

*Workshop on Informetrics and Scientometrics
16-19 March 1998, Bangalore*

Paper : AA

INFORMETRICS : Scope, Definition, Methodology and Conceptual Questions

I.K. Ravichandra Rao Professor and Head, Documentation Research and Training Centre, Indian Statistical Institute, 8th Mile, Mysore Road, Bangalore 560 059

[Terminology developments – statistical bibliography to Informetrics – are traced. Librametry and its applications, issues pertaining to bibliometric research, and methodological standards are discussed]

1. INTRODUCTION

Prof. P.C. Mahalanobis, founder of the Indian Statistical Institute, in the early 1950's argued that statistics is a "key technology" – as it is required for all socio-economic development activities and since statistical techniques are used in all development and forecasting studies. In recent past, statistics has been applied to a number of areas such as perspective planning, industrial and agricultural development, etc. Statistics and related techniques when applied in depth to a field may give rise to a new subject, econometrics is an example. Even in the field of library and information science, the applications of quantitative techniques are becoming increasingly popular. The early works in this area

goes back to mid 1920s and such studies – applications of quantitative techniques to library and bibliographical work – used to be known as “statistical bibliography”. Wittig (11), in a footnote traces the history of statistical bibliography indicating that the origin of the term could be traced to its first use by Cole and Eales (4) in 1917 and by Hulme (6) in 1922.

Ranganathan (1) suggested as early as 1948 at the Aslib conference in Leamington Spa “that it is necessary for librarians to develop “librametry” on the lines of biometry, econometry and psychometry, since many of the matters connected with library work and services involve large number”. In spite of his early attempt to define the scope of librametry, the subject hardly developed until the early 1970s. Pritchard (7) used the term ‘bibliometrics’ in 1969 to describe all studies which seek to quantify the process of written communication. Fairthorne (5) in 1969 defined bibliometrics as “ the quantitative treatment of the properties of recorded discourse and behaviour pertaining to it”. This same concept however in East Europe was then known as scientometrics. The scope of scientometrics in general is centred around quantitative aspects of science, science policy, science administration and particularly of the quantitative studies of “output of science”.

In another development, the term informetrics was proposed by Otto Nacke of West Germany in 1979 (3). In 1984, FID/IM was formed, Otto Nack as its Chairman. Rajan (8) from INSDOC, New Delhi, who was the next Chairman of the Committee, reformulated the objectives of informetrics to be the provision of reliable data for research and development; for policy-making and planning; and for the measurement of institutions, projects, programmes and activities. It is also said to be concerned with the origins and development of concepts(8). Brookes(3) in 1989 pointed out that this definition of informetrics is the widest and deepest of the three metrics terms, we are concerned with. Informetrics, basically used to connote the use and development of variety of measures to

study and analyze several properties of information in general and documents in particular.

In 1985, I mentioned in a short communication (9) on “Informetrics Vis-a-vis Bibliometrics: scope and its development” that informetrics is a field wherein the flow of

information and behaviour of information are analyzed, measured and quantitative relations are established. It is a scientific field wherein the developments of measurement of impact of Information is assessed continuously. Bibliometrics may be treated as a synonym to informetrics having a scope to analyze quantitative characteristics of information

Thus, several terms were coined to devote more or less same thing – statistical bibliography to librametry to bibliometrics / scientometrics to informetrics – over a period of time. New terms were coined partly because the parent subject has undergone a tremendous change in the last two to three decades; the change was from library science to documentation and then to information studies/information science. These changes mainly due to the impact of information technology and the inter-disciplinary nature of the subject. Also, for many years, this subject is known as informatics, particularly in Eastern Europe. Obviously, the quantitative studies in this field at different time periods were known as librametry (in 40’s), bibliometry / bibliometrics (in late 60’s), and informetrics (in mid 80’s). Of course, the most popularly known term ‘scientometrics’ is being used in Eastern Europe, since mid 60’s. If one carefully examines the scope and definition of bibliometrics / informetrics and scientometrics, one may easily conclude that they are synonymous to each other. In succeeding sections, I am discussing the scope and definition of these terms in little more detail.

The fact is that whether we call our research area as librametry, or as scientometrics or as bibliometrics or as informetrics, most of the topics we deal are :

- Quantitative aspects of library and information science, especially use and user studies.
- Quantitative studies related to book usage, acquisition, age distribution of documents etc.
- Circulation studies.
- Citation studies / analyses (impact factors and other measures)
- Journal productivity (by coverage, by use, by citation, etc.)
- Author productivity.
- Obsolescence and growth studies.
- Quantitative analysis of science (- science indicators, country-wise, language-wise, subject-wise etc.).
- Identifying relations among various disciplines
- Structure of subjects / disciplines
- Evaluation of scientific research (by institutions, by individuals, by countries etc.)

This usage of different terminologies is perhaps due to a lack of proper communication, especially before we publish our research in the learned journals. The lack of proper communication may be due to the absence of a “simple newsletter” in the field (preferably, by learned society, dedicated to informetrics/ scientometrics) or it may be due to “language barrier”. When Rajan took over as the Chairman of FID/IM, he had plans to bring out FID/IM newsletter regularly. He did publish one issue in June 1985 in which I had discussed scope and definition of informetrics and bibliometrics; afterwards, it was never published!

Acceptance of single term to define a subject and acceptance of its scope are necessary for any scientist; otherwise, it is difficult to include it in syllabi of library and information science course or in any other similar courses; and also it is difficult to get

research grants from different agencies. It further helps us in identifying the research groups especially at the international and national levels.

Time has come for bibliometricians, informetricians, and scientometricians to come-together and engage in a fruitful exchange of ideas with an objective to promote research in the area of quantitative studies in library and information science, including the quantitative studies of science in general and particularly of science policy, science programmes and science administration.

Glanzel and Schoepflin in their article “Little Scientometrics, Big Scientometrics ... and beyond” presented at the Fourth International Conference on Informetrics held at Berlin from 11-15 Sept. 93, argued that Bibliometrics/Scientometrics/Informetrics is in a crisis. Is it really in a crisis? Or is it some thing else? According to Glanzel and Schoepflin, the following are the causes of the crisis:

1. lack of integrating personalities
2. audience of bibliometrics: different target groups,
3. limited scope of Lotka’s and Bradford’s laws (according to Glanzel and Schoepflin, these laws do not “account for dynamics and flexibility of social forces which are influenced by many different factors”)
4. high cost of bibliographic data
5. publication policies and retrieval possibilities offered by the vendors which tend to put limitations on bibliometrics research
6. improper use of the research
7. breakdown of the social and political system in Eastern Europe

Some of the problems are trivial in nature and most of these issues should not be a major concern for informetrics research. However, to overcome these problems, the authors suggest:

- to reach a minimum consensus - in relation to theory and terminology,

- field must include all quantitative aspects of models of science communications and retrieval of scientific information,
- needs interactions with other disciplines and specialisties
- need for national research programmes
- need a code of ethics

Undoubtedly these are well recognized by many and one has to give a serious thought to implement them or practice them. But what is important is that I believe bibliometrics / informetrics / scientometrics is moving mostly towards theoretical work. In order to appreciate it one must apply his/her research work! One has to put a lot of efforts to apply it to the day-today activities in relation to “information handling”; it must be applied in libraries or information centres or in science administration. Measures must be developed or identified, not simply for the sake of developing or identifying; it must be used by the practicing professionals - to improve the existing systems and services, to study and understand the systems to reduce their operational costs, etc. Further, we may notice that a gap between the theoretical work in the “metrics” (whether scientometrics or informetrics or bibliometrics or librametry!) and its applications is increasing. If we carefully concentrate more on basic research in the field of quantitative studies in library and information science including science policy, science programmes, and science administration; and if the results of the research or measures which are developed/ identified are used by the practicing professionals (including information scientists, science administrators, people from information industry), soon we will have a “big scientometrics”; the “little scientometrics” (- librametry, bibliometrics, etc!) may then only be of historical importance. Shall we brand the “big scientometrics” as “informetrics”?

This workshop is designed basically to discuss these topics. Attempts are made to discuss recent advances, as far as possible. However, some elementary and fundamental topics are covered in this volume hoping that this volume will be useful for the beginners in the profession.

2. LIBRAMETRY : Some of the Applications

Ranganathan, formulated the normative principles, the Five Laws of Library science, by which he evaluated the helpfulness of otherwise of a technique, strategy or theory applied to the field; the normative principles are essentially information user-centred and help us in enhancing

- a) information use
- b) efficiency and effectiveness of information management
- c) convenience and economy of time and effort to the user

Ranganathan has intuitively applied quantification in a number of areas of library and information work. Some of them are :

- a) arrangement of books and periodicals
- b) selection of periodicals – by studying scattering and seepage
- c) budget allocation for books and periodicals
- d) physical planning of libraries
- e) workload and staff deployment
- f) reclassification and recataloguing
- g) choice and rendering of data elements in a catalogue entry

Ranganathan in the Madras university library arranged classes of books in the stack room by the frequency of their use rather than in the Colon Class number sequence. For example, the books in the class “O literature” were to be found nearest to the entrance on the ground floor, as this was the most frequently used class of books; those books which were on low demand were kept on the highest floor. Similarly, periodicals displayed by grouping, based on main subject, since a high percentage of the papers in a subject field are published in a few periodicals devoted to that subject, a smaller percentage in periodicals devoted to closely related subject fields and so on. Thus, a user

was able to find a good percentage of papers relevant to his/her interests within a relatively much smaller space on the display rack than if the periodicals were arranged alphabetically by title.

Also, study of requests for back volumes/issues of periodicals indicated that the demand was relatively much higher for issues of the two or three years preceding the current year and Ranganathan suggested that this fact be taken into consideration when sending vouchers of periodicals for binding so that users are not inconvenienced.

The other areas, where he used the quantitative techniques are related to decisions regarding the work load and staff deployment. These are all discussed in detail elsewhere (10). Further, in 1960s, DRTC Conducted several scatter and seepage studies in different subject areas with objects to :

- a) demonstrate that in interdisciplinary subjects the seepage was relatively more extensive
- b) provide a basis for periodicals selection, and
- c) understand the roots of a new subject – these studies were related to the modes of formation of subjects proposed by Ranganathan.

3. BIBLIOMETRIC RESEARCH

Bibliometric research include studies pertaining to scattering of articles over journals, growth of literature, obsolescence of documents, productivity and impact of research, distributions of scientific publications by country, by language; circulation studies, etc. Most of these studies are pertaining to countries, user groups, institutions, disciplines, documents, etc. There are very few articles in bibliometrics concerned with

theoretical and philosophical foundations. Some of the important studies on theoretical and philosophical foundations are in the area of :

1. Scattering of Articles
2. Author productivity
3. Word productivity/Law of least effort
4. Success-breeds-success phenomenon
5. Circulation theory

There are many other theoretical studies, especially in the area of circulation theory, citation analysis, sources-items relation, etc. For citations to most of these studies, one may refer to :

1. Proceedings of the Conferences on Informetrics (1987, Belgium; 1989, Canada; 1991, India; 1993, Germany, 1995, Chicago, 1997, Israel.)
2. *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*. By Leo Egghe and Ronald Rousseau (Elsevier, Amsterdam 1990).
3. *Quantitative Methods for Library and Information Science*. By I.K. Ravichandra Rao (Wiley-Eastern, New Delhi, 1983)
4. *A Bibliography of Bibliometrics and Citation Indexing and Analysis*. Stockholm. The Royal Institute of Technology 1980.

Yet another important area of bibliometric studies is related to citation analysis. It is largely concerned with examinations of references in various articles published in journals. It is centered around :

1. Which authors/journals are most cited
2. Existence of linkages (who cites whom, which journal cites which journals, what subject areas are cited in the literature of a particular discipline, etc.)

Bibliometrics studies, based on citation data however have several limitations. They are :

1. Citation studies are mostly dependent on SCI data - which cover only limited number of journals and its coverage does not remain constant, since new journals are added regularly and some are dropped.
2. Cited articles appear only under the name of the first author.
3. Problems of eliminating self citation are cumbersome.
4. Unethical practices in cited articles
5. Languages bannier
6. Biased to certain countries, etc.,

In most of the bibliometric studies, the following data are collected and analyzed.

1. Measures of Productivity (Number of publications, Number of pages, Cost-effectiveness measures, etc.)
2. Measures of impact (Reprints request, Photo copies made, Citations received, Sources of citations, Immediacy of citations, Number of reviews, Adoption rates (text books!).
3. Measures of Journals Productivity, (size: pages, papers, words, etc., Circulations/sale, Uses, Impact, Age of sources cited, Coverage in data base, Cost-effectiveness measures, number of relevant articles in a journal on a given topic, etc.
4. Growth and obsolescence – number of publications and number of citation received over a period of time.

A primary objective of bibliometric research is the development of a general and systematic set of theories from which hypotheses may be generated and tested. But, if explanations for library or science or information phenomena are based solely upon

individual (s) experience(s) and know-how then they may not be accurate and/or reliable. Hence, they would not be amenable to generalization.

4. METHODOLOGICAL STANDARDS

Bibliometric studies vary from each other on several points of view. They adopt different methods of data collection as well as different techniques. Even, there are no universally accepted terminologies. In addition, use of algebraic symbols vary from one study to another. Under these circumstances, it would be difficult to think of “bibliometric standards”, let alone formulating them.

Most of the bibliometric studies are empirical in nature. In such circumstances, to reproduce bibliometric research, one has to repeat the survey and analyze the data right from the beginning. In Natural Science, it is possible and quite common that research may be repeated in laboratories. But in Social Sciences, such a thing is not only difficult, but is not possible. Further, an important cause of over all unreliability and therefore a cause of invalidity in any basic research

in Social Sciences is due to “small sample” size. If the study is based on “large sample”, it is undoubtedly difficult to reproduce the results of research.

Under the circumstances, if the study is based on certain guidelines/methodologies, one may easily accept its validity as well as its generalization of the result. Also, if scientific methods are followed, we may satisfy ourselves with certain statistical parameters - measures of central tendency, measures of dispersions, confidence limits, etc. The general guidelines are:

1. Identify the general problem(s)
2. Conduct literature search.
3. State specific problem(s)
4. Decide the design methodology

5. Gather data either for the whole population or for the sample
6. Analyze the data
7. Report the results
8. Refine the hypotheses

(Note : Steps 2 to 8 may have to be repeated)

In Step 1, the research objectives are explicitly identified and described. Further, information about the research objectives and investigative tasks are analyzed, and relevant terms and variables are defined. The research question is stated and/or hypotheses are formulated in Step 3 with a clear-cut definitions, assumptions, suppositions, etc. The objective of the study is then carefully observed and, if necessary, causal factors associated with the observed phenomenon are identified. If sufficient data are collected, the hypothesis may be tested with some degree of certainty. If data are collected from machine readable databases, results are much more reliable than otherwise. Thus, whenever possible, one may use, databases for collecting the data. However, investigators are more certain when they select and conduct similar studies (i.e. conduct follow-up case studies in the same area.)

In most of the bibliometric studies, several assumptions and/or suppositions are made. The validity of the results are often depending on how far such assumptions and suppositions are valid. For instance, Bradford(), in his study assumed that $n_1=n_2=n$, where $n_1=r/r_1$ and $n_2=r_1/r_2$; r , r_1 , r_2 are the average numbers of articles per periodical in three different zones, in formulating his law of scattering. How far this assumption is correct? One has to critically look into it!

In order to arrive at valid results, it is thus recommended that in every bibliometric studies one must strictly follow the scientific method as enunciated in the above said general guidelines. A mere presentation of the summary of the data should not be considered as the end results of bibliometric research.

If these guidelines are followed in bibliometric studies, they are as good as following bibliometric standards. These are of course, methodological and conceptual standards. In addition to the above guidelines, depending on the problems, the following bibliometric methods and procedures may be adopted for processing data :

1. Rank and size frequency analyses
2. Correlation and regression analysis (fitting linear and non-linear models.)
3. Cluster and factor analyses.
4. Analysis of variance
5. Use of OR techniques
6. Statistical Inference, etc.

Thus what is most important is that there is no single method which is applicable for all bibliometric research. There are different methods and procedures for different problems and they are different standards. It is evident that bibliometric studies are still in empirical stages. The facts are gathered either through surveys or from published bibliographies, indexes, data bases. Based on these facts, empirical models and principles are being developed. The normative principles and standards have to evolve from the logical analyses of the empirical models. The stage is set to integrate empirical models of bibliometrics into standards. Further, bibliometric studies have to address these issues and reach the stage of normative principles.

REFERENCE :

1. ASLIB Proceedings. 1; 1949, p.102.
2. Bradford. S.C. Documentation. 1948. Crosby Cockwood. London p156.

3. Brookes, B.C. Biblio-, Sciento-, Info-, metrics? What are we talking about? (In *Informetrics* 89/90, Egghe and R Rousseau Eds. Elsevier, Amsterdam. 1990. p31-44)
4. Cole, F J and Eales, N.B. The history of comparative anatomy Part 1 : A statistical analysis of the literature (*Science Progress.* 11; 1917, 578-96).
5. Fairthorne, Robet. A. Empirical hyporbolic distributions (Bradford-Zipf-Mandelbort) for bibliometric description and predication. (*Journal of Documentation.* 25, 4; 1969; p319-43).
6. Hulme, E.W *Statistical Bibliography in relation to the growth of modern civilization.* 1923. London.
7. Pritchad. A. *Statistical bibliography and bibliometrics* (*Journal of Documentation.* 25, 4, 1969, 348-9.
8. Rajan T.N, and Sen B.K. *An essay on informetrics: a study in growth and development.* IASLIC XV. Bangalore, 1985.
9. Ravichandra Rao I.K. *Informetrics vis-a-vis Bibliometrics.* FID/IM Newsletter, No.1 June 1985. Published by INSDOC, New Delhi.
10. Ravichandra Rao. I.K. and Neelameghan A. *From Librametry to Informetrics: An overview and Ranganathan's contributions.* (*Libri;* 1992;40,3; p 242-57)
11. Wittig, Glen R. *Statistical Bibliography - a historical footnote.* (*Journal of Documentation.* 34,3; 1978; p.240-1)