MEASUREMENT, *3*(1), 50–62

For Your Citations Only? Hot Topics in Bibliometric Analysis

Anthony F. J. van Raan
Centre for Science and Technology Studies
University of Leiden

In this rejoinder, I structure my comments around a number of main topics discussed by the authors of the commentaries.

ASSESSMENT OF THE BIBLIOMETRIC INSTRUMENT

This is an important point (Veugelers, this issue): As in all empirical research, one has to know not only "what" is measured, but also "with what" one is measuring. In other words, one has to know the characteristics of the measuring instrument. In this context, it is crucial to realize that the "bibliometric instrument" can be used for two main tasks: the evaluation of research performance (Garfield, 1979), and the study of science as a knowledge-generating and communication system and its interaction with technology (Feller, this issue; Veugelers, this issue). For a proper application of bibliometric indicators, the technical system on which these indicators are based must, first of all, be sufficiently advanced and sophisticated.

A short discussion of several crucial technical issues that also illustrate the limitations of bibliometric analysis is given in Intermezzo 1. Methodological issues are under discussion in most of the further comments in this rejoinder. For a detailed, general discussion on the validity and reliability of evaluation of scholarly performance, I refer to Nederhof (1988).

Requests for reprints should be sent to Anthony F. J. van Raan, Centre for Science and Technology Studies, University of Leiden, Wassenaarseweg, 52, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: vanraan@cwts.leidenuniv.nl

INTERMEZZO 1: TECHNICAL PROBLEMS

The most central technical process, on which citation analysis is entirely based, is the matching of citing publications with cited publications. With this process, the following is meant. When in a publication (which is the “citing publication”) a reference is given to another publication (which is the “cited publication”), the reference has to be identified as an earlier “source publication” in the citation indexes. This in fact is the most important technical procedure that has to be executed by any producer of citation indexes. A considerable number of errors may occur in the citing–cited matching process, leading to a “loss” of citations to a specific publication. Frequently occurring nonmatching problems relate to publications written by “consortia” (large groups of authors); variations and errors in author names, particularly—but certainly not only—authors from non-English speaking countries; errors in journal volume numbers; errors in initial page numbers; discrepancies due to journals with dual volume-numbering systems or combined volumes; or journals applying different article numbering systems. These nonmatching citations are highly unevenly distributed in specific situations, which may cause an increase of the percentage of lost citations of up to 30% (Moed, 2002). So if the citation indexes are used for evaluation purposes, all of these possible errors have to be corrected as much as possible.

A second important technical problem relates to the attribution of publications—and with that, of the citations to these publications—to specific organizations such as institutes, university departments, and even on a high aggregation level to the main organization—for instance, universities. Very often, it is thought that the citation indexes can simply be scanned to find all publications of University X. This assumption is based on the argument that all these publications clearly mention somewhere in the address data of the publication “University X” as the main affiliation of the authors. But this assumption does not hold at all. Next to variations in the name of the same university, departments and institutes (in many variations) are to a nonnegligible extent mentioned without proper indication of the university. Furthermore, groups or institutes of a national research organization (such as the French Cxxxx Nxxxx Rxxxx Sxxxx) are quite often mentioned instead of the university where the research actually takes place. The same problems occur for “graduate schools.” Even the name of a city as such will not always be sufficient, as parts of universities may be located in suburbs.

Another difficulty with large research organizations is to distinguish the many institutes within the main organization. Very problematic in this respect is capturing the medical research of universities, as often only the medical school and/or the name of the hospital is mentioned without the university being indicated. One explicitly needs the names of all those hospitals in a specific city—and also in the suburbs of a city—that are in fact university hospitals. So again, problems related to names of cities and their suburbs play a crucial role. Again, large efforts in cleaning and rebuilding the original Institute for Scientific Information (ISI) citation indexes are necessary to solve this problem.

On top of that, further problems arise when two or more universities are within one city. In some cases, these problems are so large (e.g., in the case of Vrije Universiteit Brussel and the Université Libre de Bruxelles, both are indexed as "Free University [of] Brussels") that it is virtually impossible to distinguish both universities on citation-index based address data only. Similar problems occur for Leuven and Louvain-la-Neuve. Next, there are major differences in research systems between countries affecting the definition of a university. For instance, the University of London is not a university anymore in the usual sense. It is an umbrella organization covering several different, virtually autonomous universities. But in Paris and other French cities, no such umbrella structure exists. There we deal with completely autonomous universities that were originally part of one mother university. As a consequence, it is very cumbersome to distinguish between departments, medical schools, and hospitals of these different universities within a city.

All these affiliation-related problems require very careful and mostly cumbersome cleaning and, in addition and most important, as discussed previously, unification of addresses and attribution of publications to the proper main organization, including unification of cities with their suburbs. As in the case of the earlier mentioned technical problems related to discrepancies in cited references, large and continuous efforts are necessary to rebuild the citation indexes into a data system that is accurate enough to be used for the calculation of bibliometric indicators and to be applied for evaluation purposes.

Work by Other Groups

My article is written mainly as a tutorial based on the many practical experiences of our group in Leiden, with an emphasis on standardization. Given this approach, I can understand that, in the comments, a high number of self-citations and even a neglect of work by other groups (Lewison, this issue; Veugelers, this issue) are mentioned. A more comprehensive overview of measuring science—with ample discussion of, and references to, the advances made by other groups—is given in the first chapter of the new *Handbook of Quantitative Science and Technology Research* (van Raan, 2004a).

Journal Coverage

Any index system has, by definition, a restricted coverage. This has advantages—for instance, a focus on core literature in, at least, most of the natural and life sciences—and disadvantages (Feller, this issue; Lewison, this issue; Veugelers, this issue), for instance, a too limited view on scientific communication, particularly in the applied sciences, social sciences, and humanities. Nevertheless, also in these latter disciplines—but to a lesser extent in the humanities—researchers tend to publish more and more in core literature journals (Glänzel, 1996; Hicks, 1999).

Broader journal coverage of a publication database as compared to the Citation Index (CI) system (Veugelers, this issue) will, however, introduce new difficulties, in particular those similar to the impact dilution effects of non-English language journals covered by the CI. The reason is the (very) low impact of these journals (i.e., of most of the publications in these journals). Problems may occur even at high aggregation levels, such as entire countries—see Intermezzo 2. Extension of journal (and other media) coverage will introduce similar problems, as most of the added journals will have a considerably lower impact as compared to the already covered journals. So I think that any large-scale extension of the CI will inevitably lead to an even stronger and more status-related distinction between core and peripheral journals. A new discussion about what constitutes scientific impact in the context of a much broader index system will certainly arise.

INTERMEZZO 2: LANGUAGE PROBLEMS

Recent work (Grupp, Schmoch, & Hinze, 2001; van Leeuwen, Moed, Tijssen, Visser, & van Raan, 2001) shows that the utmost care must be taken in interpreting bibliometric data in a comparative evaluation, even in the case of national research systems (May, 1997). The measured value of impact indicators of research activities at the level of an institution and even of a country strongly depends on whether one includes or excludes publications in CI-covered journals written in languages other than English. This is due to the simple fact that the CI covers non-English language (particularly French- and German-language) journals of which the articles have a considerably lower impact than those in the English-language journals. So if researchers use these non-English language journals for publications, these publications are indeed formally entered into the citation indexes. But generally, the impact of publications of these French- and German-language journals is (very) low. So in the calculation of impact indicator, these publications count on the output side, but they contribute very little, if any at all, on the impact side.

Therefore, such non-English language publications considerably dilute the measured impact of a university or a department. We have clear empirical evidence (van Leeuwen et al., 2001) that the use of German-language journals covered by the citation indexes may lead to about 25% lower measured impact. Simply by removing the publications in these German-language journals and using only the English-language journals (which is fair in an international comparison and certainly in a comparison with, e.g., the United States and the United Kingdom), the measured impact will "improve" 25% for the whole medical faculty of a university! No doubt there will be the same effect for French-language journals. These findings clearly illustrate again that indicators need to be interpreted against the background of their inherent limitations, such as, in this case, effects of publication language, even at the macro level of entire countries, but certainly at the level of institutions.

The comments concerning the monopoly position of the Institute for Scientific Information (ISI/Thomson Scientific)—and, partly related to this, the poor coverage in fields like the social sciences and the humanities—are certainly important and directly related to the previously mentioned discussion (Veugelers, this issue). Weingart (2003) pointed to the very influential role of the monopolist citation data producer ISI/Thomson Scientific, as its commercialization of these data (Adam, 2002) rapidly increased the nonexpert use of bibliometric indicators such as rankings (Shanghai Jiao Tong University, 2003). I refer to the recent discussion initiated by Weingart (van Raan, in press-a; Weingart, 2003). Probably, the new publication database, Sxxxx Cxxxx Oxxxx Pxxxx Uxxxx Sxxxx, which was created by a consortium of publishers coordinated by Elsevier, will change this monopoly situation by, say, 5 years from now.

Skewness of Distributions, Other Statistical Properties, Scaling, Networks

Skewness of citation distributions (Veugelers, this issue) is a major element in bibliometric analysis (Haitun, 1982; Lotka, 1926; Naranan, 1971). Therefore, an extensive study of the statistical properties (Lewison, this issue; Zitt, this issue) of bibliometric measures is of crucial importance, and it is one of the central issues in the earlier mentioned point of the assessment of the bibliometric instrument. For instance, how does our crown indicator—average number of citations per publication divided by the world average in a combination of fields (*CPP/FCSm*)—which is based on calculations using mean values, relate to measures based on (the top of) citations distribution functions of journals or fields (Lewison, this issue; Zitt, this issue)?

Because of the skewness of the underlying distribution functions (Redner, 1998; Seglen, 1992, 1994), confidence intervals of bibliometric indicators are a special problem. By virtue of the central limit theorem, however, bibliometric indicators based on mean values of, for instance, the number of citations per article of research groups tend to follow a more normal distribution. This provides us with the possibility of estimating confidence intervals. In my focus article, I did not discuss such intervals, but as a general rule, we indicate in our bibliometric studies the statistical significance of indicators as presented in Table 1 of the focus article (see, e.g., recent Centre for Science and Technology Studies reports). These calculations are based on a model of the statistical reliability of citation impact comparisons developed by Schubert and Glänzel (1983).

A forthcoming article (van Raan, in press-c) addresses in detail the statistical properties of bibliometric indicators. Another forthcoming article (van Raan, in press-b) discusses the properties of networks based on bibliographic coupling. Knowledge about the statistical properties is also necessary for a deeper understanding of the network structures that can be created on the basis of citation and word-similarity linkages (Zitt, this issue). I point out that I contributed to the study of the basic statistical characteristics of bibliometrics by one of the very few—or even the only—ab initio model for the distribution of citation over publications (van Raan, 2001).

The power-law functions that characterize the correlation between number of citations and number of publications at higher aggregation levels such as countries or fields of science (Katz, this issue) strongly suggest scaling properties of the science system. Also, the networks (Zitt, this issue) in science, based, for instance, on cocitation, bibliographic coupling, coword, or coauthor relations, show self-similar, scale-free structures (Newman, 2001a, 2001b, 2001c; Peters & van Raan, 1991; Price, 1965). This poses the problem (Zitt, this issue) of which, if any, level of aggregation has a preferential position. A practical answer is that, in many natural network systems—and science is not an exception—we have both a network structure and different forms of hierarchical clustering (Albert & Barabási, 2002; Amaral, Scala, Barthélémy, & Stanley, 2000; Klemm & Eguíluz, 2002; Mossa, Barthélémy, Stanley, & Amaral, 2002). A research group is such a natural clustering, and it is generally accepted that the research group—mostly a team of junior and senior researchers around one or two leading professors—is the most important working-floor entity in science.

I agree that large differences in size of research groups may introduce problems, as discussed clearly by Zitt (this issue), in impact-normalization procedures. This problem is nicely illustrated by Figure 1 in van Raan (2000), where the (field-normalized) crown indicator *CPP/FCSm* is given for about 150 physics groups of different sizes (size approximated by the number of publications of a group). Clearly, the larger the group, the more the indicator value tends to 1, which is the asymptotic value (viz., for the case where the "group" is all physics publications in the whole world). Indeed, in interinstitutional or intergroup benchmarking, size matters. The increase of empirical data on bibliometric indicator values for a large range of entities of different sizes will provide us with the possibility of investigating this important problem more thoroughly.

The problem of size-dependent normalization is also clearly visible in the cumulative advantage effects (Katz, this issue) related to the power-law correlations between citation numbers and publication numbers. Katz's observations are most important, and I discuss these findings in more detail in my forthcoming article on statistical issues (van Raan, in press-c). I restrict myself here to one specific point. The scaling relation between citations and publications shows power-law behavior with an exponent between 1 and 2. On the basis of this finding, one may conclude that larger entities (i.e., entities with a larger number of publications) have a cumulative advantage or a Matthew effect (XXXXXX, XXXX) in receiving citations. Thus, it is argued that size-related corrections are needed to arrive at scale-independent measures (Katz, this issue). In simple words, larger groups should be punished for their size.

Immediately, a conceptual problem emerges: Why should we do that? Clearly, an increasing size—at least at the level of research groups—is in itself an indication of performance: A group that performs well will generally grow. In this respect, cumulative advantage is part of a well-deserved impact! But moreover, if we take the indicator *CPP*, then it follows immediately from the previously mentioned exponent that this indicator has a cumulative disadvantage for larger groups, as the power-law exponent will now be between 0 and 1.

So by working with the *CPP* indicator, larger groups have a disadvantage: A larger number of publications will, as it were, dilute the total impact. The earlier mentioned size-dependent behavior of the *CPP/FCSm* indicator is a striking example of this effect. We certainly need more research to better understand this remarkable coincidence of advantage and disadvantage in relation to size-related normalization (Katz, this issue; Zitt, this issue).

Bibliometric Mapping and Field Definitions

Field-definition based on a set of journals (Lewison, this issue; Zitt, this issue) is certainly not ideal. But it provides a first practical, standardized, and consistent approach to the problem of how to delineate fields, subfields, and research themes. Indeed, for a better understanding of scientific development, we need more advanced ways to define fields of science. Bibliometric mapping offers us ample opportunities, as the clusters found in the map relate to meaningful structures ("entities") in science. Bibliometric mapping as discussed in the Mapping the Structure of Interdisciplinary Research section of my focus article is based on word-similarity structures and is therefore completely independent of citation relations (Veugelers, this issue). Thus, this mapping methodology can be applied to all types of documents in any data system. For instance, the microelectronics map presented in my focus article is based on data from the physics and computer science database Ixxxx Nxxxx Sxxxx Pxxxx Exxxx Cxxxx. By using conference databases, maps can be created with a smaller time delay for those fields where conferences are used by researchers as first communication channels (Veugelers, this issue). I stress, however, that publication in top journals (*Nature, Science, Physical Review Letters*) can often be considerably faster than publication in conferences proceedings! Moreover, by using top journal publication data as elements for mapping, I focus on the forefront of science rather than on the mainstream, which is dominated mainly by followers. So we see that mapping can be performed in different modalities, all highlighting different aspects of the science system.

By applying the mapping methodology at different levels of aggregation, fields, subfields, and research themes can be found. As these fields, subfields, and research themes are characterized by specific concepts, it is possible to classify (Lewison, this issue) these entities, and thus also the publications covered by these entities, in to categories such as "experimental," "theoretical," "clinical," "instrument development," and so on. Furthermore, these entities can be used for a theme-specific normalization (Zitt, this issue) of the impact measures. And this is precisely what we did in our recent combined evaluation and mapping studies (see our mapping Web site[1]). I think it is too simple to state (Zitt, this issue) that the main progress in the mapping methodology has been brought about by natural language processing after the pioneering work in the 1980s. Word-similarity based mapping was for quite a long time almost dead, and the Leiden group has been one of the very few that revitalized this approach by the further pioneering work of Braam, Moed, and van Raan (1991x), Peters and van Raan (1993a, 1993b), and Engelsman and van Raan (1994), including the study of the dynamical properties of bibliometric maps. A fundamental problem of these dynamical properties concerns the conceptual difficulties in retrospective studies—in other words, in reconstructing the past (see Noyons & van Raan, 1998).

The importance of the role of the user of maps is particularly stressed in the Leiden group's work, and this has two crucial components. First, to navigate in the map and to retrieve specific data, the user needs additional facilities that can only be realized in advanced Internet-based user interfaces. Second, the problems of "reading" abstract representations such as bibliometric maps should not be underestimated. Collaboration with psychology researchers is necessary to reshape bibliometric maps into cognitive maps. I refer to the Leiden group's current mapping work (Buter et al., 2004; Noyons et al., 2003a, 2003b).

The discovery of "new and unexpected linkages" (Lewison, this issue, p. 31) is certainly an interesting element in mapping methodology. Whether such discoveries can be regarded as"spectacular successes" (Lewison, this issue, p. 31) is a matter of taste. In my opinion, the importance of bibliometric mapping is that it allows us to discover the structure of the scientific and technological landscape. Also, in general terms, and as a next step, it allows us to gain concrete and detailed insight into the development of fields of science and technology and their interaction—that is, the "knowledge flows" between them (Zitt, this issue, p. 38). Moreover, this mapping methodology is appropriate to visualize the position of research and development (R&D) activities of universities, institutes, and groups (Noyons, Luwel, & Moed, 1999) in the international landscape in relation to interdisciplinary R&D developments (Feller, this issue) and particularly in relation to socioeconomic problems. Furthermore, mapping allows the identification and positioning

[1]www.cwts.nl/ec-coe/home.html

of the major actors. It even provides the possibility of foresight (van Raan & Noyons, 2002).

Citation Behavior, Time Lags

We do not have an adequate a priori theory of citation behavior (Lewison, this issue; Veugelers, this issue; Zitt, this issue), let alone field-dependent components of such a theory. My guess is that there will never be such a theory and that one can only build on empirical findings to be gained by large-scale application of bibliometric indicators in many situations (different fields, different levels of aggregation, etc.). The motives for giving (or not giving) a reference to a particular article may vary considerably (Brooks, 1986; MacRoberts & MacRoberts, 1988; Vinkler, 1998). So undoubtedly the process of citation is a complex one, and it certainly does not provide an ideal monitor of scientific performance (MacRoberts & MacRoberts, 1996). This is particularly the case at a statistically low aggregation level—for example, the individual researcher. But in my opinion, again, there is sufficient evidence that these reference motives are not so different nor randomly given to such an extent that the phenomenon of citation would lose its role as a reliable measure of impact. An important empirical support is given by the correlation between the outcomes of peer judgment and bibliometric results in research performance assessments. A short discussion of this topic is given in Intermezzo 3.

INTERMEZZO 3: PEER REVIEW AND BIBLIOMETRIC ANALYSIS

The results of peer review judgment and those of bibliometric assessment are not completely independent variables. Peers take bibliometric aspects into account in their judgment of, for instance, (number of) publications in the better journals. Thorough studies of larger-scale evaluation procedures in which empirical material is available with data on both peer judgment as well as bibliometric indicators are rare. I refer to Rinia, van Leeuwen, van Vuren, and van Raan (1998) for a comparison of bibliometric assessment based on various indicators with peer review judgment in condensed matter physics, and to Rinia et al. (2001) for a study of the influence of interdisciplinarity on peer review in comparison to bibliometric assessment. In the Leidens group's current work, the relation between bibliometric assessment and peer judgment for 150 chemistry research groups in the Netherlands (Vxxxx Sxxxx Nxxxx Uxxxx, 2002) is studied. Generally, these studies show that there is considerable agreement between measurement of research performance and the results of peer review. But at the same time, remarkable differences are found in which peer judgment does not necessarily have to be considered *right* (Horrobin, 1990; Moxham & Anderson, 1992).

A frequently posed question concerns the delay problem (Veugelers, this issue): Does bibliometric analysis suffer from a substantial delay in the measurement of research performance (Egghe & Rousseau, 2000)? An answer to this question requires further refinement: delay as compared to what? To the average processing time of a publication? To the average running time of a project? Or to peer review time cycles? The entire process—starting with scientific activities and leading to publishable results, writing of an article, submission of the article, publication of the article, and citations to the article—varies considerably for the different fields of science and often within a field. Depending on type of activities and type of results, it may take years, but during that time, the work is improved. The whole processing time cannot be regarded as a delay or a waste of time. Furthermore, the average duration of a major research project is about 4 years, and the same is the case for most peer review time cycles. Also, during the publication process, the awareness of the scientific community (and peers!) evolves (e.g., average time between field-specific conferences, etc.). We also have cases where the analysis can be performed almost in real time, as illustrated by an example of a recent physics article with citing articles published in the same year as the cited publication.

The previously mentioned implies that *bibliometric awareness* does not necessarily take more time than *peer awareness.* Moreover, the bibliometric system itself proves empirically the robustness of the method simply by showing that, in many cases, citation-analysis based indicators for universities, institutes, and larger research groups are remarkably stable, as illustrated clearly by the results presented in Table 1 of the focus article. I conclude that recent past performance is a reliable predictor for near-future performance.

We also have to keep in mind that the importance of a publication does not necessarily appear immediately, even to peers, and that identification of quality may take considerable time. An interesting phenomenon in this respect is the "Sleeping Beauty in Science" (Garfield, 1980; Glänzel, Schlemmer, & Thijs, 2003; van Raan, 2004b), a publication that goes unnoticed (sleeps) for a long time and then, almost suddenly, attracts a lot of attention (is awakened by a prince).

Efficiency of Research

It is important to develop research information systems that cover more than simply publication data (e.g., number of full-time equivalents for research, mission, and objectives; age distribution data of the research personnel; data on facilities; supporting personnel; data on funding; etc.). Such an extension to nonbibliometric data or, in general terms, institutional-context data (Feller, this issue) would allow us to gain a deeper insight into important aspects of research performance, particularly the efficiency of research (Veugelers, this issue) and its social characteristics (Gläser & Laudel, 2001).

ACKNOWLEDGMENTS

This rejoinder is based in part on a recent paper by the author concerning the problem of ranking universities by bibliometric methods (van Raan, in press-a). I would like to thank the authors of the commentaries for their detailed, clear, and frank comments. I am convinced that this set of articles will contribute substantially to the further advancement of the bibliometric methodology.

REFERENCES

Adam, D. (2002). The counting house. *Nature, 415,* 726–729.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys., 74,* 47–97.

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proc. Nat. Academy of Sciences*, *97*, 11149–11152.

Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991a). 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.

Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science, 37,* 34–36.

Buter, R. K., Noyons, E. C. M., van Raan, A. F. J., Laine, T., Sutinen, E., van Mackelenbergh, M., et al. (2004). *Monitoring environment and health R&D in relation to socio-economic problems. A new approach to impact assessment* (Final Report to the European Commission*)*. Leiden, The Netherlands: Centre for Science and Technology Studies.

Egghe, L., & Rousseau, R. (2000). The influence of publication delays on the observed aging distribution of scientific literature. *Journal of the American Society for Information Science, 51,* 158–165.

Engelsman, E. C., & van Raan, A. F. J. (1994). A patent-based cartography of technology. *Research Policy, 23,* 1–26.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1,* 359–375.

Garfield, E. (1980, May 26). Premature discovery or delayed recognition—Why? *Current Contents, 21,* 5–10.

Glänzel, W. (1996). A bibliometric approach to social sciences, national research performances in 6 selected social science areas, 1990–1992. *Scientometrics, 35,* 291–307.

Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics, 58,* 571–586.

Gläser, J., & Laudel, G. (2001). Integrating scientometric indicators into sociological studies: Methodical and methodological problems. *Scientometrics, 52,* 411–434.

Grupp, H., Schmoch, U., & Hinze, S. (2001). International alignment and scientific regard as macro-indicators for international comparisons of publications. *Scientometrics, 51,* 359–380.

Haitun, S. D. (1982). Stationary scientometric distributions, 1: Different approximations. *Scientometrics, 4,* 89–104.

Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics, 44,* 193–215.

Horrobin, D. F. (1990). The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association, 263,* 1438–1441.

Klemm, K., & Eguíluz, V. M. (2002). Highly clustered scale-free networks. *Physical Review E, 65,* 036123.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *J. Washington Acad. Sci., 16,* 317–323.

MacRoberts, M. H., & MacRoberts, B. R. (1988). Author motivation for not giving citing influences—a methodological note. *Journal of the American Society for Information Science, 39,* 432–433.

MacRoberts, M. H., & MacRoberts, B. R. (1996). Problems of citation analysis. *Scientometrics, 36,* 435–444.

May, R. M. (1997). The scientific wealth of nations. *Science, 275,* 793–796.

Moed, H. F. (2002). The impact factors debate: The ISI's uses and limits. *Nature, 415,* 731–732.

Mossa, S., Barthélémy, M., Stanley, H. E., & Amaral, L. A. N. (2002). Truncation of power law behavior in "scale-free" network models due to information filtering. *Phys. Rev. Lett., 88,* 138701.

Moxham, H., & Anderson, J. (1992). Peer review: A view from the inside. *Science and Technology Policy, XX,* 7–15.

Naranan, S. (1971). Power law relations in science bibliography—A self-consistent interpretation. *Journal of Documentation, 27,* 83–97.

Nederhof, A. J. (1988). The validity and reliability of evaluation of scholarly performance. In A. F. J. van Raan (Ed.), *Handbook of quantitative studies of science and technology* (pp. 193–228). Amsterdam: Elsevier/North-Holland.

Newman, M. E. J. (2001a). Scientific collaboration networks, I: Network construction and fundamental results. *Physical Review E, 64,* 016131.

Newman, M. E. J. (2001b). Scientific collaboration networks, II: Shortest paths, weighted networks, and centrality. *Physical Review E, 64,* 016132.

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences, 98,* 404–409.

Noyons, E. C. M., Buter, R. K., van Raan, A. F. J., Schmoch, U., Heinze, T., Hinze, S., et al. (2003a). *Mapping excellence in science and technology across Europe, Part 1: Life sciences* (Report to the European Commission). Leiden, The Netherlands and Karlsruhe, Germany: The Leiden University Centre for Science and Technology Studies and the Fraunhofer Institute for Systems and Innovation Research.

Noyons, E. C. M., Buter, R. K., van Raan, A. F. J., Schmoch, U., Heinze, T., Hinze, S., et al. (2003b). *Mapping excellence in science and technology across Europe, Part 2: Nanoscience and nanotechnology* (Report to the European Commission). Leiden, The Netherlands and Karlsruhe, Germany: The Leiden University Centre for Science and Technology Studies and the Fraunhofer Institute for Systems and Innovation Research.

Noyons, E. C. M., Luwel, M., & Moed, H. F. (1999). Combining mapping and citation analysis for evaluative bibliometric purpose. A bibliometric study on recent development in micro-electronics. *Journal of the American Society for Information Science, 50,* 115–131.

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.

Peters, H. P. F., & van Raan, A. F. J. (1991). Structuring scientific activities by co-author analysis: An exercise on a university faculty level. *Scientometrics, 20,* 235–255.

Peters, H. P. F., & van Raan, A. F. J. (1993a). Co-word based science maps of chemical engineering, by combined clustering and multidimensional scaling. *Research Policy, 22,* 47–71.

Peters, H. P. F., & van Raan, A. F. J. (1993b). Co-word based science maps of chemical engineering, Part I: Representations by direct multidimensional scaling. *Research Policy, 22,* 23–45.

Price, D. J. de S. (1965). Networks of scientific papers. *Science, 149,* 510–515.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B, 4,* 131–134.

Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in The Netherlands. *Research Policy, 27,* 95–107.

Rinia, E. J., van Leeuwen, T. N., van Vuren, H. G., & van Raan, A. F. J. (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations. *Research Policy, 30,* 357–361.

Schubert, A., & Glänzel, W. (1983). Statistical reliability of comparisons based on the citation impact of scientometric publications. *Scientometrics, 5,* 59–74.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science, 43,* 628–638.

Seglen, P. O. (1994). Causal relationship between article citedness and journal impact. *Journal of the American Society for Information Science, 45,* 1–11.

Shanghai Jiao Tong University. (2003). *Ranking methodology.* Retrieved XXXXX XX, XXXX, from http://ed.sjtu.edu.cn/rank/methodology.htm

van Leeuwen, T. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics, 51,* 335–346.

van Raan, A. F. J. (2000). The Pandora's box of citation analysis: Measuring scientific excellence, the last evil? In B. Cronin & H. Barsky Atkins (Eds.), *The web of knowledge. A Festschrift in honor of Eugene Garfield* (pp. 301–319). Medford, NJ: ASIS.

van Raan, A. F. J. (2001). Two-step competition process leads to quasi power-law income distributions. Application to scientific publication and citation distributions. *Physica A, 298,* 530–536.

van Raan, A. F. J. (2004a). Measuring science. Capita selecta of current main issues. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 19–50). Dordrecht, The Netherlands: Kluwer.

van Raan, A. F. J. (2004b). Sleeping beauties in science. *Scientometrics, 59,* 461–466.

van Raan, A. F. J. (in press-a). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics.*

van Raan, A. F. J. (in press-b). Reference-based publication networks with episodic memories. *XXXXX.*

van Raan, A. F. J. (in press-c). Statistical properties of bibliometric indicators: Research group indicator distributions and correlations. *XXXXX.*

van Raan, A. F. J., & Noyons, E. C. M. (2002). Discovery of patterns of scientific and technological development and knowledge transfer. In W. Adamczak & A. Nase (Eds.), *Gaining insight from research information, proceedings of the 6th International Conference on Current Research Information Systems* (pp. 105–112). Kassel, XXXXX: University Press.

Vinkler, P. (1998). Comparative investigation of frequency and strength of motives toward referencing, the reference threshold model—Comments on theories of citation? *Scientometrics, 43,* 107–127.

Vxxxx Sxxxx Nxxxx Uxxxx. (2002). *Quality assessment of research. Chemistry and chemical engineering research in The Netherlands.* Utrecht, The Netherlands: Author.

Weingart, P. (2003). Evaluation of research performance: The danger of numbers. In X. XXXXXXXX (Ed.), *Second Conference of the Central Library: Bibliometric analysis in science and research: Applications, benefits and limitations*. (pp. 7–19). XXXXXXX, XXXXXXX: XXXXXXXX.