

LOK MAHAVIDYALAYA, WARDHA
DEPARTMENT OF LIBRARY SCIENCE

M.Lib. Sem - IV



Paper - I

System Analysis & Bibliometrics

Introduction to Bibliometrics

This chapter presents the theoretical background of bibliometrics. The emphasis is on application of the quantitative methods in solving problems of library and information science.

1. A Theoretical Background to Bibliometrics

The exponential growth of the production of all types of literature in the last few decades has caused librarians and bibliographers to look for quantitative and statistical methods to keep track of the flood of information.

Historically the beginnings of the use of statistical technique were in 1917 by Cole and Eales¹ who conducted a study by counting and categorising publications by country of origin and by field. The study analysed the data by counting the number of titles, both books and journal articles, and grouping them by country.

In 1922 Hulme² introduced the term "statistical bibliography". The term was unspecific and scarcely used. However, Hulme's purpose is clear from the explanation of the technique: "To shed light on the process of written communication and of the nature and course of development of a discipline by means of counting and analysing the various facets of written communication."

In 1948 Ranganathan³ suggested 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 numbers."

Alan Pritchard⁴ in 1969 coined the term "Bibliometrics", which conveyed the meaning of the concept in a better way. He used the term to describe all studies which seek to quantify the process of written communication. Bibliometrics has been defined by Fairthorne⁵ as "quantitative treatment of the properties of recorded discourse and behaviour pertaining to it."

Bibliometrics can be defined, in simple words, as quantitative or numerical or statistical analysis of recorded communication—printed and non-printed sources. The analysis can be done by observation, measurement, and grouping or classifying. Further, this analysis deals with intrinsic and extrinsic (implicit and explicit) characteristics and behaviour of documents. Intrinsic here includes the entire machinery that works behind the scene, i.e., those responsible in bringing the printed records to day light (authors, publishers, research agencies and so on), intention of author and place of publication of journal, as well as, those features which are microscopic and those which need reading between the lines. Extrinsic includes, anything and everything explicit, like type of journals, co-authors and bibliographical details. For the purpose of the present study, bibliometrics is defined as a branch of library science which is concerned with the techniques of measurement and evaluation of communication media.

Librametry and bibliometrics as discussed above, show that librametry primarily aims at quantitative analysis of the library work and bibliometrics is limited to recorded knowledge. On the whole in both the cases analysis relates to three important processes or the trio in the library viz., documents, personnel and users.

1.1 Scientometrics

Scientometrics is a new emerging discipline which uses bibliometric measurements for evaluation of factors like scientific progress, levels of scientific development, social relevance, and impact of the application of science and technology on society.

Under scientometrics there are many studies which move around science of science, science policy, etc.

Yuasa⁶ made a statistical study of the scientific achievements in various countries. The results revealed that a country becomes a scientific centre, if its scientific achievements exceed a quarter of the net world's scientific achievements. According to him the shifting of scientific achievement from one country to another is as follows:

Italy	1540 to 1610
Britain	1660 to 1730
France	1770 to 1830
Germany	1810 to 1920
USA	1920 to present

To promote scientometric studies the Hungarian Academy of Sciences, Budapest has in 1979 set up a Department of Scientometrics of the Information Science and Scientometrics Research Unit (ISSRU). Its objectives are : (i) to conduct basic and applied research in scientometrics and (ii) to develop novel scientometric approaches which can be of use to science policy decision makers, research managers and scientometricians.

1.2 Informetrics

It is a field in which Bibliometrics, econometrics (quantitatively testing economic theories for solving economic problems), librametrics, scientometrics and such other metric analysis studies are integrated. Thus "Informetric is a field which brings into focus the concept of 'organised complexity' of the Info Society."⁷

Informetrics formalises and consolidates many measurement studies — which focus on information productivity. Basic to all the metric studies is information/knowledge that is generated by such studies and its utility for the betterment of society. These metric studies, lead to crucial measurement of cost efficiency and cost benefits of investment, in the field of information.

Morales⁸ describes informetrics as "a kind of scientific information activity, at the same time, a component part of information science and it studies various metric aspects of its study object in order to increase the information activity efficiency of information establishments."

International Federation for Documentation's (IFD) Committee for Informetrics, was constituted in 1980, to bring under one umbrella the different groups that presently are engaged in measurement studies of information.

1.3 Qualitative and Quantitative Applications of Bibliometrics

Bibliometric studies have expanded in two dimensions, viz., qualitative and quantitative. The qualitative studies in their application of bibliometrics emphasize practical utilization of research findings and have the following functions:⁹

- (a) Tracing the spread of ideas as a study of epidemics;
- (b) Establishing "a transition point between zones of higher and lower utility";
- (c) Classifying segments of a literature through interconnection of citations;
- (d) Determining the impact value of a given document; and
- (e) Locating criticisms of published results of research and experiments.

Quantitative studies of literature have been documented since Cole and Eales work. This quantification has enabled the bibliometricians to formulate laws. These laws are derived on the basis of studies of some characteristics of the literature. The characteristics applied are: (i) measuring the productivity of authors on the basis of the number of articles contributed by them, (ii) ranking of words in a text on the basis of the frequency of occurrence of words, and (iii) scattering of articles on a subject in various periodicals. And these are linked with emergence of three respective laws namely, Lotka's law of author productivity (1926), Zipf's law of word frequency (1933) and Bradford's law

of scattering (1934). Quantitative analysis of a literature, has the following functions:¹⁰

- (a) Design of more economic information systems and networks;
- (b) Improvement of efficiency rates of information handling processes;
- (c) Identification and measurement of deficiencies in present bibliographical services;
- (d) Prediction of publishing trends; and
- (e) Discovery and elucidation of empirical laws that can provide a basis for developing a theory of information science.

The last function — i.e., the discovery of the empirical laws — resulted in combination of the quantitative and qualitative dimensions and which meets the research needs in both areas.

1.3.1 Lotka's Law

This law was put forth by Alfred J. Lotka in 1926. It relates to the productivity of scientists, in terms of number of papers published. He was interested in determining, "if possible the part which men of different calibre contribute to the progress of science."¹¹

Lotka studied the productivity of authors by publication frequency as indicated in *Chemical Abstracts* from 1907 to 1916. Similarly, he studied, the name index of *Auerbach's Geschichtstafeln der Physik*. It revealed that the productivity of scientists conformed to inverse square law such that for every 100 authors contributing one article, 25 will contribute 2 articles, 11 will contribute 3 articles and 6 will contribute 4 articles, and so on. The observed figures for single article authors were 57.9 per cent for *Chemical Abstracts* data (6,891 contributors) and 59.2 per cent for the *Physik* data (1,325 contributors).

The law states that "the number of chemists publishing in papers is proportional to $1/n^2$ in each", resulting in the

authorship of a large number of documents by a small number of writers.¹²

The original paper of Lotka¹³ gave no suggestion to show that this was a universal law with applicability to all branches of knowledge, or even for that matter, to all the branches of science. Subsequent studies have shown that this law is applicable to the subjects of history of technology and that of map librarianship. For library and information science literature, some alteration, seems necessary, in the suggested consonant like $1/n$.^{3,5}

Success breeds success theory meant that an author of many papers is more likely to publish again, than one who is less prolific. Likewise a journal which has been frequently used is more likely to be used again, than an infrequently used journal; and an article in a journal which has been cited many a time is more likely to be cited again, than the one which has been rarely cited. Extending this success breeds success theory, Derek de Solla Price¹⁴ developed his own theory that "half of the scientific papers are contributed by the square root of the total number of scientific authors."

Similarly, Ravichandra Rao¹⁵ argues that, those documents which have been borrowed frequently are more likely to be borrowed again, than those borrowed infrequently in academic library in a year. The same analogy, as a characteristic of bibliometric distribution, has also been applied to frequent borrowers in and academic library in a year.

Furthermore, there have been a number of studies on scientific productivity from different points of view, viz., relationship of quantity of publications of individuals and scientific recognition; identification of elites in science in relation to their publication; and impact of social change on productivity.

1.3.2 Zipf's Law

This law enunciated by George K. Zipf is based of frequency of occurrence of words in a text and their ranking in a descending order. The study covered James Joyce's *Ulysses*, *Beowulf* and the *Iliad*, for which indexes and concordances are available. The

analysis, of the novel *Ulysses* by using Zipf's law is given below, as an example.

The novel contains 2,60,430 totally running words, of which 29,899 unique word forms. A frequency table of words that were in the novel, arranged in the order of decreasing frequency, was available ready-made. By analysis it revealed that the product of a word (r) and its frequency (f) was a constant. For instance, the tenth most frequent word ($r = 10$) occurred 2,653 times ($f = 2,653$); the hundredth word ($r = 100$) occurred 265 times ($f = 265$), the two hundredth word ($r = 200$) occurred 133 times ($f = 133$), and so on.¹⁶

In other words, Zipf analysed the words and arranged these in a descending order of frequency and multiplied the numerical value of each rank (r) with its frequency (f) and arrived at a product (c).

Thus the Zipf's law states "that if words occurring in natural language text of sizeable length were listed in the order of decreasing frequency, then the rank of any given word in the list would be inversely proportional to the frequency of occurrence of the word."

Zipf's equation is

$$rf = c$$

where r and f are, rank and frequency of words, respectively and c is a constant.¹⁷

Zipf derived the law from a general principle of "least effort." Words whose cost of usage is small or whose transmission demands the least effort, are frequently used in a large text. This is also illustrated in a subject index using a controlled vocabulary. All the descriptions, or terms, that are in a controlled vocabulary, will not be used with the same frequency, on the contrary few terms are over used and some are rarely picked up.

Numerous studies have been carried on the areas of application of Zipf's law. Its utility lies in the fact that the law can measure an author's richness in vocabulary.

1.3.3 Bradford's Law

Of the three bibliometric laws, Bradford's law has received

the greatest attention in literature of library and information science. Bradford's law of scattering was promulgated by the British bibliographer, S.C. Bradford,¹⁸ interpreter of Universal Decimal Classification, who projected the theory of scattering of the literature of a discipline and its consequences in a chapter entitled "Documentary Chaos", in his work on documentation. "Bradford's concern was with the twin problems of overlap and omission in the coverage of primary journal articles by indexing and abstracting sources."

Much earlier than the above treatise, he described the pattern of scatter of literature on a subject in various periodicals, in a paper on Applied Geo-physics and Lubrication.¹⁹ In this study he found out that 9 journals covered 429 articles, next 59 journals covered 499 articles and the next 258 journals accounted for 404 articles, in other words first nine journals contributed for one-third of the Articles found on the subject, the next 5×9 journals accounted for another one-third, and the next $5 \times 9 \times 9$ journals for the remaining one-third. In other words, periodicals can be categorised in three separate groups; (i) those periodicals which carry four references in a year, in a given subject, (ii) those which carry between two and four in a year and (iii) those which carry one or fewer references a year. The first group thus becomes the nucleus of periodicals in a subject and necessarily contains more articles on that subject rather than periodicals that cover articles on related subjects.

"Thus if scientific journals are arranged in the descending order of productivity of articles in a subject, there will be a nucleus of periodicals contributing more on the subject, and several other groups or zones containing the same number of articles as in the nucleus, when the number of periodicals in the nucleus and succeeding zones will be $1 : n : n^2 \dots$ "²⁰ This is the law of distribution, or "Bradford's law of scattering." The refinement of Law has been made by B.C. Vickery. He found discrepancy between the verbal and graphical representations, and restated the mathematical expression. He pointed out that the application of the Bradford's law should not only be to three zones, but with suitable modifications of the values of the ratio n , to any number of zones. The practical application of Bradford's

law was first initiated by F.F. Leimkuhler, and continued by B.C. Brookes and innumerable statisticians, indexers and librarians.²¹

1.4 Citation Analysis

The combination of quantitative and qualitative dimensions can be seen in citation analysis, which qualitatively analyses the link between the cited and citing document and quantitatively studies the variables. By citation analysis one can evaluate and interpret citations received by "articles, authors, institutions and other aggregates of scientific activities."²² According to Narin and Moll²³ "The most active area of modern bibliometrics is concerned with citations."

Citation indexes like, *Science Citation Index* (Institute for Scientific Information, Philadelphia), provide for each article included in it, the citations on it by other authors. The more an article is cited, the more its value increases. Prior to its appearance, citation analysis was confined to the production of few citation counts for determining core journals. But presently citation analysis is used to study scope and structure of the subjects, to cluster the documents, to prepare a bibliography on subjects, etc.

Citations analysis has three dimensions: (a) Co-citation: two citations are cited together; (b) Direct citation: which establishes the relationship between document and the researchers who use them; and (c) Bibliographic coupling: the relation of two documents by virtue of their joint descent from the third.²⁴

Citation analysis technique can be applied for various studies, like, (a) library and information-oriented studies, instances of which are visible in: rates of use of library materials, and rank order listings of journals; (b) science-oriented studies as seen in network determination through co-citation and clustering; and (c) management-oriented studies, this can be visualised in evaluation of productivity studies—in terms of citation analysis.

ever increasing used needs, growth of disciplines and sub-disciplines, collection developmental pressures, continuously shrinking budgets, a growing sense of professional responsibility among the information managers, space shortage, advances in storage and retrieval systems, etc.

Most of these are interrelated. For instance, literature explosion is directly linked with growth of new subjects, advances both in the production, and the means of storage and retrieval of information, etc. Literature explosion, or explosion of information and explosion of the published material (which is not in any way explosion of knowledge) has occurred due to many reasons. Significantly, the twentieth century is called as information age or the age of publish or perish. Ability to publish decides the next-ability to survive, and by implication to the theory of survival of the fittest.

Literature explosion is a catch word for many studies. In an interesting study on explosion of literature, Atrion conducted a study to address three facets of this modern phenomena: (a) the concept of information dimensionality, (b) determination of common parameters to measure the amount of information within each dimension; and (c) a working definition of 'explosiveness'. The preliminary findings of the study reveal that there has not been an explosive increase in the information production area. Whereas there is an explosive grow in the information distribution. Regarding the information flow the findings reveal that there are explosive and unexplosive increases as well as unexplosive decreases at different points for most of the variables.

This literature explosion has led to a major problem of book selection. The post World War II era has totally changed the scenario, from the previous era of no selection at all, to detailed book selection polici and guidelines today. This means much of experimentation has gone in a today bibliometric is getting acceptance as means of providing support in developing guidelines for selection of materials, as well as for other more important activities, like storage, processing and weeding of materials in a library.

4.2. Bibliometric Laws

The three laws that have emerged in this field are as follows:

Bradford's Law of Scatter, Zipf's Law of word frequency, and Lotka's Law of author productivity.

4.2.1 Bradford's Law

Of the three laws, Bradford's law has received the great attention in literature of library science. Bradford's Law of Scattering was promulgated by the British bibliographer, S. C. Bradford, interpreter of UDC. His theory dealt with the scattering of the literature of a discipline and its consequences, in his now most cited work, entitle Documentation. His concern was with the twin problem of overlap and omission in the coverage of primary journal articles by indexing and abstracting services. Bradford's main concern was to identify the levels of scatter or dispersion of publications of one field into other fields. So much was this scatter and so large was the spread over of the literature that this work of Bradford opened the eyes of many, who had not really taken serious note of this matter.

Earlier, in anther study, Bradford analyzed this pattern of dispersion of literature on a subject in various periodicals, in a paper on applied ago-physics and lubrication. He found in this study that 9 journals covered 429 articles, the next 59 journals covered 499 articles, and the next 258 journals accommodated for 404 articles. This implied that first nine journals contribute for one-third (about 33%) of the articles in a subject, the next 589 journals accounted for another one-third (33%), and the remaining 5 x 9 x 9 journals for the last one-third (33%).

That is to say, periodicals are usefully categorized in three groups: (i) those periodicals which carry four references in one year, in a particular subject; (ii) those which carry between two and four reference in a year, and (iii) those which carry one or fewer references in a year. The first category forms the nucleus of periodicals in a particular subject; this goes to become therefore the most useful group with most productive and relevant titles of periodicals in the concerned subject. This nucleus is visible in the law which Bradford stated as $1: n: n^2$.

Then according to Bradford, the first nucleus of the one-third (or around 33%) of the titles, in a subject are the most highly productive title, and these are the core journals. Core journals make up the most of the required material, and hence, if identified using proper bibliometric methods, proper data sources, and counter checking the bibliometric results with other sources will provide enough basis for decision making in acquisition of library materials.

A library can decide to subscribe these only, and for the other require titles, inter-library loan, xerox supply service, etc., can be tried, as there is a shortage of budgets, there is an always increase in the prices of journals, increasing user needs, increasing of new subjects, etc.

Since that date of the Bradford's prediction of the scientific publications many bibliometricians have improved on this statement and given different perspectives. The refinement of the law was attempted by B. C. Vickery. He found discrepancy between the verbal and graphical representations, and restated the mathematical expression. He was one of those pioneers who pointed out that the application of the law should not only be to the three zones which Bradford had identified, but with suitable modifications of the values of the ratios, to any number of zones. In terms of practical applications of the law, it was F. F. Leimkuhler who initiated the process, and B. C. Brookes and other bibliometricians took up this task further.

Of all the areas in which this law is applied, the most prominent is the study of journal productivity, most specifically in science. Core titles are derived on the basis of either numerical or verbal analysis.

4.2.2 Lotka's Law

While Bradford was concerned with the scatter of publications, another scientist, Alfred J. Lotka studied, in 1926, the frequency of publications of authors. He was concerned with the number of papers published by scientists, and desired to determine the level of contributions by scientists of different caliber in the progress of science.

In his study of the literature Lotka found that the productivity of scientists conformed to inverse square law such that for every 100 authors contributing one article, 25 will contribute 2 articles, 11 will contribute 3 articles, and 6 will contribute 4 articles, etc. The observations for author's with single articles were 57.9% for Chemical Abstracts (6, 891 contributors) and 59.2% for the Physik Data (1, 325 contributors). The law as stated by Lotka is "the number of chemists publishing papers is proportional to $1/n^2$ in each", resultingly authorship of a large number of documents by a small number of authors became obvious.

Bibliometricians have further studied this pattern in other areas, and find that this law is applicable to the subjects of history of technology, map librarianship, etc. They suggest that there is a need for alteration in the constant $1/n^{3.5}$ for a field like library science.

This led to the theory of success breeds success. This means that an author of many papers is likely to publish more, as compared to one who is less productive. So also, a journal which is used more frequently is more likely to be used again, than a less frequently used journal; and in this same sense, an article which is highly cited is likely to be cited further, than an article which is rarely cited. Derek de Solla Price developed his own theory that "half of the scientific papers are contributed by the square root of the total number of scientific authors".

Price's general theory of Cumulative advantage processes require a brief explanation. Cumulative advantage process operates in situations where success breeds success. This theory proposes a statistical model that predicts the distributions they produce. These distributions fit those derived from Lotka's Bradford's, Zipf and among others like the Pareto law of income distribution. Incidentally, this also fits in the empirical results of citation-frequency analysis. It is really cumulative advantage, in a sense, it unites all the laws and citation data which come under the umbrella of bibliometrics.

In this same line of thinking, Ravichandra Rao argues, that those documents which have been frequently borrowed more likely to be borrowed again, than those least in a library.

4.2.3 Zipf's Law

As against the above two laws, dealing with scatter of publications and authors productivity patterns, Zipf undertook yet another area of counting the frequency of words in a text. He ranked the most frequently appearing words in a descending order. He analyzed the works of fiction, and found that the product of a word (r) and its frequency (f) was a constant. For instance, the tenth most frequent word ($r=10$) occurred 2,653 times ($f=2,653$), the hundredth word ($r=100$) occurred 265 times ($f=265$), the two hundredth word ($r=200$) occurred 133 times ($f=133$), etc., which means he listed ranked words in descending order of frequency, and multiplied the numerical value of each rank (r) with its frequency (f) and arrived at a product (c).

Stated in his own words, the Law is "that if words occurring in a natural language text of sizable length were listed in the order of decreasing frequency, then the rank of any given word in the list would be inversely proportional to the frequency of occurrence of the word." Zipf's equation is $rf = c$ where r and f are rank and frequency of words, respectively and c is a constant.

In this law, it is clear that Zipf made a word frequency analysis in a text. He derived the law from a general principle of "least effort." That is, words whose cost of use is small or whose transmission demanded least effort are more frequently used. The law has been successfully tested in different fields and the results have shown, among other things, the richness of an author in vocabulary.

For explaining this word count technique a search of this text on bibliometrics was conducted. It showed there are 9912 words in all; the decreasing order of terms and names (excluded are other words—though zipf would include all occurrences):

Citation (variants included)	57
study (variants included)	37
author (variants included)	24
bibliography (variants included)	19

many	18
etc.	18
bibliometric	18
article (s)	17
bibliometrics	16
Bradford (variants included)	15

Before we end this discussion of the laws of bibliometrics, it is appropriate to state that of all these three, Bradford received the greatest attention and has still much appeal to all. Attempts are also made to show the inter-relationships between the laws. For instance, Abraham Bookstein, analyzed the three laws and states: "Both Zipf's and Bradford's laws emphasize the higher-ranking entities Lotka. On the other hand, begins with the low end." (Hertzfel: 190).

4.3. Citation Analysis

Citation directs our attention to something already existing, that is, it means to cite, or quote or refer. It is a specific reference to a work from which the text is quoting, or showing its source as an authority for a statement or idea. In one way the citing author transfers the authority about an idea to someone else.

One may ask why do author's cite? The answers can be many: as a reference for further study, acknowledge the quote, appreciate a related work, suggest another work for comparison, prove a statement, and so on. In fact all this may indicate that there is a similarity between a text and the references that are cited. And this supposed similarity has led to many studies.

By this method of analyzing the cited list of references one can evaluate the citations received by "articles, authors, institutions and other aggregates of scientific activities" says Ravichandra Rao. Two clarifications are essential at this stage. One relates to evaluating the citations received by an institution. By this it means that we assess the publications of the institution, as well as acceptability of its publications, which is very easily visible in the cited sources.

The other clarification is of the terms citing author, cited author, and citations received. Citing author is the author of the text, cited author is the reference quoted in the text, and citations received refers to the number of times an author is quoted or referred. Nicholas and Ritchie, provide more clarity to other terms, "A distinction must be made between a reference made by an item (to another item) and a citation received by an item (from another is according are, of course, also citations: the distinction is according to whether the source or citing item is, being considered.

Study of the citations has been the most active area in bibliometrics. Analysis of the relationship between the cited and citing document is one of its main purpose. The first attempt to find this relationship was, in the form of citation counts, by Gross and Gross in 1927, who produced a list of journals that were highly cited by the Journal of American Chemical Society. In 1961 Eugene Garfield started the Science Citation Index, and later fulfilled the other areas: Social Science Citation Index, and Arts and Humanities Citation Index. Garfield took up another study to find the impact factor of citing and cited documents. Kessler initiated the study of bibliographic coupling, small took up the study of co-citation, and the area is progressing.

A study of citing pattern, by the National Enquiry reveals that there is an increase in the number of references in each text. The reason of this increase could be more secondary sources are providing access to citations, and that has directly resulted in increase in the access to literature.

The same Enquiry also found, that "Citation references in journal articles are preponderantly to other publications within the field.

Nearly all citation references made in classical studies are to journals in classics. Corresponding figures for the other disciplines are English, 84 percent; philosophy, 82 percent; sociology, 74 percent and history, 68 percent.

A word can be said about the relationships that is the document, as well as, with other documents. The inner links are many, like the link of the text with the references of a previously

existing document. A link with the external source is seen in different ways, such as the influence of citing and cited documents on each other, the author's sources of data in the document; similarly there are links in the place and time of the context of relationships: "All documents rely to a greater or lesser extent on information contained within previously published documents. Inside the literature bibliographic ramifications of these relationships (by matching, for instance, the bibliographic description of the host publication with that of the cited publications), the nature and strength of each relationship can be determined (pp. 10-11)."

A lot of survey and studies have been done in the area of citation analysis. Basically these are of two types: studies dealing with productivity of science literature; and, studies reflecting the use of the literature. Wherein studies based on counting entries are from the productivity category. The techniques used in analysis of citations mostly are obsolescence, clustering, citation indexes, etc., and these are discussed below.

4.3.1 Obsolescence

Basically stating it as use-less, this term obsolescence also means no life or half-life documents, or no longer in use. The half-life of a document in a subject is that period during which half the literature used today was published. That is, if one is analyzing half-life of the literature in psychology, all documents published in the subject in the year 1996 are its base. A survey of use of the publications of 1996 and the previous years is to be conducted analyzing the annual of the literature in the subject. The results may show that half of the literature used in these libraries during 1996 was published during the past ten years. In such a case half-life of the literature of this subject is ten years.

While all bibliometric studies try to find the utility of publications this technique of obsolescence has been specifically used in confirm or rejecting the idea that document use or value declines with its age. That is, a document may become of least use after some time of their appearance for reasons, which may include, ephemeral nature (e.g., pamphlets have a short life

span), out dated information (like, advances in the field), decline in the general interest (for instance, new subjects grow and old decline regularly), etc. Then obsolescence refers to no longer useful or reliable documents. This is what a group of bibliometricians predict.

All these predications are based on the idea that decline in use is related to the growth of literature. Nicholas and Ritchie are of the opinion that "Recent research would seem to indicate that the 'decline' and the reasons for it are exaggerated or misinterpreted. Two factors have combined to create this situation. The first is that decline has been exaggerated by intense and multiple purpose use of recent literature (up to two years old)...The second factor is that few studies have taken account of growth of the literature when calculating supposed rates of decline" (p. 122).

There are others who are cautious and don't declare any document as dead or useless, unless it satisfies many more tests. While Maurice B. Line defined half-life as "half the active life", this was interpreted as "meaning the time during which one-half of the currently active literature was published." This flexibility in interpretation and deriving what actually half-life or obsolete means, has led another group, which includes Brookes, who refuted the Line's approach, and concluded that it was "both questionable and impractical." He rather stressed on the point that "rate of obsolescence is a function of both the subject literature and of the local usage of that literature" (Hertzal, p. 192-193).

4.3.2 Clustering

This technique refers to the use of grouping or creating families of each category of documents. By creating a network, or cluster, of related documents, bibliometricians, like Mark P. Carpenter and Francis Narin, who found that "the process (clustering process) used to divide sets of journals into subject areas has two underlying assumptions: first, that journal referencing patterns; and second, that journals which deal with the same subject area will refer to each other" (Hertzal, p. 196). Nicholas and Ritchie state that "As each document has a number

of relationships, some of which are not reciprocated to the same degree, a very complex multi-dimensional network is created. From this a 'cluster' or related documents can be defined, free of prejudicial classification. Once clusters are identified subject descriptors can then be allotted" (p. 110).

Bibliometricians have also worked in other areas, which include co-citation, bibliographic coupling, self citation, etc. Bibliographic coupling, looks similar, but there is a difference. Bibliographic coupling links source documents and is a fixed measure for any number of coupled papers. Two works are said to be coupled bibliographically if they have a common citation. Here it links source documents.

Co-citation indicates the extent to which a citing author perceives a pair or more papers as related. Co-citation is the frequency with which two citations from earlier publications, are cited together. All these, in short, try to build networks to measure the strength of relationships. Co-citation also enables mapping of science. This activity of mapping a field using citations has been greatly facilitated by the citation indexes.

4.3.3 Mapping of Subjects

The structure of any subject is a difficult analysis, as is the understanding of the universe of knowledge, growth of new subject, etc. Garfield thus states, the same more specifically in the context of the structure of science: "The structure of science is an intriguing subject to those who study science as a system. Beyond the intuitively comfortable hypothesis that science is a mosaic of small units, rather than a structural monolith, there are many more questions than answers. What is the nature of the basic units in the mosaic structure? How do they relate to each other? Are the intellectual and social structures similar? Are they made up of the same basic units? What is the relationship between them? Is there a variety of configurations at the infrastructure level? (Garfield: 1979: 98).

Co-citation and clustering have resulted in what is called mapping the structure of subjects—macro and micro. Specialists like Small, Griffith, Price, Narin, Goffman, Jahn, Garfield, etc., tried to study the relationships in citations through these

networks, and have attempted to map the structure of specialities, to start with.

These mapping of the structure of disciplines is attempted both in natural and social sciences, and in sciences includes physics, biomedicine, cancer, etc. and in the social sciences, includes, psychoanalysis, sociology, etc. To quote from Garfield, again, these "represent the most sophisticated attempts made... to use citation analysis to define the structure of science on the scale and at the level of detail needed for science policy purposes. On the other hand, their method seems capable of depicting all the major scientific specialities in a single, coherent structure. On the other hand, it offers a range of resolution broad enough to permit the detailed examination of any substructure level that is appropriate to the questions an investigator may choose to pose. This combination of scale and variable resolution produced a functional capability, for analytical purposes, that is analogous to an automated design system in which a computer is used to display a product at all production levels, from final configuration down through the entire hierarchy of subassemblies.

4.3.4 Citation Indexes

What are these indexes, and how are these different from other indexes. The answer lies in the point that "A citation index is an ordered list of cited articles each of which is accomplished by a list of citing articles. The citing article is identified by a source citation, the cited article by a reference citation. The index is arranged by reference citations. Any source citation may subsequently become a reference citation." (Hertz, p. 194).

The traditional indexes required highly trained indexes; whereas in citation indexes the test of an article provides the indexing through its citations. A clarification is necessary to distinguish between source citation, and reference citation.

4.3.5. Science Citation Index (SCI)

This is published as a five part index. The two key parts are the Citation Index and the Source Index. The other parts are

the Patent Index, the Permuterm Subject Index, and the Corporate Index. Citation Index links items published during the year with past items they have cited in references. The Source Index is author index to the items published during the year, and it is also organized alphabetically by surnames of authors. The Patent Index, specifically, deals with patents. The Permuterm subject Index is a title-word index. The Corporate Index takes care of the journals published during the year listing organizations with which the author's are attached, or affiliated. Each organizational entry displays the articles that the staff published during the year.

Citation indexes led to the development of advances in citation studies, like Journal Citation Reports, impact factors, etc.

4.3.6. Journal Citation Reports (JCR) are part of the Science Citation Indexes (SCI), and deal with citation counts and links of the literature. JCR provides the following data on the source journals covered by SCI:

- (a) how often each journal is cited
- (b) how many items it published
- (c) how often (on an average) each item is cited, which is called impact factor
- (d) how often (on an average) each item is cited during the year of its publication, which is called immediacy index
- (e) the source journals responsible for the references to each journals, the number of references received from each, and how they were distributed by the publication years of cited source.
- (f) the number of references each journal published, to what journals, and how the references were distributed by the publication years of the cited issues.

By identifying what are the citations, and who is citing, JCR enables defining what journals cite what journals, and with what frequency, JC defines the core and non-core of the literature on any subject.

4.3.7 Impact Factor

This refers to the impact of a journal, not of an article; when a journal is analyzed for its productivity, the total number of items published by the journal influences the number of times it is cited; the more it is published, greater the number of opportunities it has of being cited. This is evident more so in large size of journals, than in the small one, as large journal has more articles. To get a correct impact factor, the analysis must show the average citation rate per item in a journal. This is derived by dividing the number of items the journal has been cited by the number of items it has been published.

4.4 Library Use Studies

Bibliometric studies have been useful in different ways in libraries. Libraries need feedback and also need constantly information about book selection, evaluation of services and facilities, etc. And in many of these areas bibliometric analyses has been found effective. Three main areas can be identified in which many research studies have been taken up: book and serial acquisitions, library use patterns, and in matters relating to decisions about less used materials, that is weeding of documents.

Many studies, have used journal articles for identification of core journals. These have applied Bradford's law and for micro and macro subject areas, like superconductivity, surgery, biochemistry, biophysics, cancer, religion, etc. While the technique is yet to receive full acceptance, nevertheless, citation indexes, impact factor and other improved measure have found some acceptability of bibliometrics. Libraries are trying to convince the users of the validity of such results.

It is necessary to mention one or two examples, to illustrate the exact meaning and scope of the bibliometric application. To begin with, we can see the impact of bibliometrics in identifying a list of core titles, using the Bradford law. Mohamed Taher studied in his doctoral work the subject of Islamic studies, and aimed at identifying, among others, core journals. The study used the verbal suggestion of Bradfordian law, which prescribes the three zones of a literature, and categorized the literature of the field into the respective zones;

the most relevant, less relevant and least relevant sources. The study, gathered data from a printed bibliography, having international coverage, for 1980s in English. 1073 journals were indexed and this accounted for 24 titles (in first zone), 118 (in second zone) and 931 (in third zone). This means 24 titles are capable of satisfying the majority of the user needs in the subject. As against this verbal analysis, a statistical analysis of the same data identified 13 journals as core titles. This difference in results, was probably due to statistical measure, and the like.

I. N. Sengupta analyzed the literature in medicine and allied fields, with an aim of identifying core journals. The data was gathered from the references provided at the end of each article of Annual Review of Medicine, for the years 1965-1969. The results showed that 35 journals, out of 975, accounted for 52.8% citations. These 35 have been identified as core journals in the field.

Bradford's law has been, as said earlier, applied in journal citation analysis only. So also Lotka's law is yet to find other areas of application. Ravichandra Rao and Burhanuddin's case studies are relevant in this aspect, to measure the author productivity as seen from an interpretation of the Lotka's law. Ravichandra Rao combines both the Bradford and Lotka laws in one study, and hence its importance. No study, on the application of Zipf's law was found in the literature, the reasons being that the study demands much hard work, if done manually and the labor that is put in this work gives less results to the researcher.

In the area of library use patterns, there is less of criticism, and even lesser number of studies, in indicating the efficiency or other wise of library services and facilities. Ravichandra Rao in his doctoral study analyzed use behavior in a library. The study focused on six automated Canadian university libraries, aimed at identifying a probability distribution of transactions over documents and also over users. The results were, that most size-frequency data of transactions follow a negative binomial distribution, which by implication is a manifestation of success-breeds-success phenomenon; the rank distribution is likely to follow Bradfordian distribution; frequent users were more likely, than infrequent borrowers, to borrow documents which were

borrowed frequently; and users in the area of arts were likely to borrow documents, irrespectively of whether the documents were borrowed frequently, or not.

Mohammed Burhanuddin in his doctoral study assessed the author productivity of the Indian writers in American studies. The study analyzed the data for the years, 1969-1987, and identified 856 serials and among others, studied the availability of these in a leading, also representative library. The study found that 72.2% percent of cited items are available in the library. The same study also checked for the availability of other cited items, such as, 1410 cited books, the availability was 71.9%; 24 reports and unpublished items, and the availability was around 20%. These figures then indicate that published materials, such as books and journals are sufficient, but reports and unpublished materials are highly inadequate in the library surveyed. The same study evaluated the dispersion phenomenon in this field, and found that: out of 325 cited journals, 33.5% or 18 titles were enough to be identified as core journals of American Studies. This represented the first zone; the second zone has 80 journals and third has as many as 227 journals.

These sample studies are illustrative of the type of use one can make of bibliometric method. In case of library use, while many studies try to conduct qualitative experiments, much effort has not been done to measure the efficiency of library services using quantitative measures and bibliometrics. So also weeding and improvement of library require much attention by the bibliometricians.

4. 5 Indian Contribution to the development of Bibliometrics

In terms of growth of this subject, India has made many attempts and can find a place as an important contributor among the world's prominent bibliometrically advanced countries. U.S., U.K., Canada and others.

India, has to begin with a FID Committee on Informetrics, established in Delhi, around 1985, which also promotes bibliometrics. Many books have appeared dealing with bibliometrics, including those by I.N. Sengupta, I.K. Ravichandra Rao, B.M. Gupta, S. Subba Rao, Mohamed Taher, etc. There are

at least dozens of Indian scholars who have published their contributes in Indian and international sources, to name a few: B. Maheshwarappa, B. Guha, R. Shalini Urs, B.K. Sen, M. Mahapatra, K. Meghanatha Reddy, K.S. Ragahavan.

Coming to teaching this subject, it is a full paper, in some universities, and as a part of research method in some universities. Whereas some universities have not yet updated their syllabus, and hence have failed to include quantitative and bibliometric methods in the program.

Extensive bibliographies have appeared in Indian Library science periodicals, like the IASLIC Bulletin, Annals of Library & Information Science, etc. IASLIC has its seminar on this topic in 1985, DRTC had its first All-India bibliometric conference in 1969 (as the theme of its Annual Seminar), again, DRTC conducted Refresher Seminar in 1981, and there have been many more programs conducted all over the country.

Indian bibliometricians do regularly attend the international bibliometric conferences, and the recent meet held at Chicago had many Indians participating in the program. And as a part of the exchange of ideas program, bibliometricians from other countries have continuously visited India, and those include: Leo Egghe, F.W. Lancaster, and others. Indians who are now in the United States, and other countries have also contributed to the development of the field, prominent among these are Chandra Prabha and K. Subramanyam, who are living in the U.S., and have published extensively.

4. 6 Difficulties in Bibliometrics

Like all evolving fields Bibliometrics to have to find ways of getting acceptability in the universe of knowledge. At present there are scholars who don't feel very easy with this quantification techniques, some out rightly refuse to give any weightage.

Take for instance, a quote from Carlos Candra, relevant in the context of this other side's view: "no amount of quantitative or mechanical manipulation of texts, bibliographies, or other

tools will enable one to identify the key contributions to information science (Hertzal: 198)."

There are many problems in using bibliometric method, such as, determining the exact meaning of definitions that are in use (terms like monograph, journal, newsletter, house journals, etc.); identifying a source which can provide maximum data for analysis (as many studies rely on indexing and abstracting services); finding a comparable source, as are existing in the developed countries, for analysis of the literature in the developing and underdeveloped areas; combining the results of bibliometric study with other research methods to ensure the validity of the results, etc.

There is much debate on the validity of these studies. On one side the critics of bibliometric studies are not ready or do not care to accept the bibliometric results. And for different reasons, library users do not accept the decisions, taken by librarians to add or drop subscription to journals based on library related studies. A list of core journals will provide maximum required items with minimum cost and effort. This is taken as a convenience, than a strategy to face the problems.

One point deserves clarification to highlight the hidden drawback of this technique. Bibliometric studies are focused on citations, that is use of documents. In this focus however, all use of books without citing it in the text, remains unnoticed. Shelf users do not always cite all that has been browsed, referred at the shelf, or even those books that are read all the time. This is to say that readership survey is also an equally important method, which can supplement and compliment the citation study, but citations as such are not sole indicators of use.

Regarding the citation studies, specifically about data gathered on highly cited titles, one view is: "citation lists do not indicate the least-used titles in a particular library, which is the information needed by librarians to determine which subscriptions to discontinue. The correlation between journals least cited and journals least used is usually less strong than the correlation between journals most cited and journals most used in a library (Martyn and Lancaster, p. 196)."

Garfield has listed many objections that are raised by critics of citation studies. Such as, "A person's rate could conceivably be inflated by self-citations. A paper may be cited frequently in refutation or as a negative example.. there is the problem of sloppy, and even biased, bibliographic citations.. Not everyone cites all the obvious, classical antecedents or is conscientious about citing all the sources actually used. Not everyone conducts an exhaustive literature search or uses all sources that should have been used. Not everyone limits references to only material that was actually read (Garfield: p. 63)."

While objections raised by one group of critics with what citation counts measure, another group deals with what citations do not or cannot measure. And the debates go on to similar other issues. And Garfield, clarifies some of these: "Any fair appraisal of citation analysis as an aid in evaluating scientists must acknowledge that there is much about the meaning of citation rates that we do not know. We are still imprecise about the quality of scientific performance they measure. We still know very little about how sociological factors affect citation rates. There is still much uncertainty about all the possible reasons for low citation rates. And there is still much to learn about the variations in citation patterns from field to field (p. 249)."

It is hoped that the field will find a right place and be hopeful, and as positive as, Hertzal feels: "At present it is composed of methods and techniques without a coordinating theory; but it seems possible that as more theoretical information is converted to fact, and behavioral patterns are established for the assessment and evaluation of the structural components of literature (i.e., authors, publications, words, laws) and their relationships, the more probable causal explanations will be evident and bibliometrics will be closer to being recognized universally as a science (p. 199)."