely available on internet.

Steps of co-word analysis

Co-word analysis helps us to structure the data at various levels of analysis: as networks of links and nodes; as distributions of interacting networks and as transformation of networks over time periods (Ding et al. 2001). It reduces and presented data into a visual representation which explain essential information of contained in data. Step wise procedures of analysis are mentioned below as it is shown in Fig. 1:

Data collection

Keyword plays a very important role in co-word analysis. There are two ways to extract keywords form the targeted sample (i.e. journals articles, conference papers, reports, chapters of book etc.): Non-parametric and Parametric. Non- Parametric deals with manual efforts of collecting keywords given by author, journal database, abstract database and citation database. Some important keywords can be extracted from abstract and title of the document. Coulter et al. (1998) selected keywords for their study which were added by indexer and selected form article and then standardized to remove the different variants form of keywords. Courtial (1994) used keywords given by authors. Coulter et al. (1998) chose descriptor provided by GUIDE database as keyword. Looze and Lemarie (1997)

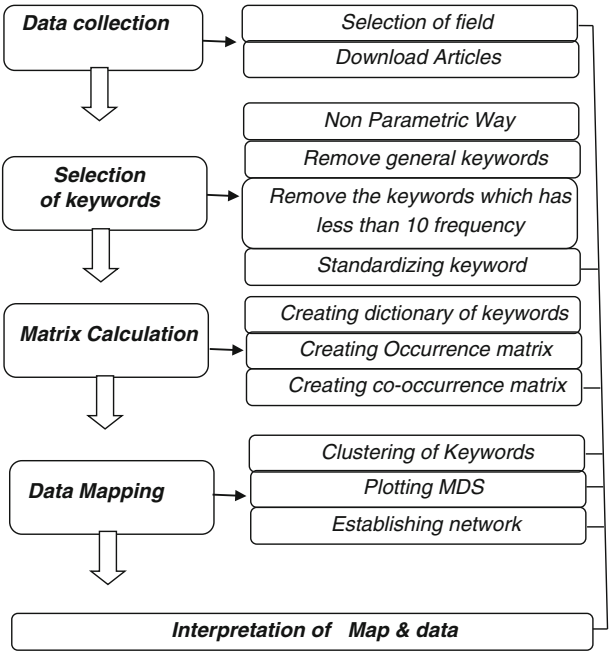

Fig. 1 Step of co-word analysis used in Scientometrics

used keywords suggested by expert of the subject. Nyams and Van Ram (1998) conducted a study using co-occurrence of classification code.

Another method uses software to extract keywords from the text such as NP tools, Ti.exe, GenEx. These softwares use an algorithm or text rank to select keywords which represent core topics of the specific subject. Mihalcea and Tarau (2004) examine different algorithm and developed a text rank- graph based on ranking model for key word extraction from full text. This method of keywords extraction eliminates the biasness of indexer and building thesauri and classification system. Ding et al. (2001) used above mention methods for their study.

Non-parametric way of key extraction has been used in the study, we have accepted not only all keywords added by author to the article but have also collected keywords provided by LISA to the same articles and some keywords extracted non parametrically from the articles' titles and abstract. Thus total 889 raw keywords were collected from 915 articles. These keywords are standardized (standardizing process given bellow under 4.2 sub-heading) and 240 unique keywords find for study. The average number of selected keywords per article is found 16.07. The range of selected keywords for each article which found out of 240 keywords varies from 1 to 51. Approx 5.84 % articles have more than 30 keywords in which 2 papers have 51 keywords, 21.74 % articles have 21–30 keywords and 45.57 % articles have 11–20 keywords. Thus 26.85 % articles have up to 10 keywords while 73.15 % articles use more than 10 keywords.

Data standardizing

After selection of keywords, these should be standardized by vocabulary tool (Ding et al. 2001; Yue 2012) as some related concept is presented by different words. Standardizations of keywords remove all synonyms, ambiguity, general term (a term which occurred mostly in a subject) and different variants form of word.

All keywords which are chosen for this study from the titles and abstract are standardized with help of the LCSH, SLSH and Bibliometric Dictionary, in order to remove the variants form of keywords such as singular/plural and synonyms words.

Keywords representing the same concept have been clubbed into standardized form. Words having low frequency are merged into broader term. Words not having broader or similar term or which are not specific or which are common words, have been ignored (Ding et al. 2001) such as author, citation, bibliometric etc. Using said selection procedure, different variants form of keywords such as singular/plural and synonyms words have been eliminated. After multiple corrections, 240 keywords are chosen as a research sample for co-word analysis; and each of them has minimum ten frequencies.

Matrix calculation

After selection of 240 keywords, a document term matrix is prepared on the basis of key word occurrence in Corpus. With the help of 'R' software, the document term matrix transformed into co-occurrence matrix (symmetrical matrix) of 240 keywords and the diagonal value of matrix is treated as missing data.

Data mapping

Co-word analysis can be done by different methods; Many scholar uses software: LEXIMAPPE (Law and Whittaker 1992; Cambrosio et al. 1993; Courtial 1994; Looze and Lemarie 1997); CAIR: Content Analysis and Information Retrieval, (Coulter et al. 1998);

BibTechMon (Bibliometric Technology Monitoring) (Kopcsa and Schiebel 1998); Kohnen's neural network algorthim (Polanco et al. 1998).

Whereas, the most commonly used methods are multidimensional scaling and clustering techniques. Ding et al. (2001) applied this method using SPSS software. This study has utilized hierarchical clustering, multidimensional scaling and social network map to visualize the co-word structures in Scientometrics.

Clustering

Clustering is a method of grouping objects into a set on the basis of similarity and dissimilarity. Single linkage is a widely used clustering technique but this technique limiting one cluster to 10 co-words only (Lee and Jeong 2008). Leydesdorff and Zaal (1988) make a study and suggested that 'ward' mode of analysis is better suited for symmetrical metrics than single linkage clustering because of the large number of zero hits which may lead to 'chaining' in the first cluster and isolates. This dissertation has used 'Complete' method of clustering and five clusters of the keywords have been made for co-word analysis.

MDS and Correlation

Following steps have been followed in order to plot two dimensional MDS map for each cluster

(a) Pearson's r calculated on the basis of co-occurrence matrix of keywords.
(b) With the help of 'R' statistical software MDS was applied to all keywords of each cluster.
(c) Five MDS map are generated for a segment of the period,
(d) Generated map displays relationship among the keywords of a cluster.

Similarity between two words (Pearson correlation coefficient) is calculated on the basis of all co-occurrence frequency that two words have in the cluster and this similarity matrix has been used for plotting MDS. Pearson's r is used to identify the relation among the keywords (Leydesdorff and Zaal 1988; An and Wu 2011). Pearson's coefficient calculates a value which indicates similarity in pair of keywords. Pearson's value lies between +1 and −1. Positive value indicates similarity between the keywords, whereas −ve value shows dissimilarity. Dissimilarity causes problem during measurement of the proximity among the keywords, this problem can be solved by linear transformation of $(r + 1)/2$. Linear transformation would transform the value between 0 and 1 (Leydesdorff and Vaughan 2006) and hence, better analysis can be done. The words with high Pearson correlation coefficient are located together in the MDS map, and those words located together in the map have high similarity in terms of co-occurrence profile within the whole matrix.

Network of keywords

Thin lines over MDS map between keywords form a network which correspond to a social network. Linkage among keywords presents the co-occurrence strength. Co-occurrence strength takes value of Salton index. Salton value can be calculated using following formula (Yue 2012):

$$Salton = C_{ij}/\sqrt{N_i N_j}$$

here, C_{ij} = Co-occurrence frequency of i and j keywords; N_i = Frequency of 'i' keyword; N_j = Frequency of 'j' keyword; Salton value lies between 0 and 1.

In this section, link strength among keywords of the network map is defined as Salton >0.2 (Ding et al. 2001; Yue 2012).

Result

In order to understand the overall co-word analysis, co-occurrence of keywords were analyzed during whole period (2005–2010). Whole period has been divided into two segments i.e. year 2005–2007 and 2008–2010, so that dynamic changes during these two periods can be observed.

Co-word analysis (year 2005–2010)

General overview structure

In order to understand the positional concept (relative position) of different clusters of scientometric field during 2005–2010, a general overview structure were generated with the help of MDS. Each cluster is selected as an input variable. As shown in Fig. 2, each cluster is represented by five keywords having highest frequency in the cluster. Position of a cluster (sub-domain) in the graph depicts the relation of its keyword with other clusters. Strongest linkage between clusters is shown by network developed by thin lines, which has value >0.35 (according to Salton index).

Detail co-word structure

In order to plot detail structure of each cluster, keywords are selected as input variable to map the sub-domain based on correlation matrix of 240 keywords. Detailed co-word structures (Figs. 3, 4, 5, 6, 7) are plotted to visualize specific characteristics of each sub-domain (clusters) in the field of scientometrics. The thin lines represent the link between

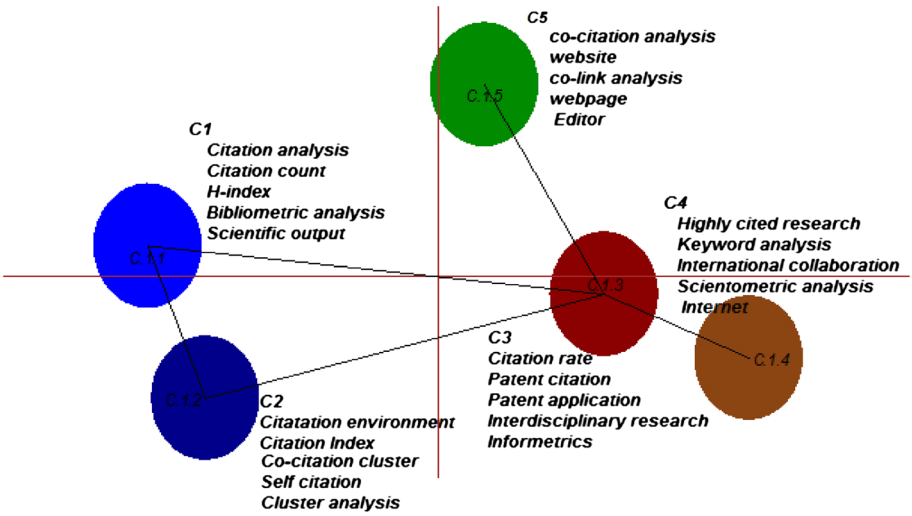

Fig. 2 General overview structure of Scientometrics in 2005–2010

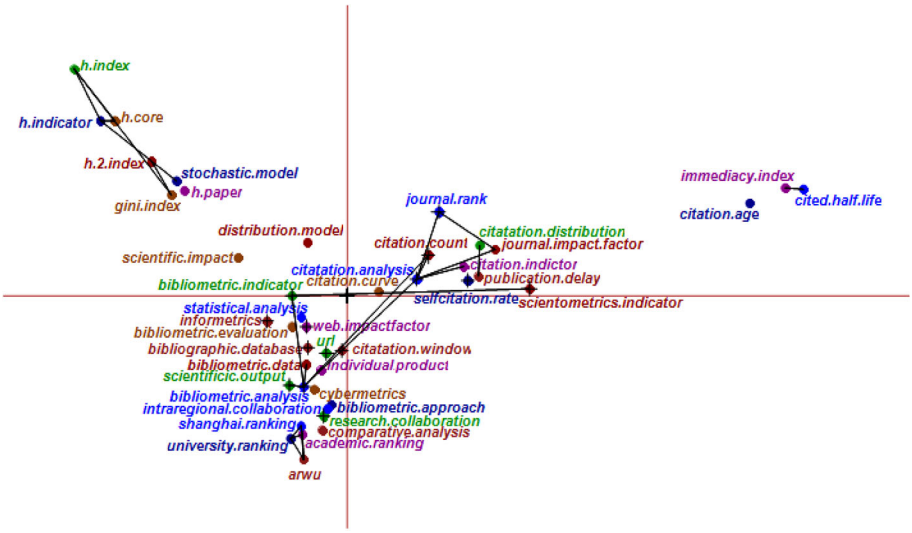

Fig. 3 Detail co-word structure of cluster 1 (C1) in 2005–2010

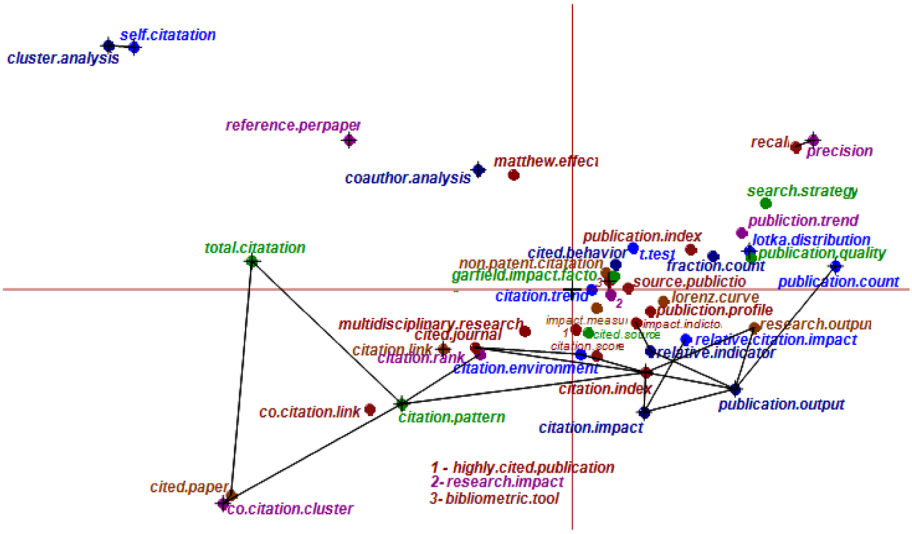

Fig. 4 Detail co-word structure of cluster 2 (C2) in 2005–2010

two keywords with the Salton Index (>0.2). The keywords having crass mark represent to outer link key in detail co-word structure map.

Custer 1 includes 43 keywords with higher frequency like Citation analysis, Scientific output, University ranking and Case study. Cluster 2 has 45 keywords like Cluster analysis, Self citation, Publication output, citation pattern, Search strategy, Statistical analysis. Cluster 3 includes 54 keywords as Scientometric analysis, Keyword analysis, Collaboration, Authorship, Bibliometric law, etc. Cluster 4 includes 51 keywords related to Interdisciplinary research, Power law, Centrality, Informetrics, and Patent citation analysis and

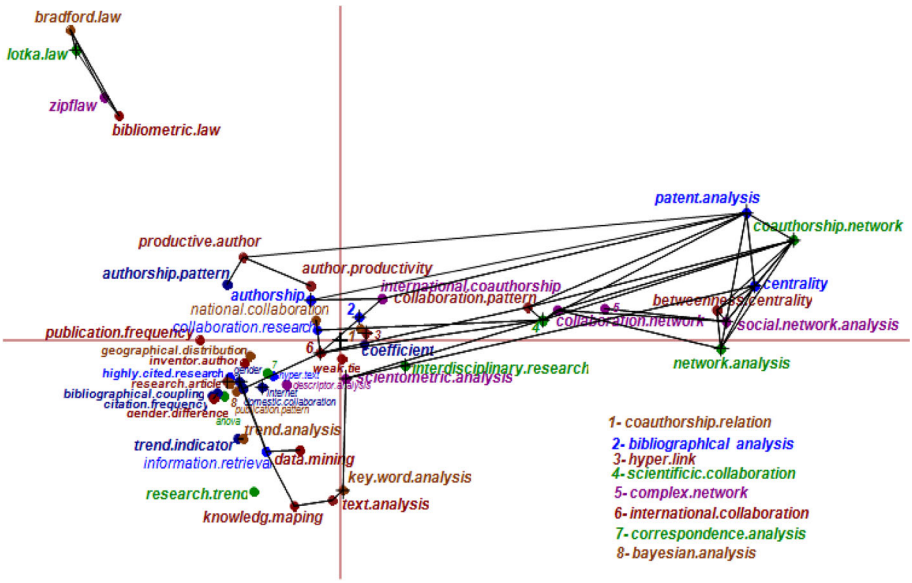

Fig. 5 Detail co-word structure of cluster 3 (C3) in 2005–2010

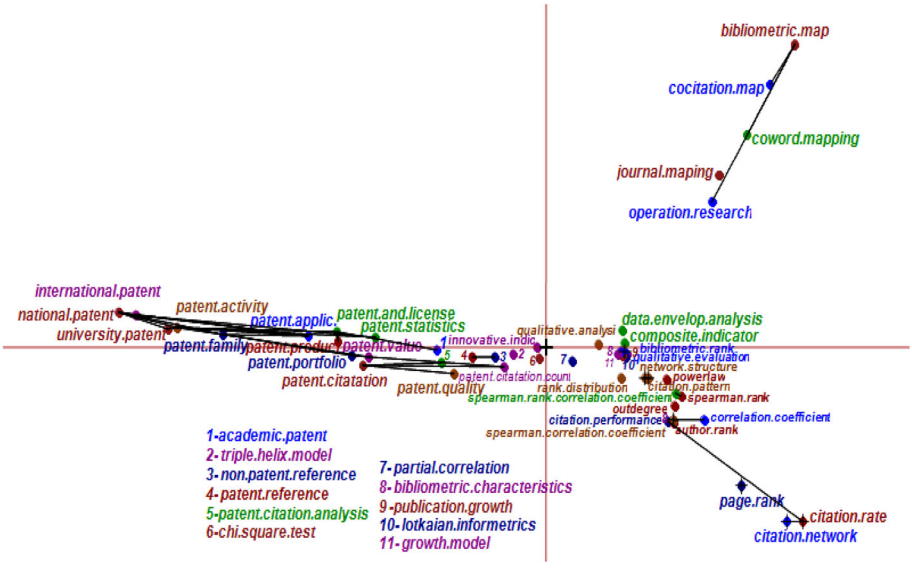

Fig. 6 Detail co-word structure of cluster 4 (C4) in 2005–2010

so on. Cluster 5 includes 47 topics relating to Website, URL, Co-link analysis, Gini-index, web-citation, co-word analysis, collaboration pattern topics on and so on.

It can be observed from above plotting network of different cluster that connection strength is not same. University ranking, Bibliometric analysis, Citation analysis in cluster 1 (C1); Citation pattern, publication output,Citation index, in C2; Patent analysis, Co-

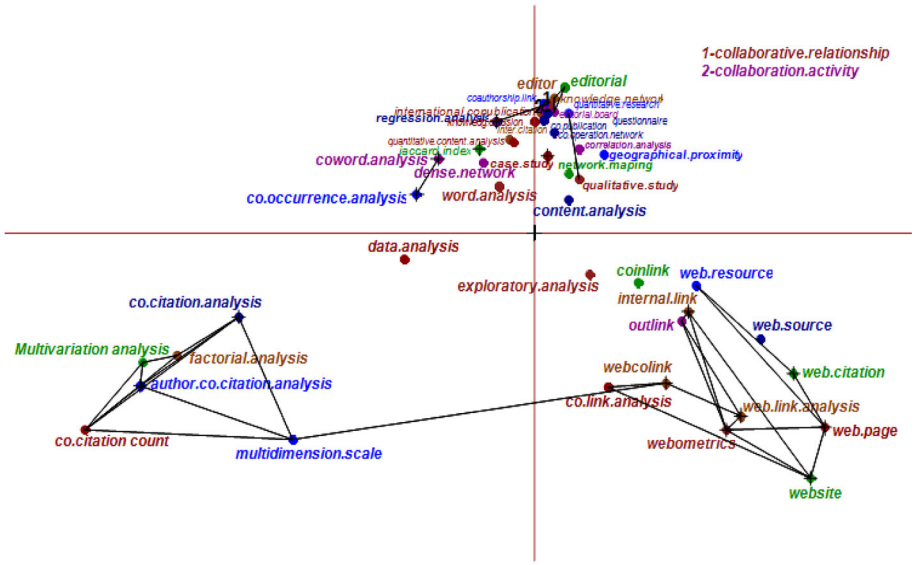

Fig. 7 Detail co-word structure of cluster 5 (C5) in 2005–2010

authorship network, Scientific collaboration, International collaboration, Lotka law in C3; International patent, University patent, Bibliometric map, and Citation performance in C4; Co-citation analysis, Co-citation count, Multidimensional scale, Website, Webometrics in C5 are active keywords which occupy more links.

For the comparative study of different clusters, following characteristics of network are tabulated below:

1. *Outer link* refer to the no. of links of a clusters which established network with other clusters (sub-domain).
2. *Inner link* refer to the no. of link which exist between keywords of a cluster. If A & B connected with an arch, it means A & B each have one link.
3. *Total link* refer to the sum of inner and outer link.
4. $Inner\ Link\ \% = \frac{Inner\ Link}{Total\ link} \times 100.$
5. $Outer\ Link\ \% = \frac{Outer\ link}{Total\ link} \times 100.$
6. *Inner Link Key* refers to the no. of keywords which has link within cluster.
7. *Outer Link Key* refers to the no. of keywords which has link from other clusters.
8. *Total Key* refers to the no. of keywords within a cluster.
9. $Inner\ link\ key\ \% = \frac{Inner\ Link\ key}{Total\ key} \times 100$
10. $Outer\ Link\ Key\ \% = \frac{Outer\ link\ key}{Total\ Key} \times 100$
11. $Average\ link\ per\ key = \frac{Total\ link}{Total\ key}$
12. *Centarlity is* defined as mean of outer link (sum of Salton index of outer links/outer links). It measures the intensity of links for a given cluster with other clusters.

$$Centrality = \frac{\sum X}{Outer\ Links}$$

Here, X = values of Salton index of outer link.

13. *Density* is defined as mean of inner link (sum of Salton index of inner links/inner links). It measures the strength of the links that tie the words making up the cluster together.

$$Density = \frac{\sum Z}{Inner\ Links}$$

Here, $Z =$ values of Salton index of inner link.

From Table 1, it can be observed that the intra-connection strength (inner link) of C1, C3, C4 and C5 is much higher than its outer link, whereas C1's inner link is slightly higher than its outer link. Overall 63 % of links are inner links; these links reflect the substantial relationship among keywords of clusters.

The number of keywords which bear inner link is much higher than outer link. Overall 56 % key bear inner link while 35 % key bear outer link only. On the average, C1 and C3 has 2 links (outer and inner links) and other cluster has only 1 link per key.

C4 owns highest centrality and highest density among the whole research networks of this period. Here, centrality is slightly higher to its density, it means this cluster owns not only connection within the cluster but also owns extensive connection with others clusters' keywords. This event indicates that this cluster's topics lies in the core of all research subject and research subject composed by its keywords are tending to mature. C2 occupies the highest density and higher centrality. The inner connection in this clusters are intense which explains that the research topics has capacity to maintain it and to develop over course of time. C1 and C3 has lowest centrality but its density is slightly higher it means its topic is immature in comparison to other clusters. C5's centrality is slighter higher than its density means its topics are in strategic position in comparison to C1 and C2.

Co-word analysis of 2005–2007

General overview structure

During this period, 188 keywords are selected as the keywords research sample. 52 Keywords (see Table 4 of Appendix 1) which does not have appropriate frequency (<10)

Table 1 Co-word structure of five clusters during year 2005–2010

S. No.	Characteristics	C1	C2	C3	C4	C5	Average
1	Inner link	52	40	102	60	70	64.8
2	Outer link	39	39	40	13	23	41.2
3	Total link	91	79	142	73	93	106
4	Inner link (%)	57	51	72	82	75	63.2
5	Outer link (%)	43	49	28	18	25	36.8
6	Inner link key	24	19	34	29	29	27
7	Outer link key	14	15	17	7	14	16.8
8	Total key	43	45	54	51	47	48
9	Inner link key (%)	56	42.22	62.96	56.86	61.7	56.25
10	Outer link key (%)	33	33.33	31.48	13.73	29.79	35
11	Average link per key	2.12	1.76	2.63	1.43	1.98	2.226
12	Density	0.261	0.301	0.284	0.285	0.273	
13	Centrality	0.254	0.267	0.25	0.3	0.275	

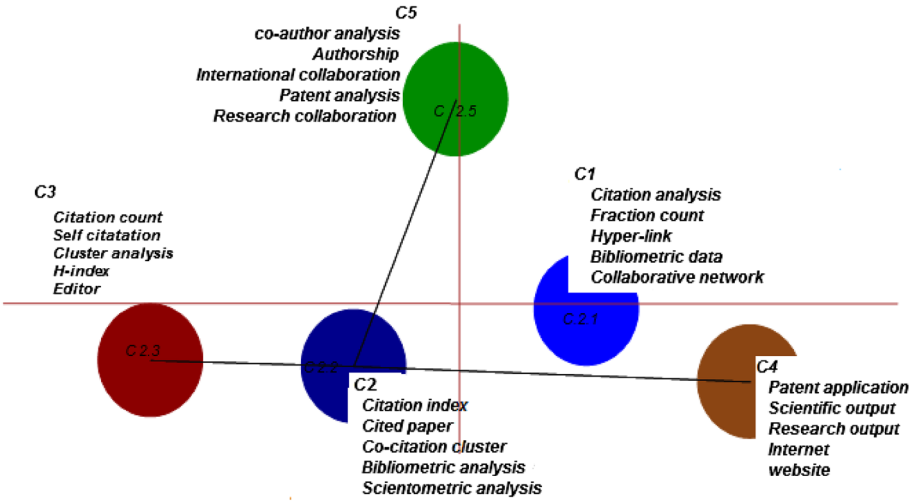

Fig. 8 General overview of the structure of Scientometrics in 2005–2007

are excluded. Previous method was applied to generate the general overview map of Scientometrics in 2005–2007 by MDS (Fig. 8) and each cluster was represented by its five highest no. of keywords and network with thin line show strongest relation between cluster according to Salton index (>0.35).

Detail co-word structure

In order to plot detail structure, keywords are selected as input variable to map the sub-domain based on correlation matrix of 188 keywords. Thus fives detailed sub-domain (cluster) structure (Figs. 9, 10, 11, 12, 13) are plotted to visualize specific characters of each sub-domains (clusters) in the field of scientometrics.

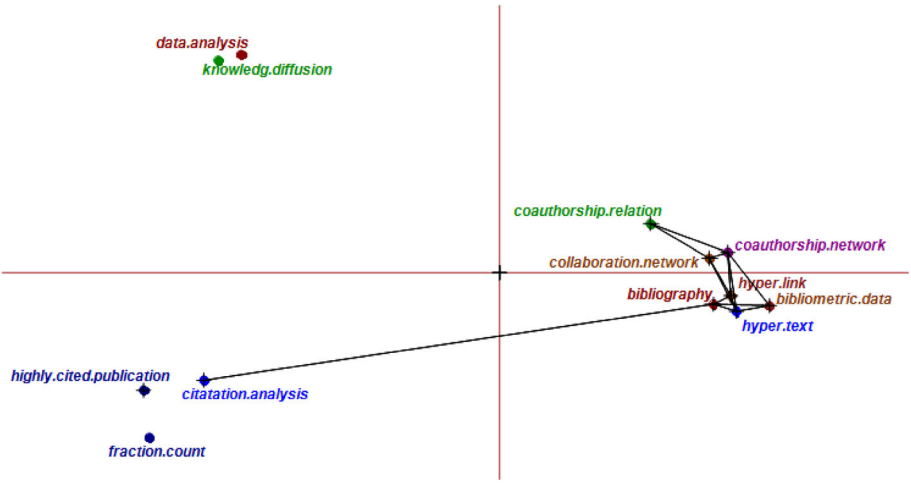

Fig. 9 Detail co-word structure of cluster 1 (C1) in 2005–2007

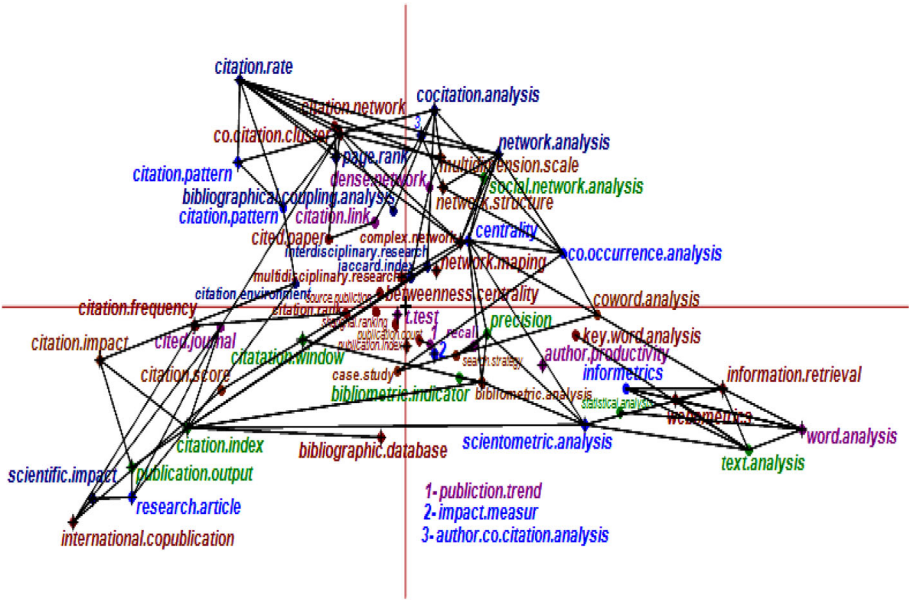

Fig. 10 Detail co-word structure of cluster 2 (C2) in 2005–2007

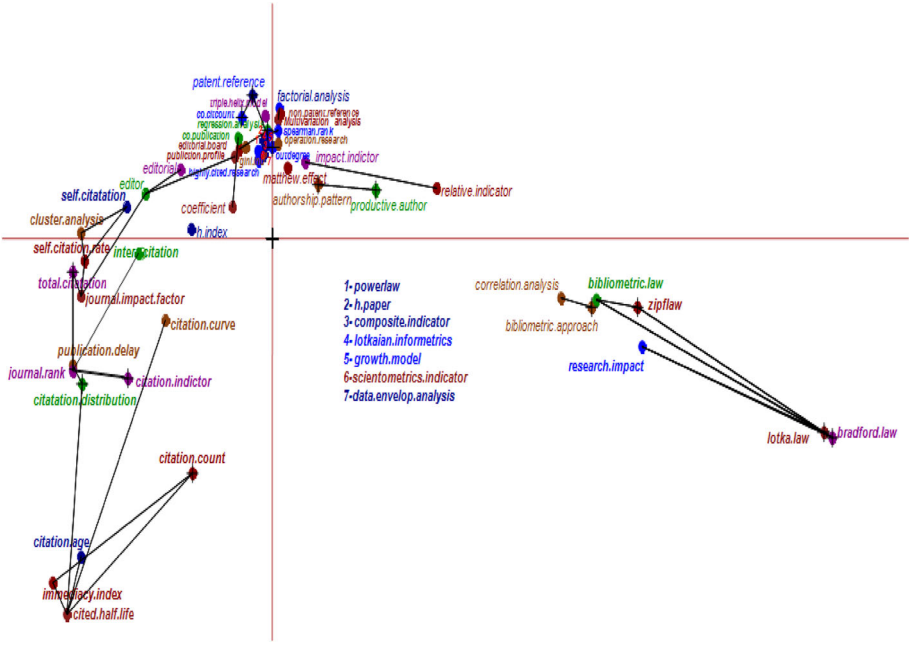

Fig. 11 Detail co-word structure of cluster 3 (C3) in 2005–2007

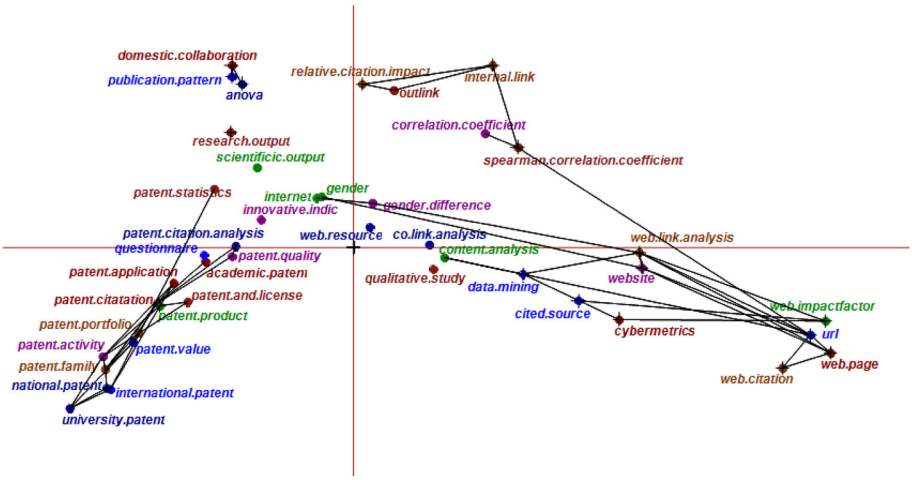

Fig. 12 Detail co-word structure of cluster 4(C4) in 2005–2007

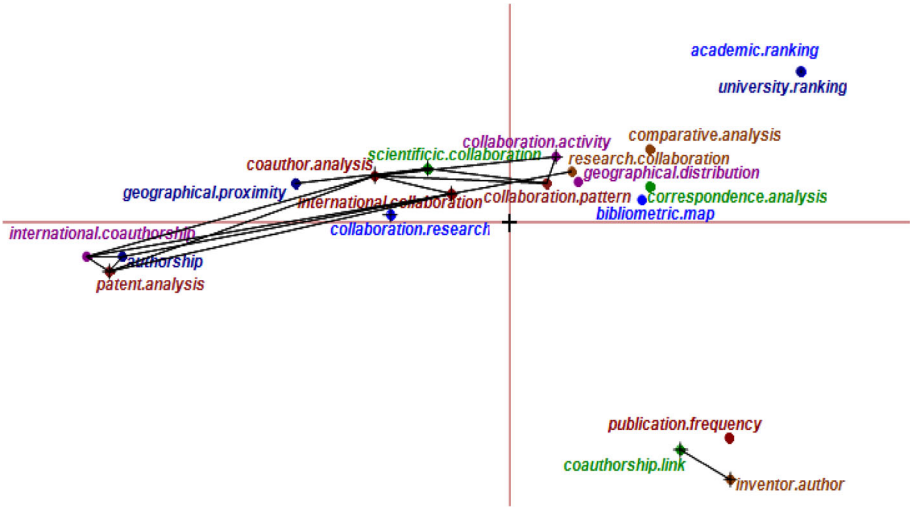

Fig. 13 Detail co-word structure of cluster 5 (C5) in 2005–2007

During year 2005–2007, each cluster has average value of 37 keywords. For plotting detailed co-word structured, keywords selected as input variable to map the sub-domain based on correlation matrix of 188 keywords. Five detailed sub-domain (cluster) structure (Fig. 9, 10, 11, 12, 13) were plotted to visualize specific characters each sub-domains (clusters) in the field of scientometrics during year 2005–2007.

Cluster 1 describes research topics related to Bibligraphical analysis, Citation analysis, Hyper link, Collaboration network, Knowledge diffusion. Cluster 2 describes research topics related to Keyword analysis, Precision, Interdisciplinary research, Case study, Informetrics. Cluster 3 describes topics on cluster analysis, self-citation, H-index, Editorial

board, Power law, Journal impact factor, Citation distribution. Cluster 4 focuses on Scientific output, Web page, Website, Internet, URL, Co-link analysis, Content analysis, Correlation coefficient. Finally, Cluster 5 appears to focus on Co-author analysis, Authorship, international collaboration, Patent analysis, Scientific collaboration, University rank, and Correspondence analysis.

Observing the co-word network of different clusters during this period (2005–2007), it found that C2 and C4 had more nodal keywords. Co-authorship network, Collaboration network, Hyper link, and Bibliometric network, in C1; Co-citation rate, Co-citation analysis, Co-citation cluster analysis, Network analysis, Centrality, Co-occurrence analysis, Citation environment, Citation index, Information retrieval, Text analysis, Scientometric analysis in C2; Citation count, Journal rank, Journal impact factor, Self-citation, Lotka law, Complex network in C3; University patent, Patent citation analysis, Patent productivity, Data mining, Domestic collaboration, Internal link, Web impact factor, Web-link analysis, URL, Cybermetrics in C4; Patent analysis, Co-authorship analysis, and International collaboration in C5 are active keywords which occupy more link with other keywords and focused main research topics of this period. In Table 2, different characteristics of five clusters can be seen.

As shown in Table 2, during this period C2, C3 and C4 have much higher than its outer link but C1 is on contrary. It shows abundant internal connection among keywords. C3 and C4 had about 88 % inner link key which is highest in whole research network and C1 had 67 % which is lowest. On the contrary C1 has highest outer link key which indicates strong relation with other clusters. Over all 81 % key are inner link key while only 42 % key has outer link. Thus these links show a stable internal composition in each cluster but week extensive relation among clusters.

On the average, C1 had highest no. of link (inner and outer link) per key which show coherent network among keywords while C2 had lowest no. of link per key. Overall average link per key is three. It is clearly advisable dense link between keywords.

In this period, C4 had highest density and higher centrality means this cluster did not only owns intense inner connection but also had good connection with other cluster. On this basis it can be said that C4's research topics had already been cared and able to

Table 2 Co-word structure of five clusters during year 2005–2007

S. No.	Characteristics	C1	C2	C3	C4	C5	Average
1	Inner link	28	184	70	98	36	83.2
2	Outer link	42	49	23	31	13	31.6
3	Total link	70	233	93	129	49	114.8
4	Inner link (%)	40	79	75	76	73	68.6
5	Outer link (%)	60	21	25	24	27	31.4
6	Inner link key	8	53	40	36	14	30.2
7	Outer link key	9	31	14	17	8	15.8
8	Total key	12	60	53	41	20	37.2
9	Inner link key (%)	66.67	88.33	75.47	87.8	70	81.18
10	Outer link key (%)	75	51.67	26.42	41.46	40	42.47
11	Average link per key	5.83	3.88	1.75	3.15	2.45	3.412
12	Density	0.266	0.277	0.342	0.315	0.301	
13	Centrality	0.256	0.246	0.246	0.272	0.27	

maintain itself to develop over the course of time and this field also considered important by Scienometrics research community. C1& C2 both has lower density and centrality. These research fields are immature and developing. These clusters did not have important network and nodal points.

Co-word analysis of 2008–2010

General overview Structure

During this period, among 240 keywords only nine keywords were not included (see Table 5 of Appendix 2) as these keywords did not appeared with appropriate frequency (it occurred less than ten times in the corpus). Thus 231 keywords were selected for research sample. The same method was used to generate the general overview structure of Scientometrics in 2008–2010 by plotting MDS (Fig. 14) and each sub-domain (cluster) was labeled by top five most frequent keywords within the cluster as before.

In order to pot detailed structured of each cluster during the period of 2008–2010, each cluster's keywords selected as input variable to map the sub-domain based on correlation matrix of 231 keywords. Thus fives detailed sub-domain (cluster) structure (Figs. 15, 16, 17, 18, 19) were plotted to visualize specific characters each of five sub-domains (clusters) in the field of scientometrics.

During this period (2008–2010) C1, C2, C3, C4 and C5 has 50, 76, 37, 37, 31 keywords respectively. Cluster 1 includes the research topics related to Bibliometrics, Scientific output, Citation index, Editor, International collaboration, Citation analysis. Cluster 2 includes keywords related to Website, Co-citation analysis, Citation rate, Patent citation, Patent application, Co-word analysis. Cluster 3 contains topic related to Co-author analysis, Cluster analysis, Interdisciplinary research, Hyper link. Cluster 4 includes topics on H-index, Co-citation cluster, Web page, Cited paper, Citation impact, Internet. And cluster 5 includes research topics on Scientometrics, Collaboration pattern, Keyword analysis, Centrality, Social network analysis, Network analysis, Patent analysis, Knowledge mapping.

It was found to observe the plotted figure that Bibliometric indicator, Bibliometric analysis, Iinternational collaboration, and Scientific output in C1; Patent related topics, Co-

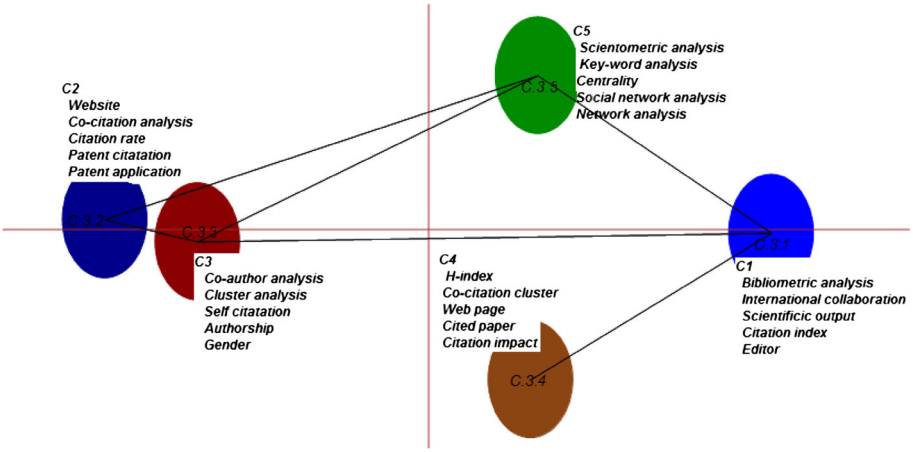

Fig. 14 General overview structure of Scientometrics in 2008–2010

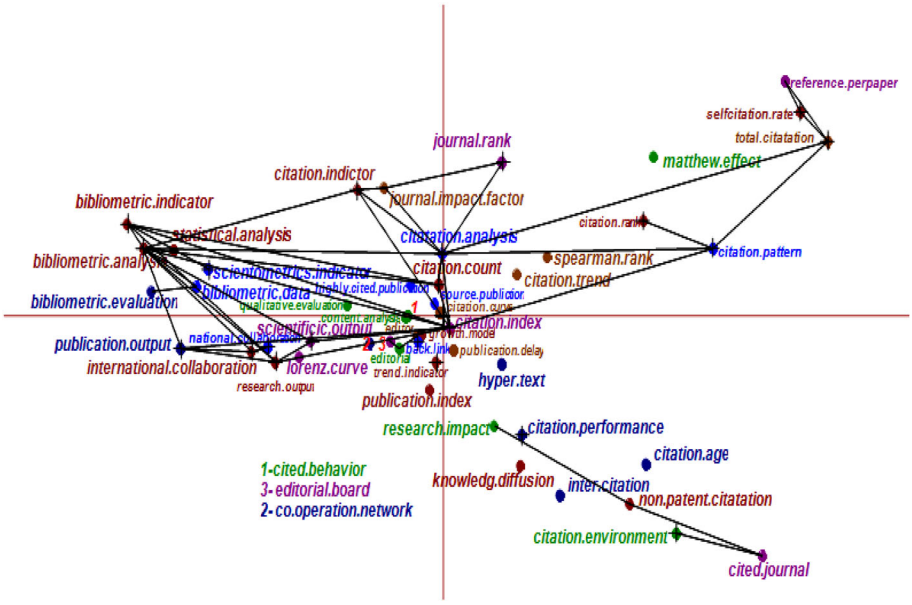

Fig. 15 Detail co-word structure of cluster 1 (C1) in 2008–2010

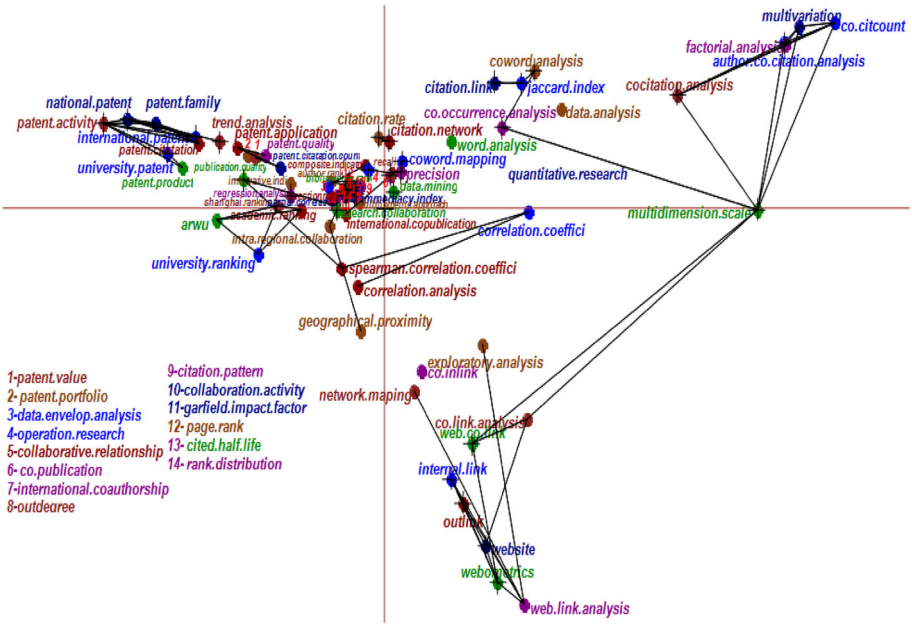

Fig. 16 Detail co-word structure of cluster 2 (C2) in 2008–2010

citation,Websites, Web link, Out link, Web link analysis, academic ranking in C2; Lotka law in C3; h-index, web citation, URL, and topics related to internet in C4; Informetrics, Scientometrics, Network analysis, Co-authorship network, Scientific collaboration,

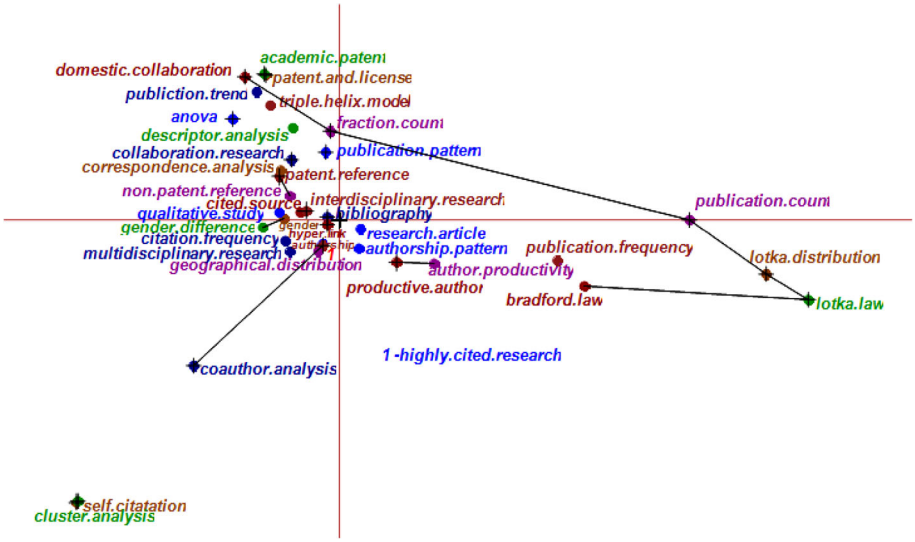

Fig. 17 Detail co-word structure of cluster 3 (C3) in 2008–2010

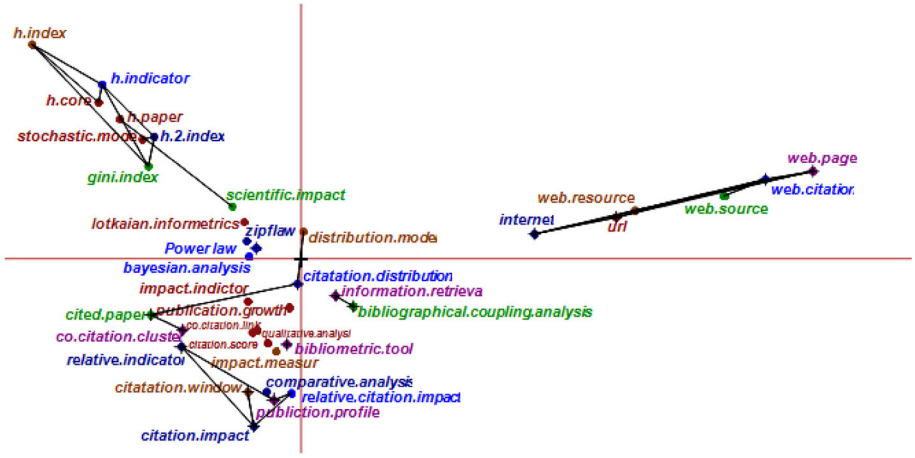

Fig. 18 Detail co-word structure of cluster 4 (C4) in 2008–2010

Collaboration pattern, Network structure in C5 are active keywords which associated with more keywords in the cluster and focused main research themes of this period.

From Table 3, it can be noticed that each cluster had inner link more than its outer link. Cluster 2 (C2) has highest no. of outer link but cluster 3 is on contrary while this cluster has 16 % outer link key only. On average each cluster has 67 % inner link key strong relationship between keywords. Overall around 59 % link are located within different clusters are inner link. These links reflect not only intense network within the cluster but also show loose linkage with other clusters.

Cluster 5 and 4 has highest 5 and 3 link per key respectively which is evidence of abundant links among keywords and coherent network.

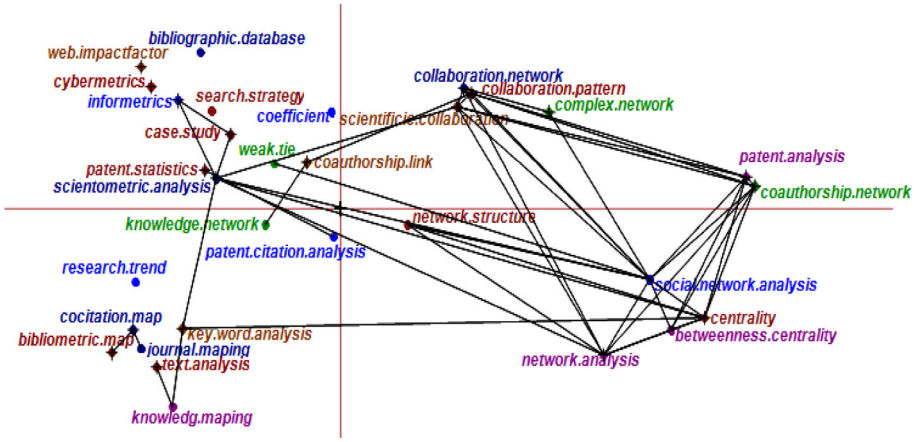

Fig. 19 Detail co-word structure of cluster 5 (C5) in 2008–2010

Table 3 Co-word structure of five clusters during year 2008–2010

S. No.	Characteristics	C1	C2	C3	C4	C5	Average
1	Inner link	88	140	22	54	94	79.6
2	Outer link	62	64	35	34	47	48.4
3	Total link	150	204	57	88	14	128
4	Inner link (%)	59	69	39	61	67	59
5	Outer link (%)	41	31	61	39	33	41
6	Inner link key	28	59	18	25	25	31
7	Outer link key	21	34	19	15	19	19
8	Total key	50	76	37	37	31	46.2
9	Inner link key (%)	56	77.63	48.65	67.57	80.65	67.1
10	Outer link key (%)	42	44.74	16.22	40.54	61.29	41.13
11	Average link per key	3	2.68	1.54	2.38	4.55	2.83
12	Density	0.254	0.283	0.402	0.3	0.279	
13	Centrality	0.247	0.267	0.26	0.258	0.248	

During this period C3 has highest density and centrality among whole research network, here centrality is more less than its density, it means the topics which embedded in the C3 had already form their own subfields with strong internal composition. C1 owns lowest density and centrality in whole research network which indicates its immaturity during this period. C2 is one with highest centrality indicating its strong linkage with other clusters. C4 and C5 also has more density than its centrality which indicates its intense connection between keywords which explains that its topic are already cared and developed.

Discussion and conclusion

Co-word analysis is powerful tool to identify the linkage and association among different themes of subject through the analysis of co-occurrence frequencies of keywords and

phrases. It is used to detect the themes, relationship between these themes, the extent to which these themes are central to and the degree to which these themes are internally structured in a given research area (He 1999). This technique is mainly used for quantitative study of literatures. It is used for the better understanding and to communicate the development of scientific field (Borner et al. 2003) as well as for evaluation purpose (Noyons 2005). Analysis can be performed by different open source and proprietary software like Word Stat, Text Mapping, Python, R Software and T-Lab etc.

'R' is a very sophisticated statistical analysis open source software (used in this study), as it helps us to find out the frequency (TM package) and co-occurrences of keywords from a document in a corpus. It can be used for MDS plotting, creation of network among keywords and calculating its different characteristics. This can also be used to reduce a large space of related descriptors to multiple related smaller spaces that are easier to comprehend.

This study provides a description of intellectual structure of the field of scientometrics from the perspective of frequently appearing keywords and phrases using co-word analysis, hierarchal clustering and link analysis. As we know, increasing number of new keywords and its association within or outside of the cluster is the evidence of micro level study and conception of new themes. Keyword associations provide a detail structure through which trend, and formed pattern of a publication can be analyzed.

Hence, we can conclude about the structure of scientometrics using co-word analysis of scientometrics literature in the journal Scientometrics. These field rapidly evolving as demonstrated by appearance in increasing no. of keywords in the publication. Research topics related to the data analysis of the period of 2005–2007 shows a research trend focusing on hyper-text, hyper-link, citation analysis, co-citation cluster, network analysis, URL, web-link analysis, authorship, text analysis, co-word analysis, patent citation, co-occurrence analysis, cluster analysis, scientific output, international collaboration, information retrieval, collaboration network. Some marginal topics which do not have good association and frequency like keywords analysis and web resource but get popularity in 2008–2010.

Consistent themes such as citation analysis, network analysis, web-link analysis, Lotka law, collaboration network have been noticed in second phase (year 2008–2010), but the focuses are moving towards h-index, website, citation rate, co-author analysis, author co-citation analysis, web resource, keyword analysis, scientometric analysis, scientific collaboration, bibliographic coupling, information retrieval, webometrics, informetrics. Whereas, cluster analysis, content analysis, self-citation, university patent, Bradford law, Zipf law, university patent are being noticed dimmer at the same time. Some new keywords like ARWU (academic world ranking of universities), H-indicator, Back-link, Knowledge mapping, web Co-link analysis, Garfield impact factor, trend analysis, Lotka distribution, Intra-regional collaboration, Co-citation analysis and Distributional model, H(2)index, Citation performance and have been found during analysis (see Table 9 of Appendix 2).

Scientometrics themes have well defined genealogies such as, Citation analysis, Author productivity, Bibliometric analysis and other appear to emanate from multiple preceding themes as, H-indicator, Co-citation map, and Co-citation link. Some topics emerge quickly with little evidence of ancestry as knowledge mapping, Back-link, Weak-tie, Bayesian analysis, Stochastic model and others. Thus it can be concluded that during year 2005–2010 scientometrics has good number of established themes and new themes had emerged due to dynamic nature of the subjects.

Acknowledgments We are grateful to Dr. Ying Ding (Associate professor, Department of information science, Indiana University, Netherlands) for her valuable suggestion.

Appendix 1

See Tables 4 and 5.

Table 4 List of keywords which are not noticed during year 2005–2008

S. No.	Keywords	S. No.	Keywords
1	ARWU (academic ranking of world university)	28	Journal maping
2	Author rank	29	Knowledg maping
3	Back link	30	Knowledge network
4	Bayesian analysis	31	Lorenz curve
5	Bibliometric characteristics	32	Lotka distribution
6	Bibliometric evaluation	33	National collaboration
7	Bibliometric rank	34	Non-patent citatation
8	Bibliometric tool	35	Partial correlation
9	Chi square test	36	Patent citatation count
10	Citation performance	37	Publication growth
11	Citation trend	38	Publication quality
12	Cited behavior	39	Qualitative analysi
13	Co-citation link	40	Qualitative evaluation
14	Cocitation map	41	Quantitative content analysis
15	Co-inlink	42	Quantitative research
16	Collaborative relationship	43	Rank distribution
17	Co-operation network	44	Reference perpaper
18	Coword mapping	45	Research trend
19	Descriptor analysis	46	Spearman rank correlation coefficient
20	Distribution model	47	Stochastic model
21	Exploratory analysis	48	Trend analysis
22	Garfield impact factor	49	Trend indicator
23	H(2)index	50	Weak tie
24	H-core	51	Web co-link
25	H-indicator	52	Web source
26	Individual product		
27	Intra-region collaboration		

Table 5 List of keywords which have not been noticed during year 2008–2010

S. No.	Keywords
1	Bibliometric characteristics
2	Bibliometric law
3	Chi square test
4	Coauthoship relation
5	dense network
6	individual product
7	inventor author
8	quantitative content analysis
9	Spearman correlation coefficient

In this study, keywords having <10 frequencies were not noticed

Appendix 2: Status of keywords' frequency during year 2005–2007 and 2008–2010

See Tables 6, 7, 8 and 9.

Table 6 List of 50 most occurred keywords during year 2005–2007 and year 2008–2010

S. No.	Year 2005–2007		Year 2008–2010	
	Keywords	Frequency	Keywords	Frequency
1	Cluster analysis	1,186	H-index	2,355
2	Self citation	958	Scientometric analysis	1,742
3	Bibliometric analysis	831	Key-word analysis	1,422
4	Co-citation cluster	678	Co-author analysis	1,231
5	Co-author analysis	642	Cluster analysis	1,171
6	Patent application	515	Co-citation cluster	1,163
7	Scientific output	489	Bibliometric analysis	1,155
8	Centrality	475	Centrality	1,100
9	H-index	472	Self-citation	1,019
10	Web page	456	Scientific output	850
11	Citation index	448	Website	832
12	Coefficient	409	Web page	762
13	Website	399	Citation index	706
14	Cited paper	398	Citation count	678
15	Scientometric analysis	384	Cited paper	677
16	Internet	372	Editor	640
17	Authorship	345	Coefficient	620
18	International collaboration	339	International collaboration	586
19	Editor	322	Authorship	535
20	Citation count	283	Citation impact	500
21	Key-word analysis	261	Publication output	490
22	Precision	255	Gender	463
23	Publication delay	233	Research output	455
24	Co-citation analysis	230	Citation analysis	449
25	Publication output	228	Co-citation analysis	437
26	Citation impact	220	Internet	423
27	Research output	213	Interdisciplinary research	371
28	URL	213	Citation rate	360
29	Self citation rate	210	Patent citation	358
30	Bibliography	190	Social network analysis	358
31	Correlation coefficient	190	Patent application	349
32	Patent citation	188	Bibliometric indicator	319
33	Interdisciplinary research	188	Network analysis	314
34	Citation analysis	181	URL	300
35	Patent activity	177	Precision	288
36	Journal impact factor	171	Research collaboration	271
37	Gender	164	Patent analysis	270

Table 6 continued

S. No.	Year 2005–2007		Year 2008–2010	
	Keywords	Frequency	Keywords	Frequency
38	Citation distribution	163	Co-link analysis	265
39	Patent analysis	162	Gini index	264
40	Questionnaire	161	Bibliography	257
41	Citation rate	160	Total citation	249
42	Research article	158	Correlation coefficient	244
43	Bibliometric indicator	157	Editorial	242
44	Editorial board	154	Journal impact factor	238
45	Case study	154	H-indicator	236
46	Editorial	147	Informetrics	234
47	University patent	146	Collaboration network	233
48	Informetrics	146	Multidimensional scale	231
49	Research collaboration	136	Scientific collaboration	229
50	Co-link analysis	118	Citation window	220

Table 7 List of most increased keywords during year 2005–2007 to year 2008–2010

S. No.	Keywords	Increment in %	Frequency in year 2005–2007	Frequency in year 2008–2010
1	International co-publication	1,180	10	128
2	Knowledge diffusion	867	12	116
3	Web resource	742	12	101
4	Web impact factor	710	11	89
5	Author co-citation analysis	629	14	102
6	Betweenness centrality	600	22	154
7	Key-word analysis	445	261	1,422
8	Patent quality	440	20	108
9	Network structure	431	23	122
10	Network analysis	415	61	314
11	H-index	399	472	2,355
12	Collaboration network	396	47	233
13	Academic ranking	391	21	103
14	Cybermetrics	391	11	54
15	Social network analysis	384	74	358
16	ANOVA	369	16	75
17	Scientometric analysis	354	384	1,742
18	Scientific collaboration	333	53	229
19	Operation research	330	20	86
20	Scientific impact	323	22	93
21	Citation curve	300	11	44
22	Co-authorship network	276	45	169
23	Gini index	267	72	264

Table 7 continued

S. No.	Keywords	Increment in %	Frequency in year 2005–2007	Frequency in year 2008–2010
24	Scientometrics indicator	257	39	139
25	Patent value	255	20	71
26	Collaboration research	254	30	106
27	Complex network	250	10	35
28	Publication count	227	49	160
29	Composite indicator	222	14	45
30	Data mining	193	14	41
31	Bibliometric approach	188	16	46
32	Correlation analysis	183	17	48
33	Gender	183	164	463
34	Growth model	173	11	30
35	Total citation	163	95	249
36	Citation rank	162	21	55
37	Multidimensional scale	157	90	231
38	Research impact	154	30	76
39	Citation analysis	149	181	449
40	Citation window	148	89	220
41	Co-authorship link	146	11	27
42	Hyper-text	146	37	91
43	Citation network	145	59	144
44	Citation count	140	283	678
45	Highly cited publication	138	16	38
46	Centrality	132	475	1,100
47	Citation impact	128	220	500
48	Citation rate	125	160	360
49	Co-link analysis	125	118	265
50	Bibliometric map	124	26	58

Table 8 List of all decreased keywords during 2005–2007 to 2008–2010

S. No.	Keywords	Decrement in %	Frequency in year 2005–2007	Frequency in year 2008–2010
1	Inventor author	84	49	8
2	Non-patent reference	78	114	26
3	Academic patent	75	107	27
4	Zipf law	70	40	12
5	Patent reference	62	59	23
6	Dense network	60	15	6
7	Geographical distribution	59	72	30
8	Bradford law	58	56	24
9	Bibliometric law	57	16	7

Table 8 continued

S. No.	Keywords	Decrement in %	Frequency in year 2005–2007	Frequency in year 2008–2010
10	Network mapping	52	25	12
11	Publication delay	51	233	116
12	Co authorship relation	50	12	6
13	Citation pattern	47	28	15
14	Self citation rate	47	210	113
15	Correspondence analysis	46	24	13
16	Chi square test	43	7	4
17	Innovative indic	43	19	11
18	Patent activity	42	177	104
19	Web-link analysis	42	39	23
20	Jacquard index	36	34	22
21	Source publication	35	20	13
22	International co authorship	34	96	64
23	Patent and license	34	24	16
24	Out degree	33	31	21
25	Patent application	33	515	349
26	Page rank	31	83	58
27	Questionnaire	29	161	115
28	Bibliographic database	28	50	36
29	Patent portfolio	28	25	18
30	Lotkaian informetrics	24	13	10
31	Publication frequency	23	22	17
32	University patent	23	146	113
33	Citation environment	20	40	32
34	Citation score	17	59	49
35	Immediacy index	16	83	70
36	Citation distribution	16	163	138
37	International patent	11	58	52
38	Citation age	10	11	10
39	Inter-citation	8	25	23
40	Publication pattern	8	52	48
41	Publication index	5	20	19
42	National patent	5	41	39
43	Impact measure	5	21	20
44	Spearman correlation coefficient	5	21	20
45	Citation link	3	40	39
46	Cluster analysis	2	1,186	1,171

Table 9 List of keywords which were newly noticed during year 2008–2010

S. No.	Keywords	Frequency
1	H-indicator	236
2	H-core	181
3	Co-inlink	101
4	Knowledge mapping	96
5	Knowledge network	95
6	ARWU (Academic ranking of world university)	89
7	Research trend	52
8	National collaboration	51
9	Citation performance	50
10	Back link	48
11	Weak tie	45
12	H(2)index	39
13	Reference per paper	35
14	Partial correlation	35
15	Collaborative relationship	34
16	Patent citation count	29
17	Publication growth	28
18	Cited behavior	26
19	Lorenz curve	25
20	Rank distribution	24
21	Garfield impact factor	24
22	Non-patent citation	23
23	Bibliometric evaluation	23
24	Publication quality	22
25	Author rank	21
26	Web co-link	21
27	Bibliometric rank	20
28	Bibliometric tool	20
29	Co-citation map	19
30	Co-operation network	19
31	Co-citation link	18
32	Journal mapping	18
33	Web source	17
34	Trend analysis	17
35	Lotka distribution	14
36	Trend indicator	14
37	Bayesian analysis	14
38	Exploratory analysis	14
39	Distribution model	14
40	Intra-region collaboration	14
41	Qualitative analysis	13
42	Descriptor analysis	12
43	Qualitative evaluation	11

Table 9 continued

S. No.	Keywords	Frequency
44	Citation trend	10
45	Co-word mapping	10
46	Stochastic model	10
47	Quantitative research	10

References

An, X. Y., & Wu, Q. Q. (2011). Co-word analysis of the trends in stem cells field based on subject heading weighting. *Scientometrics, 88*(1), 133–144.

Bhattacharya, S., & Basu, R. K. (1998). Mapping a research area at the micro level using co-word analysis. *Scientometrics, 43*(3), 359–372.

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

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics, 22*(1), 153–203.

Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometircs? Mapping over 70 years of biological safety research with co-word analysis. *Scientometrics, 27*(2), 119–143.

Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science, 49*(13), 1206–1223.

Courtial, J. P. (1994). A coword analysis of scientometrics. *Scientometrics, 31*(3), 251–260.

Courtial, J. P., Callon, M., & Sigogneau, A. (1993). The use of patent titles for identifying the topics of invention and forecasting trends. *Scientometrics, 26*(2), 231–242.

Culnan, M. J. (1986). The intellectual development of management information systems, 1972–1982: A co-citation analysis. *Management Science, 32*(2), 156–172.

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management, 37*(6), 817–842.

He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends, 48*(1), 133–159.

Hinze, S. (1994). Bibliographical cartography of an emerging interdisciplinary discipline: The case of bioelectronics. *Scientometrics, 29*(3), 353–376.

Hou, H.Y., Liu, Z.Y., Chen, Y., Jiang, C.L., Yin, L.C., & Pang, J. (2006). Mapping of science studies: the trend of research fronts. Science Research Management, 22(3), 90–96.

Kopcsa, A., & Schiebel, E. (1998). Science and technology mapping: A new iteration model for representing multidimensional relationships. *Journal of the American Society for Information Science, 49*(1), 7–17.

Kostoff, R. N., Stump, J. A., Johnson, D., Murday, J. S., Lau, C. G. Y., & Tolles, W. M. (2006). The structure and infrastructure of the global nanotechnology literature. *Journal of Nanoparticle Research, 8*(3–4), 301–321.

Law, J., Bauin, S., Courtial, J.-P., & Whittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics, 14*(3–4), 251–264.

Law, J., & Whittaker, J. (1992). Mapping acidification research: A test of the co-word method. *Scientometrics, 23*(3), 417–461.

Lee, B., & Jeong, Y. I. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics, 77*(1), 3–19.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *Journal of the American Society for Information Science and Technology, 57*(12), 1616–1628.

Leydesdorff, L., & Zaal, R. (1988). Co-words and citations. relations between document sets and environments. In L. Egghe, & R. Rousseau (Eds.), *Informetrics 87/88* (pp. 105–19). Amsterdam: Elsevier.

Looze, M. D., & Lemarie, J. (1997). Corpus relevance through co-word analysis: An application to plant proteins. *Scientometrics, 39*(3), 267–280.

Mihalcea, R., Tarau, P. (2004). TextRank: Bringing order into texts. In *Proceedings of the conference on empirical methods in natural language processing* (*EMNLP 2004*). Barcelona: Spain.

Mohammadi, E. (2012). Knowledge mapping of the Iranian nanoscience and technology: A text mining approach. *Scientometrics, 92*(3), 593–608.
Mooghali, R., Alijani, M. S., Karami, N., & Khasseh, A. (2011). Scientometrics analysis of scientometrics literature. *International Journal of Information Science and Management, 9*(1), 19–31.
Musgrove, P. B., Binns, R., Page-Kennedy, T., & Thelwall, M. (2003). A method for identifying clusters in sets of interlinking web spaces. *Scientometrics*, *58*(3), 657–672.
Noyons, C. (2005). Science maps within a science policy context. In *Handbook of quantitative science and technology research*, (pp. 237–255).
Noyons, E. C. M., & van Raan, A. F. J. (1994). Bibliometric cartography of scientific and technological development of an R&D field. *Scientometrics, 30*(1), 157–173.
Noyons, E. C. M., & van Raan, A. F. J. (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science*, *49*(1), 68–81.
Polanco, X., Francois, C., & Keim, J. P. (1998). Artificial neural network technology for the classification and cartography of science and technology information. *Scientometrics, 41*(1–2), 69–82.
Ponzi, L. J. (2002). The intellectual structure and interdisciplinary breadth of knowledge management: A bibliometric study of its early stage of development. *Scientometrics, 55*(2), 259–272.
R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/. Accessed 4 Sept 2013.
Rikken, P., Kiers, H. A. L., & Vos, R. (1995). Mapping the dynamics of adverse drug reactions in subsequent time periods using INDSCAL. *Scientometrics, 33*(3), 367–380.
Rip, A., & Courtial, J. P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics., 6*(6), 381–400.
Sedighi, M., & Jalalimanesh, A. (2014). Mapping research trends in the field of knowledge management. *Malaysian Journal of Library and Information Science, 19*(1), 71–85.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between documents. *Journal of the American Society for Information Science, 24*(4), 265–269.
Small, H., & Griffith, B. (1974). The structure of scientific literature I: Identifying and graphing specialities. *Science Studies, 4*(1), 17–40.
Van Raan, A. F. J., & Tijssen, R. J. W. (1993). The neural net of neural network research. *Scientometrics, 26*(1), 169–192.
Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608–1618.
Yue, H. (2012). Mapping the intellectual structure by co-word: A case of international management science. *Web Information Systems and Mining, 75*(29), 621–628.