MEASUREMENT, *3*(1), 1–19

FOCUS ARTICLE

Measurement of Central Aspects of Scientific Research: Performance, Interdisciplinarity, Structure

Anthony F. J. van Raan
Centre for Science and Technology Studies
Leiden University, The Netherlands

This article presents an overview of measuring science based on a bibliometric methodology. The 2 main lines of this methodology are discussed. First, the measurement of research performance is addressed, including aspects such as interdisciplinarity, collaboration, and knowledge users. It is demonstrated that advanced bibliometric methods are an indispensable element next to peer review in research evaluation procedures, particularly at the level of research groups, university departments, and institutes. Problems involved in the application of bibliometric analysis in the social sciences are discussed.

Next, an introduction to the mapping of science is presented, focusing on basic concepts and issues of practical application of science maps. These maps are unique instruments to discover patterns in the structure of scientific fields, to identify processes of knowledge dissemination, and to visualize the dynamics of scientific developments. Their potential for unraveling interdisciplinary developments and interfaces between science and technology is discussed. The article concludes with an indication of the limitations and possible problems of bibliometric analysis.

Keywords: bibliometric analysis, bibliometric indicators, bibliometric spectroscopy, citation, citation index, institutional quality, interdisciplinarity, science mapping

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

BIBLIOMETRIC MEASUREMENT OF SCIENTIFIC PERFORMANCE

Basic Concepts

The rationale of the bibliometric approach to measuring scientific performance presented in this contribution is as follows. *Scientific progress* can be defined as the substantial increase of our knowledge about everything. We discern basic knowledge (understanding) and applicable knowledge (use). Both types of knowledge can be tacit (craftsmanship) or codified (archived and publicly accessible). Scientists have communicated (and codified) their findings in a relatively orderly, well-defined way since the 17th Century. Particularly crucial is the phenomenon of serial literature: publications in international journals. Thus, communication—that is, exchange of research results—is a crucial aspect of the scientific endeavor. Publications are not the only elements—but certainly are very important ones—in this process of knowledge exchange.

Each year, about 1,000,000 publications are added to the scientific archive of this planet. This number (and also numbers for subsets of science, e.g., fields, institutes) is, in many cases, sufficiently high to allow quantitative analyses yielding statistically significant findings. Publications offer usable elements for measuring important aspects of science: author names, institutional addresses, journal (which indicates not only the field of research, but also status!), references (citations), and concepts (keywords, keyword combinations). Although not perfect, we adopt a publication as a building block of science and as a source of data.

Thus, bibliometric assessment of research performance is based on one central assumption: Scientists who have something important to say do publish their findings vigorously in the open international journal (serial) literature. This choice unavoidably introduces a bibliometrically limited view of a complex reality. For instance, journal articles are not in all fields the main carrier of scientific knowledge; they are not equivalent elements in the scientific process; they differ widely in importance; and they are challenged as the gold standard by new types of publication behavior, particularly electronic publishing (van Raan, 2001).

However, the daily practice of scientific research shows that inspired scientists in most cases, and particularly in the natural sciences and medical research fields, go for publication in the better and—if possible—the best journals. A similar situation is developing in the social and behavioral sciences, in engineering, and, to a lesser extent, in the humanities. This observation is confirmed by many years of experience in peer review-based research evaluation procedures.

Work of at least some importance provokes reactions of colleagues. They are the international forum, the "invisible college," by which research results are discussed. Often, these colleagues play their role as a member of the invisible college by referring in their own work to earlier work of other scientists. This process of citation is a complex one, and it certainly does not provide an ideal monitor of scientific performance. This is particularly the case at a statistically low aggregation level—for example, the individual researcher. But the application of citation analysis to the work, the oeuvre of a group of researchers as a whole over a longer period of time, does yield, in many situations, a strong indicator of scientific performance.

Citation analysis is based on reference practices of scientists. The motives for giving a reference to a particular article may vary considerably. There is, however, sufficient evidence that these reference motives are not so different or randomly given to such an extent that the phenomenon of citation would lose its role as a reliable measure of impact (van Raan, 1998).

Why bibliometric analysis of research performance? Peer review undoubtedly is and has to remain the principal procedure of quality judgment. But peer review and related expert-based judgments may have serious shortcomings and disadvantages (Horrobin, 1990; Moxham & Anderson, 1992). Subjectivity—that is, dependence of the outcomes on the choice of individual committee members—is one of the major problems. This dependence may result in conflicts of interests, lack of awareness of quality, or negative bias against younger people or newcomers to the field. Basically, the methodological problem of determining the quality of a specific subject is still far from solved, as illustrated by the results of rereview of previously granted research proposals (see, e.g., Nederhof, 1988). I do not plead for a replacement of peer review by bibliometric analysis. Subjective aspects are not merely negative. In any judgment, there must be room for the intuitive insights of experts. I claim, however, that for a substantial improvement of decision making, an advanced bibliometric method such as presented in this contribution has to be used in parallel with a peer-based evaluation procedure (Rinia, van Leeuwen, van Vuren, & van Raan, 1998).

The most crucial parameter in the assessment of research performance is international scientific influence. I consider international influence as an important, measurable aspect of scientific quality, and, therefore, the Leiden group has developed standardized bibliometric procedures for assessing research performance within the framework of international influence. Undoubtedly, the bibliometric approach is not an ideal instrument that works perfectly in all fields under all circumstances. But the approach presented in this contribution works very well in the large majority of the natural and medical sciences—and in most cases works quite well in the applied and behavioral sciences (see Implications for the Social Sciences). For a recent application of bibliometric research performance assessment in a typical applied field such as food and nutrition research, I refer to van Raan and van Leeuwen (2002). The application of bibliometric analysis in the humanities has been discussed by Moed, Luwel, and Nederhof (2002).

Citation-based bibliometric analysis provides indicators of international impact and influence. This can be regarded as at least one crucial aspect of scientific quality and thus a proxy of quality as follows from a longstanding experience in bibliometric analysis. A first and good indication of whether bibliometric analysis is applicable to a specific field is provided by its publication characteristics—in particular, the role of international, refereed journals. If international journals are a dominating or at least a major means of communication in a field, then, in most cases, bibliometric analysis is applicable. Therefore, it is important to study the publication practices of a research group, department, or institute to establish whether bibliometric analysis can be applied. A practical measure here is the share of publications in citation index (CI)-covered[1] journals in the total number of publications. In most of the natural and medical science fields, this share is between 80% and 95%. In the social and behavioral sciences, engineering science fields, and particularly in the humanities, this share may be much lower. A major exception is psychology, where the share of publications in CI-covered journals may vary between 50% and 70%, and the engineering science fields, with a coverage around 50%. For publications not covered by CI, a restricted type of bibliometric analysis is possible, insofar as these publications are cited by articles in journals covered by CI (Visser, Rons, van der Wurff, Moed, & Nederhof, 2003). A more detailed discussion of the problems encountered in the application of bibliometric analysis to the social and behavioral sciences is given in the Mapping the Structure of Interdisciplinary Research section.

METHOD

I use the results of our recent analysis of a German medical research institute (over the period of 1992–2000) as an example. Thus, I present the details of the bibliometric methodology at the institutional level. *Research output* is defined as the number of articles of the institute, as far as covered by the CI indexes. As *article,* I consider the following publication-types: normal articles (including proceedings articles published in journals), letters, notes, and reviews (but not meeting abstracts, obituaries, corrections, editorials, etc.).

Table 1 shows the number of articles published, *P,* which is also a first indication of the size of an institute. This number is about 250 per year. Next is the total number of citations, *C*, received by these *P* publications in the indicated period and corrected for self-citations. For articles published in 1996, citations are counted during the period of 1996–2000; for 1997, citations in 1997–2000; and so on. The establishment of these two basic indicators is far from trivial. Verification is crucial to remove errors and to detect incompleteness of addresses of research organizations, departments, and groups. In citation analysis, an entire range of pitfalls and

[1]The Science Citation Index, the Social Science Citation Index (SSCI), the Arts & Humanities Citation Index, and the specialty citation indexes (CompuMath, Biochemistry & Biophysics, Biotechnology, Chemistry, Material Science, Neurosciences) are produced and published by the Institute for Scientific Information (ISI/Thomson Scientific) in Philadelphia. Throughout this article, I use the term *CI* for the previously mentioned set of indexes.

TABLE 1
Bibliometric Analysis of a Medical Research Institute, 1992–2000

Period	*P*	*C*	*CPP*	*%Pnc*	*CPP/JCSm*	*CPP/FCSm*	*CPP/D-FCSm*	*JCSm/FCSm*	*%SCit*
1992–2000	**2,245**	**43,665**	**19.45**	**21**	**1.26**	**1.95**	**1.85**	**1.55**	**18**
1992–1996	1,080	11,151	10.33	36	1.27	2.02	1.95	1.58	22
1993–1997	1,198	12,794	10.68	34	1.24	2.03	1.92	1.63	21
1994–1998	1,261	12,217	9.69	32	1.19	1.85	1.72	1.55	22
1995–1999	1,350	13,709	10.15	31	1.21	1.89	1.76	1.56	21
1996–2000	1,410	14,815	10.51	30	1.20	1.91	1.76	1.59	21

Note. P = number of articles published; C = total number of citations; CPP = average number of citations per publication; *%Pnc* = percentage of noncited articles; JCSm = average citation rate of a journal set; FCSm = world average in a combination of fields; D-FCSm = Germany-specific field-based world average; %SCit = percentage of self-citations.

sources of error is lurking. I refer to Moed, De Bruin, and van Leeuwen (1995) and van Raan (1996) for the many methodological and technical problems that have to be solved to conduct an analysis properly. There is ample empirical evidence that in the natural and life sciences—basic as well as applied—the average peak in the number of citations is in the 3rd or 4th year after publication. Therefore, a 5-year period is appropriate for impact assessment. A trend analysis is then based on moving and partially overlapping 5-year periods, as presented in Table 1.

The third and fourth indicators are the average number of citations per publication (*CPP*), again without self-citations, and the percentage of not-cited articles, *%Pnc*. I stress that this *%Pnc* concerns, like all other indicators, the given time period. It is possible that publications not cited within such a time period will be cited after a longer time. This is clearly visible when comparing this indicator for the 5-year periods (e.g., 1996–2000: 30%) with that of the whole period (1992–2000: 21%). The values found for this medical research institute are quite normal.

How do we know that a certain number of citations, or a certain value of citations-per-publication, is low or high? To answer this question, we have to make a comparison with (or normalization to) a well chosen international reference value and thus establish a reliable measure of relative, internationally field-normalized impact. Another reason for normalizing the measured impact of an institute (*CPP*) to international reference values is that, overall, worldwide citation rates are increasing.

First, the average citation rate of all articles (worldwide) in the journals in which the institute has published (*JCSm*, the mean Journal Citation Score of the institute's "journal set," and *JCS,* for one specific journal) is calculated. Thus, this indicator, *JCSm,* defines a worldwide reference level for the citation rate of the institute. It is calculated in the same way as *CPP*, but now for all publications in a set of journals. A novel and unique aspect is that the Leiden group takes into account the type of article (e.g., letter, normal article, review) as well as the specific years in which the articles were published. This is necessary because the average impact of journals may have considerable annual fluctuations and large differences per article type (see Moed & van Leeuwen, 1995, 1996).

With help of the *CPP/JCSm* ratio, we may observe whether the measured impact is above or below the international average. However, comparison of the institute's citation rate (*CPP*) with the average citation rate of its journal set (*JCSm*) introduces a specific problem related to journal status. For instance, if a research group publishes in prestigious (high-impact) journals, and another group in rather mediocre journals, the citation rate of articles published by both groups may be equal relative to the average citation rate of their respective journal sets. But generally, one would argue that the first group evidently performs better than the second.

Therefore, the Leiden group developed a second international reference level, a field-based world average (*FCS*), and *FCSm* in the case in which more fields are involved. This indicator is based on the citation rate of all articles (worldwide) published in all journals of the field(s)[2] in which the institute is active and not only the journals in which the institute's researchers publish their articles. Thus, for a publication in a less prestigious journal, one may have a (relatively) high *CPP/JCSm* but a lower *CPP/FCSm*, and for a publication in a more prestigious journal, one may expect a higher *CPP/FCSm* because publications in a prestigious journal will generally have an impact above the field-specific average.

The same procedure is used as applied in the calculation of *JCSm*. Often, an institute is active in more than one field. In such cases, a weighted average value is calculated, the weights being determined by the total *P* by the institute in each field. For instance, if the institute publishes in journals belonging to genetics as well as to cell biology, then the *FCSm* of this institute will be based on both field averages. Thus, the indicator *FCSm* represents a world average[3] in a specific (combination of) field(s). It is also possible to calculate *FCSm* for a specific country or for the European Union. The example discussed in this article concerns a German medical research institute, and for this institute I calculated the Germany-specific *FCSm*-value, *D-FCSm*.

As in the case of *CPP/JCSm*, if the *CPP/FCSm* ratio is above 1.0, the impact of the institute's articles exceeds the field-based (i.e., all journals in the field) world average. As shown in Table 1, the *CPP/JCSm* is 1.20, *CPP/FCSm* is 1.91, and *CPP/D-FCSm* is 1.76 in the last period, 1996–2000. These results show that the institute is performing well above the international average. The *JCSm/FCSm* ratio is also an interesting indicator. If it is above 1.0, the mean citation score of the institute's journal set exceeds the mean citation score of all articles published in the field(s) to which the journals belong. For the institute, this ratio is around 1.59. This means that the institute publishes in journals with, generally, a high impact. The last indicator shows the percentages of self-citations. About 30% is normal, so the self-citation rates for this institute are certainly not high (about 20%).

I regard the internationally standardized impact indicator *CPP/FCSm* as the *crown indicator.* This indicator enables us to observe immediately whether the performance of a research group or institute is significantly far below (indicator value <0.5), below (indicator value 0.5–0.8), about (0.8–1.2), above (1.2–1.5), or far above (>1.5) the international impact standard of the field. I stress, however, that for the interpretation of the measured impact value, one has to take into account the aggregation level of the entity under study. The higher the aggregation level, the

---

[2]I here use the definition of fields based on a classification of scientific journals into categories developed by ISI. Although this classification is not perfect, it provides a clear and "fixed" consistent field definition suitable for automated procedures within our data system. A more real-world, user-oriented definition of fields can be provided by the mapping methodology discussed in the Mapping the Structure of Interdisciplinary Research section.

[3]About 80% of all CI-covered articles are authored by scientists from the United States, Western Europe, Japan, Canada, and Australia. Therefore, our "world average" is dominated by the Western world.

larger the volume in publications, and the more difficult it is to have an impact significantly above the international level.

Based on my long-standing experiences, I can say the following. At the *meso-level* (e.g., a university, faculty, or large institute with about 500 or more publications per year), a *CPP/FCSm* value above 1.2 means that the institute's impact as a whole is significantly above the (Western) world average. With a *CPP/FCSm* value above 1.5, such as in the example, the institute can be considered to be scientifically strong, with a high probability of finding very good to excellent groups. Thus, the next step in an analysis of research performance is a breakdown of the institution into smaller units—that is, research groups and/or programs. Therefore, the bibliometric analysis has to be applied on the basis of institutional input data about personnel and composition of groups. The algorithms then can be repeated on the lowest but most important aggregation level—the research group or research program. In most cases, the volume of publications at this level is 10 to 20 per year.

Particularly at this lower aggregation level, the verification of the data is crucial (e.g., correct assignment of publications to research groups, completeness of publications sets).The Leiden group has developed standardized procedures for carrying out bibliometric analysis as conscientiously as possible. These procedures are discussed thoroughly beforehand with the client institutes.

At the group level, a *CPP/FCSm* value above 2 indicates a very strong group, and a value above 3 indicates the groups can be generally considered to be excellent and comparable to the top groups at the best U.S. universities. If the threshold value for the *CPP/FCSm* indicator is set at 3.0, excellent groups can be identified with high probability (van Raan, 2000). As an additional indicator of scientific excellence, the number of publications within the top 10% of the worldwide impact distribution of the field concerned is determined for the target entity (see Noyons et al., 2003a, 2003b). In the calculation of this indicator, the entire citation distribution function is taken into account, thus providing a better statistical measure than those based on mean values.

For all of the previously mentioned indicators, I also perform a breakdown into types of scientific cooperation according to the publication addresses: work by only the unit itself, in a national collaboration, or in an international collaboration. Generally, one observes the highest impact for publications in international collaboration (see, e.g., van Raan & van Leeuwen, 2002).

IMPLICATIONS FOR THE SOCIAL SCIENCES

The increasing use of bibliometric indicators is a matter of achieving a more balanced and thus more objective assessment. Particularly in the social sciences, more than in the natural and medical sciences, local and national orientations (Kyvik &

Larsen, 1994; Nederhof & Zwaan, 1991) are present, where less consensus exists on what successful scientific approaches are. But also within the social sciences, we see differences in orientation, and, therefore, publication and citation characteristics vary widely among the different fields of social sciences. For instance, the internationally oriented publication culture of experimental psychologists is in sharp contrast to the much more locally oriented sociologists.

I already noted that bibliometric assessment of research performance is based on one important assumption: The work to be evaluated must be published in the open, international journal literature. This means that bibliometric indicators are very applicable in the natural and life sciences. However, in the applied and engineering sciences as well as in the social and behavioral sciences (and even more so in the humanities), international journals are often not the primary communication channel. In the section on basic concepts, I indicated the share of publications in CI-covered journals (particularly in the SSCI) in these fields. Meertens, Nederhof, and Wilke (1992) found that non-CI publications play an important role in social psychology. They established that books and book chapters constitute about a third of all Dutch social psychology publications. As discussed in earlier, these non-CI publications, however, can be cited quite well in articles in CI-covered journals. With appropriate analytical routines, their impact can be estimated.

Thus, I caution against an all too easy acceptance of the persistent characterization of the social sciences (and the humanities) as being bibliometrically inaccessible. The idea that the previously mentioned features—such as the less important role of journals, the local orientation of many research fields, and also the dominant role of older literature—are general characteristics of all social sciences and humanities is refuted by recent empirical work. For instance, nowadays linguistics and experimental psychology are more and more approaching the publication behavior of the so-called hard sciences, evidenced by the dominant role of international "core" journals and the increasing citation of recent work. Furthermore, a European benchmark is an effective means of coping with a possible Anglo-Saxon bias in the SSCI. Such a European reference standard could be based on a selected group of European journals covered by the SSCI. Recent experience shows that bibliometric analysis can be applied successfully in the social sciences (Vereniging van Samenwerkende Nederlandse Universiteiten, 1999).

Application of bibliometric indicators at the level of individual scientists is often statistically problematic, but it is especially so in the social sciences, where the number of citations is generally an order of magnitude lower than in the natural and medical sciences. A major methodological problem concerns the time dimension. Citations are given after publication. How long must we wait? In other words, what is an acceptable length for the "citation window"? For the social sciences, this window should be longer than in the natural sciences—around 5 to 6 years. This unavoidable time lag—impact is mainly received after the work has been published—is often misused by critics (even in the natural sciences, where it is about 2 to 3 years) as a general objection (the delay argument) against bibliometric analysis. Yet even peers need time to see whether research results will take root!

An important general conclusion for the application of bibliometric methods is that performance measurement, particularly in the social sciences, must cover a wider range of years. Snapshots are useless; even periods of 5 years are too short. So an important lesson is learned from bibliometric analysis: Research groups need time to establish their position; it is incorrect to judge research performance on the basis of just a few years.

BIBLIOMETRIC SPECTROSCOPY: MEASURING INTERDISCIPLINARITY

A further important step is the breakdown of the institute's output into research fields. This provides a clear impression of the research scope, or profile, of the institute. Such a spectral analysis of the output is based on the simple fact that the researchers publish in journals of many different fields. My example, the German medical research institute, is a center for medicine-oriented molecular research. The researchers of this institute are working in a typical interdisciplinary environment.

By ranking fields according to their size (in terms of numbers of publications) in a graphical display, I construct the research profile of the institute. Furthermore, the field-normalized measured impact values of the institute's publications in these different fields is given with the help of *CPP/FCSm*.

Figure 1 shows the result of this bibliometric spectroscopy. Thus, it becomes immediately visible in which fields within its interdisciplinary research profile the institute has a high (or lower) performance. The institute's articles are published in a wide range of fields: biochemistry and molecular biology, genetics and heredity, vascular diseases, neurosciences, oncology, cell biology, and so on.

We can observe the scientific strength of the target institute: Its performance in the top four fields is high to very high.[4] If we find a smaller field with a relatively low impact (i.e., a field in the lower part, the "tail," of the profile), this does not necessarily mean that the (few) publications of the institute in this particular field are bad. Often, these small fields in a profile are those that are quite remote from the institute's core fields. They are, so to say, peripheral fields. In such a case, the institute's researchers may not belong to the dominating international research community of those fields, and their work may not be cited as frequently as the work of these dominating (card holding) community members.

---

[4]In Figures 1 and 2, the shading of the bars is a first indication of impact: "Low" means *CPP/FCSm* < 0.8, "high" means *CPP/FCSm* > 1.2, and "average" is a *CPP/FCSm* value in between. Fields with fewer than 10 publications are not shown in the figures.

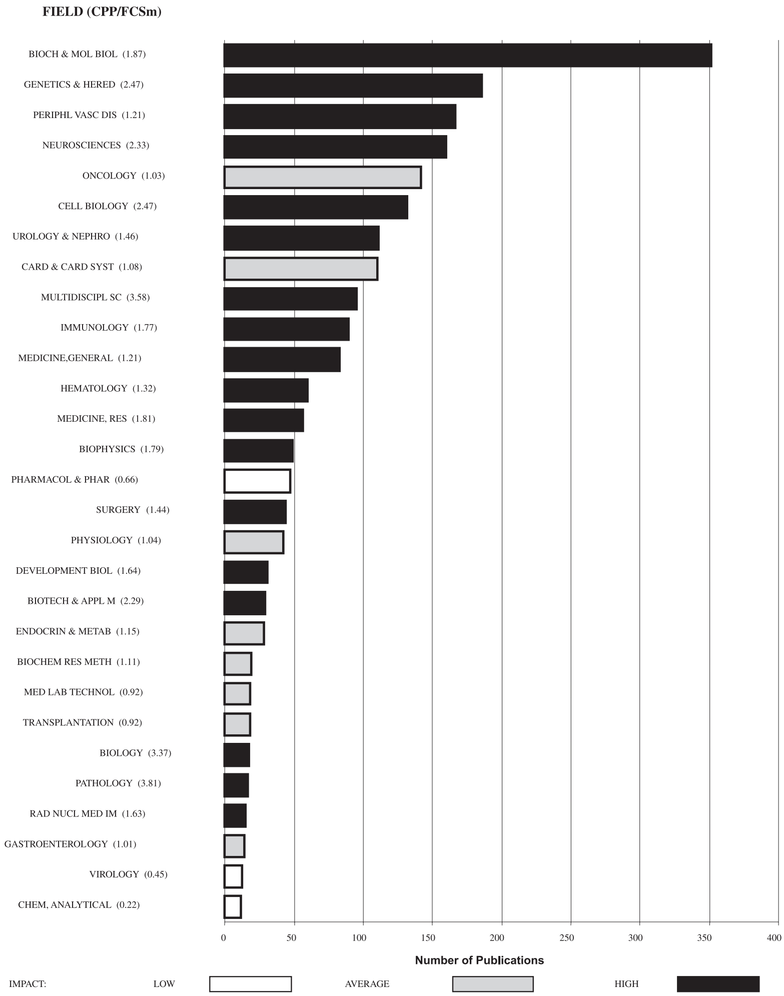

FIGURE 1 Research profile of the output of the medical research institute, 1992–2000.

In a similar way, a breakdown of the citing publications into fields of science is made, which yields a profile of the users of scientific results (as far as represented by citing publications). Also, the impact indicator *CPP/FCSm* of these citing publications is calculated (by a similar citation analysis as discussed previously, but now applied to the citing publications, thus creating a kind of second order citation analysis). This knowledge users profile for the same medical research institute is shown in Figure 2. We can immediately observe some major characteristics. First, the citing publications are themselves of (very) high impact. This clearly means

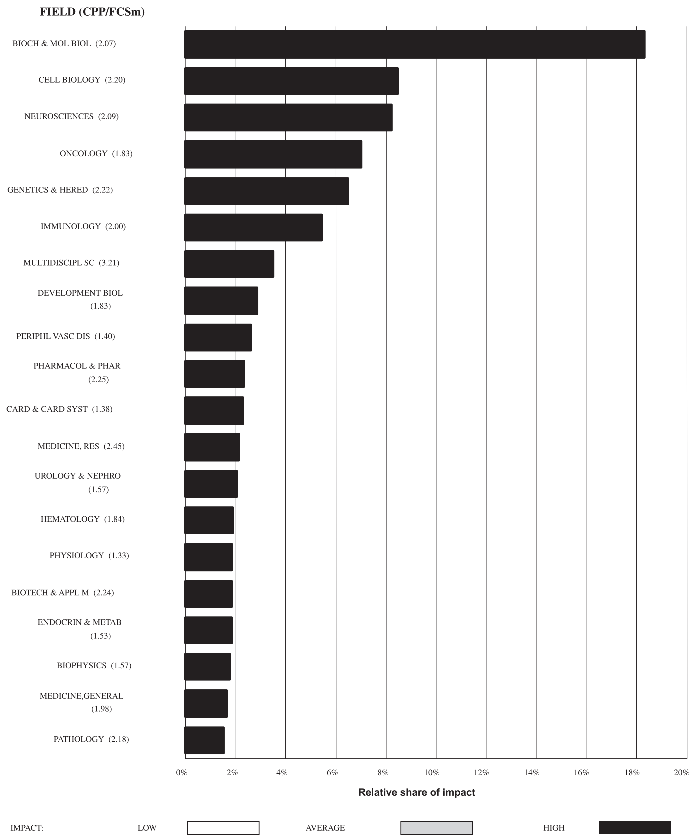

FIGURE 2 Research profile of the knowledge users of the medical research institute, 1992–2000.

that the research work of the institute is cited by generally very good groups. This is also observed in other similar cases: The research of good groups is often predominantly cited by other groups of high scientific performance.

Next, it is interesting to compare the ranking of fields in the knowledge users profile (Figure 2) with the field ranking of the institute's output profile. Figure 1 shows, for instance, that the fields cell biology and developmental biology take a higher ranking on the citing side as compared to the output profile of the institute. Given the data behind the knowledge users profile, it is possible to take a further step in the analysis (beyond the scope of this article): Who is using which research results, where (in which fields), and when. Thus, it analyzes knowledge diffusion and knowledge use, and it indicates further interdisciplinary bridges, potential collaboration, and possible markets in the case of applied research (van Raan & van Leeuwen, 2002).

The construction of these profiles can be considered also as an empirical method to study interdisciplinary aspects of research. For instance, the distribution of the lengths of the field-specific bars in the profile can be used as a measure of interdisciplinarity.

MAPPING THE STRUCTURE OF INTERDISCIPLINARY RESEARCH

Each year, about a million scientific articles are published. How should one keep track of all these developments? Are there specific patterns hidden in this mass of published knowledge at a metalevel, and if so, how can these patterns be interpreted (van Raan & Noyons, 2002)?

A first and crucial step is the definition of a research field. In the Method section, I applied a field definition that is properly applicable in performance analysis. For the mapping of a scientific field, however, we need a more extended field definition than journals sets only. The reason for this is that the mapping procedures described should also be applicable to data from databases other than the CI system. There are several approaches: on the basis of selected concepts (keywords) and/or classification codes in a specific database, selected sets of journals, a database of field-specific publications, or any combination of these approaches. Along these lines, titles and abstracts of all relevant publications can be collected for a series of successive years, thus operating on many tens of thousands of publications per field. Next, with a specific computer-linguistic algorithm, titles and abstracts of all these publications can be parsed. This automated grammatical procedure yields all nouns and noun phrases (standardized) that are present in the entire set of collected publications (Noyons, 1999).

An additional algorithm creates a frequency list of these many thousands of parsed nouns and noun phrases while filtering out general, trivial words. The most frequent nouns/noun phrases can be considered to be the most characteristic concepts of the field (this can be 100 to 1,000 concepts—say, N concepts). The next step is to encode each of the publications with these concepts. In fact, this code is a binary string (yes/no) indicating which of the N concepts is present in title or abstract. This encoding is, as it were, the genetic code of a publication. As in genetic algorithms, the encoding of each publication can be compared with that of any other publication by calculating the genetic code similarity (here, concept similarity) of all publications in a specific field pair-wise.

The more concepts two publications have in common, the more these publications are related on the basis of concept similarity, and thus they can be regarded as belonging to the same subfield or research theme. To use a biological metaphor:

The more specific DNA elements two living beings have in common, the more closely they are related. Above a certain similarity threshold, they will belong to a particular species.

The previously mentioned procedure allows clustering of *information carriers*—the publications—on the basis of similarity in *information elements*—the concepts (publication similarity analysis). Alternatively, the more specific concepts are mentioned together in different publications, the more these concepts are related. Thus, information elements are clustered (coconcept analysis). Both approaches—the publication similarity and the coconcept analysis—are related by matrix algebra rules. In practice, the coconcept approach (Noyons & van Raan, 1998) is most suited to *science mapping*—that is, *the organization of science according to concepts.*

Intermezzo

For a supermarket, client similarity on the basis of shopping lists can be translated into a clustering of either the clients (information carriers, in which the information elements are the products on their shopping lists) or the products. Both approaches are important: The first gives insight into groups of clients (young, old, male, female, different ethnic groups, etc.), and the second is important for the spatial division of the supermarket into product groups.

Procedures and Implications

In outline, the clustering procedure is as follows. First, for each field, a matrix is constructed that is composed of co-occurrences of the *N* concepts in the set of publications for a specific period of time. This raw co-occurrence matrix is normalized in such a way that the similarity of concepts is no longer based on the pair-wise co-occurrences, but on the co-occurrence profiles of the two concepts in relation to all other concepts. This similarity matrix is the input for a cluster analysis. Standard hierarchical cluster algorithms, including statistical criteria, can be used to find an optimal number of clusters. The identified clusters of concepts represent, in most cases, recognizable subfields, or research themes. Each subfield represents a subset of publications on the basis of concept-similarity profiles. If any of the concepts is in a publication, this publication will be attached to the relevant subfield. Thus, publications may be attached to more than one subfield. This overlap between subfields in terms of joint publications is used to calculate a further co-occurrence matrix, now based on subfield publication similarity.

To construct a map of the field, the subfields (clusters) are positioned by multidimensional scaling. Thus, subfields with a high similarity are positioned in each other's vicinity, and subfields with low similarity are distant from each other. The size of a subfield (represented by the area of a circle) indicates its share of publica-

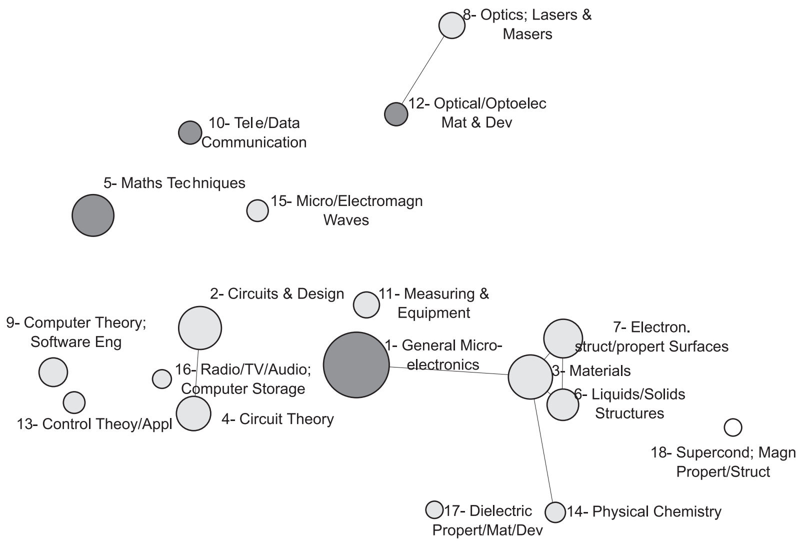

FIGURE 3 Bibliometric map of microelectronics research. The map essentially represents a relational structure of clusters of publications based on cluster-similarity measures. The clusters can be identified as research fields. The closer the clusters are, the more related the fields concerned. White clusters (here, only Cluster 18) are characterized by decreasing publication activity (worldwide); dark gray clusters (e.g., Cluster 1) are characterized by increasing activity.

tions in relation to the field as a whole. A two-dimensional structure is not sufficient to cover all relations embedded in the underlying matrix. Particularly strong relations between two individual subfields are indicated by a connecting line.

In Figure 3, the result of one of our first exercises on microelectronics research is shown. The map clearly shows 18 subfields, represented by these clusters. Major subfields such as general microelectronics, circuits and design, materials, circuit theory, mathematical techniques, liquids, and structure of solids can be observed. Meanwhile, the Leiden group further developed our mapping procedure so that updates of maps can be constructed. Recent developments can be found via the Centre for Science and Technology Studies web site.[5]

A next step (Noyons, Luwel, & Moed, 1999) is the integration of mapping and performance assessment. It enables us to position actors—such as universities, institutes, research & development (R&D) divisions of companies, research groups—on the worldwide map of their field and to measure their influence in relation to the impact level of the different subfields and themes. Thus, a strategic map is created: Who is where in science, and how strong are these actors?

[5]www.cwts.leidenuniv.nl

A series of maps of successive time periods reveals trends and changes in structure and even may allow prediction of near-future developments by extrapolation. Such changes in maps over time (field structure, position of actors) may indicate the impact of R&D programs, particularly in research themes around social and economic problems. In this way, our mapping methodology is also applicable in the study of the socioeconomic impact of R&D (Airaghi et al., 1999).

Bibliometric maps provide an instrument that can be used optimally in an electronic environment. Moreover, there is a large amount of detailed information behind the maps. Hence, it is of crucial importance that this underlying information, particularly about research performance, can be retrieved in an efficient way to provide the user with a possibility of exploring the fields and judging the usefulness of maps against the user's own expertise. Advanced Internet-based user-interface facilities are necessary (Noyons et al., 2003a, 2003b) to enable this further exploration of the maps. Thus, bibliometric maps will enable users to compare the scientific performance of groups or institutes with other "benchmark" institutes. Likewise, the maps can be used for the selection of benchmark institutes—for instance, institutes chosen by the experts.

Cocitation analysis provides an alternative type of mapping, but it unavoidably depends on the availability of citation (reference) data, and thus its applicability is less general than in the case of concept similarity-based mapping. I refer for recent work on cocitation analysis to Braam, Moed, and van Raan (1991a, 1991b) and to Schwechheimer and Winterhager (2001).

CONCLUDING REMARKS

I presented an overview of advanced bibliometric methods for the objective and transparent assessment of strengths and weaknesses in research performance and for monitoring scientific, particularly interdisciplinary, developments. First, I focused on the detailed analysis of research performance in an international comparative perspective. I applied our approach at the institutional level and showed that this level is the crucial starting point of the search for excellence.

I demonstrated that our recently developed indicators are useful and informative, and I concluded that advanced bibliometric methods are, particularly at the level of research groups, university departments, and institutes, an indispensable method in evaluation studies.

In the second application—monitoring of scientific (basic and applied) developments—I showed that recent advances in bibliometric mapping techniques are promising. They are unique instruments to discover patterns in the structure of a research field. By adding communication linkages based on the extent to which publications in a specific subfield cite publications in other subfields, we are able to identify processes of knowledge dissemination (van Raan & Noyons, 2002).

Time-dependent analysis reveals the dynamics of scientific developments, and offers the ability to focus on interdisciplinary developments. This is important, as we know that interdisciplinary crossroads of basic and applied scientific fields are often the loci of discovery and technological innovation.

It is important, however, to know the limitations of bibliometric analysis. As already discussed, the method will fail to assess adequately those research activities where CI-covered journals play a minor role in the communication of research results. In fields where bibliometric analysis can be applied, it is important to study possible negative effects of the method on publication practices—for instance, an increasing pressure on researchers to publish only in CI-covered journals and thus neglect other important channels of communication. Also, we should take care that bibliometric analysis is not used as a stand-alone tool, but always as a support to peer review-based assessments. Furthermore, given the skewness of the citation distributions over publications, statistical issues remain important and need further investigation.

ACKNOWLEDGMENTS

This article is a revised and extended version of earlier recent publications by the author (van Raan 2000, 2003, 2004).

The author thanks his CWTS colleagues Dr. Ed Noyons, Dr. Thed van Leeuwen, and Martijn Visser for important parts of the analytical work, stimulating discussions, and suggestions.

REFERENCES

Airaghi, A., Viana Baptista, J., Bush, N. E., Georghiou, L., Ledoux, M. J., van Raan, A. F. J., et al. (1999). *Options and limits for assessing the socio-economic impact of European RTD programmes* (ETAN Working Paper No. EUR 18884). Luxembourg: Office for Official Publications of the European Communities.

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.

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

Kyvik, S., & Larsen, I. M. (1994). International contact and research performance. *Scientometrics, 29,* 161–172.

Meertens, R. W., Nederhof, A. J., & Wilke, H. A. M. (1992). Social psychological research in the Netherlands, 1980–1988. *European Journal of Social Psychology, 22,* 93–100.

Moed, H. F., De Bruin, R. E., & van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics, 33,* 381–422.

Moed, H. F., Luwel, M., & Nederhof, A. J. (2002). Towards research performance in the humanities. *Library Trends, 50,* 498–520.

Moed, H. F., & van Leeuwen, T. N. (1995). Improving the accuracy of the Institute for Scientific Information's Journal Impact Factors. *Journal of the American Society for Information Science, 46,* 461–467.

Moed, H. F., & van Leeuwen, T. N. (1996). Impact factors can mislead. *Nature, 381,* 186.

Moxham, H., & Anderson, J. (1992). Peer review. A view from the inside. *Science and Technology Policy, 5*(1), 7–15.

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.

Nederhof, A. J., & Zwaan, R. A. (1991). Quality judgments of journals as indicators of research performance in the humanities and the social and behavioral sciences. *Journal of the American Society for Information Science, 42,* 332–340.

Noyons, E. C. M. (1999). *Bibliometric mapping as a science policy and research management tool.* Leiden, The Netherlands: DSWO Press.

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: Leiden University Centre for Science and Technology Studies/Karlsruhe, Germany: 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: Leiden University Centre for Science and Technology Studies/Karlsruhe, Germany: 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 purposes. A bibliometric study on recent development in micro-electronics. *Journal of the American Society for Information Science and Technology, 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 and Technology, 49,* 68–81.

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.

Schwechheimer, H., & Winterhager, M. (2001). Mapping interdisciplinary research fronts in neuroscience: A bibliometric view to retrograde amnesia. *Scientometrics, 51,* 311–318.

van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36,* 397–420.

van Raan, A. F. J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics, 43,* 129–139.

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: American Society for Information Science.

van Raan, A. F. J. (2001). Bibliometrics and Internet: Some observations and expectations. *Scientometrics, 50,* 59–63.

van Raan, A. F. J. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung, Theorie und Praxis, 12*(1), 20–29.

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

van Raan, A. F. J., & Noyons, E. C. M. (2002, August). 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). Germany: Kassel University Press.

van Raan, A. F. J., & van Leeuwen, T. N. (2002). Assessment of the scientific basis of interdisciplinary, applied research. Application of bibliometric methods in nutrition and food research. *Research Policy, 31,* 611–632.

Visser, M. S., Rons, N., van der Wurff, L. J., Moed, H. F., & Nederhof, A. J. (2003). *Bibliometrische studie van onderzoeksdisciplines aan de Vrije Universiteit Brussel 1992–2001* [Bibliometric study of research disciplines at the Free University of Brussels 1992–2001] (Report No. CWTSbv 2003-03). Leiden, The Netherlands: Leiden University Centre for Science and Technology Studies.

Vereniging van Samenwerkende Nederlandse Universiteiten. (1999). *Quality assessment of research: Psychological research in the Netherlands.* Utrecht, The Netherlands: Author.