

CHAPTER 3

Bibliometrics: Conceptual Framework

3.0. INTRODUCTION

With the innovative developments in information communication technology there is a significant increase in literature in every discipline. In this era of information revolution, information plays a vital role in all spheres of individuals' life may be in social, economic or political. Traditionally libraries are considered as storehouse of information but this concept has been changed due to the technological discoveries. Now libraries have their versatile role for information retrieval to serve their clientele to satisfy their information needs. Due to information explosion there has been innumerable increase in literature in different subjects. Gathering and re-arrangement of data are essential activities for information developments. New subjects are added to the existing areas like science and technology, social science and humanities. So, there is a need of statistical data analysis for planning, policy making and collection development in the library and information centres for providing services its user.

It is very essential to pay attention to the scientific methods. It is high time to arrange ideas, theories and experiments in different fields for which scientific methodology is very important. In the digital era, the sources of information are scattered all around in different forms of literature. So, there must be an appropriate methodology for arranging and compiling this information literature to make it available to the users. There is a need to design and develop need based

system and services to fulfill the information need of users. Bibliometrics is a branch of scientific methods which attempts to analyze the quantitative parameters of recorded information and knowledge. Citation analysis and content analysis are the branches of bibliometric study. Citation analysis is the key methods for quantitative assessment of scientific activity in the different areas of research. Bibliometrics studies are needed to identify the current of publication, authorship and citation analysis which can give an insight into the dynamics of the area under consideration. The aim of bibliometric studies was to measure national research performance in the international context or to describe the development of a science field with the help of bibliometrics means. (Braun, Bujdoso and Schubert 1987).

3.1. MEANING AND DEFINITIONS OF BIBLIOMETRICS

Bibliometrics is a set of methods used to analyse and quantify the scientific literatures. Bibliometric was first coined by Alan Pritchard in a paper entitled “Statistical Bibliography or Bibliometrics?” published in 1969. He defined the term as “the application of mathematics and statistical methods to books and other media of communication.” Bibliometric methods are the most significant methods to be used in the field of library and information science research. These techniques implement statistical and mathematical tools to measure the data that measure researcher’s contributions to scientific and technical development. Raising in 1962 defined “Statistical Bibliography is the assembling and interpretation of statistics relating to books and periodicals to demonstrate historical movements, to determine the national or universal research use of books and journals, and to

ascertain in many local situations the general use of books and journals. The term statistical bibliography was replaced by bibliometrics. Bibliometrics is recognized as a method to analyse and quantify the bibliographic data and offers a powerful set of methods and measure for studying the structure and process of scholarly communication.

Following are the some of the definitions given by different Bibliometricians:

Nicholas and Ritchie (1978), in their book entitled “Literature on Bibliometrics”, stated “bibliometrics provide information about the structure of knowledge and how it is communicated?”

Fairthorne (1969) defined bibliometric as “quantitative treatment of properties of recorded discourse and behavior appearing to it.”

British Standard Institution (1976) defined the bibliometric as the “application of mathematical and statistical methods in the study of the use of documents and publication paterns.”

Hawkins(1977) defined bibliometricsas “the application of quantitative analysis of the bibliographic references of body of literature.”

Herubel explained bibliometrics as “a quantitative analysis of publications for the purpose of ascertaining specific kind of phenomenon.”

Egghe (1988) defined and explained bibliometric as “the quantitative study of literatures as they are reflected in bibliographies. Its task, immodestly enough, is to provide evolutionary models of science, technology, and scholarship.”

Sengupta had defined this term as the “organization, classification and quantitative evolution of publication patterns of all macro and micro

communications along their authorship by mathematical and statistical calculations.”

It is understood from the above definitions given by different authors that the bibliometric is a research methodology which deals with bibliographies structured in a scientific manner by the researcher in their research publications. It aims to organize, classify and quantitatively assess the publication patterns of all kinds of cited documents.

3.2. HISTORY OF BIBLIOMETRICS

A brief history of bibliometrics is represented in the Table 3.1 below:

Table 3.1. Evolution from Bibliography to Bibliometrics

Sl. No.	Year	Author	Term	Title of the Paper
1.	1896	F. Campbell	Bibliography	Theory of national and international Bibliography.
2.	1917	F. J. Cole and N. B. Eales,	Statistical Bibliography	The History of Comparative Anatomy Part 1. A statistical analysis of the literature.
3.	1922	E. WydhanHulme	Statistical Bibliography	Statistical Bibliography in relation to the growth of Modern civilization.
4.	1938	H. H. Henkle	Statistical Bibliography	The periodical literature of biochemistry.
5.	1943	C. F. Gosnell	Statistical Bibliography	The rate of obsolescence in College Library Books as determined by an analysis three select list of books for college libraries
6.	1944	C. F. Gosnell	Statistical Bibliography	Obsolescence of Books in college libraries.
7.	1948	H. H. Fussler	Statistical Bibliography	Characteristics of research literature used by Chemists and Physicists in the United States.

Sl. No.	Year	Author	Term	Title of the Paper
8.	1962	L. M. Raising	Statistical Bibliography	Statistical Bibliography in the health sciences.
9.	1966	D. L. Barker	Statistical Bibliography	Characteristics of the scientific literature cited by Chemist of the Soviet Union.
10.	1968	Alan Pitchard	Statistical Bibliography	Computers, Statistical bibliography and abstracting services.
11.	1969	Alan Pitchard	Statistical Bibliography	Statistical Bibliography: An interim Bibliography.
12.	1969	Alan Pitchard	Statistical Bibliography	Statistical Bibliography or Bibliometrics?
13.	1969	R. A. Fairthorn	Bibliometrics	Emperical hyperbolic distributions for bibliometric description and prediction.
14.	1970	Alan Pitchard	Bibliometrics	Computers, Bibliometrics and abstracting services.

3.3. BIBLIOMETRICS: GENESIS OF BIBLIOMETRICS

The word ‘Bibliometrics’ is derived from two words ‘biblio’ and ‘metrics’. The word ‘biblio’ is derived from combination of a Latin and Greek word ‘biblion’, which means book, paper. On the other hand, the word ‘metrics’ indicates the science of meter i.e., measurement.

Cole & Eale’s (1917) used the expression ‘Statistical Analysis’ for the first time in his study entitled “The History of Comparative Anatomy”. It was considered the first bibliometric study done for the contribution in the field of anatomy during the period 1543-1838.

The phrase “Statistical Bibliography” was used by Hulme in the year 1923 to describe the study of use and non-use of information in Scientific Literature. He had done analysis of author and journal entries in the International

Catalogue and found that Germany ranked the first position according to the country of origin during the period 1900-13.

In 1938, the term of 'Statistical Bibliography' was again used by Heckle in his paper entitled "The periodical literature of bio-chemistry," in 1943 by Gosnell in his dissertation and later in 1984 in his article. Later, the term of 'Statistical Bibliography' was used in 1948 and 1949 by Fusilier, in 1962 by Raisig, in 1966 by Baker and in 1969 by Pritchard. This term was considered very clumsy, not so descriptive and can be confused with statistics itself or bibliographies on statistics.

Dr. Ranganathan coined the term "Librametry" in the year 1948 to maintain the day-to-day statistics of different library operations. This technique provided him necessary data for deducing staff formula. In later days 'Librametry' was known as 'Librametrics'. All India IASLIC Conference on Bibliometrics held at Bangalore University in 1985 where librametrics defined as "information processes and information handling in libraries and information centres by quantitatively analyzing the characteristics and behavior of documents, library staff and library users."

Dr. Ranganathan suggested the use of librametry for determining the optimum size of a service library, design of library building, size of reading stack area, judicious distribution of staff at the different service points. (Srivastava, 2010)

V. V. Nalimove in 1969 coined the Russian word "naukometriya" (first translated as science metrics). Nalimove and Mulchenko wrote his basic book, 'Scientometrics: Studying Science as an Information Process,' explaining that,

“Scientometrics is a complex of quantitative (mathematical and statistical) methods, which are used to investigate the processes of science.”

The word 'Bibliometrics' was first coined by Alan Pritchard in 1969 in his article “Statistical bibliography or bibliometrics?” He defined bibliometrics as “All studies which seek to quantify the process of written communication”. Fairthorne defined statistical bibliography as “the quantitative treatment of the properties of recorded discourse and behavior pertaining to it.

The analogous term to ‘Bibliometrics’ are ‘Librametrics,’ ‘Scientometrics,’ ‘Econometrics,’ ‘Psychometrics,’ ‘Sociometrics,’ ‘Biometrics’ which are used to solve the problems in the field of library science, history of science, information science, economics, psychology, sociology and biology respectively by analyzing through mathematical and statistical methods systematically.

The term 'Informetrics' introduced by Professor Otto Nackle in 1979, as a branch of science, using mathematical and statistical methods to investigate scientific and technical information on theoretical level and practical information activities.

In summary, the bibliometrics is the use of complex mathematical and statistical methods to quantitatively analyse the scientific and technical research output with bibliographic data. J. M. Barllan defines bibliometrics as “the study of nature, use and non-use of documents only. It deals only with the document that is the unit of analysis of the documents and its characteristics. It does not deal with the user and his needs.” Ravichandra Rao states, "Bibliometrics is understood to

cover the study of statistical distribution of the process relating to the activities of library staff and readers."

3.4. SCIENTOMETRICS

Alan Prichard and Nalimove Mulchenko introduced the terms "Bibliometrics" and "Sceintrometrics" in 1969. Scientometrics is the new emerging concept in the field of Library and Information Science research. Scientometrics is the analogous to bibliometrics "used to evaluate quantitatively recent of any basic scientific discipline and the factors responsible for the steady growth in research activity in that area of knowledge in the post-war period (Sengupta 1985). Dereck de Solla Price asserted that the major aim of scientometrics is to do scientific analysis of science mathematically. He stated that "since science and scientific activity is peculiarly measurable and peculiarly regular in its behavior, even compared with the models of scholarship." Nalimove and Mulchenko (1989) defined scientometrics as "the application of those quantitative methods which are dealing with the analysis of science viewed as an information process". S. D. Haitun defines scientometrics "as an approach the science of science aimed at a reproducible measurement of science reveals its objective in quantitative regularities." According to J. P. Courtial "Scientometrics is a hybrid field made of invisible college and a lot of user, thus controlled by both scientific research, and final uses."

Sengupta outlined the objectives of scientometric as: "to evaluate quantitatively recent of any basic scientific discipline, and the factors responsible

for the steady growth in research activity in that area of knowledge in the post-war period.”

Quantitative evaluation of science along with the inter and intra comparison of scientific activities, including its productivity, and progress may be of great help to the science administrators in planning proper utilization of available financial and manpower resources more economically for the betterment of socio-economic structure of a country. One of the basic aims of scientometric is to measure the organizational and management aspects of science.

3.5. INFORMETRICS

Informetrics is a term synonymous and analogous to bibliometrics, scientometrics and librmetrics. The concept was first introduced by Germany Prof. Otto Nacke in 1979 and was later adopted by Vserossiisky Institut Nauchnoii Tehnicheskoi Informatsii and International Federation of Documentation (FID). Informetrics is a well-practiced measuring technique of Information Science. It is mostly through the techniques of statistical and mathematical calculation. Information deals with all aspects of information. In the era of literature explosion, informetrics is considered as a vital commodity for generation of human knowledge, and this commodity is a transactable one. The information technique are applied to interpret various complex characteristics of information related to media, computer, cybernetics, telecommunication and networking.

M. Morales has laid great importance, to informetrics and its application in scientific activity relating to it application. He views informetrics as a scientific

activity relating to information, and thus as Morales identified some potential area, where informatics studies can be very profitably undertaken. These are as follows:

- (i) Quantitative growth of literature,
- (ii) Obsolescence and scattering of information,
- (iii) Efficiency in information product and services, Science and Technology, and their overall production,
- (iv) Efficiency of information and information establishment in general.
- (v) The role of different kind of document as means of scientific communications.
- (vi) Information pertinence and relevance,
- (vii) Ranking of periodical and other serials by various parameters,
- (viii) The role of information channels in scientific communications,
- (ix) Overlapping of subject content between periodicals, and serials,
- (x) Citations habits of scientists and the growing role of citation analysis and
- (xi) Intra-disciplinary and inter-disciplinary relation determined on the basis of bibliographical references.

It is concluded that, informatics today grown to a well defined concept involving applied statistics, modeling, simulation, cluster analysis etc. the trend in

“Informatics” is most towards discovery of theory and/or a generalized mathematical model, which can help in achieving efficiency both in the management of information systems and services as envisaged in Ranganathan’s five law of library sciences.

3.6. WEBOMETRICS

The term Webometrics was first coined by Almind and Ingwersen in 1997. Bjorneborn and Ingwersen in 2004 defined webometrics as “the study of the quantitative aspects of the construction and use of information

“Webometrics is the quantitative analysis of web phenomena, drawing upon information methods, and typically addressing problems related to bibliometrics.” They also said 1997 “Webometrics was triggered by realization that the web is an enormous document repository with many of these documents being academic-related. Moreover, the web has its own citation indexes in the form of commercial search engines, can also deliver their result automatically to investigators’ computer programs, allowing large-scale investigations. According to Thelwall, “webometrics is (a) a set of quantitative techniques for tracking and evaluating the impact of web sites and online ideas and (b) the information science research field that developed these ideas. Webometric techniques include link analysis, web mention analysis, blog analysis and search engine evaluation, but from the perspective of digital library evaluation the main method is link analysis...” One of the most visible outputs of webometrics is the rank of the world universities based upon their web sites and online impact. Webometrics

includes link analysis, web citation analysis, search engine evaluation and purely descriptive studies of the web.

3.7. CYBERMETRICS

Cybermetrics is the most recently developed concept for science of measure for cyber materials. Cybermetrics term derived from two word 'Cyber' and 'Metrics' where metrics came from 'metre' which was derived from latin word 'metrun' means measure. By combining the two words we get the meaning of cybermetrics as science of measurement involving cyber objects.

Cybermetrics is a relatively new information discipline that aims to quantitatively measure web phenomena. By web phenomena we mean the number of hyperlinks on a web site, their structure and patterns of use by web users, and their reciprocal links by other websites.

3.8. SCOPE OF BIBLIOMETRICS

Many scholars have quoted different scope of Bibliometric research which are discussed as below:

Alan Prichard in 1969 emphasized that the purpose of bibliometrics is

- (i) to shed light on the processes of written communication and of the nature and course of development of a discipline (in so far as this is displayed through written communication), by means of counting and emphasizing the various facets of written communication.

- (ii) The assembling and interpretation of statistics relating to book and periodicals...to demonstrate historical movements, to determine the national or universal research of books and journals and to ascertain in many local situations the general use of books and journals.

According to Nicholas and Ritche (1978), the scope of bibliometrics was to provide information about the structure of knowledge and how it is communicated. Bibliometrics is classified into two groups: descriptive bibliometrics and behavioural bibliometrics.

Daniel O' Connor and Henry Voos (1981) expressed the same view and said that bibliometrics aim at consistent patterns, involving authors, monographs, journals etc. According to them the descriptive bibliometrics describes the characteristic features of a literature whereas, behavioural bibliometrics examines the relationships formed between components of literature.

Borgman in 1989 indicated the scope of bibliometrics by mentioning that scholarly communication could be studied by bibliometric methods using one or more of three theatrical variables: producers of communication, artifacts of communication and communication concepts.

3.9. USES OF BIBLIOMETRICS

Bibliometrics have major importance in the present era of information explosion. It was evident from studies that "Bibliometrics measures have been applied to evaluation of scientists, academic departments and scientific publications. Alan

Pitchard and Glenn Witting advice some of the uses to which bibliometrics may be put are:

- i) The use as a visibility sign of an underlying problem of social structure relating to individuals, e.g. sex differences promotion policies, creativity.
- ii) The use for the evaluation of organization-research sponsorship, government policies, standing in the academic community etc.
- iii) The use for the evaluation of countries, i.e. science policy studies, either looking at the situation within a country comparing countries.
- iv) The use for the examination of general growth and the social structure within a subject or discipline.
- v) The use for evaluating individual journals or group of journals, especially for acquisition decisions
- vi) The use as raw data for operations research and other mathematical models.
- vii) The study of bibliometrics distribution in their own right or as member of larger families which are of interest to social science generally.

3.10 APPLICATION OF BIBLIOMETRICS

Senguptain 1985 stated the following general application of the subject bibliometrics:

- (i) Identifying research trends and the growth of knowledge in different scientific disciplines;

- (ii) Estimating the comprehensiveness of secondary periodicals;
- (iii) Identifying the users of different subjects;
- (iv) Identifying authorship and its trends in documents on various subjects;
- (v) Measuring the usefulness of ad hoc and retrospective SDI services;
- (vi) Forecasting past, present and future publishing trends;
- (vii) Developing experimental models correlating or bypassing the existing ones;
- (viii) Identifying core periodicals in different disciplines;
- (ix) Formulating an accurate need-based acquisition policy within limited budgetary provision;
- (x) Adapting an accurate weeding and stacking policy;
- (xi) Initiating effective multilevel network systems;
- (xii) Regulating the inflow of information and communication;
- (xiii) Studying obsolescence and the dispersion of scientific literature;
- (xiv) Predicting the productivity of publishers, individual authors, organizations,
- (xv) Countries or an entire discipline;
- (xvi) Designing automatic language processing for auto-indexing, auto-abstracting

(xvii) Auto-classification; and developing norms for standardization.

3.11. BIBLIOMETRICS TECHNIQUES

The bibliometric techniques are discussed below:

3.11.1 Citation Analysis

The most important part of the research output especially in the dynamics of science subjects, are the citations/references of referred articles. Citation analysis is the examination of the frequency, patterns, and graphs of citations in articles and books. Rubin (2010) and Garfield (1983). It uses citations in scholarly works to establish links to other works or other researchers. Citation analysis is one of the most widely used methods of bibliometrics. Ziman (1968) states a scientific paper does not stand alone; it is embedded in the 'literature' of the subject. A scientific paper does not stand alone but it is embedded in the literature of the subject. The nature of this embedding is specified by the use of foot-notes. A reference is the acknowledgement that one document gives to another; a citation is the acknowledgement that one document receives from another (Narin, Francis et al., 1976). Citation analysis is that area of bibliometrics which deals with the study of these relationships (Osareh, 1996). In general, a citation implies a relationship between a part or the whole of the cited document a part or the whole of the citing document (Malin, 1968).

3.11.1.1. Citations or References

For scientific communication there exists a network between their scholarly publications. Citation play an important role for showing link between the articles

of different scientists. Krauze 1971 says “by reference we mean a footnote in an article which gives bibliographic information about an earlier article. By a citation we mean a footnote referring to a given article and necessarily contained an article written at a later date. It is useful to think of reference as a backward looking concept and citation as a forward looking concept. A footnote is of course, both a reference and a citation.”

“Scientists writing article refer to earlier articles which relate to the theme of the paper. These references are supposed to identify those earlier researchers whose concepts, methods, equipments, opinion, etc. inspired or were used by the author in developing own article. We should distinguish between the notions ‘reference’ and ‘citations.’ If article A contains a bibliographic footnote using and describing paper B, then A contains a reference to B and B has a citation from A. We can say that, a reference is the acknowledgement that one document gives to another while a citation is the acknowledgement that one document receives from another document” (Tiwari, 2006).

A citation represents a relationship between the cited and citing document. Citations are used to assess the value, quality, impact, penetration, and originality, visibility of individual and corporate performance within and across disciplines. Due to following reasons authors cite each other as enumerated by Weinstock (1994).

- (i) Paying homage to pioneers;
- (ii) Giving credit for related work;
- (iii) Identifying Methodology, equipment etc;

- (iv) Providing background reading;
- (v) Correcting one's own work;
- (vi) Correcting the work of others;
- (vii) Criticising previous work of other;
- (viii) Substantiating claims;
- (ix) Ascertaining researchers to forthcoming work;
- (x) Providing leads to poorly disseminated, poorly indexed or uncited work;
- (xi) Authenticating data and classes of fact-physical constants etc;
- (xii) Identifying the original publications in which an idea or concept was discussed;
- (xiii) Disclaiming work or ideas of others;
- (xiv) Disputing priority claims of others; and
- (xv) Identifying the original publication describing an eponymic concept or term.

According to Fulton and Pastel in 1945 bibliographic citation serve two basic functions which are as follows:

- (i) Identify the source of a given statement, and
- (ii) Describe the nature and scope of the printed document in which the statement is found.

Brittain and Line in 1972 discussed importance of citation as for:

- (i) Identification of key documents and creation of core of journals

- (ii) Study of coverage of primary journals and other materials in secondary service.
- (iii) Clustering of documents according to common reference and citations.
- (iv) Study of attributes of literature including growth rate, obsolescence, citation practices.
- (v) Study of structures of scientific literature according to language, country of origin, age subject, form, authorship or any combination of these attributes.
- (vi) Study of historical and sociological aspects of scholarly communication in science and technology.

From the above discussion it is understood that citation are very vital source for writing articles refers to earlier articles which are related in respect ideas, concepts, methods etc.

3.11.1.2. Techniques Using Citation Analysis

Some of the techniques used for citation analysis are discussed below:

3.11.1.2.1. *Half-life and Obsolescence*

By 'obsolescence', we mean the document which are no longer of use. Gross and Gross determine what materials were vital for maintenance of a dynamic college and library, they were projecting needs. The flip side of this positive thinking is what is no longer needed or of use. This can be termed as 'obsolescence'. Obsolescence is the process whereby materials become no longer useful or reliable

(Gosnell 1943). Gosnell, whose dissertation was on obsolescence, started “the causes of book mortality or obsolescence are many, varying from pure fad through extension of scientific knowledge, technological changes, to fundamental changes in our civilization” (Gosnell 1943).

Robert E. Burton and R. W. Kebler in 1960 wrote “The ‘Half-life’ of some Scientific and Technical Literature” when they compared the rate of obsolescence of scientific literature with that of radioactive substances.

According to Maurice B. Line, *Half-life* (1970) “half the active life” and this was usually interpreted “as meaning the time during which one-half of the current active literature was published.”

The results' of many of the half-life studies varied so greatly that they did not project a possible general application. There were also discrepancies as to meanings. This was Maurice Line's explanation of the process that if the average number of articles in each journal remained constant, and the number of journals doubled, "the probability of citation is twice as great. “The so-called 'half-life' of a literature is therefore compounded of its obsolescence rate and its growth rate.”

The citation half-life of different subjects provides useful information which provides guidance for collection development. However it depends on the subject and nature of study and use made and the worth of the publication. The calculation of half-life in libraries helps to weeding out documents not useful for the readers.

3.11.1.2.2. *Citation Indexes*

A citation index is an ordered list of cited articles each of which is accompanied by a list of citing article. The 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 (Garfield, 1964). A citation index keeps track of which articles in scientific journals cite with other articles. The index arranged by reference citation. Any index is arranged by a list of citing article by a reference citation. The citations are listed alphabetically drawn from reference and footnotes of source item arranged under the first author.

Arrangement by author is favoured in the citation index and the source index because the research scientist usually approaches the literature first by author. A complete source index containing full source article title and certain additional data is similar to an upgraded conventional author index would not be valid. In his "Citation Indexing: A Natural Science Literature retrieval system for the Social Sciences", Garfield differentiated between the dissemination and retrieval of information which had been stored. Information which was current was disseminated but information which had been stored was retrieved. The problems associated with both dissemination and retrieval, he believed, "are largely overcome by citation indexing."

The first citation index is Science Citation Index (SCI) developed by Eugene Garfield in 1963. There are two other citation indexes, Social Science Citation Index (SSCI) and Art and Humanities Citation Index (A&HCI) published in the year 1973 and 1978 respectively.

The Journal Citation Report (JCR), a companion volume to the Citation Index, was also first published by ISI in 1973. The JCR includes ranking of journals by citations and by impact factor, and also covers two ranked lists for each journal: cited journal listing and citing journal listing. Peritz (1992) assert that these search tools (i.e citation indices) are often found to be more informative and efficient than traditional subject indexes.

3.11.1.2.3. Clustering

Mark P. Carpenter and Francis Narin reported their result of a cluster analysis in “Clustering of Scientific Journals based on the journals based on the Journal Citation Report (JCI) citations from 1821 different journals. The clustering process used to divide sets of journals into subject areas has two underlying assumptions: first that journals which deals with the same subject area will similar journal referencing patterns, and second that journals which deals with same subject area will refer to each other.

3.11.1.2.4. Bibliographic Coupling

The term ‘Bibliographic coupling’ was first introduced by M. M. Kessler in 1963 of the Massachusetts of Technology. He defined a unity of coupling between two papers as an item of reference used by these two papers. Bibliographic coupling is the linking of two papers which cite the same article. For example, if the papers A and B both cite the same article C then they are said to be bibliographically coupled. The more number of papers, they both cite the stronger their relationship will be. Egghe and Rousseau, 1990 states that “the strength of bibliographic

coupling is measured by the number of references that the two papers have in common.”

3.11.1.2.5. Co-citation Analysis

Small and Griffith in Gilbreth 1977 have proposed the analysis of citation for examining the structure of science. They states that co-citation “identifies relationships between papers which are regarded as important by authors in the specialty and provides a natural and quantitative way to group or cluster the cited documents.” Osareh, 1996 defines Co-citation analysis is a methods used to establish a subject similarity between two documents. For example, if papers A and B are both cited by article C, they may be said to be related to one another, even though they are not related to each other. Bellardo, 1980 states that “the greater the number of times that a pair of documents is cited together, the more likely it is that they are related in content.” If paper A and B are both cited by many other papers, they have a stronger their relationship. So the more papers they are cited by, the stronger their relationship is.

Co-citation pattern indicates the probable relationship that exists when the two citations are cited together. The frequency of their occurrence together determines the strength of linkage. It represents the degree of relationship among the paper as observed by the practicing scientists. The fluctuations may take place in a particular field as and when the interests and intellectual patterns of that field undergo change with time. This further provides the clues to understand the mechanism of speciality development.

3.11.3. Publication Counts

One of the easiest techniques of Bibliometrics is counting the total number of publications a scientist or a group of them have published. While the publication count gives a quantitative measure of the total volume of research output, the quantitative aspect of the published work remains to be assessed. The publication count, however, has its other limitations also. The multiple authorship, which has shown sharp increase during past few decades owing to the prevalence of collaborative research, pose problem in the process of publication count. Jones “studied the correlation between the publication counts and other measures of scientific merits such as research fundingscitedness, etc. and found a reasonable correlation.”

3.11.4. Direct Citations

Citations are considered signpost left behind after information has been utilized and as such provide data by which one may build pictures of user’s behavior. The direct citation count is the easiest technique to determine the number of citation received by a given document or set of document over a period of time from a particular set of citing documents, where from citation data for analysis was taken.

3.11.5. Self-Citation

Self-Citation is a common practice noticed in the literature of all kinds. The citation analysis study is able to derive the type, rate, and trend of self-citation in the different kind of literature. The self-citation of an author refers to those cited

references that have been authored or coauthored by the same citing author. The rate of self-citation in different subjects varies a lot.

3.12. APPLICATION OF CITATION ANALYSIS

Citation analysis can be applicable in two broad areas. In one aspect it is helpful to the library professional to analyse various features of literature use like type of documents, languages, age, country of origin, subject distribution, most prolific authors, highly cited journals, citation rate, authorship pattern etc. mainly for collection development and in another side, the citation analysis help to drive the research productivity in a area of study.

A number of bibliographic parameters like total number of pages, words, letters, articles chapters, etc. have been used effectively for measuring knowledge and its growth. The phrases like “Exponential Growth” and “Information Explosion” used widely even today are the result of analyses conducted decades ago.

3.12.1. Literature of “Studies”

In this case citations in a particular subject area are studied to describe the pattern of citations. The source for this kind of analysis may be limited to a single journal in a field, or they may cover many sources including types of materials in addition to the journals. Smith in 1981 reviewed the literature on applications of citation analysis and identified the following areas of applications. Small for the first time translated the purely bibliometric structure of a co-citation map into a language

structure by substituting frequently occurring terms or phrases from the citations using context analysis for each document on the map.

Small in his article, "Co-citation in the Scientific Literature: A New Measure of the Relationship between Two Document" used co-citation analysis to explore the relationship of Information Science to the Social Sciences. Co-citation provide a natural and quantitative way to group or cluster the cited documents since the patterns of co-citation changes with time as new discoveries are made and introduced through the literature resulting into generation of new linkages. Several studies have since been made to locate networks to define scientific specialities and the nature of activities at the research front.

3.12.2. Type of literature Studies

Citation analysis has been used to measure use of publications in certain type of literature such as government publications and patents. Chambers and Healey analysed the citations given in the theses to evaluate the local collections. Kubota analysed citations to graduate theses submitted to the School of Library and Information Science, University of Keio, Japan. Goehlert studied the use of government publications by international organisations by analysing the references cited by their publications.

3.12.3. User's Study

The application of citation analysis to study the user need is very useful in dissemination of required service. The user's study has implications for collection development and design of services for a particular library. Preference list of

articles written by a user or a group of users of a specific collection / library were analysed to determine the types of materials, age, subject, language and whether the referred material locally owned or borrowed. The citation analysis can be used to compare behaviour of users in different fields in different time span.

3.12.4. Historical Studies

Citations can be used to trace the chronology of events, relationships among them, and their relative importance. The citation counts have also been used to identify important contributions to a discipline. The use of citation indexing for historical research was suggested by Garfield. Methodology for use of citation indexing in sociological and historical research was, however, standardized after publication of Science Citation Index. The Citations have been used to trace the chronology of events relationships amongst them, and relative importance. In 1964 Garfield compared the history of DNA research according to Asimov's "Genetic Code" with the citation network of key citation derived from the Science Citation Index. He concluded that citation studies are valid for creating accurate historical descriptions of scientific fields. Ellis using citations in patents generated the citation networks to display the history of technical developments. The pattern observed through such studies can be validated through interaction with specialist and questionnaire surveys.

3.12.5. Communication Patterns

It is assumed that citation analysis help to trace scientific communication. Contemporary science depends almost entirely on the journal literature, which in turn, provides an effective channel for scientific communication. Citations may be

used to trace patterns in scientific communication. Although citation linkages do not necessarily reflect social contacts. The assumption can be helpful to identify problem areas in communication, including linguistic isolation, limited dissemination of new ideas and barriers between basic and applied science or between specialists and the public at large. Typical studies that have been carried out include formal and informal communication, publication vs. communications, co-authorship and inter-organisational mobility

3.12.6. Collection Development

Citation analysis can be used as a tool for collection development. Citation analysis has been used as a tool for collection development though it has attracted several criticism. Journal ranking by citation counts may bear little relation to the frequency of their use in a particular library as citation analysis and use analysis measure different activities. Besides, non-cited journals like trade and technical journals and professional magazines are not considered in the analysis. Gross and Gross used citation analysis to measure journals of significance.

In spite of these criticisms, the rank derived from citation counts can still be used to (i) identify highly ranked journals not available locally and within subject scope as worth examining in more details; and (ii) low-ranked journals available at the library may be examined; (iii) peripheral journals which may not be available in the collection may be of use and can be acquired for local use on inter-library loan.

3.12.7. Obsolescence

Obsolescence has been defined as the decline in the validity of the information. The obsolescence studies are of interest to the collection managers to help them decide as to which item to keep and which to discard. The obsolescence study culminates in simple mathematical formulae which could be applied with equal success to any library. Line and Sandison had in a comprehensive review of 200 papers, summarised their data and made some clarifying distinctions between synchronous and diachronous studies and their results. Menard discusses at length the relationship between the growth rate of a field and bibliometric measures, Clark studied the obsolescence of patents.

3.12.8. Predicting Nobel Prize Winners

Garfield produced a list of 50 most cited authors for 1967 out of about one million scientists from the Science Citation Index database. Two out of the list of 50 were recipient of 1969 Nobel Prize. Another dozen from the predicted list were the recipient of the Nobel Prize in subsequent years.

The approach is based on purely mechanical and objective methods which did not require reading and evaluating the work of the scientists included in the list. The technique of citation indexing thus can be used to identify those individuals who might make a major impact on science. Increasingly scarce intellectual and financial resources for supporting research can be managed more efficiently with such an identification tool. Garfield observed that 14% of the 1000 citation classics were written by Noble laureates and that the Noble laureates were cited 30 times as often as average authors.

3.12.9. Automatic Classification

A document classification aims at grouping the similar documents together. Citation indexing is a natural or automatic system of classification; here the material to be classified orders itself through the changes observed may be depicted and analysed by various graphing techniques. Garfield developed a computer based system for automatic classification of scientific article determined by citation and co-citation patterns from Science Citation Index database.

3.12.10. Information Storage and Retrieval

The citations are considered as useful supplements to keywords in identifying relevant documents. A cited document is related in content to the citing documents, Barlup in 1969 in an experiment in which authors were asked to assess the degree of relatedness of citations to their own work, found that 72% of articles to be definitely related and only 5% to be definitely not related.

Citation links have been used in developing, document representation in automatic classification and various retrieval algorithms which make use of the ability to find related documents in the file irrespective of words and languages. Citation as a retrieval tool have the advantage that they are unaffected by changing terminology and they reveal relevant papers which cannot be searched using conventional indexing. Yermish described an interactive information retrieval system which he developed to manipulate citation relations existing among bibliographic records efficiently. Gerard Salton, based on the results of several studies conducted by him, concluded that the bibliographic citations are generally useful for document analysis and information retrieval systems should assign a

bibliographic citation code to the documents in addition to standard content indicators. Bichteler and Eaton developed a linkage similarity measure for combined use of bibliographic coupling of documents and their co-citations for document retrieval.

3.12.11. Evaluative Bibliometrics

It is defined as evaluation and interpretation of citations received by articles, scientists, universities, countries and other aggregates of scientific activity, used as a measure of scientific influence and productivity. This area of study is now being highly used as an evaluative tool by science administrators.

3.13. EMPIRICAL LAWS OF BIBLIOMETRICS

Bibliometrics is the branch of research to study the bibliographic data of scientific activity to resolve the practical problem by the application of Empirical laws of bibliometrics. There are three most important laws of bibliometrics. These are Lotka's law of scientific productivity, Bradfords law of scattering and Zipf's law of word occurrence.

3.13.1. Lotka's Law of Scientific Productivity

The Lotka's definite work called Lotka's Law was introduced in 1926 which is concerned with productivity analysis. He showed that the number of authors with n publications in a bibliography is described by a power law of the form C/n^a , where 'C' is a constant. The exponent 'a' is often close to 2. Rewriting this equation as a statistical distribution (so that the sum over all n becomes 1), he showed that in the case that a is exactly equal to two, 'C' must be $6/(\pi)^2$, or approximately 0.61. This

implies that is a bibliography can be described by Lotka's Square Law, approximately 61 percent of all authors have contributed just one article to this bibliography.

Lotka's Law describes the frequency of publication by authors in a given field. It states that "...the number (of authors) making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that makes a single contribution, is about 60 per cent." This means that out of all the authors in a given field, 60 per cent will have just one publication, and 15 per cent will have two publications ($1/2^2$ times 60), 7 per cent of author will have three publications ($1/3^2$ times 60), and so on. According to Lotka's Law of Scientific productivity, only 6 per cent of the authors in a field will produce more than ten articles. Lotka's Laws, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact.

Since the publication of Lotka's original article in 1926, much research has been done on author productivity in various subject fields. The publications arising from this research have come to be associated with Lotka's work and are often cited as proving or supporting his findings. However, a review of this literature reveals that Lotka's article was not cited until 1941, that his distribution was not termed "Lotka's law" until 1949, and that no attempts were made to test the applicability of Lotka's law to other disciplines until 1973.

In that year Murphy did his study of "Lotka's Law in the Humanities." Larry, J. Murphy chose for his work the first decade of the publication (on history of technology) *Technology and Culture* and found "the actual number of single au-

thors” to be 130. He rounded off his numbers and plotted the results, leading him to conclude, without any statistical test for goodness of fit, that "The tabulation essentially verifies that Lotka's Law applies to this field within the humanities.

T. Radhakrishnan and R. Kernizan had done their experiment to verify the satisfaction of Lotka’s Law, with the papers published in the area of computer science. In the first experiment the assumption was made that, “author publishes exclusively through one scientific journal but it was found that this assumption was not valid.” From the chi-square value it is clear that Lotka’s original suggestion of x/n^2 does not apply very well. But, the prediction by x/n^3 is quite close.” In the second experiment all the papers published by the authors, irrespective of Journals, were considered. The deviation from the Lotka’s Law was high. In the judgment of authors:

- (i) Lotka’s Law, in its generalized form, seems applicable when we consider the publications of authors in one periodical and,
- (ii) When we consider all the publications of the authors in various journals, the observed values deviate considerably from the predictions of the law. This may be due to the differences between pure and applied science. They suggested further testing the law using “the service tapes such as Engineering Index, INPEC, etc

In 1981, K. Subramanyan did a study with library literature, and Valerie L. Richardson collected data from the Catalogue at State College of Victoria at Frankston,” Subramanyan found that the pattern did not conform to Lotka’s results and suggested more work be done. Richardson stated from her investigation, “It

can be seen that the Frankston datado fit Lotka's Law in its general form, but since the slope of the Frankston line is greater than 2, the data do not correspond to the inverse square law.

The compilation "Frequency Distributions of Scientific Performance: A Bibliography of Lotka's Law and Related Phenomena," by Jan Vlachy, although not a substantive work, is included here because of its excellence. This bibliography of 437 papers is international for it lists articles in English, from England, Canada, and the United States; and in Czech, Russian, Danish, French, and German. Lotka, Bradford, and Zipf are included as well as many other areas of bibliometrics. Garfield in an analysis of 1000 articles most cited in science Citation Index from 1961 to 1982, found that 14% of 1000 citation classics were written by Nobel laureates. He estimated that Nobel laureates were cited 30 times as often as the average authors. In a paper entitled "Do Nobel Laureates write Citation Classics," Garfield observes that all Nobel laureates have written at least one citation classics. He however maintained that the relationship between publication productivity citations, influence, and recognition in the form of prizes such as Nobel is complex. Garfield also observed that Nobel Prize winning work continues to be cited over a long period of time.

3.13.2. Bradfords law of scattering

In 1934, Saumuel Clement Bradfords formulated that if a group of journals are arranged in an order of decreasing productivity, i.e., the journal that yields the most relevant article coming first and the most unproductive coming last, then journals will be grouped into a number of zones each producing a similar number of

relevant articles. However, the number of journals in each zones will be increasing very rapidly. This is shown in Table3.2 as extracted from his law.

Table3.2. Bradfords Zones

Zone	No. of Journal Times	No. of papers
1	9	429
2	56	499
3	256	404

Bradford presented his paper entitled ‘Source of information on specific subjects’ for which he examined two bibliographies of journal articles in ‘Applied Geophysics’ (1928-31) and ‘Lubrication’ (1331-32).

One of the Bradford’s hypotheses in 1934 was that “Reference are scattered throughout all periodicals with a frequency approximately related inversely to the scope. On this frequency aggregate of periodicals can be divided into classes according to relevance of scope to the subject concerned, but the more remote classes will, in the aggregate, produce as many reference as the more related classes.”

The law of scattering may thus be stated. If periodicals containing articles on the given subject are arranged in decreasing order of the number of such articles they contain and divided into a nucleus of journals more specifically devoted to the subject and zones of periodicals containing the same number of special articles as

the nucleus, the numbers of journals in the nucleus and succeeding zones are as $1:N:N'..$ where N is about 5 or 6.

In 1934, Bradford's classic article containing his law was printed. Again he wrote about the number of articles abstracted and indexed by a number of abstracting and indexing journals showing his continued concern about wasted money and skills brought about by duplication of effort. He stated that "although the 300 abstracting and indexing journals notice 750,000 articles each year, only 250,000 different articles are dealt with and 500,000 are missed." His concern was shared by Ernest Lancaster-Jones, assistant keeper in the Science Museum. Consequently, "A statistical analysis was made of (the references in two quarterly bibliographies compiled in the Science Library, i.e., the Current Bibliography of Applied Geophysics and the (Quarterly Bibliography of Lubrication" by Lancaster-Jones, a trained mathematician who also "had received recognition as an expert in applied geophysics." Bradford, using the data, evolved his "law."

Bradford came to the conclusion that "A standard classification, must be adopted, so that references to the same subject would be brought together by the classification, irrespective of source or abstracting bureau, when, without increase of labour, a complete index to scientific literature would be achieved."

In a paper on abstracting and indexing periodicals, presented in 1937 at the ASLIB Proceedings 14th Conference, Bradford told that the details of the analysis "to determine the extent to which scientific papers are scattered in periodicals devoted to other subjects," published in Engineering in 1934, were reprinted in Publication No. 1 of the British Society for International Bibliography, 1934.

In 1973, *Library Trends* published an issue, ‘Analyses of Bibliographies,’ to which B.C. Brookes contributed the lengthy but excellent “Numerical Methods of Bibliographic Analysis.” He listed five main objectives of numerical analyses,” based on the belief that quantification is a necessary component of the design of economic information systems and that measurement of the key processes of an information system is a necessary component of management control.” These are:

- (i) The design of more economic information systems and networks;
- (ii) The improvement of the efficiencies of information handling processes;
- (iii) The identification and measurement of deficiencies in present bibliographical services;
- (iv) The prediction of publishing trends; and
- (v) The discovery and elucidation of empirical laws which could form the basis for developing a theory of information science.

Brookes in 1977 contributed his “Theory of the Bradford law.” In this superb paper, Brook did a complete revaluation of the Bradfords Law and concluded, “Bradford was a pioneer in social mathematics.” “Within the paper Brook Stated: “The analysis of the Bradfords Law has hitherto been applied to theoretical models which are too static, too deterministic and too physical.” All Bradford data are derived by the activities of a set of sources over some appropriate periods and by noting these activities, as measured in the numbers of

items each source accounts for in that times. Thus the Bradford law is concerned with:

- (i) A finite set of active sources (an ensemble) whose activities are made manifest by the generation or consumption of specified type of item.
- (ii) Observation of those activities over a specific sampling period.
- (iii) Items of some homogenous kind which are discrete and countable.

One of the points that has so far escaped analysis is that the statistical distributions of such an ensemble of activity must depend on relationships, and the period of observation which provides the sample data. All Bradford distributions are samples of some on going activity but, all too often, the sample data have regarded as constituting a total population.

3.13.2.1. *Sengupta's Extension of Bradfords Law of Scattering*

I.M. Sengupta's study fostered "an extension of Bradford's law of scattering: that during phases of rapid and vigorous growth of knowledge in a scientific discipline, articles of interest to that discipline appear in increasing number in periodicals distant from that fields."

3.13.3. Zipf's Law of Word Occurrence

George Kingsley Zipf formulated an empirical law to predict the relation between the rank of a word and the frequency of its appearance in a long text. The Law states that, "In a relatively lengthy text, if you list the words occurring within that

text in order of increasing frequency, the rank of a word on that list multiplied by its frequency will equal a constant.” The equation for this relationship is:

$$r \times f = k$$

where, r = the rank of the word,

f = the frequency, and

k = the constant.

Zipf's in 1949 introduced a book "Human Behaviour and the Principle of Least Effort appeared" based on introduction to human ecology. The purpose of this book "is to establish the Principle of Least Effort as the primary principle that governs our entire individual and collective behavior of all sorts." Zipf stated that, "An investigator who undertakes to introduced any such primary scientific principle of human behavior must discharge three major obligations, that is, have a large verifiable number of observations, be consistent, and have an understandable presentation."

"In simplest terms, the Principle of Least Effort means, for example, that a person in solving his immediate problems will view these against the background of his probable future problems, as estimated by himself." In other words, a person will tend "to minimize the probable average rate of his work-expenditure (over time), meaning use the least amount of effort." The connection between this principle and the rank-size law is by no means clear, and Zipf's theoretical arguments now have at most only historical interest.

In writing the chapter "On the Economy of Words," Zipf applied his principle to speech development. Using Miles L. Hanley's Index of Words for James Joyce's *Ulysses*, he found in the rank frequency word distribution "a clear cut correlation" between R (rank) and F(frequency) of the words 'in the sense that they approximate the simple equation of an equilateral hyperbola: $rx = C$." This, he believed, gave "clear evidence of the existence of a vocabulary balance."

More analysis showed "N" number of different words of the same integral frequency of occurrence (under the conditions of the equation, $rx = C$) will be inversely proportionate to the square of their frequency (approximately)—or, stated somewhat more precisely in equation form, that: $N(F^{-1/4}) = C$ ". This was sometimes considered Zipf's second law and has been, called his "weak" law.

Zipf, applying his "Principle of Least Effort" to many different areas of language and to other disciplines, cited many noted scientists, including Lotka, Davis, Willis, and Pareto, all of whom worked with distributions, but in various areas. Zipf's very wordy, philosophical, and detailed explanations are beyond the scope of this paper.

Many scientists have analyzed, refined, and evaluated Zipf's endeavors, but Ronald E. Wyllis, who had made a special study of Zipf's Law, called it "One of the most puzzling phenomena in bibliometrics."

Anatole Rapaport, in writing about the "Principle of Least Effort," stated Zipf's arguments are vague appeals to the recognition of the Principle in a great variety of situations simply on the basis of its plausibility. And even these appeals

are often stretched far beyond ordinary credibility.” However, other serious researchers, Simon and Mandelbrot in particular, endeavoring to put Zipf's investigations on a more scientific basis, have also done some very serious evaluating of his work. “Benoit Mandelbrot has published several studies of generalizations of Zipf's law dealing both with the question of whether the slope is -1 and with the deeper problem of explaining why the r^d products should be relatively constant.

Bruce M. Hill, a professor of statistics, was responsible for a group of papers regarding Zipf's law. Hill's first paper was “Zipf's Law and Prior Distributions for the Composition of a Population” (1970) in which he cited Wills for first noting the law and discussed the fact that the “theoretical models yielding such a law have been proposed by Yule and Simon.” “Rank-Frequency Form of Zipf's Law” (1974), published next, “presents a theoretical derivation of the rank-frequency form of Zipf's law based on a Bose-Einstein form of the classical occupancy model.” “Of course, other models have been formulated, and other approaches taken, in regard to justifying the Zipf (or Pareto) law, most notable those of Mandelbrot and Simon.” Hill's next paper was “Stronger Forms of Zipf's Law” (1975), which is really an extension of his 1970 paper.

3.14. CONCLUSION

In the age of information explosion and budget crunch, there is a need of a suitable methodology which will help the librarian and library managers in decision making for collection development, financial planning and weeding out of obsolete materials from the library. All the major aspects in the library function can be

easily maintained with application of bibliometric studies. This chapter explains the theoretical background of bibliometrics especially related to bibliometric techniques, its applications and the empirical laws of bibliometrics. Application of bibliometric techniques makes the library service dynamic and efficient by exploring the research activities in the academic disciplines.