ISSN: 2319-8435 Impact Factor :3.1028(UIF) Volume - 5 | Issue - 1 | November - 2016

e-Library Science Research Journal

A SURVEY ON BIBLIOMETRICS TOOL AND TECHNIQUE

Dr. Shiv Singh
Assistant Librarian, Central Library, Lovely Professional University, Punjab, India.

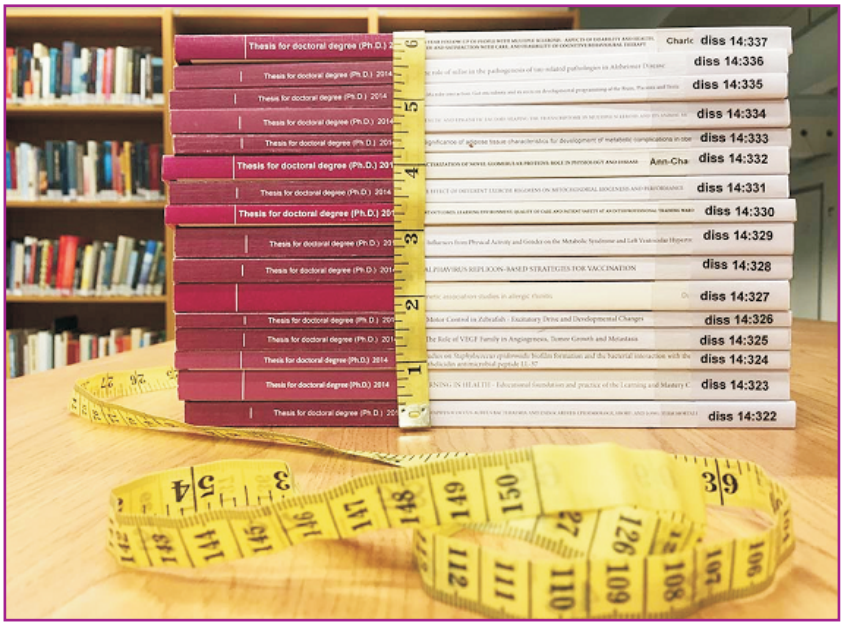
ABSTRACT

Scholarly communication is remarkably essential for the understanding of the genesis and progress of disciplines. There are various statistical tools used to explore the databases in library science i.e. Bibliometrics, Scientometrics and Infometrics, etc. The Bibliometrics and Scientometrics are very much related, their different roles are distinguished by their different context. Bibliometrics is used to get better scientific documentation, information and communication activities using quantitative analysis of library collections. In the present article the importance and the limitations of Bibliometrics are focused. In addition to that laws of bibliometrics are also discussed along with their implications.

KEYWORDS:*Bibliometrics, Lotka's law, zipf law, Bardford laws.*

1.INTRODUCTION:

The Bibliometrics and Scientometrics are very much related, their different roles are distinguished by their different context. Bibliometrics is used to get better scientific documentation, information and communication activities using quantitative analysis of library collections. On the other hand scientometrics, gives a better understanding of mechanism of scientific research as a social active, a quantitative analysis of generation, transmission and utilization of scientific information.[1] In this perspective L. Egghe[2] stated that the notion of scientometrics deals more with science policy research and therefore is more related with citation analysis studies. However, it is important to mention here that the father of library science i.e. Dr. S.R. Ranganathan in 1948 introduced the term 'Librametry', which is analogous to bibliometrics.[3] The term informetrics was introduced by O. Nacke4, it focuses on information productivity, it integrates the information technology and complex connections of information theory, cybermetrics and decision theory etc. Thus, the three terms discussed are nearly synonymous to each other and they have much common factors, so it depends on the researcher which they may like to select and use. However, the Bibliometrics is still used as a stable term because it is considered as the one of the most attractive subject of information research for the librarians and the other information professionals. This subject has attracted a considerable attention because its helps to improve scientific documentation, information and communication activities by quantitative analysis of library collections and services. This subject is



treated as one of the techniques to evaluate and study the scientific works. The word bibliometrics is a grouping of two words i.e. Biblio and Metrics. Biblio is derived from Latin/Greek word "Biblion" means "Books" and Metrics is derived either from Latin/Greek word means "Matricus" (in Latin) or "Matrikos" (in Greek) i.e. measurement.

Bibliometric study has many practical applications; few of them are as follows:

- It helps to improve the bibliographical control because bibliometric analysis helps to know the character of literature in various fields.
- It derives the subject relationships.
- The citation data analysis and the volume of publication in year wise can be useful in planning retrospective bibliographies.
- It also determines statistics of literature relating to the country of origin, subject, and language distribution of documents as well as their incidence of translation.
- The bibliometric data also helps in taking some management decisions.
- It also provides information regarding the structure of knowledge and pattern of communication.
- The citation data also helps to determine the list of highly cited journals or books, which can be used in taking decision while discarding the stock of the library.

In spite of its strengths and potential contributions to the understanding of the scientific enterprise, it has also received a fair amount of criticism. F.W. Lancaster[5] shed light on the limitations of bibliometric study. Some of them are listed below:

- This Study does not include the informal publications and communications. Hence scientific development cannot be predicted properly.
- The bibliographical references, taken for citation analysis study, are not always standardized. This makes the difficulty while ranking the authors on the basis of the frequency of their getting citations.
- Self citation is another limitation of citation analysis study.

Besides these limitations, bibliometric study is one of the best tool to acquire knowledge of scientific productivity of individual authors, institutions and journals and to study the pattern of growth of literature and nature of research publications, age of library used, information needs of scientist, etc.

In the present paper, we have tried to explore the concept and importance of bibliometrics. Section 2 contains the chronological outlook of bibliometrics. The Lotkas's laws of bibliometrics is presented in section 3. The section 4 and section 5 focus on Bradford's and Zipf law respectively. The summary is presented in section 6.

2. Bibliometrics: Chronological Outlook

"Statistical analysis" was used by Cole and Eales6 in their articles "the history of comparative anatomy". They had graphically represented the activities of comparative anatomists for a period of time. They reported that "it seemed possible to reduce to geometrical form of activities of the corporate body of anatomical research, and the subject". A description of literature, using publication counts and graphic illustrations by year and country has been reported by Cole and Eales[6]. They reported the publications dealing with animal anatomy for the period 1543 to 1860, in chronological charts. The influence of contemporary events, public bodies and individuals on the history of anatomical thoughts have also been reported by them.

After many years in 1969 A. Pritchard[7] coined term "bibliometrics" in a paper entitled as "Statistical Bibliography or Bibliometrics" which was published in the Journal of Documentation. He has used this term to replace the earlier world "Statistical Bibliography". He emphasized that the use of this term is "to shed light on the processes of written communication and of the nature and course of development of a discipline (in so far as this is displayed through written communication), by means of counting and analyzing the various facets of written communication". Different authors have given different perspectives on bibliometric study likewise D.E. Price8 proposed a cumulative distribution, which models statistically the situation in which success breeds success. The explanation given by him differs from the negative Binomial distribution in that lack of success, being a non-event, is not punished by increased chance of failure. He has also shown that such a stochastic law is

governed by the Beta Function, containing only one free parameter, and this is approximated by a skew or hyperbolic distribution of the type that is widespread in bibliometrics and diverse social science phenomena. He also concluded that the Beta function is clearly well-designed for these multiple purposes since it gives both the actual and the cumulative distribution in simple form, and contains a limiting case of an inverse square law. Nicholas and Ritche[9] have divided the bibliometric studies in two broad groups' i.e. descriptive studies and behavioral studies. They also stated that both studies are complimentary to each other. I.N. Sengupta[10] affirmed that this study lies between the broader areas of the social science and the physical science. In similar trend W.G. Potter[11] has also divided the bibliometric studies into two categories. Firstly descriptive (means to study the body of literature) and secondly, more evaluative (attempts to study the use of a body of literatures with help of citation analysis).

3 Lotka's Law: Inverse Square Law, Relation of Authors to Papers

A.J. Lotka12, in his classic paper published in the Journal of the Washington Academy of Sciences on the frequency distribution of scientific productivity has given an analysis of the number of publications listed in Chemical Abstracts from 1907 to 1916 with the frequency of publications by particular authors. He expelled the names of mutual authors, but has only considered the names of authors whose names start with A and B as listed in the index. He neither adopted a sampling design nor computed a sample size by using scientific methods. He has also made a similar study in the field of physics. He applied the same process to the name index of Auerbach's Geschichtstafel der Physics which covers the entire range of history up to the year 1900. He developed an inverse relationship among number of producers and production. This relationship is summarized in a form of power law referred to as Lotka's Law. He found that once he knew the number of authors writing one paper he could predict the number of authors writing two, three or more. Thus, he formulated a general equation for his law.

The relation between the frequency Y of person making X contribution.

$$X^n \propto \frac{1}{Y}$$

or $X^n Y = C$

Here, Y is the Frequency of authors contributing X papers each, and C is a constant.

He has given the graph between the percentage of authors making 1, 2, 3…n contributions and the number of contributions with both variables on a logarithmic scale. After that the least square method was used to determine the slope of the line that best fit the plotted data, and reported that the slope was approximately 2. Putting n=2, he then found the value of constant C.

For larger value of X, the variation was too high. Therefore, he considered only first 17 point of data in physics and first 30 point in Chemistry. Based on this data, the inverse square law of scientific productivity was given as.

$$Y_X = \frac{K}{x^{\alpha}} \qquad x = 1, 2, 3 \ldots\ldots \alpha$$

Y_x is the relative frequency of authors publishing x numbers of papers. The value of α is found to be 2 for physicists and 1.89 for chemists. The reason behind this difference is due to sampling error. Thus Lotka's equation is determined in its general from by 3 important parameters.

- The number of scientists with minimal productivity (authors with single paper each-NY1)
- The maximal productivity of scientists (X max.)
- The Characteristics exponent X.

4 Bradford's Law: Scattering Observations Published

In 1934, S.C. Bradford13, the Director of Science Museum Library thought that abstracting and indexing services might be lacking as much as two-third of the total literature on a given subject. The two hypothesis given by him were on the structure of journal and subject.

The classical statement given by him is as follows "If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus; when the succeeding zones will be as $1:n:n^2$ (i.e. 1:5:25)". In 1948, Bradford searched for papers related to applied Geophysics and Lubrication. He considered an example that the first region contains only a small number of highly productive journals, says n1; the second region contains a large number of moderately productive journals, say n2; and the third region a still large number of journal of low productivity say n3. Then as per as the Law of scatter:

$$n1 : n2 : n3 = 1:\alpha:\alpha^2$$

Where a is a constant

He suggested the Empirical Law of Scattering in a different way which states:"If periodicals contributing to a subject are ranked and then grouped in such a way that each group contributes the same number of articles, the number of periodicals in each group increases geometrically". Thus the Bradford's law has universal importance in application from the basic use in scientific field.

5 Zipf's Law

The influential work in the area of distribution of frequency of appearance of words in texts has been performed by G.K. Zipf.14 Zipf's first Law, the principle of Least Effort, has been shown to be relevant to almost any distribution in which a variety of items can be ranked in order of frequency of occurrence. He has given an equation correlating high and low frequency words of long textual matters of natural language and showed that there exists a reciprocal correlation among rank word and frequency of words. The statement of Zipf's law is:

"In long textual matters if the words are arranged in their decreasing order of frequency, then the rank of any given word of the text will be inversely proportional to the frequency of occurrence of the word". Mathematically it was written as follows

$$r \propto \frac{1}{f}$$

or, $rf = c$

Where r stands for rank words, f stands for frequency and c is a constant.

$$\log f + \log r = \log c$$

This law gives only an approximation of relationship between rank and frequency (which is hyperbolic). A more generalized picture of Zipfs law was:

$$f = \frac{c}{r^{\alpha}} \qquad (\alpha > 0)$$

This form was adopted by Booth in his model by J. Fedorowicz15.

A. Booth16 however, suggested a method for estimating the Zipf constant c (in the equation rf = c) for a given author. He expressed the Zipfs law as:

$$\frac{l_1}{l_n} = \frac{n(n+1)}{2}$$

Where, I_n is the number or proportion of distinct word, each of which appears in the text exactly n times and I_1 is the number or proportion of distinct word; each of which occurs only once. This statement was also known as Zipf's second law. Booth16 argued that:

• His estimation provide a measure of author's richness in vocabulary.

• The general form of the law of occurrence for low frequency word is independent of the detailed validity of Zipfs law for the distribution as a whole.

6 SUMMARY:

Bibliometric is one of the most prominent tool to analyze the productivity count of literature for a given field. It is used for comparing the amount of research in different countries, during different periods, by the different authors, on different subjects, relationship of quantity of publication and their production of literature. It is most widely used for finding the core journals, ranking of journals, growth of publications, impact of one journals on the other, prominent authors and the pattern of their contribution etc for the any literature or field.

REFERENCES

1.Braun, T. et al. (1985). Scientometric indicators, a 32-country comparative evaluation of publishing performance and citation impact. Singapore: World Scientific.

2.Egghe, L. (1988). Methodological aspects of bibliometrics. Lib. Sc., 25, 179-191.

3.Aslib proceeding (1949). 1, 103.

4.Nacke, O. (1979). Informetrie: Eine Neuer Name für eine neue Disziplin. Nachrichten für Dokumentation, 30(6), 219–226.

5.Lancaster, F. W. (1991). Bibliometric methods in assessing productivity and impact of research. Sarada Ranganathan Endownment for Library Science (pp. 1-15). Bangalore.

6.Cole, F. J., & Eales, N. B. (1917). History of comparative anatomy. Pt. 1. A statistical analysis of the literature. Science Progress, 11, 578-596.

7.Pritchard, A. (1969). Statistical Bibliography or bibliometrics? Journal of Documentation, 25(4), 348-349.

8.Price, Derek de Solla (1976). General theory of bibliometric and other cumulative advantage processes. Journal of the American Society For Information Science, 27(5-6), 292-306.

9.Nicholas, David, & Ritche, Maureen (1978). Literature and bibliometrics (pp. 18). London: Clive Bingley.

10.Sengupta, I. N. (1985). Bibliometrics: A birds eye view. IASLIC Bulletin, 30(4), 167-174.

11.Potter, W. G. (1981). Introduction to bibliometrics. Library Trends, 30(1), 5-7.

12.Lotka, A. J. (1926). Statistics–The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16, 317-325.

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

14.Zipf, G. K. (1965). The Psycho-Biology of language: An introduction to dynamic philogy. Cambridge: MIT Press, MA, 1935; 3, 7, 18, 25, 41-44, 47, v, preface, reprint ed., 12.

15.Fedorowicz, Jane (1982). A Zipflian modle of an automatic bibliographic system: An application of MEDLINE. Journal of American Society of Information Science, 223-32.

16.Booth, Andrew (1976). A 'Law' of occurrences for words of low frequency. Information and Control. 10, 386-93.

Dr. Shiv Singh
Assistant Librarian, Central Library, Lovely Professional University, Punjab, India.