Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 68, No. 3 (2006) 377–393

Mapping research topics using word-reference co-occurrences: A method and an exploratory case study

PETER VAN DEN BESSELAAR,[a,b] GASTON HEIMERIKS[c]

[a] *Science System Assessment Center, Rathenau Instituut, Den Haag (The Netherlands)*
[b] *Amsterdam School of Communication Research (ASCoR), Universiteit van Amsterdam (The Netherlands)*
[c] *Institute for Prospective Technological Studies (IPTS), Seville (Spain)*

Mapping of science and technology can be done at different levels of aggregation, using a variety of methods. In this paper, we propose a method in which title words are used as indicators for the content of a research topic, and cited references are used as the context in which words get their meaning. Research topics are represented by sets of papers that are similar in terms of these word-reference combinations. In this way we use words without neglecting differences and changes in their meanings. The method has several advantages, such as high coverage of publications. As an illustration we apply the method to produce knowledge maps of information science.

Introduction

Science mapping aims at revealing the structure and dynamics of science using attributes of communications, most importantly scientific publications. Mapping, however, can be done at several levels of granularity. For example, one may want to map the disciplinary structure of science, the development of research fields, of subfields or of the research front (in terms of specific research topics). In this paper we distinguish the various levels of granularity at which the structure and dynamics of the sciences can be mapped. The highest level is the *discipline*, such as sociology or

Received December 20, 2005

Address for correspondence:
PETER VAN DEN BESSELAAR
Science System Assessment, Rathenau Instituut
P. O. Box 95366, 2509 CJ Den Haag, The Netherlands
E-mail: p.vandenbesselaar@rathenau.nl

0138–9130/US $ 20.00

physics. At the next level we distinguish *research fields* within disciplines, such as science and technology studies, or particle physics. An even more detailed level is the *subfield*, such as scientometrics, and this again is subdivided into *research topics*, which is the smallest unit under consideration. Within scientometrics we consider patent studies, co-word analysis, or science mapping as research topics.

Mapping science and technology has a long tradition that started with the work of SMALL (1973) and of others such as MARSHAKOVA (1973) and KESSLER (1963). Since then, a variety of methods have been proposed, such as journal citation analysis, co-citation analysis, bibliometric coupling, and co-word analysis. More recently also visualization tools have improved, to make the maps more informative and easier to understand.[*] In this paper we will introduce a method for mapping research topics, based on co-occurrences of word-reference combinations. These co-occurrences are used to cluster papers as representations of research fronts. We will describe the method, and apply it to map the field of *information science*. The results will be used to evaluate the method, and we will indicate practical and theoretical advantages.

Mapping research fields

Researchers in a (sub)field share a common knowledge base, and this is reflected in the selection of references. Therefore, a research (sub)field can be defined as a network of journals covering a coherent set of research questions and methodologies and referring to a largely overlapping literature. As a consequence of this last characteristic, we expect journals belonging to the same research field to exhibit similar aggregated citation patterns. Through their local citing behavior, researchers reproduce the identity of the field at an aggregated level – and inter-journal citations can be used to map this identity in terms of sets of journals with similar citing patterns.

An example is the field of science and technology studies that can be defined as the journal network around five core scholarly journals in the field: Social Studies of Science, Scientometrics, Research Policy, Science & Public Policy, and Science, Technology & Human Values. The selection of the core journals should be based on consensus among specialists in the field. However, analysis of the STS-field indicates that at a lower level, this network is composed of three relatively weakly related subfields, with only a small shared knowledge base (VAN DEN BESSELAAR, 2000; 2001). Indeed, as we will see, scientometrics is becoming a subfield of information science, much more than being a part of science and technology studies. This indicates that the delineation on the higher level of research fields is an essential step when one wants to map the research topics in the field. If one would add unrelated or weakly related

[*] For overviews: MOED et al. (2004), CHEN (2003), SHIFFRIN & BÖRNER (2005), WHITE & MCCAIN (1989).

journals to the set, the map of the research topics would be influenced considerably. By introducing large differences between unrelated fields, the smaller differences between topics *within* a field become less visible.

We described the method of journal-journal citation analysis elsewhere in detail and showed that it provides us with an operational definition of a research (sub)fields (VAN DEN BESSELAAR & LEYDESDORFF, 1996). The method has been developed and used for delineating disciplinary research fields (COZZENS & LEYDESDORFF, 1993), but have we showed that the method is also well suited for delineating interdisciplinary fields (VAN DEN BESSELAAR & HEIMERIKS, 1991). In this paper it forms the context for a detailed mapping of research fronts.

Mapping research fronts

Various methods for the mapping of the fine-structure of research fields are discussed in the literature, and especially co-citation analysis and co-word analysis are among the most widely used techniques.

MARSHAKOVA (1973) and SMALL (1973) independently developed co-citation analysis by noting that if two references are cited together in a publication, the two references are themselves related. The greater the number of times they are cited together, the greater their co-citation strength. Co-citation analysis can be done with papers, but also with authors. Author co-citation analysis results in clusters of authors, who are linked through co-occurrences in reference lists. MCCAIN (1990) presented a comprehensive technical review of mapping authors in intellectual spaces, and she applied it among others to map information science for the period 1972–1995 (WHITE & MCCAIN, 1998). This method creates maps of research fields in terms of clusters of authors rather than topics per se, and these have to be derived from the known research interests of these authors. As most researchers cover various topics in their active life, the clusters generally do not exhibit a micro picture of the (changing) research topics that dominate a discipline, but place the authors in an intellectual space which one needs to map independently, e.g., through article co-citation analysis.

Earlier, KESSLER (1963) suggested a technique known as bibliographic coupling. Bibliographical coupling is measuring similarity between papers by the number of references two papers have in common. He showed that a clustering based on this measure yields meaningful groupings of papers as "a number of papers bear a meaningful relation to each other when they have one or more references in common". The major difference between bibliometric coupling and co-citation is that while coupling measures the relationship between source documents, co-citation measures the relations between cited documents. The former is based on conscious behavior: an author purposefully decides to relate two articles, whereas the latter is used merely as an association with hindsight between two articles. This implies that bibliometric coupling is better suited for mapping the *research front*, whereas co-citation analysis maps the *past structure* of a research field, as it is perceived now by researchers.

Another approach to map research topics is co-word analysis, resulting in clusters of words jointly appearing in titles, abstracts, or in full texts (CALLON et al., 1986). Co-word analysis does not lead to clusters of authors, but should give more of a direct access to the research topics in terms of used concepts. Different research topics are expected to use different words, and sets of co-occurring words may indicate the specific research topics within larger specialties. The literature presents many meaningful examples of co-word mappings (BHATTACHARYA & BASU, 1998), and various directions for improving methods for co-word analysis have been proposed (NOYONS & VAN RAAN, 1998).

The methods discussed here have also been criticized. For example LEYDESDORFF (1997) argues that "words and co-words cannot map the development of the sciences", because words are not specific enough, and do have different meanings in different (textual) contexts. In other words, one needs to know the boundaries of a research topic before co-word analysis should be carried out. Co-citation analysis has been criticized too, for example for the loss of many relevant papers, the inclusion of non-relevant papers, an over-representation of theoretical papers, time lag, and a subjectivity in threshold setting (KING, 1987). Apart from that, the behavioral foundations of the methods are also weak, as they both refer only to a single behavioral dimension: referencing strategies, or choice of words (RIP, 1988).

Some authors have applied a combination of methods. For example, BRAAM et al. (1991a; 1991b) combined co-citation analysis with co-word analysis. In this way, they used the information about the stock of knowledge in research fields (the references) as well as the current research front represented by the concepts (the words) used by researchers to describe their work. The combination of sets of title-words and sets of cited references is expected to identify a research topic in a better way than title-words alone. The way researchers draw on earlier works, and their sharing of a set of exemplars is considered to be reflected in the referencing practices of the specialty members. On the other hand, the shared interest in a set of research problems and concepts is expected to be reflected in the word patterns. The congruence in both mapping approaches is presupposed in many scientometric studies, but also criticized. BRAAM et al. (1991a; 1991b) showed that the mapping of science by combining and comparing co-citation and co-word analysis is a useful tool to map the subject matter of research specialties in a given period. However, as the analysis of Braam et al. is based on a sequential application of citation relations and words, it therefore not only combines the strengths of the two methods but also their weaknesses.

An alternative approach: co-occurrences of word-reference combinations

We propose here a different method which is based on co-occurrences of word-reference combinations. We use papers as the units of analysis within the set of journals that together define a research field. One can view a scientific field as a communication network. Scientific publications in journals allow us to map these communication systems. Knowledge is produced by combining and extending existing papers, and new knowledge is related to previous research by cited references. Consequently, the development of research topics can be observed in the form of an evolving paper system.

The idea behind the approach is the following: Researchers *simultaneously* select words to describe their research subject and refer to specific literature to indicate the tradition in which they do their work. The words acquire their specific meaning within the context of the cited references. In this way we can account for different ways of using words, and for changes over time in the meanings of words. In other words, we expect the cited references to provide more of a context for the words than the words do for each other. The advantage of using word-reference pairs is that it combines two relevant attributes of documents in determining the fine-grained structure of the specialty under study. This combined indicator reflects the subject of the research topic through the title words, and its position within the specialty through the references (VAN DEN BESSELAAR & HEIMERIKS, 2000).

Operationally, a research topic is defined as a set of papers that are similar in terms of word-reference combinations. The similarity measure is the number of word-reference combinations two papers share. If we define the boundary of a research topic in a restrictive way, the maps of the research fields become more detailed: many small topics; if we take a more relaxed criterion, the research topics will be of a broader nature. More specifically, the questions addressed in this paper are:

- Can we identify research topics using co-occurrences of word-reference combinations?
- Can we aggregate the research topics into meaningful subfields?
- Can we trace the development of the field?
- Is the map comprehensive?

An exploratory case study

We will use the field of *information science* between 1986 and 2002 as case study to illustrate the method. We limit ourselves to the cartography of the *cognitive dimension* of information science: what is the research front and how does it change over the years? As the research front is represented by journal articles, the method can easily provide social and geographical maps of a research field: e.g., who are conducting this research, and how are the field and the topics distributed over countries? This, however, is not within the scope of the current paper.

A variety of maps of information science have been made in the past, based of different methods. Some of the maps use a restricted definition of research fields as a single journal. For example, COURTIAL (1994) produced a co-word map of scientometrics, using the journal *Scientometrics* as representation of the field. PERSSON (1994) did something similar, but for the *Journal of the American Society for Information Society (JASIS)* between 1986 and 1990, whereas using bibliometric coupling. As a systematical delineation of the field shows, this is too restrictive. It creates an artificial image of coherence in our view, which makes mapping easier but also less adequate. A wider domain was used by WHITE & MCCAIN (1998) for their visualization of information science. However, as their map was based on author co-citation analysis, the map does not represent the research front, but similarity among authors and through this an image of the research past. NOYONS (1999) extended the field of information science more, by including also the field of science and technology studies. Of course, both social studies of science and information science (including scientometrics) study the sciences and their development, but from such different perspectives that they hardly have any (citation) relations at all (VAN DEN BESSELAAR, 2000), and they also lack a substantial shared intellectual base (VAN DEN BESSELAAR, 2001). Consequently, words and noun phrases can be expected to have different meanings in the two fields, and using both to map a research front may merge topics that are in fact unrelated (but use similar word combinations).

Data

The analysis of the topology of research topics is based on the documents published in the set of journals that define the research field. We have delineated information science using journal-journal citations, and previous analysis showed that this journal system has differentiated into three smaller but related subfields (VAN DEN BESSELAAR & LEYDESDORFF, 1996; VAN DEN BESSELAAR & HEIMERIKS, 2000). The first and largest cluster is a set of journals around the *Journal of the American Society of Information Science (JASIS)* and the *Journal of Documentation.* A second set of journals includes *Scientometrics* and the *Journal of Information Science*. Finally one finds one or more sets of journals on libraries and library research. The first two clusters are strongly related (in terms of journal inter-citation relations) and in some of the years they actually coincide. Therefore, we will base our cartography on these two clusters.

As this journal set has a stable core but also does change over time, one needs to decide whether to include also the more marginal journals in the map of the research field. In this study, we will focus on the research topics within the field of information science that is defined in terms of the stable core set of eight journals around *JASIST* in the period 1986 to 2002. Only the journals *Journal of Documentation, Information Processing and Management, Annual Review of Information Science and Technology, Proceedings Asist, Canadian Journal of Information Science, Journal of Information Science*, *Scientometrics* and of course *JASIST* itself are present in each year within a somewhat larger and changing set of journals. By using this stable set of journals, we are able to map the developing research front of information science without having to take into account the more peripheral changes indicated by the more marginal journals.

As usual in this type of analysis, we only use articles and reviews, as they form the core set of publications that constitute the field. We omitted the other publication types such as letters, notes, book reviews, meeting reports, editorials, and the like. The bibliographical information about these documents has been downloaded in the dialog format, for all of the years under study (1986, 1992, 1996, 2000 and 2002), using the CD-Rom version of the SCI and the SSCI. Table 1 summarizes the data we have used for this paper.

Special software was used for preprocessing, in order to transform the data into a format appropriate for further analysis. Frequency lists of title words (excluding a list of stop words) and of cited references were calculated. A database with all possible combinations of title words and cited references was created, and we linked those to all publications. Of course, only combinations that occur in more than one publication are included, because the relationship between publications is defined as sharing the same word-reference combination. This means that we did not have to set a threshold but include as many publications as possible.

Table 1. Description of the data

Journals*	Year	Total**	Included***
Journal of the American Society of Information Science			
Journal of Documentation	1986	204	47%
Information Processing and Management	1992	294	72%
Annual Review of Information Science and Technology	1996	314	76%
Proceedings ASIST	2000	299	73%
Canadian Journal of Information Science	2002	440	58%
Journal of Information Science			
Scientometrics	total	1551	67%

* Not included as citing source in the SCI and the SSCI: the *Annual Review of Information Science and Technology* in 1992 and 2000; the *Canadian Journal of Information Science* in 1996; the *Proceedings ASIST* in 2000.
** Number of articles and reviews in the mentioned journals
*** Percentage of documents included in the analysis

Studying large and complex networks, like the relationships between all publications a research field, requires a relational analysis that determines the emerging clusters of similar papers. We use the so-called cosine algorithm for determining the association (proximity) of two papers. We used BibTechMon, a software tool for analyzing and visualizing large networks in various dimensions. It is based on a 'mechanical spring model' (KOPCSA & SCHIEBEL, 2000), and enables a transformation of the multi-dimensional space into a two-dimensional map. Visual inspection enables us to divide a relational map into clusters of papers. The underlying database of documents provides the titles of documents that cluster together. We evaluated the titles within a cluster in terms of homogeneity, and in case of doubt we went back to the original article. If a cluster was not homogeneously, we repeated the procedure using different boundaries. The resulting homogeneous sets of titles indicate the nature of the various research topics. In this paper, we did determine the boundaries through visual inspection of the network representations, but in the future we aim to apply formal network measures to determine the boundaries of the clusters.

How should one read the following maps? The nodes in the network represent documents. The size of the node is proportional to the total number of ties (word-reference combinations) the node shares with other nodes. The more ties between two nodes, the stronger the tie, and the closer they are in the graph. The thickness of the line between two nodes indicates the strength of the link. Overlapping nodes strongly related.

Findings

Figure 1 shows the set of related publications in information science in 1986. In that year, the journals contained together 204 articles and reviews, the basic material for producing the map. About 47% of these 204 documents share a word-reference combination with at least one other document in the set. The core of the 1986 map is formed by information retrieval (subfield 1) in the form of a few dense (but unrelated) clusters of papers on topics like general information retrieval (1.1), information searching (1.2), catalogs and indexes (1.3) and social and political aspects of the information society (1.4). The second subfield in the map is scientometrics (subfield 2), which consists of a few mutually unrelated topics. The map shows a cluster of documents around indicators and mapping (2.1), and other topics such as scientometric distributions (2.2), science evaluation (2.3) and methods for quantitative analysis (2.4). An example of a third and small subfield visible in the map is library studies (3).

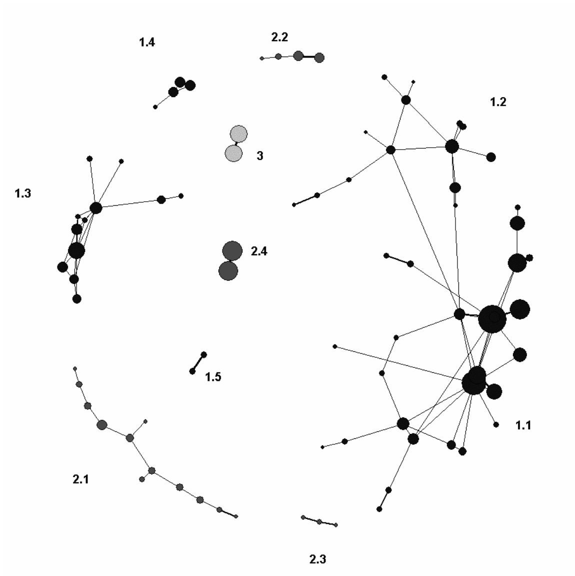

Figure 1. Information Science 1986

In 1992, the journals included 294 documents belonging to the categories of articles and reviews. Some 72% share a word-reference combination with another publication in the set. The map of 1992 (Figure 2) revealed that the subfields information retrieval (subfield 1) and scientometrics (subfield 2) again form the core of information science. The subfield of information retrieval is a large densely connected set of papers around topics such as IR methods (1.1), databases retrieval (1.2), IR and hypertext (1.3) and data-compression (1.4). The subfield of scientometrics is better represented than in 1986. It consists of the topics scientometric indicators (2.1), informetrics (2.2), citation analysis (2.3), and scientometric distributions (2.4).

In 1996, the map consists of 240 nodes, out of in total 314 articles and reviews (Figure 3). Information retrieval is again at the core of the field. Information retrieval is a large and densely connected set of papers around topics like information seeking and retrieval (1.1), methods in IR (1.2), retrieval performance and evaluation (1.3), information systems (1.4) and retrieval algorithms (1.5). Scientometrics remains the second subfield, and up to this year it is to a large extent based on papers published in the journal *Scientometrics*. In later years also other journals in the set increasingly publish about scientometric topics. Within this subfield, the most important topics are impact-factors (2.1), scientific collaboration (2.2), and citation analysis (2.3). The largest other research topic in 1996 is distance learning (3).

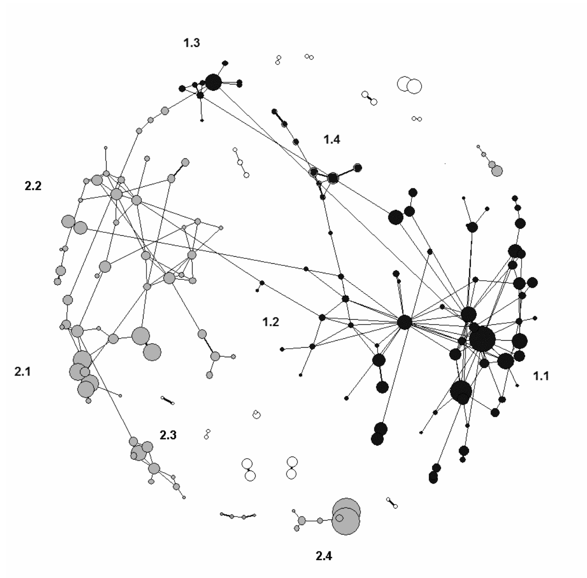

Figure 2. Information Science 1992

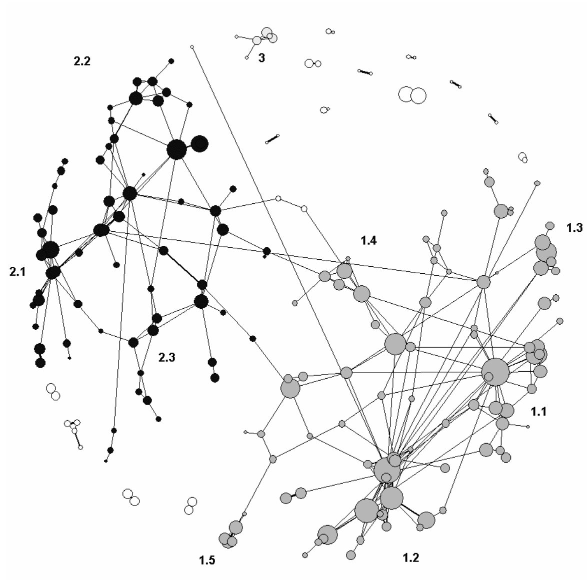

Figure 3. Information Science 1996

In 2000, the network consisted of 218 nodes (out of a potential set of 299). Figure 4 shows that several new research topics emerged, most importantly in the new subfield of web studies, such as web use (3.1) and scholarly communication on the web, such as e-publishing (3.2). The subfield of information retrieval consists of a cluster of papers about information searching (1.1), one on retrieval algorithms (1.2), and one on online searching and digital libraries (1.3). The latter is closely related to the topic of web use. The subfield of scientometrics consists of several connected clusters of papers around the topics like patentometrics (2.1), citation analysis (2.2), domain analysis (2.3) and scientometric distributions, (2.4). The resulting map is much more densely interlinked than in previous years, mainly because the clusters of papers within the new subfield of web studies are positioned in between the traditional subfields of information retrieval and scientometrics.

Also in 2002, the subfield of web studies is important (see Figure 5). It consists of clusters of papers around web use and searching (3.1), and digital libraries (3.2, two clusters). Next to web use research we find the related topics within the subfield of IR such as information searching (1.1), information retrieval (1.2), and methods for IR (1.3). The subfield of scientometrics consists of impact analysis (2.1) citation analysis (2.2), domain analysis (2.3), scientific communication (2.4), and methodologies (2.5). Finally a set of other topics can be discerned, such as theory of information (4), library research (5), and a new topic that has become fashionable: knowledge management (6, two unconnected clusters).

Summarizing, the core of information science consists of the subfields information retrieval and scientometrics. Information retrieval is still the largest subfield, although its relative importance has declined over the years. In 2002 it still accounted for about one third of all publications. Scientometrics has grown over the years up to about 30% of all papers. Recently, we witnessed the emergence of web studies, accounting for about one sixth of all papers in 2002. Finally, the other topics cover in average 20% of the papers in the journals we analyzed here.

On the level of research topics we see stability and change. A number of topics are present in all years, whereas others have disappeared. New topics emerge as a recombination of existing topics and in interaction with new (technological) developments. Examples are topics around the internet (such as CMC, digital libraries, distance learning) and more specifically around the WWW (such as web use, webometrics, web publishing, retrieval on the web). Table 2 lists the main topics in information science over the 1986–2002 period.

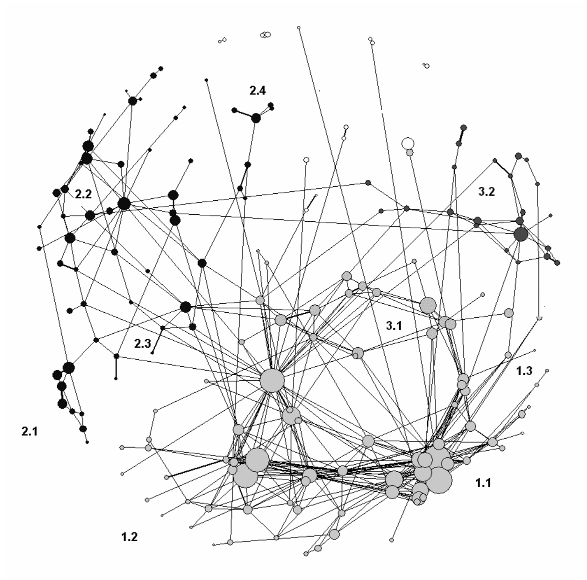

Figure 4. Information Science 2000

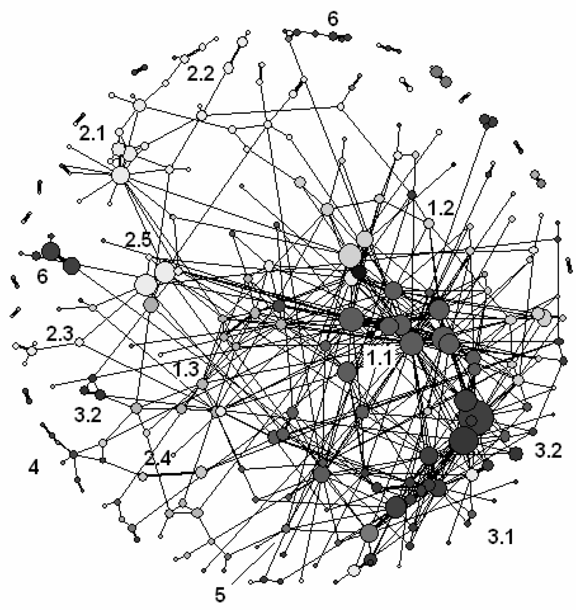

Figure 5. The network of topics of Information Science 2002

Table 2. Central topics in the main three subfields

	Information retrieval	Scientometrics	Web studies	Other (included in the maps)
1986	– information searching – social aspects of the information society – general IR – catalogs and indexes	– indicators and mapping – scientometric distributions – quantitative methods – evaluation		– libraries
1992	– database retrieval – methods in IR – data compression. – IR and hypertext	– indicators – citation analysis – informetrics – scientometric distributions		
1996	– information seeking – methods in IR – retrieval performance and evaluation – information systems – retrieval algorithms	– citations analysis – impact factors – scientific collaboration – (performance) indicators		– distance learning
2000	– retrieval algorithms – information searching – online searching	– citation analysis – patentometrics – scientific distributions – domain analysis	– webometrics – web use – scientific publications on the web (e–publishing)	– digital libraries
2002	– information searching – information architecture – document retrieval	– impact analysis – citation analysis – domain analysis – scientific communication – research methods	– web use – web searching	– libraries – theories of information – knowledge management – (digital) library services

In order to gain more insight in the development of the field over time, we created a map of all articles and reviews in the 1986–2002 period under study (Figure 6). The figure shows 1030 nodes (of a total of 1551 articles and reviews in this period) as well as the strongest 2000 of in total the 4454 relations between them.[*] It becomes visible that information science developed towards a tri-polar network, based on the subfields of information retrieval, scientometrics, and web studies. It is also clear that the field has changed over time, as the years (represented by different grey tones) are not evenly distributed over the map. The subfield of information retrieval is represented by a denser network than scientometrics. An example of a stable topic over the years within IR is retrieval methods, in the map indicated by a dense cluster of nodes from all years. Research on evaluation of retrieval systems was popular in the early years, but not so much any more in the more recent period: The cluster representing this topic only contains light colored nodes, indicating 1986 and 1992. On the other hand, empirical studies of information seeking are represented by a cluster of dark colored nodes:

[*] In this map one sees isolates, but this is not really the case: in order to improve readability, we do not show all links – the map would become very full. In the other maps all links are shown.

A more recent topic. The subfield of scientometrics also shows stable topics, declining topics and new topics, indicating a shifting research focus. The topic scientific collaboration is indicated in the map, and as the light colors of the nodes show, this topic is less prominent in the recent years. On the other hand, impact studies and scholarly communication are relatively stable over the years, whereas network analysis emerges as new topic within scientometrics. Finally, related to the recent topic of web use studies, we find the older topic of online information search.

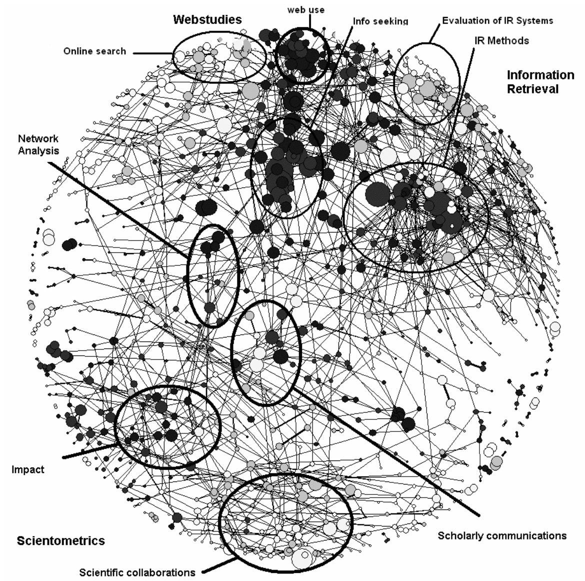

Figure 6. The changing network of topics in Information Science
(1986: white; 1992: light grey; 1996: medium grey; 2000: dark grey; 2002: black)

Conclusions and discussion

We introduced a method based on a combined use of title words and cited references, and this seems promising. The advantage of the method is that it combines both attributes of documents in determining the fine-grained topical structure of the research field under study. Consequently, the method is not founded in a single behavioral mechanism, but in two: citing strategies and word choice. The two-dimensional indicator reflects the subject of the research topic (through the title words) and its position within the specialty (through the references). The resulting maps give an insightful representation of the research topics within information science, and also the relationships between the different topics within subfields are made visible. The maps also give a fairly complete picture of the topics and subfields, as in average almost 70% of the published articles and reviews are represented in the maps. Only in the first year under consideration the share of included papers was low (47%), mainly due to the in that year still small size of the field. Elsewhere we tested the approach for other – rather turbulent – research fields, such as Artificial Intelligence, and also there it results into detailed maps of the research front (HEIMERIKS et al., submitted).

PETERS & VAN RAAN (1993) suggest that mapping should be based on abstract words instead of title words, as this improves the representation of research fields – but also results in a lower coverage of publications. As title words are less specific, a higher coverage can be expected when using these. Although this is a correct observation, it does not effect our results as we use *title words – cited reference combinations* which are much more specific than title words only. However, using abstract words may have another advantage when applied in our method: As abstracts are much longer than titles, they increase not only the quality of the representation, but also the number of possible co-occurrences of word-reference combinations. This may result in an increase (and not decrease) of the coverage of publications.[*]

The method we propose in this paper does not suffer from the problems of mapping techniques mentioned in the literature. Firstly, threshold setting is not a problem, as all papers that share at least a one word-reference combination with at least one other paper are included. Secondly, a higher percentage of source documents are included in the analysis than co-citation mappings usually have. Because of the large share of publications included in each year, we do not need to use data over relatively long periods to produce a map. This enables us to study also the short term changes within research fields. The case study illustrates this: The representation of information science is relatively stable over time, as it should be, and at the same time it shows the changes that are taking place. The change from the bipolar to a tripolar structure is nicely demonstrated. Thirdly, because of the high inclusion of papers, the proposed method is less vulnerable for over-representing specific categories of (theoretical) papers.

Several improvements can be thought of. First of all, we did not do any cleaning of the data (removing spelling errors) or preprocessing (reducing words to their stem). Doing this is expected to increase coverage as well as the quality of the representation. Secondly, it is worthwhile to repeat the analysis using also abstract words. This may result in a better and more complete map. Finally, we aim to extend the method by more formal ways of delineating the boundaries of the research topics, using techniques from social network analysis. Also the empirical base may be improved by using overlapping two years periods (e.g., 2000–2001, 2001–2002, 2002–2003, etc.). By doing this the maps become less dependent on the selection of the period, and this may also improve

[*] The *Social Science Citation Index* does not include the abstracts for the earlier years. Therefore we restricted ourselves in this paper to title words.

the visibility of the smaller research topics. Also the probability that papers share a word-reference combination increases, rising the already high percentage of included documents. However, the maps based on data from one year only (as we show in this paper) already are rather informative.

*

This paper is based on chapter 3 of the PhD thesis of Gaston Heimeriks (2005). A preliminary version was presented at ISSI 2005 (HEIMERIKS & VAN DEN BESSELAAR, 2005) – this version is thoroughly revised and corrected. The work was done in the context of two project partly funded by the European Commission: the SOEIS project and the EICSTES project. Loet Leydesdorff developed the software for preprocessing the data. The Bibtechmon software is developed by our colleagues at the Austrian Research Centres Seibersdorf (ARCS). Arie Rip, Eleftheria Vasileiadou and three anonymous reviewers provided us with useful comments.

References

BHATTACHARYA, S., BASU, P. K. (1998), Mapping a research area at the micro level using co-word analysis. *Scientometrics*, 43 : 359–372.

BRAAM, R. R., MOED, H. F., VAN RAAN, A. F. J. (1991 a), Mapping of science by combined co-citation and word analysis, I: Structural aspects. *Journal of the American Society for Information Science*, 42 : 233–251.

BRAAM, R. R., MOED, H. F., VAN RAAN, A. F. J. (1991b), Mapping of science by combined co-citation and word analysis, II: Dynamical aspects. *Journal of the American Society for Information Science*, 42 : 252–266.

CALLON, M., LAW, J., RIP, A. (1986), *Mapping the Dynamics of Science and Technology*. London: MacMillan.

CHEN, C. (2003), *Mapping Scientific Frontiers.* Springer

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

COZZENS, S. E., LEYDESDORFF, L. (1993). Journal systems as macro-indicators of structural change in the sciences. In: A. F. J. VAN RAAN, R. E. DE BRUIN, H. F. MOED, A. J. NEDERHOF, R. W. J. TIJSSEN (Eds), *Science and Technology in a Policy Context*, Leiden: DSWO/Leiden University Press, pp. 219–233.

HEIMERIKS, G. (2005), *Knowledge Production and Communication in the Information Society. Mapping Communications in Heterogeneous Research Networks*. Unpublished PhD thesis University of Amsterdam.

HEIMERIKS, G., VAN DEN BESSELAAR, P. (2005), Mapping research topics through word-reference co-occurrences. In: P. INGWERSEN, B. LARSEN (Eds), *Proceedings of ISSI 2005*. Stockholm: Karolinska University Press, pp. 575–584.

HEIMERIKS, G., VAN SOMEREN, M., VAN DEN BESSELAAR, P. (submitted), Structure and development of AI: a bibliometric analysis.

KESSLER, M. M. (1963), Bibliographic coupling between scientific papers. *American Documentation*, 14 : 10–25.

KING, J. (1987), A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, 13 : 261–276.

KOPCSA, A., SCHIEBEL, E. (2000), Science and technology mapping: A new iteration model for representing multidimensional relationships. *Journal of the American Society for Information Science*, 49 : 7–17.

LEYDESDORFF, L. (1997), Why words and co-words cannot measure the development of the sciences. *Journal of the American Society of Information Science*, 48 : 418–427.

MARSHAKOVA, I. V. (1973), A system of document connection based on references. *Scientific and Technical Information Serial of VINITI,* 6 (2) : 3–8.

MCCAIN, K. W. (1990), Mapping authors in intellectual space: a technical review. *Journal of the American Society for Information Science*, 41 : 433–443.

MOED, H. F., GLÄNZEL, W., SCHMOCH, U. (Eds) (2004), *Handbook of Quantitative Science and Technology Research,* Kluwer, Dordrecht.

NOYONS, E. C. M. (1999), *Bibliometric Mapping as a Science Policy and Research Management Tool*. Leiden: DSWO Press.

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 : 68–81.

PERSSON, O. (1994), The intellectual base and research front of JASIS 1986–1990. *Journal of the American Society for Information Science*, 45 : 31–38.

PETERS, H. P. F., VAN RAAN, A. F. J. (1993), Co-word-based science maps of chemical engineering. Part 1: Representations by direct multidimensional scaling. *Research Policy*, 22 : 23–45.

RIP, A. (1988), Mapping of science, possibilities and limitations. In: A. F. J. VAN RAAN (Ed.), *Handbook of Quantitative Studies of Science and Technology*. Amsterdam: Elsevier Science, pp. 253–273.

SHIFFRIN, R. M., BÖRNER, K. (Eds) (2004), *Mapping Knowledge Domains*. Supplement of *PNAS* 101 (April 6, 2004)

SMALL, H. (1973), Co-citation in scientific literature: a new measure of the relationship between publications. *Journal of the American Society for Information Science*, 24 : 265–269.

VAN DEN BESSELAAR, P. (2000), Communication between science and technology studies journals: A case study in differentiation and integration in scientific fields. *Scientometrics*, 47 : 169–193.

VAN DEN BESSELAAR, P. (2001), The cognitive and the social structure of Science and Technology Studies. *Scientometrics*, 51 : 441–460.

VAN DEN BESSELAAR, P., HEIMERIKS, G. (2000), *Codification and Self-Organization in the European STI System.* (Deliverable 2.6 – SOEIS project). University of Amsterdam.

VAN DEN BESSELAAR, P., Heimeriks, G. (2001), Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In: M. DAVIS, C. S. WILSON (Eds), *Proceedings 8th International Conference on Scientometrics and Informetrics - ISSI2001*. Sydney: UNSW, pp. 705–716.

VAN DEN BESSELAAR, P., LEYDESDORFF, L. (1996), Mapping change in scientific specialties; a scientometric case study of the development of artificial intelligence, *Journal of the American Society of Information Science,* 47 : 415–436.

WHITE, H. D., MCCAIN, K W. (1989), Bibliometrics. *Annual Review of Information Science and Technology*, 24 : 119–186.

WHITE, H. D., MCCAIN, K. W. (1997), Visualization of literatures, *Annual Review of Information Science and Technology*, 32 : 99–168.

WHITE, H. D., MCCAIN, K. W. (1998), Visualizing a discipline: an author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, 49 : 327–356.