

Meeting the micro-level challenges: bibliometrics at the individual level

Erik Sandström and Ulf Sandström

Paper to the 12th Conference on Scientometrics and Informetrics July 14-17, 2009, Rio de Janeiro, Brazil

Abstract

The aim of this paper is to demonstrate a method for the evaluation at the individual level of research staff currently employed within a university department or other knowledge organisations with research purposes. Based on methods for citation analysis and methods for clustering of papers into research lines (bibliographic coupling) we present an analysis of one researcher in three dimensions: 1) publication and citation indicators; 2) publication profile, and 3) research lines. One of the features of the method is the benchmark against other researchers within the same research line, i.e. researchers that use the same references and, accordingly, are active in the same field of research. The paper suggests this method as a means to avoid the fallacies of evaluation based on sub-field categories in the Web of Science. The method was used in a Research Assessment Exercise accomplished in the autumn of 2008 at Royal Institute of Technology.

Introduction

Nowadays performance assessment usually includes the use of publication and citation counts. The reward system in science is based on recognition, and this emphasizes the importance of publications to the science system. Bibliometric analysis is a fundamental asset for evaluative studies. Beside publication and citation analysis the bibliometric mapping techniques enable visualizations of the positioning of research groups in relation to their research communities. Bibliometrics makes it possible to visualize performance for the author and the subject clusters relevant for the units of assessment. In this paper we try to illuminate the activities of the researcher by making use of several different methods. These methods are, to some extent, not that transparent to the reader, e.g. clustering techniques rests on thresholds and several other variables, but at the same time the resulting maps are often easier to understand and carry more contextual information of interest to all parties in an evaluation.

Theories of citing

The choice of citations as an important indicator calls for a theory of citing: a theory that makes it possible to explain why author x cited article a at time t . What factors should be considered when we discuss why researchers cite back to former literature? The need for a theoretical underpinning of citation analysis has been acknowledged for a long time and several theories have been put forward. In short, there are three types of theories: normative, constructive and pragmatic. Normative theories are based on a naïve functionalist sociology, and constructivist theories are opposed to these assumptions. The third alternative, the Nordic pragmatist school (e.g. Seglen, 1998, Luukonen, 1997, Amsterdamska & Leydesdorff, 1989; Aksnes 2003), emphasizes utility in research as an important aspect, and cognitive qualities another, and together they are criteria for reference selection. Based on Cole (1992) the Norwegian Aksnes (2003b) have introduced the concepts of quality and visibility dynamics in order to depict the mechanisms involved.

Factors like journal space limitations prevent researchers from citing all the sources they draw on; it has been estimated that only a third of the literature base of a scientific paper is rewarded with citations. Therefore, citation does not implicate that the cited author was necessarily “correct”, but that the research was useful. We should not forget that negative findings can be of considerable value in terms of direction and method. If a paper is used by others, it has some importance. In retrospect, the idea or method may be totally rejected; yet use of the citation is clearly closer to “important contribution to knowledge” than just the

publication count in itself. The citation signifies recognition and typically bestows prestige, symbolizing influence and continuity.

From the view of the pragmatist citation school, a discussion of the limits of citation counting is necessary. As stated above, not all works that “ought” to be cited are actually cited, and not all works that are cited “ought” to be. As a consequence, the validity of using citation counts in evaluative citation analysis is contestable. Even if the quality of the earlier document is the most significant factor affecting its citation counts, the combined effect of other variables is sufficiently powerful and much too complex to rule out positive correlations between citation count and cited-document quality.

The view taken in this paper is that authors cite earlier work in order to substantiate particular points in their own work. Therefore, on a general level the citation of a scientific paper is an indication of the importance that the community attaches to the paper, or the pragmatic value connected to the paper. If it is used by others there is a cognitive quality in the paper and this quality is measurable by citations, preferably with normalization according to field.

The field and sub-field problematique

In bibliometric studies the definition of fields is generally based on the classification of scientific journals into more than 250 categories, developed by Thomson Reuters. Although this classification is not perfect, it provides a clear and consistent definition of fields suitable for automated procedures. However, this proposition has been challenged by several scholars (e.g. Leydesdorff, 2008; Bornmann et al. 2008). Two limitations have been pointed out: (1) multidisciplinary journals (e.g. Nature; Science); and (2) highly specialized fields of research. The Thomson Reuters classification of journals includes one sub-field category named “Multidisciplinary Sciences” for journals like PNAS, Nature and Science. More than 50 journals are classified as multidisciplinary since they publish research reports in many different fields. Fortunately, each of the papers published in this category are subject specific, and, therefore, it is possible to assign a subject category to these on the article level – what Glänzel et al. (1999) calls “item by item reclassification”.

We have followed that strategy in this paper. Thus, an automated procedure was put into place: papers in multidisciplinary journals are assigned to a new field based on the field representation of the citing journals and cited journals. If the majority of the citations to a paper published in a multidisciplinary journal come from neuroscience journals and the majority of the cited references in the paper are to neuroscience journals, the paper will be assigned to neuroscience. In short, a paper is assigned to the field in which the largest number of its references and citations are classified. The benefit of reclassification is that statistics for fields, including author, institution, country, journal and paper rankings, more accurately reflect all papers in these fields, including those found in multidisciplinary journals, some of which publish influential, highly cited research reports.

The second issue of highly specialized fields within sub-fields is less easy to resolve. Lewison, in a debate with van Raan, pinpoints some of the reasons for questioning journal-dependent classifications; e.g. traditional delineation does not distinguish between specialist and non-specialist journals (Lewison, 2005). Bornmann et al. (2008) reports a case study of one neurology group and different relative citations scores depending on if they used Thomson Reuters sub-fields or used MEDLINEs Medical Subject Headings (MeSH) assignation item-by-item. The latter methodology seems more appropriate than the classification from journals. MEDLINEs hierarchical structure is, of course, more fine-grained than the Thomson Reuters classification, but there is no such database for other scientific fields. Therefore, as analyst’s we are thrown upon the existing resources i.e. the Thomson Reuters.

However, contributors to the debate like Leydesdorff have expressed their doubts: As there are no agreed-upon alternatives, the Thomson Reuters subject categories are often used for “comparing like with like”. In Leydesdorff interpretation these categories are assigned by the Thomson Reuters staff on the basis the journal and its citation patterns. An obvious problem is that the classification matches poorly with classifications derived from the database itself on the basis of clustering analysis. Using such a methodology researchers have found that in more than 50% of the cases the Thomson Reuters categories corresponded closely with the clusters based on inter-journal citation relations (Leydesdorff, 2008).

While preparing for this paper we investigated whether a bibliographic coupling performed in two or three steps would produce more coherent field delineation. Results indicated that there is no simple method, e.g. bibliographic coupling, that would be suited for developing a new and better classification. This is mainly due to the fact that there are many relations between fields depending on different roles for basic research on the one hand and applied research on the other. A reclassification based on bibliographic coupling would mix basic and applied in such a way that it would harm the results. Therefore, our conclusion is that, with small fine-tuning, the field definitions and boundaries used by the Thomson Reuters are very well adapted to the needs of a pragmatic evaluative approach. This is confirmed by the Leiden group “Groupings based on citation relationships, however, are less stable, because journals in related fields and multidisciplinary journals often have changing citation relationships with the ‘core’ journals of a field over the course of time.” (CWTS, 2007:63)

Still, there are concerns regarding the field delineations. They have been challenged and sometimes we find groups of researchers that are in between of fields; these groups might come into a citation shadow of dominate research areas within the field, e.g. a group in Mathematics that publishes in Physics and hence is compared with other publications in physics. If we compare the groups within this specific area of research and separate the addresses from physics departments and those from mathematics we find that articles with the latter addresses systematically have a lower citation rate. Clearly, we should be cautious when it comes to field normalization. The dynamics of science constantly opens up new research lines that often are combinations at the border between research areas (or disciplines). The Thomson Reuters journal classification has to be quite stable over time; consequently this will produce inconsistencies.

In their article on dimensions of citation analysis, Leydesdorff & Amsterdamska (1990), showed that comparing “like with like” (Martin & Irvine 1983) is very hard to achieve. Although they had similar research teams with similar performances the detailed analysis gave a conclusion that pointed in a critical direction:

Our analysis suggests, however, that the criterion of what constitutes sufficiently similar groups is itself problematic. Despite the fact that the four papers we compared originated from the same small laboratory, they were clearly addressed to four distinctly structured communities. The local institutional identity of the papers’ origin is not sufficient to assure the similarity of their audiences, nor does it provide clues for explaining the papers’ perceived significance and reception. (pp. 324-5)

In order to handle these above mentioned problems we have developed a method that makes it possible to compare the groups production and citations score in a research line with all papers in that specific research line: is the group performing better, in line with or less well than their colleagues in that specialized area of research? We continue to use the fields proposed by Thomson Reuters, but we present the results in such a way that it possible to see whether the group performs according to the standards in their respective research lines. This can be seen as an alternative, and more fine-grained, method for normalization. If the two methods produce the same results, the figures are obviously more significant. The method is presented below under the heading “Clusters and Research Lines”.

Clusters and research lines

We have seen that there are no perfect measures for evaluations of research. Probably, the best way to handle this is to develop complementary indicators or to develop other indirect measures that illuminate the performance of researchers and research groups in different ways. One of the challenges in visualization of bibliometric data is to find a method that suits both the highly productive (>500 papers) and the lowly productive (<20 papers). In the following we have chosen to focus on visualizations as a means to enhance the evaluative properties of the analysis.

Thereby we used clustering and mapping techniques for research team activities over the period 2000-2006. The goal of cluster analysis is to divide data into a number of subsets (clusters) according to some given similarity measure (Chen, 2006). As already indicated it is quite problematic to delineate a research field with journals or journal categories; evidently, a researcher is often active in several of the more than 250 subject fields. Why is mapping important to evaluative bibliometrics? The obvious answer is, firstly, that mapping provides us with a profiling of the research group or the individual researcher. Accurate maps of the actual articles and their related research lines give us a description of what is going on in that field of research; how articles are related to each other; which research lines that are more close to each other and how they are connected. Secondly, and as important as the first, it gives a complementary illumination of the activities of the research group in relation to their closest colleagues. Mapping techniques makes it possible to compare the individual/group – regarding citation performance – with its nearest neighbours; i.e. other researchers working on the same topics and in the same research lines. In the maps covering articles from each unit of assessment there is also an indication of highly cited articles (and clusters of highly cited documents). For the reader this should be quite easy to interpret and analyze.

Our methods owe much to the work of Chaomei Chen at Drexel University and to the work of Dick Klavans and Kevin Boyack at SciTech Strategies Inc. We have developed a methodology based on bibliographical coupling in order to build accurate and coherent maps of areas and research lines (Boyack et al., 2005; Chen, 2006; Klavans & Boyack, 2006a; Klavans & Boyack, 2006b). For a detailed description see the technical appendix.

The mapping of article networks is a procedure to identify clusters of articles in correspondence to the underlying thematic groupings; hopefully as they are perceived by scientists themselves. The core of our method relies on bibliographical coupling of documents (articles, not journals) into research lines; groups of documents that cite the same base documents. This is an established method within bibliometrics and has been shown, by Jarneving (2007), to work well for the purpose of clustering of related papers into coherent groups. We consider research as an activity conducted by “small groups of people attacking equally small and intensely focused sets of shared problems” (Morris et. al., 2003: 413). The highly cited papers of these research lines are the research fronts of that specific research area. The immediate nature of bibliographic coupling makes it convenient to use for constructing research lines over a time period (Morris et al. 2003: 414). In our presentations each research line is described, to the left, with the most frequent terms (keywords), and to the right, the most frequent individual authors and their number of articles. We consider this method a valuable asset for evaluative bibliometrics as it visualizes the development of the research line over time – growth, decline or stability. But, we should mention that the time line is rather short in the reported analyses. A period of seven years is in most cases quite limited.

In this paper we propose that the most frequent research lines, in which the research groups are active, should be considered as a representative for their overall research. We try to avoid measuring too many of the single and more casual lines of research that a group performs. Instead, we focus our interest on the larger and more stabile research lines. Moreover, in addition to the general performance of the research line our analysis also gives the field

normalized citation score for the specific articles from the research group under consideration. Those figures can be compared with the score of the research line as a whole.

Professor Åke Bergman

In order to illustrate our method we have chosen to present the work of Professor Åke Bergman at Stockholm University. Bergman is active at the Department of Environmental Chemistry, one of the six departments in chemistry at the university. Bergman has his basis in chemistry with strong interdisciplinary and multidisciplinary links to research, teaching and international activities on chemicals and the environment. His and the department work is in particular focused on chemical synthesis, characteristics, exposure of chemicals, human health and wildlife effects. The research encompass the transport, fate and biological effects of environmental pollutants and trace substances, including the development and use of chemical, biological and physical methods as well as numerical models. We have chosen an outstanding researcher as it much more pleasant and interesting to present excellent research than the other way around. On the following three pages there is information showing results at individual, or personalized, level.

Page 1 BIBLIOMETRIC INDICATORS

The first page gives the bibliometric indicators for the individual. At the bottom there are two graphs, to the left number of papers per year, to the right vitality (reference age). It is hard to find relevant data for time series on the individual level. Vitality might be a viable indicator as it has other features than the citation indicators. Number of references are more or less stable within research areas and the reference age might not deviate over time that much. But, if we see a lowering of the reference age it might indicate that the researcher is getting closer to research at the front.

Page 2 PUBLICATION PROFILE

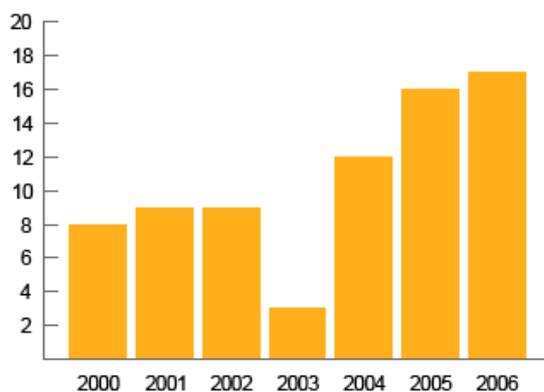
This map includes all articles that belong to the research lines (see next paragraph) to which the individual has been assigned by his or her articles. Papers written by the individual under consideration are marked yellow. Relations are based on similar references (bibliographic coupling). Papers characterized by high impact scores (>4.00) are highlighted by pink color rings; a group with a number of high impact papers will have many pink nodes in the picture. Pink lines between nodes indicate a low reference age (high vitality). We interpret higher vitality as tendency to be closer to the research front and a higher probability to impact on knowledge production at this front. At the bottom, the reader find tables of most frequent journals, most frequent collaborating institutions, and most frequent ISI sub-fields. Unification of names of the most frequent institutions has been applied. Below the map is found the most frequent keywords on articles from the individual, the most frequent co-author names, and the most frequent co-authoring institutions.

Page 3 RESEARCH LINES

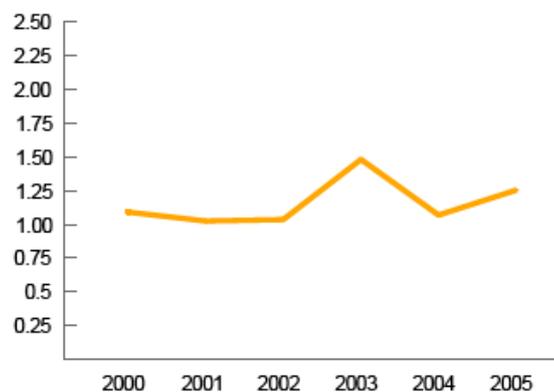
Research lines are built via a clustering based on the bibliographical coupling of all articles in the Thomson Reuters database from 2000–2006. These small communities of articles use the same references and should, therefore, deal with more or less the same topic. The idea is to demonstrate the distribution of articles (grey fields) and citations (pink lines; 2yr citation window). It is possible thereby to see the development over time of the research line.

BERGMAN, ÅKE - BIBLIOMETRIC INDICATORS

NUMBER OF PAPERS (P)	97
Number of papers (articles, letters and reviews) published by Bergman, Åke during 1998-2006.	
NUMBER OF FRACTIONALIZED PAPERS (Frac P)	18.3
Sum of author fractionalized papers.	
CITATIONS PER PAPER (CPP)	29.5
Number of citations per paper (31 December 2007).	
JOURNAL NORMALIZED CITATION SCORE (NCSj)	2.16
CPP normalized in relation to the Bergman, Åke journal set (average=1.00).	
NORMALIZED JOURNAL CITATION SCORE (NJCS)	1.53
The impact of the journal set normalized in relation to its sub-fields (average=1.00).	
FIELD NORMALIZED CITATION SCORE (NCSf)	3.42
CPP normalized in relation to the Bergman, Åke sub-field set (average=1.00).	
SUM OF FIELD NORMALIZED CITATION SCORE (Sum NCSf)	62.6
NCSf times Frac P.	
STANDARD FIELD CITATION SCORE (SCSf)	1.02
Z-score standardized citation score in relation to the Bergman, Åke sub-field set (N.B! average=0.00).	
TOP 5% (TOP5%)	31
Percentage of papers above the 95th citation percentile.	
VITALITY	1.17
Mean reference age normalized in relation to the sub-field set (average=1.00, higher=younger).	
PERCENTAGE SELF CITATION (SelfCit)	6
Percentage self-citation.	
PERCENTAGE NOT CITED PAPERS (Pnc)	5
Percentage of not cited papers during the period.	
HIRSCH INDEX (H-INDEX)	29
The h number papers that have at least h citations each.	
AUTHOR MEAN (AUm)	5.3
Mean number of authors per paper.	
INTERNATIONAL COLLABORATION MEAN (IntCOLLm)	1.8
Mean number of countries per paper.	

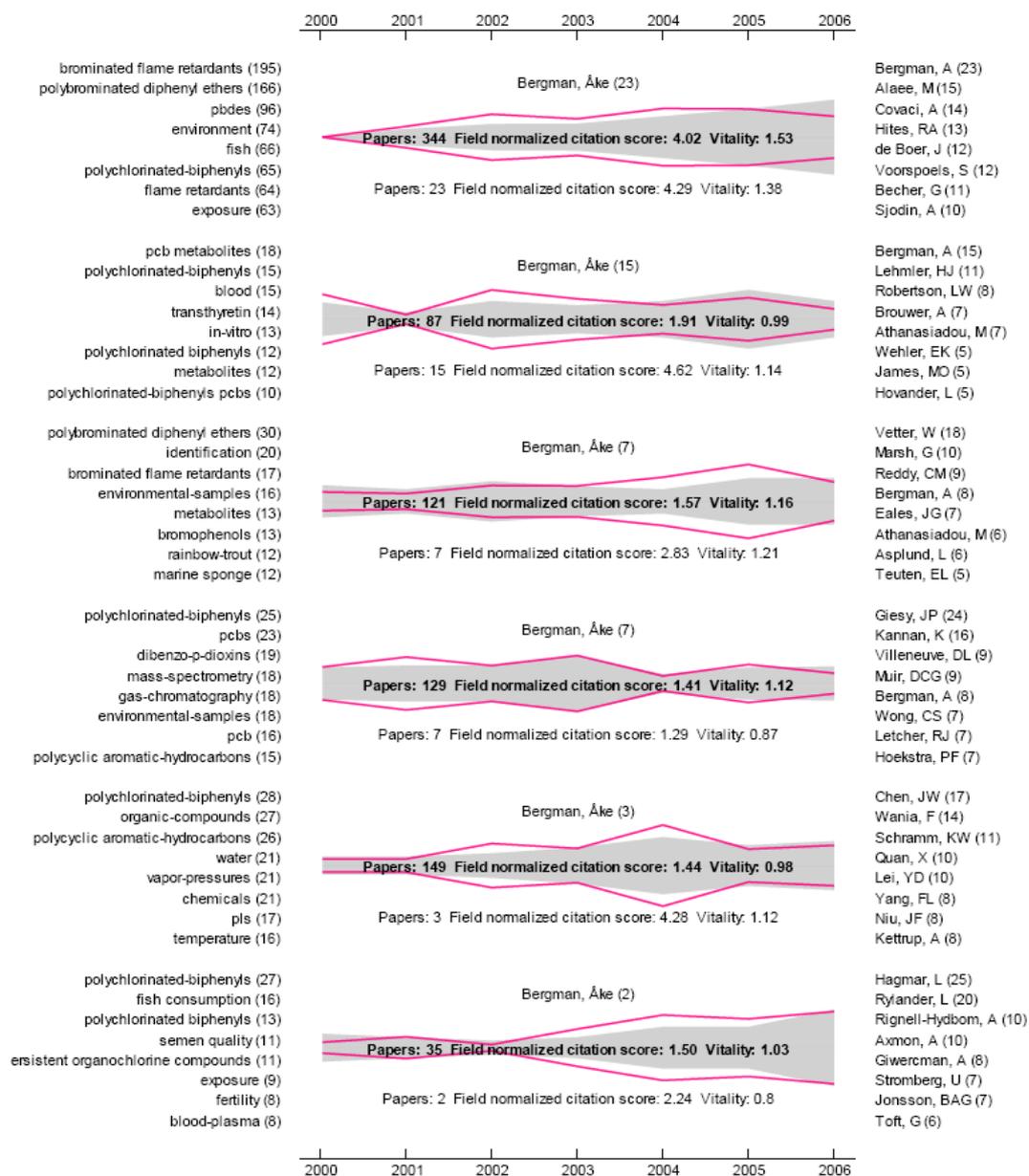


Number of papers per year



Vitality per year

ReproSafe // Stockholm University
BERGMAN, ÅKE - RESEARCH LINES



Research lines are clusters of similar papers based on bibliographic coupling. Most frequent keywords are shown to the left and most frequent authors to the right. Grey fields shows the proportion of papers each year. Pink lines shows the proportion of citations (2 year citation window) each year. Number of publications, field normalized citation score and vitality for each research line are shown at each line. Performance indicators for Bergman, Åke are shown below the line.

The total number of articles, the field normalized citation score, NCSf, and the vitality score, are displayed on the line. Researchers from the UoA (top three) active in that specific line of research are shown above the line. Presented under the line are the field normalized citation score and vitality for the UoA publications in that line of research.

As a rule of thumb we estimate that a positive difference of factor 2 or more indicates that the person is influencing the research front of that specific research line, given, of course, that there are a sufficient number of publications. To the left we show the most frequent keywords in the research lines and to the right the most frequent authors. Above the line is shown the individuals number of papers within the research line and below is displayed the field normalized citation score.

The innovative function of these research lines are that researchers can compare and benchmark themselves with their closest colleagues, those that are in the same field of research and are utilizing the same references in their research. The publication profile gives the researcher his or her positioning within the field of research. From the information given it is possible to discuss strategic action, e.g. on directions of future research.

Going deeper into the research lines the visualization show the growth of the line over time. If the grey area is expanding there are a growing number of papers per year. Inversely, there is a diminishing number of papers if the grey area is becoming smaller over time. Analogously, the red (or pink) lines show the development of number of citations over time. Several of the Bergman research lines seem to be expanding over time.

As stated above, the most important feature of the method is that we can compare the citation score of Åke Bergman with his colleagues. We find that Bergman is well above the mean citation rate in all lines except one. The most frequent research line for Bergman is a very highly cited line and he has co-authored a high number of papers. Obviously, Bergman is active in a number of fields which receive citations from other lines/areas.

Discussion

Acknowledging the influence from Boyack, Klavans and Morris we underline that this is the first version of the proposed method. Certainly, there is room for further methodological (and theoretical) development. The method seem to work well in productive environments but results from social scientist or other areas with few papers per person indicate that the method does apply only if there is cumulative knowledge production.

More on weaknesses and shortcomings of the method will be discussed in the conference presentation.

References

- Aksnes DW (2003a). A macro study of self-citations. *Scientometrics* 56(2):235–246.
- Aksnes DW (2003b). Characteristics of highly cited papers. *Research Evaluation* 12 (3): 159–170.
- Amsterdamska, O & Leydesdorff, L (1989). Citations: indicators of significance? *Scientometrics* 15 (5-6):449–471.
- Bornmann, L, Mutz, R, Neuhaus, C, Daniel H-D (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics Sci Environ Polit*, 8: 93–102.
- Boyack KW (2007). Using detailed maps of science to identify potential collaborations. *Proceedings of ISSI 2007*, edited by Torres-Salinas & Moed. Madrid, Spain June 25-27, 2007. Vol 1, s. 124-135.
- Boyack KW, Börner K (2003) Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and citation counts of research papers. *JASIST*, 54 (5): 447–461.

- Boyack KW & Klavans R (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology* 57 (2): 251–263.
- Boyack KW, Klavans R, Börner K (2005). Mapping the backbone of science. *Scientometrics*, 64: 351–374.
- Chen, CM (2006). *Information Visualization: beyond the horizon*. 2nd Edition. London: Springer Verlag.
- Cole S (1992). *Making science: between nature and society*. Cambridge, Mass.: Harvard University Press.
- CWTS (2007). Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions. Report to HEFCE by the Leiden group. November 2007. [http://www.hefce.ac.uk/pubs/rereports/2007/rd18_07/rd18_07.pdf]
- Garfield, E (1979). *Citation indexing - its theory and application in science, technology, and humanities*. New York: Wiley.
- Glänzel W, Schubert A, Schoepflin U, et al. (1999). An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis. *Scientometrics*, 46 (3): 431–441.
- Jarneving B (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics* 65 (2): 245–263.
- Jarneving B (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics*, 1 (4): 287–307.
- Kamada, T & Kawai, S (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15.
- Kessler, MM (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14, 10- 25.
- Klavans, R & Boyack, KW (2005). Mapping world-wide science at the paper level. *Proceedings of ISSI 2005*, pp. 426–436.
- Klavans, R & Boyack, KW (2008) Thought leadership: A new indicator for national and institutional comparison. *Scientometrics* , 75 (2): 239–252.
- Lewis, G. (2005). Commentary on “Measurement of Central Aspects of scientific Research”. *Measurement*, 3 (1): 29–32.
- Leydesdorff, L. (2008) Caveats for the Use of Citation Indicators in Research and Journal Evaluations. *Journal of the American Society for Information Science and Technology*, 59(2): 278–287.
- Leydesdorff, L & Amsterdamska, O (1990). Dimensions of citation analysis. *Science, Technology & Human Values* 15 (3): 305–315.
- Martin B.R. & Irvine J. (1983). Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12: 61–90.
- Martin, SB, Brown, WM, Klavans R & Boyack, KW (2008 submitted). DrL: distributed recursive (graph) layout. *Journal of Graph Algorithm and Applications*.
- Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology*, 54(5), 413–422.
- Seglen, PO (1998) Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthop Scand* 69 (3): 224–229.
- Small, H. (1993). Macro-level changes on the structure of co-citations clusters: 1983–1989. *Scientometrics* 26 (1): 5–20.
- Visser, M.S. Nederhof, A.J. (2007) Bibliometric study of the Uppsala University, Sweden, 2002–2006. In: *Quality and renewal 2007: An overall evaluation of research at Uppsala University 2006/2007*. Uppsala: Uppsala University.