CHAPTER 2

REVIEW OF LITERATURE
Bibliometric analysis has now become a well established part of information research. Both in research and practice the bibliometric analysis is gaining ground. The most obvious use of bibliometric data is to improve bibliographic control, as it is clearly not possible to provide efficient secondary services without knowing the size and character of a literature. Bibliometrics furnishes possible methods by which significant features of a literature may be delineated and its working monitored. In fact bibliometrics has grown out of the realisation that literature is growing and changing at a rate with which no librarian or information worker equipped with traditional bibliographic methods and skills could keep abreast.

Bibliometric techniques are now being used for a variety of purposes like determination of various scientific indicators, evaluation of scientific output, selection of journals for libraries and even forecasting potential Nobel Laurates. The popularity in the adaptation of bibliometric techniques in various disciplines stimulated stupendous growth of literature on bibliometrics and its related areas. The fact that the articles published in bibliometrics and allied topics constitute one fourth of all the articles published in the journals of Library and Information Science indicates the extent of on going research and utilisation of the technique.
2.1. ORIGIN AND DEFINITION OF BIBLIOMETRICS

Cole and Eales' [1] study 'The History of Comparative Anatomy, part-1: A Statistical Analysis' published in 1917 is considered to be the first bibliometric study, where for the first time the term 'Statistical Analysis' was used in the literature. Hulme [2] in 1923 published the book titled, 'Statistical Bibliography'. The term 'statistical bibliography' continued to be used for statistical measures till the end of 1960's.

Alan Pritchard [3] in 1969 coined the term 'bibliometrics' and defined it as "the application of mathematical and statistical methods to books and other media of written communication". Fairthorne [4] defined bibliometrics as, "the quantitative treatment of the properties of recorded discourse and behaviour pertaining to it". Potter [5] defined bibliometrics as "the study and measurement of the publication pattern of all forms of written communication and their authors". According to Sengupta [6] bibliometrics is "the organisation, classification and quantitative evaluation of publication patterns of all macro and micro communication along with their authorships by mathematical and statistical calculus".

The term 'Librametry' was coined by Ranganathan in 1948 to denote "measurement of various library activities and services using mathematical and statistical techniques". The terms like 'Scientometrics' and 'Informetrics' are also in use to describe the application of quantitative methods to the history of science, and information respectively. Bibliometrics is analogous to Ranganathan's 'Librametrics', Russian concept 'Scientometrics', FID's 'Informetrics' and also to some other well established disciplines like 'Econometrics', 'Psychometrics', 'Sociometrics' and 'Biometrics' where mathematical and statistical calculus have been systematically applied to study and solve problems in the fields of library science, history of science, information science, economics, psychology, sociology and biology.
respectively [7]. Kopolock [8] pointed out that fifty two different terms were coined during the period 1969-1977 to describe bibliometrics.

2.2. TYPES OF BIBLIOMETRICS

Bibliometrics is classified as i) Descriptive bibliometrics and ii) Behavioural bibliometrics. The descriptive bibliometrics describes the characteristic features of a literature and those that examine the relationships formed between the components of literature are known as behavioural studies. These two major branches of bibliometric activity complement each other. The descriptive study furnishes a background of basic statistics against which the strength and significance of behavioural patterns can be assessed.

2.3. LAWS OF BIBLIOMETRICS

Three laws became the cornerstones of bibliometrics. They are:

1. Lotka’s law dealing with productivity of authors of scientific papers;
2. Zipf’s law dealing with frequency occurrence of words in a long text; and
3. Bradford’s law about the scattering of articles.

2.3.1. Lotka’s law:

Lotka [9] in his article "The frequency distribution of scientific productivity", published in 1926 in the journal of "Washington Academy of Sciences" stated that "the number of authors making 'n' contributions is about $1/n^2$ of those making one; and the proportion of all contributors that make a single contribution, is about 60 percent". To derive his 'inverse square law', Lotka used comprehensive bibliographies in Chemistry and Physics and plotted the percentage of authors making 1, 2, 3, ..., n contributions against the number of
contributions with both variables on a logarithmic scale. Based on the data, he proposed the following inverse square law of scientific productivity.

\[ Y_x = \frac{K}{X^\alpha} \]

where \( x = 1, 2, 3, \ldots \) and \( \alpha = \text{constant} \)

\( Y_x \) is the relative frequency of authors publishing \( X \) number of papers. The value of \( \alpha \) is found to be approximately 2.

2.3.2. Zipf's law:

Zipf [10] developed and extended an empirical law, governing a relation between the rank of a word and the frequency of its appearance in a long text. If \( r \) is a rank of a word and \( f \) is its frequency the Zipf's law is stated as follows:

\[ rf = c \]

where \( c \) is a constant.

The law states that in a long textual matter 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. The Zipf's model is useful in index language planning and in administrative areas of library planning.

2.3.3. Bradford's law of scattering:

Bradford first formulated his law in 1934, but it did not receive wide attention until the first publication of his book 'Documentation' in 1948. Bradford examined all of the journal titles contributing to a bibliography on applied Geophysics. He divided the list into three 'zones' each containing roughly equal number of references. He observed that the
number of journals contributing references to each zone increased by a multiple of about five. The first zone contained nine journals which contributed 429 references. The second zone contained 59 journals producing 498 references and in the third zone 258 journals contributed 404 references. Bradford found a similar pattern of reference scatter in the field of Lubrication. On the basis of these observations Bradford deduced his law:

“If scientific periodicals 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 where the number of periodicals in the nucleus and the succeeding zones will be as 1:n:n²” [11].

For describing the scattering phenomena, he gave the following formula.

\[ F(x) = a + b \log x \]

where \( F(x) \) is the cumulative number of references as contained in the first \( x \) most productive journal and ‘a’ and ‘b’ are constants.

Bradford’s law has been found to be applicable to bibliographies as well as to larger aggregates of literature.

All the above three laws are interrelated. According to Bookstein [12] Lotka’s law, Zipf’s law and Bradford’s law are the different versions of single theoretical distribution.

2.4. REVIEW OF LITERATURE

The literature of bibliometrics is growing rapidly. Several review articles and books on development of bibliometrics have been published. The first review of bibliometric empirical laws was written by Fairthorne [13] in 1969. The second important one was
published by Hjerppe [14] in 1980, which contained more than 200 items on bibliometrics. The most comprehensive historical review was published by Hertzei [15] in 'Encyclopedia of Library and Information Science' in the year 1987. In view of the large amount of literature available in this field, an attempt has been made to review only the significant, relevant, and recent literature relating to the various aspects of bibliometrics in the following sections

2.4.1. Citation Analysis

Citation analysis is one of the important areas of bibliometrics. It is useful to study the extent of collaborative research, the different forms of documents used, the scattering of information, the obsolescence of literature, distribution of literature by country, language and subject, and for preparing rank list of journals. The conclusions drawn from such studies are helpful in the design and management of library and information services.

Citation analysis is defined as "an activity involving analysis of the citations or references which forms a part of the primary scientific paper. When one scientific paper (B) makes mention or makes reference to another scientific paper (A), the latter has been cited by the former as a source of information, as the support for the point of view, as authority for the statement of facts etc."[16].

Citations in papers are provided for a number of reasons. The dominant reason for citing references is that they provide identification of those earlier researchers whose concepts, methods etc., have inspired and influenced the researcher. Weinstock [17] has given fifteen reasons for citations. Some of them are: paying homage to pioneers, identifying the methodology, showing similar or parallel work, providing background reading, correcting one's work, correcting the previous work, disclaiming ideas of others etc. In addition to these reasons, various other factors influence the choice of references an
author cites. They include loyalty to colleagues or guides, familiarity with particular papers, restriction put up by editorial policy of the publisher. In some cases it is like mutual back-scratching, etc. Yet it may be concluded that when an author cited, the influence of others that should be worthy to be cited, though the degree of relevance may vary.

The article by Gross and Gross in 1927 [18] was perhaps the first on citation analysis. The authors suggested the use of citation counts in measuring the adequacy of a college library. The subsequent significant study was by Fussler [19] in 1949. The objective of the study was to find the characteristics of literature used by chemists of American origin and to rank the journals according to their use. The methodology adopted was the counting of bibliographic references and footnotes cited by American scientists in their research papers. Fussler, for the first time, calculated the median citation age for the literature of chemistry.

Studies in citation analysis received a considerable attention after the publication of the 'Science Citation Index' in 1969. The index compiled by Eugene Garfield was published by the Institute for Scientific Information, Philadelphia, USA.

Garfield [20] emphasised the importance of citation analysis as a tool in journal analysis. Smith [21] presented a review of the literature on citation analysis concentrating on its development as a research method, uses and abuses of the method and prospects for the future.

Many of the citation analysis works are based on journals. Some of the important studies are as follows:

Nagappa and Maheswarappa [22] analysed the citations appended to the journal 'Indian Phytopathology' and identified the different bibliographic forms used, and journals most frequently cited by Indian phytopathologists. Chakraborty's [23] study based on an
analysis of three journals in the field of petrology revealed a definite pattern of growth, which will be of help in the acquisition of old journals for ready reference. Midorikawa [24] in his study of journals in physics compared the subfields of physics.

Selten [25] in his study analysed the references in two American Library and Information Science Journals, 'College & Research Libraries' and 'The Journal of Academic Librarianship'. Kapoor [26] examined the citations in 'Annual Review of Earth and Planetary Sciences' and compiled a rank list of 82 periodicals and studied country and subject wise distributions. Defendick [27] made a similar study of the citations from three periodicals in systematic botany. Vaishnav and Dharmapurikar [28] analysed the citations appended to the articles in the journal, 'Herald of Library Science' and observed that eighty percent journal citations were from library science journals and each journal received on an average seven citations. The source journal, 'Herald of Library Science' ranked first among the cited journals.


Mahendra and Deshmukh [36] analysed the citations from the journal 'Annals of Library Science and Documentation' and elucidated the different citation patterns,
obsolescence rates and nature of scattering of literature in the field of Library and Information Science. Kajberg [37] used citation analysis to investigate the characteristics of the discipline of Library and information science in Denmark during the period 1957 to 1986.

The array of bibliometric investigations based on theses and dissertations are reviewed as under:

Labore and Halpern [38] analysed citations from the dissertations in Library and Information Science to determine the adequacy of Drexel University Library to support a doctoral program in Library Science. Maheswarappa and Prakash [39] examined doctoral theses to find out the literature use pattern by the researchers in the field of Botany. Sangam [40] made a citation analysis of references appended to doctoral theses in Social Sciences. The author presented a detailed description of citation patterns, ranked list of journals etc., in the different fields of social sciences. Dorban and Vandevenne [41] made a similar study based on the theses in Economic Sciences. Kannappanavar's [42] study was based on the doctoral theses in Library and Information Science. Similar studies have been carried out in various fields by Miles and Frederik [43], Verma and Murthy [44], Chambers and James [45], Brace [46], Omoriyi [47] and Devarajan and Vijayalakshmi [48].

2.4.2. Authorship Pattern and Collaborative Research

Authorship implies an active involvement and participation in the investigation leading to the publication of the paper. Research in science was once an individualistic involvement. Presently the individual activity gave way to team activity where each of the contributors being a member of the team, shares the co-operative achievement. Collaborative research has become popular since 1950. Correspondingly the proportion of scientific papers published by multiple authors increased remarkably. The nature and magnitude of collaboration vary from one discipline to another depending on such factors as the nature
of the research problem, the research environment and demographic factors. Bibliometric methods offer a convenient and non-reactive tool for studying collaboration in research.

Price [49] in 1963 noticed that the proportion of multiple authored papers has accelerated steadily since the beginning of the 20th century, and that if the same trend continued, there would be no single-authored paper by 1980. Obviously, this has not happened, although the extent of collaboration in research has steadily increased, individual research effort is nowhere near extinction.

The important studies on authorship pattern and collaborative research are dealt here under:

Subramanyam [50] in his paper on Research Collaboration, identified several types of collaboration and reviewed the earlier works on research collaboration. Further, he has developed a formula to calculate the degree of collaboration in research papers.

Maheswarappa and Mathias [51] examined the trends in authorship pattern in different disciplines of Applied Sciences in India and found increasing trend towards collaborative research in all disciplines. However, the actual rate of increase differed from one discipline to another. Maheswarappa and Savadatti [52] analysed the authorship pattern in Plant Breeding and observed that the average number of authors per paper has increased from 1.27 to 2.49 and the degree of collaboration from 0.21 to 0.76.

Karisiddappa et al., [53] studied the authorship pattern and collaborative research in Psychology and found that the degree of collaboration in Psychology as a whole was 0.60. Begum and Rajendra [54] found an increasing trend towards multiple authorship in Zoological Sciences.
Kannappanavar and Nuivi [55] made an attempt to elucidate the authorship pattern in Rural Transformation literature and inferred that the solo research is more favourable than the collaborative research. Mishra and Mishra’s [56] study based on Medicinal and Aromatic plants revealed that papers with two and three authors constitute a major part (52.42%) of literature and the degree of collaboration ranges between 0.85 to 0.89 during 1979-89.

Munshi et al., [57] investigated the pattern of research from amongst the scientists working in six Agricultural Universities in India, and determined the degree of collaboration in Agricultural Sciences as 5.51.

2.4.3. Ranking of Journals

Journals which are the primary sources of information, are the major communication media among the scientists. The shrinking budgets and the escalating costs of journals pose problems to the librarians and information scientists in the selection of the need based collection. Citation analysis is an effective tool to develop a ranked list of cited journals.

The basic technique in citation analysis is that the cited journals are ranked in the order of decreasing frequency. The ranked list thus prepared is useful not only to identify the journals devoted to a discipline but also to pinpoint the core journals in the discipline.

The concept of measuring the eminence of journals by citations, originates from Gross and Gross [58] work of 1927. They used the technique of citation counting in the selection of journals in the field of Chemistry. Later, Broadman [59] prepared a list of journals in Physiology in 1944. Brown [60] worked out the most cited publications in Mathematics, Physics, Chemistry, Geology, Physiology, Botany, Zoology and Entomology.

2.4.4. Scattering of Journal Literature

The problem of growth is complicated by the scattered nature of literature. The proliferation of journals and its consequent exponential growth of journal literature have literally jeopardized their capacity to transmit information efficiently and rapidly [72]. The problem of literature growth is directly related to the scattering of articles on a given subject in great many journals. Citation studies revealed that in any given subject, a substantial portion of references (around two thirds of the total) is found concentrated in a relatively small number of journals, and the rest scattered in a very large number of journals, peripheral to the subject. Bradford in 1934 [73] was the first to provide a mathematical expression to the scattering of citations from the journals.

Bradford's law of scattering is based on the basic principle that every scientific field is related, however remotely, to every other field. Bradford analysed a four year bibliography of journal articles in Applied Geophysics. A total of 1332 articles were collected from 258 journals and listed the journals in the descending order of productivity of articles. The ranked list of journals was divided into three zones, each containing roughly the same number of articles. The first zone contained 9 journals which contributed 429 articles, the second contained 59 journals contributing 499 articles, and the third contained 258 journals providing 404 articles. Bradford observed that the number of journals contributing articles to each zone increased by a multiple of about five. Bradford found a similar pattern of
scattering in the field of Lubrication. Based on these observations Bradford concluded that 'the number of periodicals in the nucleus and the succeeding zones will be \(1:n:n^2\)...'. The number of journals in each zone in applied geophysics is 1:5:25. The author, thus, suggested the following linear relation to describe the scattering phenomenon:

\[ F(x) = a + b \log x \]

where \(F(x)\) is the cumulative number of references contained in the first \(x\) most productive journal; \(a\) and \(b\) are constants.

This formula has several shortcomings. The most serious problem is that the phenomenon is described in terms of groups of journals. These rather large aggregations of titles seem to be an artifact of the statement of the law. That is it appears that the dispersion of articles over rank titles is mathematically regular rank by rank rather than being regular only for groups. There is also no hint in the formula or its derivation as to what kind of underlying probabilistic process created the scattering. This formula leaves unanswered questions such as how does one establish the size of the core? What is the best value of 'n' for any particular set of data?

Some obscurities in Bradford's law were clarified by Vickery [74] who also corrected and simplified Bradford's technique of deriving the ratio 'n'. Further it has been shown that the Bradford's law is independent of the number of zones chosen although this affects the value of the ratio multiplier.

Leimkuhler [75] in 1987 suggested a model based on Bradford's law of scattering and it is called the Bradford's distribution. According to this, the distribution of references in a collection of pertinent source documents can be described and predicted by the relation:
where the parameter $\beta$ is related to the subject field and the completeness of the collection.

The most profound impact on the theoretical foundation of Bradford's law has come from the efforts of Brookes [76]. He further, derived the following formula that does not depend on groupings of journal titles:

$$R(n) = k \log(n)$$

where 'n' is the rank of each journal and $R(n)$ the total number of articles contributed by the first 'n' journals. The value of $R(1)$ is simply the number of articles contributed by the top title. The value of $R(2)$ is the cumulative number of articles contributed by the first journal and the articles contributed by the second ranked title. $k$ is the constant which may be different for each search and is related to the document collection.

The above formula can be used to calculate the number of articles contributed by a journal at any rank. Brookes later modified the above formula as

$$R(n) = k \log \left( \frac{n}{s} \right)$$

where $k$ and $s$ are constants.

Some of the most important works based on Bradford's law are reviewed below:

Eto [77] tested the applicability of the Bradford's law to the R & D expending of firms, and successfully identified core firms, peripheral firms and minor firms. Bonitz [78] in his study applied the Bradford's law to the single scientist's journal rank distribution and
found it was invalid. Gupta [78] in his study applied the Bradford's law to the citations in 'Ethiopian Medical Journal' and established that the law is applicable to the literature. A recent study by Deshmukh [60] also conformed the applicability of Bradford's law to their data gathered from the journal 'Annals of Library Science and Documentation'.

2.4.5. Obsolescence of Literature

It is evident that some material in libraries becomes out of date as time progresses. This is known as 'obsolescence' of the literature. Obsolescence has been defined by Line and Sandison [81] as the 'decline overtime in validity of information'. Obsolescence studies are useful to practical librarians who administer growing collections in finite spaces.

Analysis of citations by age of the cited document can indicate the useful life of a document. Burton and Kebler [82] in 1960 introduced the term 'half-life' to quantitatively describe the rate of obsolescence of the scientific literature (in analogy with the term 'half-life' used in nuclear physics to describe radioactive decay). Line [83] defined half-life as 'Half the active life - the time during which one-half of the current literature was published'.

Brookes [84] developed a model which provides a means of interpreting the differences between the slowly aged and quickly aged journals. Griffith et al., [85] in their study showed that the citation data conforms well to the Brookes model. They further stated that the aging rates vary among journals, and it is relatively easy to identify journals which age at about the rate at which the literature grows and journals which exhaust their utility within a few years.

The study of Burton-Kebler [88] on the subject of aging of Physics literature calculated the half-life as 4.6 years. Chen [87] published a detailed user study of 220 Physics journals at the MIT Science Library. Rouse and Rouse [88] analysed the requests for monographs generated within an inter-library loan network for half-life statistics.
Brown [89] studied the half-life of the chemical literature. Gupta [90] in his study found that the average half-life of periodical list in Exploration Geo-Physics as 9.4 years. Sangam [91] studied the obsolescence of literature in Economics. He determined the obsolescence factors such as Annual aging factor, Half-life, Mean life, utility factor for journal literature and books in the field of Economics as indicated by citations in doctoral theses.

Gupta [92] determined the obsolescence of 'Physical Review' articles. The density of citations to 'Physical Review' has been found to decrease exponentially with a half-life of 4.9 years.

REFERENCES:

1. COLE (EJ) and EALES (NB). The history of comparative anatomy. Social Science Progress. 11; 1917; 10-17.


11. BRADFORD (SC). Sources of information on specific subjects. Engineering. 137; 1934; 85-86.


16. GUPTA (EM) and NAGPAL (MPK). Citation analysis and its applications: A review. Herald of Library Science. 18; 1979; 86-93.


18. GROSS (PK) and GROSS (FM). College Libraries and Chemical Education. Science. 61; 1927; 385-399.


21. SMITH (Linda C). Citation Analysis. Library Trends. 30; 1981; 83-106.

22. NAGAPPA (B) and MAHESWARAPPA (BS). Most frequently cited journals by Indian Phytopathologists. Journal of Library and Information Science. 6; 1981; 1-17.

23. CHAKRABORTY (R). Citation Analysis of Petrological literature. Lucknow Librarian. 15; 1983; 99-104.

24. MIDORIKAWA (N). Citation Analysis of Physics Journals: Comparison of subfields of Physics. Scientometrics. 5; 1983; 361-374.


26. KAPOOR (SK). Citation Analysis of earth science literature. Annals of Library Science and Documentation. 31; 1984; 56-62.


28. VAISHNAV (AA) and DHARMAPURIKAR (RG). Citation analysis of the Herald of Library science. Herald of Library Science. 29; 1990; 252-260.


34. Mahapatra (G) and Kaul (R). Bibliometric analysis of citation classics in life sciences. Library Science with a Slant to Documentation and Information Studies. 31; 1994; 129-134.


38. Laborie (Tim) and Halperin (M). Citation patterns in Library and Information Science Dissertations. Journal of Education for Librarianship. 16; 1976; 271-283.


40. Sangam (SL). Citation analysis of doctoral dissertations in social sciences. Gulbarga University, Ph.D. 1985 (Unpublished).

42. KANNAPPANAVAR (BU). Citation analysis of doctoral dissertations in Library and Information Science accepted by the universities in Karnataka, Dharwad. Karnataka University, Ph.D. 1992 (Unpublished).

43. MILES (RL) and FREDERIK (K). The use of Medical theses as demonstrated by journal citation. College and Research Library. 26; 1964; 7.

44. VERMA (SC) and MURTHY (AT). Use of literature by research scholars in Political Science and Economics. Library Herald. 12; 1970; 58-65.


48. DEVARAJAN (G) and VIJAYALAKSHMI (V). Applied Economics: Citation study. IASLIC Bulletin. 27; 1982; 19-22.


51. MAHESWARAPPA (BS) and MATHIAS (SA). Research collaboration in different disciplines of Applied sciences in India (1965-83): A bibliometric study. IASLIC Bulletin. 32; 1987; 105-114.

52. MAHESWARAPPA (BS) and SAVADATTI (SG). Authorship pattern and collaborative research in Plant Breeding (1934-1989). IASLIC Bulletin. 35; 1990; 119-123.
53. KARISIDDAPPA (CR) etc. Authorship pattern and collaborative research in Psychology. IASLIC Bulletin. 35; 1990; 73-78

54. BEGUM (KJ) and RAJENDRA (N). Research Collaboration in Zoological Sciences. IASLIC Bulletin. 35; 1990; 79-82.

55. KANNAPPANAVAR (BU) and NULVI (CN). Authorship trend and solo research in Rural Transformation: A Bibliometric study. ILA Bulletin. 27; 1991; 38-41.

56. MISHRA (Sanjay) and MISHRA (Manoj K). Collaborative Research in Medicinal and Aromatic Plants. Library Herald. 30; 1991; 30-34.

57. MUNSHI (Usha M) etc. Research Collaboration in Agricultural Sciences. ILA Bulletin. 28; 1993; 57-60.


62. SENGUPTA (IN). Recent growth of the literature of Biochemistry and changes in the ranking of periodicals. Journal of Documentation. 29; 1973; 210-211.

64. **SENGUPTA (IN).** Choosing Microbiology Periodicals: Study of the growth of literature in the field. *Annals of Library Science and Documentation.* 21; 1974; 95-111.


66. **BHAT (VG) and ELISHARAJU (R).** Ranking of Periodicals in the field of Astronautics and Aeronautics. *Annals of Library Science and Documentation.* 24; 1977; 114-118.

67. **NAGAPPA (B) and MAHESWARAPPA (BS).** Op. Cit.


70. **DHAWAN (SM) and YADAV (SR).** Study on most frequently cited journals by Indian scientists in Physics. *Annals of Library Science and Documentation.* 20; 1973; 39-57.


73. **BRADFORD (SC).** Sources of Information on specific subjects. *Engineering.* 137; 1934; 85-86.


80. MAHENDRA (VM) and DESHMUKH (PP). Op. Cit.


86. BURTON (RE) and KEBLER (RW). Op. Cit.

88. ROUSE (Sandra H) and ROUSE (William B). Analysis of Monograph Obsolescence at two levels of an interlibrary loan network. Information Processing and Management. 15; 1979: 219-225.


