Information Processing & Management Vol. 28, No. 1, pp. 1-3, 1992
Printed in Great Britain.

0306-4573/92 $5.00 + .00

AN INTRODUCTION TO INFORMETRICS

JEAN TAGUE-SUTCLIFFE
School of Library and Information Science, University of Western Ontario,
London, Ontario N6G 1H1, Canada

Abstract – The scope and significance of the field of informetrics is defined and related to the earlier fields of bibliometrics and scientometrics. The phenomena studied by informetricians are identified. The major contributors to the field in the past are described and current emphases are related to the contributions in this Special Issue.

Among information scientists in Western Europe and North America, the term *informetrics* has become common only in the past five years, as a general field of study which includes the earlier fields of bibliometrics and scientometrics. Its acceptance really dates from the 1987 International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval in Diepenbeek, Belgium, when B.C. Brookes suggested that the term *informetrics* be included in the name of the following conference, scheduled for London, Canada, in 1989. This meeting was thus named International Conference on Bibliometrics, Scientometrics, and Informetrics. The name of the third meeting in the series, held in Bangalore, India, in 1991, signals the final acceptance of this term: International Conference on Informetrics.

According to Brookes, the term *informetrics* was first proposed by Otto Nacke of West Germany in 1979 (Brookes, 1990). An FID committee with very broadly defined objectives in the provision of research and technical data was subsequently given this name. However, use of the word in this sense was not widely adopted.

Bibliometrics is the study of the quantitative aspects of the production, dissemination, and use of recorded information. It develops mathematical models and measures for these processes and then uses the models and measures for prediction and decision making. It appears to have been first used by Pritchard (1969), replacing the earlier term *statistical bibliography*. Some writers, notably White and McCain (1989), limit it to the quantitative study of literatures as they are reflected in bibliographies. Brookes (1990) sees it as now connected primarily to library studies.

The term *scientometrics* came to prominence as the name of a journal founded by T. Braun in 1977, originally published in Hungary and now in Amsterdam. Scientometrics is the study of the quantitative aspects of science as a discipline or economic activity. It is part of the sociology of science and has application to science policy-making. It involves quantitative studies of scientific activities, including, among others, publication, and so overlaps bibliometrics to some extent.

Informetrics is the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists. Thus it looks at the quantitative aspects of informal or spoken communication, as well as recorded, and of information needs and uses of the disadvantaged, not just the intellectual elite. It can incorporate, utilize, and extend the many studies of the measurement of information that lie outside the boundaries of both bibliometrics and scientometrics.

The scope of informetrics is both practical and theoretical. A primary emphasis has been the development of mathematical models, and a secondary emphasis the derivation of measures for the diverse phenomena studied. The value of a model lies in its ability to summarize, in terms of a few parameters, the characteristics of many data sets: the overall shape, concentration, and scatter, the way data sets change over time. In addition, models allow predictions of future behaviors and the isolation of the effect of different factors on variables of interest. Thus, together with the measures that have been derived from them, they provide a firm basis for practical decision making.

Although in practice the scope of informetrics is very broad, in the past, bibliometricians and scientometrics have concentrated their studies of mathematical models and measures in a few well defined areas. These include the following:

- statistical aspects of language, word, and phrase frequencies, in both natural language text and indexes, in both printed and electronic media;
- characteristics of authors—productivity measured by number of papers or other means, degree of collaboration;
- characteristics of publication sources, most notably distribution of papers in a discipline over journals;
- citation analysis: distribution over authors, papers, institutions, journals, countries; use in evaluation; cocitation-based mapping of disciplines;
- use of recorded information: library circulation and in-house book and journal use, database use;
- obsolescence of the literature, as measured both by use and citation;
- growth of subject literatures, databases, libraries; concomitant growth of new concepts.

Two phenomena that have not, in the past, been seen as a part of bibliometrics or scientometrics, but fit comfortably within the scope of informetrics are:

- definition and measurement of information, and
- types and characteristics of retrieval performance measures.

The former has been studied in electrical engineering and mathematics since Shannon and Weaver (1949), and versions of the original Shannon and Weaver measure have, indeed, been developed in many fields. A recent book by Losee (1990) provides an excellent overview of this aspect of informetrics. Retrieval performance measures have been studied by information retrieval theorists in information and computer science, and both theoretical and practical aspects are presented at the annual conference series, International Conference on Research and Development in Information Retrieval.

Three names are identified with the early development of the field we now call informetrics: Lotka, Zipf, and Bradford. To some extent, these investigators are identified with particular phenomena: Lotka with author productivity, Zipf with word frequency, and Bradford with journal productivity. In addition, though, they pioneered particular theoretical approaches.

Lotka (1926) provided the first model for the size-frequency distribution of items (in this case papers in chemistry) over sources (in this case authors). His model became known as the inverse square law.

Zipf (1949), from his studies of word frequencies in a variety of texts, developed both a size frequency and a *rank frequency* distribution for the distribution of word tokens over types. The latter (that the frequency of a word is inversely proportional to the rank) became known as Zipf's law.

Bradford (1934) contributed two theoretical approaches, both somewhat ambiguously called Bradford's law. He developed a cumulative loglinear form of the rank frequency distribution (for the distribution of papers in a discipline over journals). As well, he introduced the idea of a geometric series that represents the increasing number of journals in the nucleus and succeeding zones for a subject area, where the nucleus and the zones each contain equal numbers of papers but decreasing papers per journal.

Later investigators, notably Brookes (1968), Leimkuhler (1967), and Mandelbrot (1961), generalized these models and derived relationships among them. Brookes, in particular, emphasized the importance of the logarithmic and rank approaches in modelling information phenomena. Price (1976) attempted to derive these and other models from the general characteristics of *success-breeds-success* phenomena. Haitun (1982) has provided what is perhaps the most general description of the models found appropriate to informetric phenomena in his description of the non-Gaussian distributions.

At the present time, the field of informetrics seems to be more strongly polarized into theoretical and applied components than it was in the pioneering days of Lotka, Zipf, and Bradford. The theoreticians are putting informetric models and measures on a more solid mathematical and statistical basis than existed in the past. This issue contains several examples of this kind of development. Sichel provides a clear exposition of how a very general model, the GIGP distribution, can be used to fit many distributions of informetric phenomena. Three authors—Burrell, Egghe, and Rousseau—investigate mathematical properties of a measure of concentration, the Gini index. Glanzel, similarly, investigates mathematical and empirical properties of an indicator of citation speed. Tabah looks at nonlinear dynamics and chaos theory as models for the growth of literature. Bookstein examines the appropriateness of the loglinear model for qualitative data in informetrics. In all cases, applications are presented; however, the focus is on the mathematical development.

At the other pole are several papers where the emphasis is on the detailed investigation of a particular application. Bierbaum, Brookes, and Brookes look at the rise and fall of subject terms in the literature of AIDS. Pao presents a study of the relationships of funding and degree of collaboration using a large dataset of research studies about schistosomiasis. Bonzi looks at changes in productivity of academics at a single university over time. Wolfram summarizes the informetric distributions relating to information retrieval from indexed databases and shows how these models can be used, in a simulation, to determine the optimal file structure, given specified characteristics of the database and queries. In all cases, the papers have theoretical import, but the emphasis is on the application.

This introduction has given a very cursory overview of the field of informetrics. The papers that follow will expand the overview in several directions, as we have just described. For those seeking to learn more about this rapidly developing area, I recommend the proceedings of the international conferences mentioned earlier, and the two chapters on bibliometrics in the *Annual Review of Information Science and Technology*. These publications, together with some of the classic papers of the field, are listed in the *Selected References*.

I would like, in closing, to express my appreciation to Mark Kinnucan, Michael Nelson, and Paul Nicholls, for their valuable assistance in the editorial tasks for this special issue. I would like, also, to thank the editor of *Information Processing and Management*, Tefko Saracevic, for his advice and help in assembling the issue and interfacing with the publisher.

SELECTED REFERENCES

Bradford, S.C. (1934). Sources of information on specific subjects. *Engineering, 137*, 85-86.

Brookes, B.C. (1968). The derivation and application of the bradford-zipf distribution. *Journal of Documentation, 24*(4), 247-265.

Brookes, B.C. (1990). Biblio-, sciento-, infor-metrics??? What are we talking about? In *Informetrics 89/90*. Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (Eds.) (1988). *Informetrics 87/88; select proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval, Diepenbeek, Belgium, 25-28 August 1987.* Amsterdam: Elsevier.

Egghe, L., & Rousseau, R. (Eds.) (1990). *Informetrics 89/90; selection of papers submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics, London, Ontario, Canada, 5-7 July 1989.* Amsterdam: Elsevier.

Haitun, S.D. (1982). Stationary scientometric distributions. *Scientometrics, 4*(1), 5-25; *4*(2), 89-104; *4*(3), 181-194.

Leimkuhler, F.F. (1967). The Bradford distribution. *Journal of Documentation, 23*(3), 197-207.

Losee, R.M. (1990). *The Science of Information: Measurement and Application.* San Diego, Calif.: Academic Press.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science, 16*(12), 317-323.

Mandelbrot, B. (1961). On the theory of word frequencies and on related Markovian models of discourse. In *Structure of Language and its Mathematical Aspects: Proceedings of Symposia in Applied Mathematics* (p. 12).

Narin, F., & Moll, J. (1977). Bibliometrics. *Annual Review of Information Science and Technology, 12*, 35-58.

Price, D.J. de S. (1976). A general theory of bibliometric and other cumulative disadvantage processes. *Journal of the American Society for Information Science, 27*(5), 292-306.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 25*, 348-349.

White, H.D., & McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology, 24*, 119-186.

Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort.* Reading, MA: Addison-Wesley.