BIBLIOMETRICS: AN ANALYSIS

Manisha S. Tandale
Librarian
H.J. College of Education, Mumbai
E mail: manisom20@gmail.com

Abstract: This paper deals with bibliometrics study on improving scientific documentation, information & communication activities by quantitative analysis of library collections & services. Bibliometrics is recognized as method to analyze & quantify the bibliographic data & offers a powerful set of methods. This is importance to measures for studying the structure & process of scholarly communication. The aim of this study was to identify importance, applications, & limitations of bibliometrics technique.

Keywords: Bibliometrics, Lotka's law, Bradford's law, Zipf's law, Webometrics.

1.0 Bibliometric

A term first appeared in 1969, to replace statistical bibliography (E. W. Hulme in 1923). It refers to study of the use of document & patterns of publication, by mean of mathematical & statistical methods. The purpose of statistical bibliography is to shed light on the process of written communications & course of development of discipline (is so far as this is displayed through written communication) by means of counting & analyzing the various facets of written communication.

The word Bibliometrics is combination of two words i.e. Biblio & Metrics. Biblio derived from Latin/ Greek word 'Biblion' means 'Books' on the other hand 'Metrics' is derived from latin word 'Metricus' the meaning of this word 'Measurement'.

2.0 Librametrics

Librametrics first stated by Dr. S.R. Ranganathan in ASLIB conference in London 1948. Sen Gupta he was defined 'It is competitive analyses various facets of library activities & library documents by applied mathematical & statistical calculation to solved library problems.

3.0 Informatics

Informatics deals with measurement of information, it is first proposed by Okktonack in 1979. Informetrics formal study of information, including its structure, properties, uses & functions in society, in particular the technology used to record, organised, store, retrieve & disseminate. Ravichandra Rao (1993) stated that "Informetrics connotes the use & development of variety of measures to study & analyze several properties of information in general & document in particular. He further added that the scope of informetrics covers both bibliometrics & scientomrtrics

4.0 Scientometrics

Scientometrics deals with the measurement & self organisation of scientific communication. Scienometrics refers to graphic representation. Bookstein (1995) defined scientometrics as 'the science of measuring science'. This is also considering as bibliometrics measurement for evaluation of scientific development, social relevance, & impact of application of science & technology. The term Scientometrics has been more widely used in Europe & other former Soviet Union.

5.0 Webometrics

The term Webometrics was first coined by Almind & Ingwersen (1997). They described the use traditional informetrics methods as a starting point for analysis on the web as general conceivable for any kind of statistical aspect. It is something called as Internetrometrics, Cybermetrics; Webometrics is a study of quantitative aspect of construction & use of information resources, structure & technologies on the World Wide Web, drawing on the Bibliometrics, Informetrics approaches.

6.0 Definition of Bibliometrics

The word bibliometrics first appeared in print in 1969 in **Alan Pritchard's** article statistical bibliography. The term statistical bibliography is clumsy, not very descriptive, & it can be confused with statistics itself. As a result, the term BIBLIOMETRICS was introduce i.e. the application of mathematics & statistical methods of book & other media of communication be substituted for statistical bibliography.

British Standard Institution (1976) described the bibliometrics as the "application of mathematical & statistical methods in the study of use of documents & publication patterns.

7.0 Bibliometrics: Scope & Purpose

Bibliometrics studies are generally based on quantitative measurements without any quantitative evaluation. They are therefore considered only as partial indicators of scientific progress.

- Bibliometrics data can be used to compare the productivity & impact of research conducted in various countries & regions.
- It sheds light on the process of written communication & on the nature & course of development by a descriptive means of counting & analyzing the various facts of written communication.
- It provides information about structure of knowledge & how it is communicated.
- The scope of bibliometrics includes studying the relationship within a literature (citation studies) or describing literature, Typically, these descriptions focus on consistent patterns, involving authors, monographs journals, or subject/ language.
- It is quantitative science & it is divided into two categories

1. Descriptive bibliographic (productivity count)

   a. Geographic
   b. Time Period
   c. Discipline

2. Evaluative bibliometrics (literature usage count)

   a. Reference Count
   b. citation count

8.0 Bibliometrics: Its Applications

The bibliometric study besides its theoretical content has various practical applications in library management & helps in deciding science policy on researchers.

- The bibliometrics study aims to improve the bibliographical control because bibliometrics analysis helps to know the character of literature in different fields. The volume & growth of primary literature has a direct effect on structure of secondary literature Therefore, the computed growth rates & direction of change may be of considerable assistance to editors of secondary services in determining their future approach & coverage
- A major area of bibliometrics study is to determine statistics of literature relating to the country of origin, subject, form & language distribution of documents as well as their incidence of translation. These data will provide useful information determining the scope of work, & can suggest weakness in the coverage of possible improvement of secondary services. The bibliometrics analysis helps in comparative assessment of the secondary services particularly when related to overall figures on size of literature & to subject links. This may help to the publisher in getting an idea of their achievements & competitions & could be useful for marketing purpose.
- The citation data also determines the list of highly cited journals or books, which can   be used in taking decision while discarding the stock of the library.
- Citation analysis can find out subject relationships which help in suggesting titles of journals relevant to a given discipline in a particular library.
- Bibliometric study also provides information about the structure of knowledge & pattern of communication. Analysis of the size & growth of literature can identify the developing & declining areas of literature over a time & trend of literature growth.
- To develop norm for standerdasation.

9.0 Bibliometrics in Indian Scenario

There is very little awareness among the Indian scientific communities about the bibliometrics & scientometrics techniques & studies. Institutes like National Institute of Science, Technology and Development Studies (NISTADS), New Delhi, National Institute of Science Communication and Information Resources (NISCAIR), New Delhi; & Bhabha Atomic Research Center (BARC), Mumbai are the pioneers in this field in India.

There are some Universities carrying out research in this field. The southern part like Dharwad, Banglore, Karnataka, and Madras use bibliometrics technique more than Northern India. The library & information Science departments should be well equipped to deal with the new techniques & have become tool to evaluate the productivity of research institutes, individual's researcher & to map the growth of subject

10.0 Bibliometrics Proposal for Inclusion at Master Degree Level

Bibliometrics can best be taught to the Library & Information Science at the Master Level. The content can be designed:

- To teach students the basic principals of bibliometrics as related to scholarly literature.
- To work toward the construction of adequate theory of bibliometrics.
- To review the practical application of bibliometrics method for information retrieval system.
- Student should be familiarizing themselves with the landmark & the recent trends in bibliometrics research & literature to make themselves sufficient equipped for advanced research in future.

11.0 Empirical Laws of Bibliometrics

Lotka's law of scientific productivity- **Alfred J. Lotka** was a mathematician, supervisor of mathematical research in Statistical Bureau of the Metropolitical Life Insurance Company from 1924 to 1933. During this time in 1926, that the definitive work, later called Lotka's law it is also known as "Author productivity" law. This law pertains to the distribution of productivity in scientific community. It shows that there is considerable difference among scientist as for as productivity is concerned, for example as measured by the number of publication.

Lotka's law describes the frequency of publication by author in given field. It stated the number of person making 2 contributions is about one forth of those making one, the number making 3 contributions is about one night etc, the number making n contributions is about $1/n^2$ of those making one, & the proportion of all contributors, that make a single contribution, is about 605

Now, considered a real application of Lotka's law according to the Library & Information Science Abstract 1992.

No. Articles (x)	No. of Authors observed (y)
1	7229
2	771
3	198
4	50

Lotka's law is not an arithmetically accurate law. It is an approximation, but there is great logical thinking behind it. Very few people can write a number of articles. Many individuals can at most write one article among them only a few venture to write the second then the number declines.

Bradford's law of scattering (1934)- **Samuel Clement Bradford** was a very determined & dedicated person is very evident in his writing. According to this law of Bradford, journals in a field can be divided into three equal parts, each with about $1/3^{rd}$ of the article such that the first group has core of few journals, a second zone contains more journals & the third zone is bulk of journals. The main contribution of the law is that the number of journals in each group is given by the ratio $1: n: n^2$.

Bradford's law of scatter is used in library management as a guideline for journal selection. Particularly in interdisciplinary subject like instrumentation, environment science & nutrition etc. this method is very useful in journal selection. Subscribing to a few core journals is more cost effective decision that obtaining all journals cited in a study. This law is also an approximation, & it has no arithmetical precision. Bradford formulated his law after analysing bibliography of geophysics covering 359 journals in the field. These were grouped into the following three groups. (One should view these as group data failing in the range 401-500. Therefore, it is an approximation. Even the law states that it is "about 1/3"

	No. of Journals	Actual No. of articles
	9	429
	59	499
	258	404
Total	326	1332

Expected Number (1:n:n ^{2}) should have been 9x6=54 & 9x6=324. But as stated earlier the law is not statistically accurate, but used as a good approximation

Zipf's law (1933- The third law of the bibliometrics introduced by **George Kingsley Zipf**. This law states that the probability of occurrence of words or other items starts high & topers off.

Zipf's law is often used to predict the frequency of words within a text. In any communication, the words and, the, a, an, on, of…. etc occur many times in a text, whereas 'measurable' is rare, Therefore, in discussion about

Zipf's law such frequency appearing common words are excluded, for eg CDS/ISIS know that such words are in 'stop words' file. Such a file is created in the CDS/ISIS while using indexing technique. The equation for this relationship is r/f=k, where *r* is the rank of word, *f* is the frequency & *k* is the constant.

Consider the following table to understand the law:

Words	Rank	Probability of Occurrence	
-	1	1	1
-	2	½	0.5
-	3	1/3	0.33
-	4	¼	0.25
-	5	1/5	0.2

Zipfian distribution of words & corresponding distribution of probabilities of occurance have a great application in machine indexing. Zipf's law, again, is not statistically perfect, but it is very useful for indexer.

Along with these three laws, there are concepts that are integral to the bibliometrics. These are,

Immediacy Index- Immediacy index is a ratio of number of citation a journal receives in the current year to the number of articles published during that year.

II=   Number of citations a journal receives in the current
      Number of articles published that year

Immediacy Index for 2007 is ratio of number of received in 2007 to the number of articles published in 2007.

Half- Life- Cited Half-Life is a measure of the rate of decline of the citation curve.

Impact factor- Impact factor is a quantitative measure of the frequency with which the average article published in a given scholarly journals has been cited in a particular year or period.

Impact Factor has huge but controversial influence on the way a published scientific research is perceived or evaluated. It is measure of relative size of the citation curve between second & third year of its publication

Number of current citations a journal receives to

Impact Factor   =   Articles published in two previous year
                Number of articles published in those same years

In other words, Impact Factor for 2007 refers to citations in 2007 to the articles published in 2005 & 2006.

Impact factor depends on several factors. These are;

✓ Subject area of the journal.
✓ Type of journal (Letters, full papers, revises) If there are more full papers then there can be better impact factor.
✓ Average number of authors/ paper
✓ Size of journal
✓ Size of citation measurement window (2 & 5 year)

Based on these three laws & concepts, relevant data collection techniques are developed under the umbrella of bibliometrics for different research problems.

Bibliometrics studies includes studies of the growth of literature in some subject, how much literature is contributed by various individuals, groups, or countries; how much exists in various language; how the literature on some subject is scattered; & how quickly the literature on some subject become out of date. One very important group of bibliometrics studies related to what sources author cite. Citation analysis is concerned with such phenomena as: which author are most cited which journals are most cited & what linkages exist through citation & so on.

12.0 What is Citation Index?

Scientific journals published papers, notes, reviews, corrections etc. Each of these items has title, authors, affiliations, text & lastly citations or foot notes or bibliography. In authors reference to previously recorded information identifies much of the earlier work that is pertinent to the subject of his/her document.

A citation index is an ordered list of cited articles each of which is accompanied by a list of citing articles. The citing article is accompanied by a source citation. The index is arranged by reference citation & under by first author. The best know example of Citation Index is the Science Citation Index (SCI) developed by Eurgen Garfield in 1963. This publication is very much useful for:

- ✓ Conducting research activities in the field of information science,
- ✓ Studying the characteristics of scientific literature &,
- ✓ Conducting citation behaviour studies in the history & sociology of science.

13.0 What is Citation Analysis?

Everyday scientist reports the results of experiments, studies, exploration, & invention in thousands of publications. How to keep track of these research developments of ideas when one author cites another author, a relationship is established. Citation analysis is a bibliometrics techniques in which work cited in publications are examined to determine pattern of scholarly communication.

13.1 Citation analysis useful for-

- ✓ Evaluative bibliometrics- It is defined evaluation & interpretation of citation received by articles, scientists, universities & used as a measure of scientific influence & productivity of information.

- ✓ Information retrieval- In information retrieval citation play important role & considered as useful supplements of keywords in identifying relevant document.
- ✓ Collection development- The citation analysis will help in deciding titles of journals to be acquired, to continue or discontinue subscription. A citation analysis can be used as tool for collection development.
- ✓ User studies- The application of citation analysis technique in determining user needs is very useful for collection development & design library services. Citation analysis can also be used to compare user behaviour today with user behaviour several years ago.

Problems & limitations of bibliometrics methods Bibliometric like any other technique is not free from criticism. Some of the major problems & limitations of bibliometrics:

- ▪ Bibliometrics study does not include the informal publication & communications. Therefore, the scientific development cannot be predicted properly.
- ▪ The biblical structure, references, which is taken for citation analysis study, is not always standardize. This causes problem while ranking the authors on the basis of frequency of there getting citations. For eg the author S. R. Ranganathan has been cited variously as Siyali Ramamrita Ranganathan this may cause scattering of citations some person works. At the same time, it may also happen that there may be more than one author under the same name & it may be difficult to distinguish them.
- ▪ In case of collaborated author, the cited articles appear only under the name of the first authors as listed in each article. Here one must determine the names of individual's contributors who have collaborated the work to get the actual rank of the authors. This is very difficult & tedious process.
- ▪ While the citation indexes are completely open ended as far as what is cited is concerned, they cover only limited numbers of citing journals.
- ▪ Another problem to be faced, & perhaps the most serious one, is that the coverage of the Science Citation Index (or any other database for that matter) does not remain constant over time new journals are added & other are dropped (or cease publication) & even new types of material may be added for the first time. Unless one can control for this, which is by no means easy to do, serious errors of interpretation can occur from use of bibliometrics data.

- Self- citation is another limitation of citation analysis study. Although the case of self citation of single author publications can be determined & limited to find out the self citations of authors specially for author in second or later positions.
- Another problem of citation analysis is that in some cases the citation it self is not proper to the context of the citing articles. In some cases, the citations are also incomplete.

14.0 Conclusion

Bibliometrics is the quantitative presentation of study. This study is the analysis of the structure of literature using various tools, counting, rank frequency, distribution, & citation analysis, & although the structure of literature is basic to all disciplines, it is particularly important in the area of information retrieval. It seems evident that bibliometrics & other metrics was first used in attempts to evaluate journals for collection development, user study & for effective decision making for libraries & library management, it has been recognized & expanded to the study of the structure of literature in the larger encouraging field of library & information science.

In 21st century the electronic or online environment is introduced in libraries & information centres, hence the mode of electronic publications with the popular usage of Internet is need of environment. Accordingly, the libraries are also become digitized. But in spite of this bibliometrics studies still continued investing field of study among researcher, librarians, & library & information scientist thus the some bibliometrics has been emerging on the World Wide Web popularly known as Cybermetrics or Webometrics. It can be defined as the application of order of bibliometrics, informetrics & scientometrics techniques to the web in order to study the relationship of different sites on the web. Such techniques may also have used to map out (called scientific mapping in traditional bibliometrics research) area of the web that appear to be most powerful, useful based on the number of times they are hyperlinked to other web sites.

15.0 References

1. Rao, Ravichandra, Informetrics. Banglore: Sarda Ranganathan Endowment, pp.133-142.
2. Kawatra, P. S, Bibliometrics Scientometrics & Informetrics in: Textbook of Information science. New Delhi: A.P.H Publishing Corporation, 2000, pp. 35-61.
3. Mahapatra, Gayatri, Bibliometrics studies in Internet era. New Delhi: Indiana Publishing House, 2009.
4. Reddy, Rama, technology Management in Libraries. Allied Publishing.
5. Savanure, S. K, Research methodology for information science. Pune: Universal Prakashan, 2008, pp.98-107.
6. Wolfram Dietmar, Applied informetrics for information retrieval