

MEASUREMENT OF SCIENTIFIC LITERATURE: A REVIEW OF METHODS

ALEXANDRE DO ESPÍRITO SANTO*
RICHARD D. WALKER**

ABSTRACT

Measurement of scientific literature is here considered a substantial part of the study of interdisciplinarity. It is argued that the determination of interdisciplinarity can be made through the use of scientometrics and bibliometrics. Distinction between these two methods was made. Scientometrics is viewed as the quantitative form of "science of science" and bibliometrics is viewed as the quantitative form of information science. Citation analysis is presented and discussed as the tool most used by scientometricians to determine interdisciplinarity.

INTRODUÇÃO

Since the 1960s, quantitative studies of science have developed mainly in response to Derek de Solla Price's research on patterns of scientific growth (*Science Since Babylon*¹⁶ and *Big Science, Little Science*¹⁷). Price has concluded that science has grown at a rate proportional to its current size.

Such an exponential growth may be largely due to interdisciplinary relationships. Therefore, the search for determination of these relationships, whatever the aims are or the methods used, constitute a quantitative study of sciences; scientometrics or interdisciplinarity study.

In this paper we shall summarize a review of the literature and concepts of the methods and techniques that have been employed for determination of interdisciplinarity.

BIBLIOMETRICS VERSUS SCIENTOMETRICS

Bibliometrics and scientometrics are the methods that have been most used in the studies of interdisciplinarity. Scientometrics has some other applications which will not be considered here. The application of bibliometrics is even more diffuse, but only those applications that deal directly with the study of dispersion will receive attention herewith.

To consider the two methods it is necessary to make a distinction between them as methods of study of interdisciplinarity. Only in the most recent literature and very *en passant* have these two methods been brought together. An attempt to distinguish one from the other has not yet been made.

At first, one is led to believe that it is a question of western and eastern terminology for the same thing: what we call bibliometrics is called scientometrics or "naukometriya" by the Russians. However, after a closer examination of the studies that have used the methods one can see that in many cases the difference lies in the conceptual approaches employed.

Price, who probably was first western author to use the word "scientometrics" to mean the method devoted to quantitative studies of science, in a two-part work, "Studies in Scientometrics", employed the method within the context of sociology of science:

.....many of the richest areas for research in the sociology of science depend upon some understanding of what may be called the actuarial statistics of the scientific community. One needs to know the dynamic processes which govern emergence, survival and disappearance within that community⁽¹⁸⁾.

Eric de Grolier, writing on the use of quantitative data in information science defines **scientometry** as the measure of scientific activities or the

quantitative form of "science of science"⁽⁶⁾. He also states that scientometry largely overlaps the domain of bibliometry.

Narin and Moll, in their extensive review of the literature of bibliometrics, do not mention the term "scientometrics" but devote one section to the studies of "science and scientists" where they cover the name and works related to scientometrics¹¹.

Ignorance of the Russian language by western information scientists is probably responsible for the absence of studies comparing these two quantitative approaches to the study of the development of science.

The world's only scientometric serial, *Naukovedenie i Informatika*, has been published in Kiev since 1969, which in its very name makes a connection between information science and science studies, clearly within the domain of scientometrics or the fusion of the two.

It is this connection between information science and science studies that makes scientometrics distinct from, if not more specific than, bibliometrics. They both lie between the social and the physical sciences, as they use units (books, serials, articles) that can be measured, and are in fact means for measuring the phenomena to be investigated by information scientists.

As in many other disciplines within the social sciences, the application of

* Doctor of Philosophy (Ph.D.). Professor in Centro de Educação, Comunicação e Artes; FUEL.

** Doctor of Philosophy (Ph.D.). Professor in College of Arts and Sciences; University of Wisconsin, U.S.A.

mathematical methods to the investigation of information phenomena gives information science the quantitative approach that makes it a science.

Other social disciplines such as sociology and economics, avail themselves of mathematical methods, i.e., sociometrics and econometrics. Although these two methods are more distinct from each other than are scientometrics and bibliometrics they also overlap in their endeavors to measure social phenomena.

Pritchard, the coiner of the term "bibliometrics" defined it as "the application of mathematical methods to books and other media of communication¹⁹". With such a large scope it is no surprise that the term encompasses so many things. Aiyepoku, reduces its scope a little bit by adding "..... in order to shed light on the characteristics of formalized information transfer and to understand the course and development of a discipline¹".

But, even with this addition the term covers too much. Thus, bibliometrics does not have a unified face to present to the bibliometrician, let alone to the outsider.

Scientometrics, as used in this paper, is not just another term encompassed by bibliometrics; it is the quantitative form of "science of science". It specifically aims at the measuring of relationships among sciences, i.e., it is a major component of interdisciplinarity studies.

As a consequence, scientometrics is more theoretical than bibliometrics, while the former more frequently works with models, the latter more frequently works with techniques. These differences may not as yet be quite evident in the current literature, because most studies in the area of information science using a quantitative approach have been labeled "bibliometric" studies.

However, as one goes deeper into the works that have used the two methods, one begins to see that bibliometrics has more pragmatic objectives, since it is at the service of pragmatic disciplines (library science, information science, documentation) while scientometrics is at the service of less pragmatic disciplines (philosophy of science, sociology of science, science, and history of science).

Therefore, bibliometrics and scientometrics are distinct and are used with different purposes, because the disciplines of which they are components

have different purposes. While the first group of disciplines is basically concerned with characteristics of the information carrier and user's behavior, the second group is concerned with interdisciplinarity and the producer's behavior. To date, the only thing the two methods really have in common is the use of the same technique, i.e., citation analysis.

CITATION ANALYSIS

Until recently, "citation analysis" was used as a collective expression to mean "analysis" of any datum in one publication that would indicate use of somebody else's work. Narin and Moll used the concept of a difference between reference and citation (after Price), thus adding another quantifiable unit of analysis¹². They called "reference" the source unit (an article in a journal, a chapter or a division in a book, etc). and called citation the receiving unit, or a reference given by a source. These two units are believed to display different indications of interdisciplinarity.

Narin, Carpenter, and Berlt probably thought of the distinction when they attempted to measure "interrelationships of scientific journals¹³". Many of the results obtained by citation analysis through citation counts are not affected by this distinction.

Citation counts are said to be reliable and valid. As a unit of measure, citations are stable, consistent, and precise. They neither change with time nor lose properties. They are precise in the sense that they are distinguished from every other citation, but their accuracy depends on what one wants to measure.

A measure is only accurate when it is valid, i.e., citations do not always measure what the researcher believes citations are measuring. Because we do not know why citations are used by authors, citation counts cannot have **predictive validity**.

For example, one cannot predict that an author who has cited the literature of certain disciplines in one of his articles is going to cite literature of the same disciplines in his next article. However, citation counts do have **content validity**, if the sample is adequate.

One can sample the literature of a discipline and find out about certain characteristics of the literature of that discipline. The higher the frequen-

cy of a given characteristic, the better inference one can make about it as a description of the literature under study.

Citations are counted as one of the steps of citation analysis technique, which involves a fuller study of citations than simple counting. Citation analysis is based on certain assumptions which support some of its inferences: (a) all authors have similar citation habits; (b) an author cites those materials he has used as sources; and (c) all papers cited contribute to the content of the paper that cited them.

Many citation analysis studies incorporate these three assumptions into their deductions. Despite many attacks from several research fronts, researchers continue to employ this technique in current research studies of bibliographic characteristics.

Critics of the citation analysis technique argue that there are many social and individual reasons why an author cites materials he uses, therefore, they say no pattern can be consistent enough to use as a basis for the formulation of a hypothesis. They have many arguments:

- (a) citations are a formality dictated by social customs;
- (b) citations are used to gain favor with important scholars in a discipline;
- (c) authors cite themselves for self-promotion; and
- (d) authors often cite what they do not use and do not cite what they do use.

Besides these four criticisms which try to impugn specifically the assumptions mentioned above, but which are also vague and lack specific evidence through any kind of research, there are other criticisms from members of the research community that use the technique.

Norman Kaplan concluded that scientists use citations in order to (a) establish property rights, (b) confer intellectual and scientific respectability on the paper, and (c) curry favor with influential colleagues¹⁰.

Crane does not think that the use of citations constitute an exact measure of intellectual debts, nor that the citations made in particular paper have contributed equally. Of some importance to many research results is her observation that:

....the extent to which publications are cited is related to the stage of development of a research area and

cannot be used as a simple measure of its growth².

Garfield, a conspicuous worker in the field of citations analysis, wrote about some specific limitations of the technique:

... citation frequencies are a function of other variables besides scientific merit. Some of them are authors' reputations, controversiality of the subject, journal circulation, reprint dissemination, coverage by secondary indexing and abstracting services, availability in library collections, and so on³.

These limitations, however, are only a few examples of an abundant literature on the subject. These authors have observed that from most of the analytical studies in which the technique was used one or more limitations emerged. Availability and types of materials, for instance, are very important functions of citation frequency, as demonstrated by Walker et al²¹.

Parallel to the question of availability one can indicate as an important function the number of reprints that are distributed. Hagstrom pointed out that biologists are far more likely to distribute large numbers of reprints than those in other fields⁷.

One wonders how much review articles influence citation frequency. In Hagstrom's opinion references in review articles are more important than references in research articles⁸. Recent studies of scientists activities in the realm of sociology of science and of science of science (Crane, previously cited, and especially Cole and Cole³) have shown numerous factors that influence citation frequency and have indirectly demonstrated the ineffective-

ness of the technique as a measure of merit.

However, the limitations of the technique do not seem to restrain the impetus to using it. Its use has been extended from more or less simple description of disciplinary literature to analytical studies of relationships among sciences. In addition, recent studies have used the technique as a tool for evaluation of scientific performance, productivity, and quality of research¹⁴.

Reactions¹⁵ from the scientific community to the use of citation analysis as a tool for science administrators as proposed by Wade²⁰ need be considered by any researcher contemplating the use of the technique.

While there seems to be nothing wrong with the assumptions and methods employed by many of the citation studies, the reliability of the technique has seldom been tested and its validity has often been questioned. Some important conclusions by various investigators require replication and confirmatory studies.

Kanasy's dissertation⁹ presents a long list of conclusions on the subject that requires further confirmation. Although another recent dissertation³ can be considered an updated contribution to the testing of Kanasy's first out of five most important conclusions, all the other four need additional validation:

1. In all sciences there is considerable borrowing from the literature of other disciplines.
2. In the developing disciplines the cited literature is less than ten years old.
3. The most frequently cited litera-

ture in all scientific disciplines appears in a few journals.

4. In the literature of the social science, serials play a less significant role.

5. Title dispersion among social science serials is high.

CONCLUSIONS

1. It was suggested that the measurement of scientific literature is a substantial part of the study of interdisciplinarity.
2. It has also been argued that the determination of interdisciplinarity can be made through use of such methods as scientometrics and bibliometrics.
3. It was contended that the most important goal of scientometrics should be the determination of interdisciplinarity between sciences, rather than to measure only growth.
4. Distinction was made between scientometrics and bibliometrics in terms of their purposes. While scientometrics can be seen as the quantitative form of "science of science" dealing specifically with interactions among sciences, bibliometrics is seen as the quantitative form of library and information science frequently dealing with less theoretical problems.
5. Citations were said to be at the heart of interdisciplinarity studies. They were recognized as the most common, not as the only means by which interdisciplinarity can be determined.

RESUMO

Mensuração da literatura científica, parte substancial do estudo de interdisciplinaridade. Argumento sobre a viabilidade da determinação de interdisciplinaridade através do uso de cientometria e bibliometria. Distinção entre esses dois métodos e indicação de que o principal objetivo da cientometria deve ser a determinação de interdisciplinaridade entre ciências e não apenas a mensuração de suas expansões. Sugere-se também que a cientometria deve ser vista como a forma quantitativa da "ciência da ciência", interessando-se especificamente pela interação entre ciências, enquanto a bibliometria é vista como forma quantitativa da ciência da informação e da biblioteconomia, interessando-se frequentemente por problemas menos teóricos. Ênfase sobre as características da "análise de citações" como técnica empregada em bibliometria e cientometria e como um dos meios usados para determinação de interdisciplinaridade.

REFERENCES

1. AIYEPEKU, Wilson O. Bibliometrics in Information Science Curricula. *The Information Scientist*, 9: 29-34, Mar., 1975.
2. CRANE, Diana. *Invisible colleges: difusion of knowledge in scientific communities*. Chicago, University of Chicago Press, 1972.
3. COLE, Jonathan & COLE, Stephen. *Social stratification in science*. Chicago, University of Chicago Press, 1973.
4. ESPIRITO SANTO, A. *A measure of the dimensions of interdisciplinarity of two applied sciences: a scientometric model*. University of Wisconsin, 1978. Dissertation (Ph.D.) Univ. of Wisconsin
5. GARFIELD, Eugene. *What scientific*

- journals tell us about scientific journals. *IEEE Transactions on Professional Communication*, PC-16: 200-03, Dec., 1974.
6. GROLIER, Eric de. Le système des sciences et l'évolution du savoir. In: *Conceptual basis of the classification of knowledge. Proceedings of the Ottawa Conference*, 1971. Munchen, Verlag Dokumentation, 1974. p.20-118.
 7. HAGSTROM, Warren O. Factors related to the use of different modes of publishing research in four different scientific fields. In: NELSON, C.E. & POLLOCK, D.K., eds. *Communication among scientists and engineers*. Lexington, Heath Lexington Books, 1970. p. 85-124.
 8. Idem ibidem p. 88.
 9. KANASY, James E. *citation characteristics and bibliographic control of the literature of microbiology*. University of Pittsburgh, 1971. p. 33-34. Dissertation (Ph.D.) Univ. of Pittsburgh.
 10. KAPLAN, Norman. The norms of citation behavior: prolegomena to the footnote. *American Documentation*, 16: 179-84, Jul., 1965.
 11. NARIN, Francis & MOLL, Joy K. Bibliometrics. *Annual Review of Information Science and Technology*, 12: 38-8, 1977.
 12. Idem Ibidem p. 38
 13. NARIN, Francis; CARPENTER, M.; BERLT, Nancy C. Inter-relationships of scientific journals. *Journal of the American Society for Information Science*, 23: 323-331, Sep.-Oct., 1972.
 14. NARIN & MOLL - Op. Cit. p. 43.
 15. NEWS and comments. *Science*, 188: 1063-65, Sep., 1975.
 16. PRICE, J.D. de Solla. *Science since Babylon*. New Haven, Yale University Press, 1961.
 16. ---. *Big science, little science*. New York, Columbia University Press, 1963.
 17. ---. Citation measures of hard science, soft science, technology and nonscience. In: NELSON, Carnot E. & POLLOCK, Donald K. eds. *Communication among scientists and engineers*. Lexington, Heath Lexington Books, 1970. p.2-22.
 19. PRITCHARD, Alan. Statistical bibliography of bibliometrics. *Journal of Documentation*, 25: 348-49, Dec., 1969.
 20. WADE, Nicholas. Citation analysis: a new tool for science administrators. *Science*, 188: 429-32, May, 1975.
 21. WALKER, Richard D.; ZUCK, G. J.; DURRANCE, R.E.; LUEDTKE, J.R.; *Awareness of the relevant water resources literature by the personnel of the Wisconsin Department of the Natural Resources*. Madison, University of Wisconsin, Water Resources Center, 1975. (Technical Report Wis. WRC 75-09).