Chapter 2
Citation Analysis: A Review
CHAPTER – 2
CITATION ANALYSIS: A REVIEW

2.1 Introduction

The purpose of the present chapter is to focus on scanned literature on citation analysis.

Review of related literature forms an integral part of any research. It helps to find out what is already known and what is still unknown and untested. It helps to avoid duplication of research. Selecting the topic for research several books, journals, bibliographies were scanned in order to identify the previous similar studies. Search on Internet Bibliography of (AIU, 2010) database of theses and dissertation on (Inflibnet and Vidyanidhi, 2010 was searched to formulate the topic for research and to check whether the research has been completed on the formulated (problem) topic.

On “Citation analysis” there are many previous studies were done at MLISc, M Phil. and Ph.D. level. (Kannappanava, 1991; Berhanuddin, 1992; Sangam, 1986; Thoidingjam, 1997; Mishra, 1997; Chikate, 2008) studies were conducted at Ph.D level. Where as (Pradhan, 2009; Thillar, 2009; Khandare, 2010; Jawale, 2011; Kalbande, 2011; Hiwale, 2011) studies were conducted at M Phil. Level. While (Mutkule, 2002; Dharmapurikar, 1987; Onguso, 1995; Jadhav, 1990; Hambarde, 1989; Horshil, 1992; Deo, 1990; Surve, 1992; Bhand, 1993; Tupkar, 1991) studies were conducted at MLISc Level.

2.2 Definitional analysis

2.2.1 Citation analysis

Paul and Roy (1983) defined citation analysis as, “Citation is one branch of bibliometrics where the unit of analysis is a document, that is being cited as a bibliographic reference or as a footnote in a citing document”.(p.226)

Martyn (1976) has rightly defined citation analysis as, “The analysis of citations or references or both which form part of the scholarly apparatus of primary communications. The technique is used for putting items of references in some kind of rank or order, whether they are journals of authors cited”.

17
Ziman (1968) defined citation analysis as, “A scientific paper do not stand alone. It is embedded in ‘literature’ of the subject”.

2.2.2 Ph.D. Theses

Sengupta (1991), defined the Ph.D. theses or dissertation as, “A thesis is a statement of research presenting the author’s findings and any conclusions reached, submitted by the author in the support of his candidate for Ph.D. Degree in Science”.

In the present study the term thesis is confined for which, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad awarded Ph.D. in Economics, History, Political Science, Sociology, Psychology and Public Administration.

2.3 Bibliometrics

Bibliometrics is the most active field of Library and Information Science. Citation analysis study is the major portion of it. Bibliometric is a sub field in the information science. A bibliometric size, rate of citation, citedness of paper, index of immediacy, international collaboration and paper and citation distribution by research centers and universities. Bibliometrics is the study of documents and their bibliographic reference and citation structure. It is a study of quantitative aspect of production, dissemination and retrieval of scientific information.

Bibliometric methods have been successfully applied to examine the intellectual structure of several disciplines (Schneider, 2004). The quantitative study of publications, that have been written by people who have been working with the Bibliometric Research Group (BRG) at City University during the years 2001-2003. Bibliometrics is now very much an International Activity and the Biennial Conferences of the International Society for Scientometrics and Informetrics (ISST) bringing together researchers from 40 countries (Aslib Proceedings, 2005). Bibliometrics involves the quantitative analysis of the literature of a subject domain, as represented by bibliographic entries such as keywords, classification codes, authors and citations, purposes of the bibliometrics study is to find out the growth and characteristics of digital library literature. The major objectives of the bibliometrics study are to find out authorship pattern, author productivity, prolific authors, core journals in subject area, indexing terms frequency, Bradford distribution of articles, year-wise distribution of articles,
language-wise distribution of articles and country-wise distribution of journals (Singh, 2007). Bibliometrics is traditionally associated with the quantitative measurement of documentary materials. It refers to a variety of regularities taken from different fields and exhibiting a variety of forms. Although bibliometrics distributions differ greatly in appearance, they can be conceptualized as versions of a single regularity, so that we can properly speak of bibliometrics laws and their manifestations. Bradford’s law of scattering, Zipf’s law and Lotka’s Law are the best known laws dealing with important phenomena or “regularities” found in science communication. It is such as orientated for application to science policy, library and information science retrieval. “Bibliometrics”, “Informetrics”, “Scientometrics” and “Technometrics” are unfortunately not very clear, and there is choice in the terminology (Wormell, 1998).

The popularity of bibliometric techniques is because of easy data availability and the myriad possibilities of mapping linkages between authors, institutions, geographic, location and funding sources (Rangnekar, 2005). Bibliometrics in particular the extent to which published research is cited by other published research can certainly be used as a means of assessing research quality. A metrics based system that had no previous set of results to compare it with will always be judged less critically than a new system of metric which is haunted by the perpetual approach of the old. Academic information management departments have long had a rare field interest in bibliometrics which often defies simple understanding and makes little impact on real life information practice. But if it is acceptable to speak on behalf of the mythical “typical librarian”, then it is likely that the everyday LIS practitioner would probably have some simple an useful insights into real life applications of bibliometrics and citation analysis, the application of bibliometrics as a determining factor at a more narrow level is not justified (Joint, 2008). Bibliometrics is a well-known evaluation method that has been applied to numerous subjects throughout the years. Not only does it help researchers to identify the characteristics of subject literature, but it is also an aid to librarians in the areas of collection development and evaluation by giving them quantifiable figures on journal frequency, use and impact on the field. From these data, librarians can more comfortably make remote storage weeding and acquisitions decisions. Weakness is that, time-consuming to carry out. Broadly, another weakness is that bibliometrics only gives the user a string of numbers
which may not be the most important part of a journal's importance and impact on a field of study (Lawe, 2003). Bibliometric analysis of cancer articles may furnish some useful data that will not only help the librarians and information scientists of the concerned field but also to the working specialists of this field it is, therefore, necessary to make a thorough survey of cancer literature by adopting the technique of bibliometry covering different sub-fields of cancer so that the working scientists may get all authentic and relevant data on the subject (Ghosh and Sengupta, 1989). Bibliometrics is an emerging thrust area of research involving researchers from different branches of human knowledge. The term “bibliometrics” is of recent origin and was coined only in 1969. It lies in the border area of social and physical sciences (Sengupta, 1990). The assembling and interpretation of statistics relating to books and periodicals to demonstrate historical movements, to determine national and universal research use of books and journals, and to ascertain in many local situations the general use of books and journals (Raising, 1962). The purpose of bibliometrics as to shed light on the process of written communication and of the nature and course of development of descriptive (in so far as this is displayed through written communication) means of counting and analyzing the various facets of written communication, its purpose is analysis and control of the (transfer) process ‘measurement is the common theme through definition and purpose of bibliometrics’ and ‘the things that we are measuring when we carry out bibliometric study are the process variable in the information transfer process (Pritchard, 1972).

Bibliometric provide information about the structure of knowledge, and how it is communicated. They further point out that bibliometrics studies fall mainly into two broad group: (i) descriptive studies and (ii) behavioural studies. Descriptive studies include those describing the characteristic or features of a literature while behavioural studies are those examining the relationship formed between components of literature (Nicholas and Ritchie, 1978). The scope of bibliometrics includes studying the relationship within a literature (e.g. citation studies) or describing a literature. Typically these descriptions focus on consistent patterns, involving authors, monographs, journals or subject/language (O’Connor, 1981). It is a quantitative science and divides its scope into two basic categories; (i) Descriptive bibliometrics (e.g. productivity count) and (ii) evaluative bibliometrics (e.g. literature usage count). The descriptive bibliometrics i.e. the productivity
count can again be subdivided into (i) geographic, (ii) time period, and (iii) disciplines. The evaluative bibliometrics, i.e. the literature usage count may be subdivided into (i) reference count and (ii) citation count (Stevens, 1990). For bibliometrics studies, it would be possible to treat the literature of a subject as more than a ‘black box’ about which only rudimentary statistical fact can be ascertained. Bibliometrics could truly become a branch of epistemology (Small, 1980). Bibliometrics is a well-established discipline for quantitative study of the various aspects of literature of a given subject. It is a branch of information theory, which analyses quantitatively the recorded knowledge to know its properties and behavior. The recorded knowledge may be in the form of books, journals etc. In almost all subject disciplines, bibliometrics research has developed a body of theoretical knowledge and a group of techniques based on bibliographic data elements (Sahoo, 2001). The bibliometrics is an application of statistical and mathematical methods to bibliographical studies. The word bibliometrics is derived from Latin and Greek word i.e. biblio and metric, etimologically bibliometric. Dr. Ranganathan had narrated the term librametric in 1948. Alan Pritchard suggested the term bibliometric in 1969 and become popular (Kogamuramath and Pothare, 2001).

Bibliometrics is the study and measurement of the patterns of all forms of recorded information and the producer (Satish and Kabir, 2001). Application of bibliometric technique in selecting most important journals on any field of knowledge is now-a-days, a well established fact. Statistical analysis of the hundreds of bibliographies at the end of the scientific communications in the source journals selected for a particular field has provided an authentic and important medium for selecting the most important journals of that field. Gross and Gross initiated the work in the 2nd decade of this century Sengupta suggested an objective approach to citation analysis and developed a new methodology in selecting source journals for bibliometric study (Sengupta, 1980). Using bibliometric analysis Garfield conclude that the impact and quality of Indian research is yet to reach international standards. Garfield’s findings were based on data from science citation index (SCI) (Kumari, 1990). Recent development in library and information science and of science of science may be looked into.
These developments are being manifested through the so called scientometrics, bibliometrics and informetrics.

The literature in these branches of science is growing rapidly and now accounts for several thousand publication. The situation is reflected in the abstracting journals of various subjects. It may said that these ‘metrics’ are rapidly forming the methodology or system of methodologies which may put them as candidates to be considered as forming a new scientific branch (Price, 1976; Garfield, 1970; Small, 1980; Lawant, 1981). The term “bibliometrics” as a complex of mathematical and statistical methods used to analyse bibliographical characteristics of documents (Pritchard, 1969). Bibliometrics is a formed scientific subdiscipline including the complex of mathematical and statistical methods, used to analyse bibliographical characteristics of documents. Whether bibliometrics becomes the structural part of methodology of library and information science (Voverine, 1994). Bibliometric is the use of mathematical and statistical methods to study document and patterns of research publications. It is core methodology of information science (James, 2008). Bibliometrics is simply put, the study and measurement of the publication patterns of all forms of written communication and their author (Potter, 1981). Bibliometrics has commanded the attention of numerous individuals in library and information science. The measurement of bibliographic information offer the promise of providing a theory that will resolve many practical problems. It is claimed that patterns of author productivity, literature growth rates and related statistical distributions can be used to evaluate authors, assess disciplines and manage collections. Yet, it is unclear if bibliometrics is merely a method or if it meet the test of a theory in its ability to explain and predict phenomena (O’Connor, 1981).

The widespread application of practical bibliometric method useful to library managers will continue to be limited until a more general, unified theory is developed. Another limitation of bibliometric distributions is the use of unidimensional descriptions of consistency in author productivity or journal citation patterns. A sizable body of literature dealing with bibliometric models has developed. The early models were proposed because they were observed to fit graphically certain specific empirical frequency distribution. In many cases their
functional forms were identical, the similarity only noted by other writers years later. In each case depending on the subject field they applied to, there was a proliferation on papers which modified, extended, clarified, applied and generalized the initial model. Almost bibliometric models relate, in a simple functional form, one variable with another variable (Hubert, 1981). Price considers his model to be quite general: It provides a sound conceptual basis for such empirical laws as the Lotka’s distribution for scientific productivity, the Bradford’s law for journal use, the Pareto law of income distribution, and the Zipf’s law for literary word frequencies. It is therefore an underlying probability mechanism of widespread application and versatility throughout the social science (Price, 1976).

Bookstein proposed to find an expression for the expected number of author in a discipline producing articles over a defined period of time, subject to sociological factors influencing productivity and other constraints (Price, 1970).

2.3.1 Laws of bibliometrics

2.3.1.1 Bradford’s law (1934)

The law provide the fundamental theory of bibliometrics. Bradford’s law made in 1934, describe the scattering pattern of journal in the area of applied geophysics lubrication. “Law of scattering describe the distribution of documents usually journal in the discipline or problem area. Bradford’s law is based upon an observation that journal article in any specific topic show a particular pattern that a fairly large number of articles (Bandyopadhyay, 1999).

Bradford law state that if a large collection of articles is ranked in order of decreasing productivity of journal (source) relevant to a given topic, three zones can be marked off so that each zone produces one third of the total relevant papers (Singh, 2007). Bibliometrics scattering was measured on the basis of Bradford mathematical model (Bhattacharjee, 1975; Afolabi, 1983). Literature on Bradford’s law incorporates both theoretical and empirical aspects. These aspects are each coherent and developing areas of scientific inquiry. This law represents an elusive phenomenon. On an hand, it is easy to observe in real situations and can be represented with a fairly simple mathematical formula (Garfield, 1970).
2.3.1.2 Lotka’s law (1926)

The law provides fundamental theoretical base for bibliometrics study involving authorship. The scientific productivity is frequently major in terms of the publish output, mostly because the data on the number of publication by authors can be easily collected and are quite reliable. In other words, the law correlates the number of contribution. The original of what has come to be known as Lotka’s law was made in Lotka’s 1926 journal article, “The frequency distribution of scientific productivity” (Potter, 1981). Lotka’s “inverse square law” of scientific productivity has since been shown to fit data drawn from several widely varying time periods and disciplines (Singh, 2007).

2.3.1.3 Zipf’s law (1933)

“Describing word frequency ranking”, it develop an extended an empirical law, governing a relation between the rank of the word and frequency occurrence in a long text. The words are arranged in their decreasing order of frequency then the rank of any word of the text will inversely proportional to the frequency of occurrence of the words. It is the puzzling phenomena in bibliometric and, more broadly, in quantitative linguistics – is Zipf’s law. As one commentator, the statistician Gustav Herdan, has put it: Mathematicians believe in Zipf’s law because they think that linguists have established it to be a linguistic law, and linguists believe in it because they, on their part, think that mathematicians have established it to be a mathematical law” (Gustav, 1966; Wyllys, 1981).

2.4 Citation

A citation may be defined as a reference to a text identifying the document in which it may be found. The citations have more importance that what is generally assumed. An important use of citation lies in bibliographical work. The bibliography on a recent topic can easily be compiled by noting down the citation (or reference as is generally called) given at the end of relevant article and then arranging them in manner (Bose, 1986, p.90). Garfield use the term “citationology” for the study of theory and practice of citation and citation analysis, according to Leydes Dorff a variety of contexts for citation analysis exist, but no comprehensive theory of citation itself has been formulated. A citation is a complex unit and the citation process is a very complex one (Schneider, 2004). Citation represent the
pool of archival knowledge from which authors retrieve established ideas and, in turn, generate new research ideas. This knowledge may be disseminated within an area and across disciplinary boundaries (Sharif, 2004). Citation count and impact factors can be easily manipulated (Gorman, 2005). One fundamental component of the research process is the acknowledgement of previously published research the function of a bibliographic citation is to establish a relationship between the citing and cited document.

The association of concepts or ideas through bibliographic citations has been described by some scholars as intellectual transactions or formal acknowledgement of “intellectual debt” to earlier sources of information. Bibliographic citations symbolize the concepts or ideas the authors discuss (Waytowich and Onwuegbuzie, 2006). The validity of impact factor use in making selection decisions, while high local use might correlate well with impact factions, the correlation breaks down as the level of local use of particular serials declines. It has been argued that data from local use studies are more accurate than citation data for deselection decisions (Altmann, 1999). The decrease in the number of citations of a scientific paper with age has been the subject of much research. A difficulty is that the rate of decay of citation depends on several factors: the specific field chosen for study (a physical science might have a different citation decay rate from a biological science), the journal or journals selected (i.e. the coverage of the field). The rate of decay may also depend on the particular country, or language (Meadows, 2004). Referring to “citations as proxies”, citation context studies have tried to devise a classification or taxonomy based on a text analysis in order to find out the inter-document relationship in the presence of reference citations, while content analysis has tried to characterize the cited work by analyzing the semantic content of the citing papers. Garzone and Mercer’s classifying the citations according to the functions: citations of the affirmational type, citations of assumptive type, citations of conceptual type, citations of constrastive type, citations of the methodological type, citations of the negational type citations of the perfunctory type, citations of the persuasive type (Bornmann, 2008).

Citations/ references will be measured in five areas (i) total citations (ii) age of the citations (iii) language of the citations (iv) the format of the citation (v) the number of authors per citations. Citation studies are based on fairly simple
premise. References by one scholar to the work of another are part of the process of scholarship. An analysis of references, therefore, might be expected to yield information about the working habits of scholars (Lowe, 2003). Two distinct relationship between the JIFs (Journal Impact Factor) of the citing and cited journals can be found in the data material. One relates to the JIF of the cited journal, and the other relates to the difference between the JIFs of the citing and cited journal (Frandsen, 2005). Citation can be used to evaluate the scholarly output of researchers, departments whole universities or even whole nation, with caution (Li, 2003). Citation is widely used in bibliometrics researchers usually “cite” the ideas and papers of pioneers. This behavior is called “citation”. A citation includes two objects one is the source work the other is the cited work, the collection of information about cited work is a treasure for citation analysis (Chen, 2004). Citation study which provides helpful guidance in the process of collection development of the library. Citation study ascertains the half life of literature which helps in deciding how long these documents are useful and should be preserved in library (Barroah and Sharma, 1999). Citations can be used to identify the personalities that are central to a subject field. Citations are an adequate way to measure the quality of scientific work or intellectual influence on it (Kretschmer, 1994).

2.4.1 Textual property of citation

The textual characteristics of cited reference, for the purpose of identifying key indicators of document relatedness. Possible indicators were the location of the citations, the frequency of the cited references, and the actual treatment of the citations. In studying the placement of citations, confirmed that the highest density of citations was in the “introduction” section of the citing articles. Coding citation types using Moravesik and Murugesan’s scheme, can found that the largest group of citation was perfectly located in the beginning section of the text, whereas organic citations were mainly located in the middle section of an article. In addition to location and functional type, the value of citations was shown to be associated with frequency of a work cited in text. Voos and Dagaev, Herlach, Bonzi found that sources of cited work, source of citing work, frequency of work cited text, and type of citing article were good indicators of how a citation was used in the citing document (Tang, 2008).
2.4.2 Citation database

A citation database stores key information of technical publications including the paper’s title, keywords, journal name, author name(s), article type, language and references. Useful information can be extracted from a citation database based on the relationship between source and cited papers for example, to identity emerging research trends. The challenge is to extract relevant information from the citation database using effective knowledge discovery technique.

Two approaches toward injecting semantic understanding into the retrieval of relevant publications stored in citation database. In the first approach, combine knowledge based techniques with the learning capabilities of artificial neural networks (ANN) and an automatic thesaurus conceptual clustering. In second approach, we study the application of coword analysis for classification of documents by establishing associations between documents (Hui, 2004).

2.4.3 Evaluation of citation enhanced databases

When evaluating citation enhanced services for bibliometric purposes, one must consider that bibliographic databases may contribute in two distinct ways to bibliometric analysis. One is, as a data sources and second is, as a platform providing the analytical tools for bibliometric analysis. Both contributions are beset with several methodological and technical difficulties, including limited coverage of the scholarly literature, inconsistent and inaccurate data, and limited facilities for browsing, searching and analyzing data (Neubaus and Daniel, 2008).

2.4.4 Citation Impact

Open access journals are not necessarily new publications. Many established journals make only a few recent years of content available online, while the majority of content is accessible only through traditional access paths. Other established journals, having moved to open access distribution, offer access to much older content as well. The important factor for all authors is the impact of their work. If author can see improvement in the impact of their work due to open access, they will be willing to use open access routes. Access to content of journals via the web provides a new metrics for measuring the impact of articles, electronic citations which can be considered to be comparable to bibliographic citations in assessing the impact of published works. The correlation between
citation count provides a measure of the usefulness of selected articles called the “citation impact”. Citation impact can be used as a measure of the impact an article has within its particular field. An article being widely read and cited is an indication that it has had influence with other researchers within the field (Turk, 2008).

2.4.5 Citation error

The investigation also identified some characteristics of students who commit citation errors. Specifically, the canonical correlation analysis revealed that graduate students with relatively high level of self oriented perfectionalism and other oriented perfectionism tended to commit the least citation error and construct reference lists that closely resemble the publication manual of the American Psychological Association. The analysis revealed that (1) students with the lowest expectation levels tended to commit the highest citation error rate and (2) students who have taken the most courses in their graduate programs tended to receive the lowest scores pertaining to the quality of the reference lists. Moreover, that expectation in an important predictor of the frequency of citation error suggests that social cognitive theory, in general and self efficacy theory in particular, is pertinent in reference lists because expectation is a manifestation of self-efficacy (Waytowich and Owuegbuzie, 2006).

2.5 Citation Analysis

Citation analysis is a technique of bibliometrics. It is an important research tool understanding the subject, which we analyse the structure and direction of the subject. It measures the utility of documents and relationship between documents in the subject and relationship between authors and their documents. It is a tool for identifying the core journals in a subject based on counting the citations given at the end of each article from a group of primary journals. Afterwards, a number of such studies were carried out on citations in dissertation/theses, book, primary journals and reviewing journals (Doreswamy, 2001).

A citation is a reference to a document given by a more recently published document. The document citing is the citing document, and the document that receives the citation is the cited document. Citation analysis involves counting the number of citations to a particular document for a period of time after its
publication this is sometimes called direct citations. The traditional understanding of the citation function is that the frequency with which a document is cited can be taken as a measure of the impact or influence of that document on the citing literature. Citation analysis leads to more sophisticated methods, such as cocitation analysis, mapping of the literature, bibliographic coupling and co-word analysis. These methods, individually and in combination, seek to find information patterns by analyzing reference and citation patterns as well as word use frequencies, combined with statistical analysis. Citation analysis is applied to identify highly cited documents (Schneider, 2004). Citation analysis in general and the bibliometrics indicator journal impact factor (JIF) in particular play a dominant role in sciences for the perception of quality and prestige even if some journal analyses are considered so important that some journals try to manipulate their way to a higher JIF. It is widely used for evaluations even though the fairness of such analyses is debatable why it is important to attempt to qualify such analyses. Researchers, politicians and publishers often use citation analysis as the basis for statements on research quality or impact (Frandsen, 2005).

Analysis focuses on the investigation of the relationship between the citing and cited documents or “the links expressed in the references (Tang, 2008; Sharif, 2004). Citation studies and citation impact analysis to satisfy ‘useful purposes’ in supporting the interests of the research environment (Wormell, 1998). The citation analysis described are amenable as the basis for web linkage studies. In Pakistan citation analysis of documents is not common, one can hardly find one or two articles that have used this technique (Sharif, 2004). Citation analysis is an important tool used to trace scholarly research, measure impact, and justify tenure and funding decisions. It allows a researcher to follow the development and impact of an article through time by looking backward at the references the author cites, and forward to those author who then cite the article (Bauer and Bakkalasi, 2005). The application of citation analysis to research evaluation is founded on this tradition. It aims to estimate the varying contributions of scholarly work to the advancement of knowledge. Citation analysis has proved to be an important assessment tool for research evaluation conducting citation analysis using the citation indexes produced by Thomson scientific provides an obvious starting point in assessing research performance, but is bibliometrically restricted to a small
fraction of the journal literature. The availability of citation data in other bibliographic databases opens up the possibility to extend the data source for performing citation analysis, particularly to include other publication types of written scholarly communication such as books, chapters in edited books and conference proceedings (Neuhaus and Daniel, 2008). Citation analysis is one methodology adopted by libraries in evaluating their collection as well as a tool for building an effective journal collection (Aina and Mabandouln, 1997; Asundi and Kabin, 1996). Citation analysis is based on counts of the number of citations that article or monograph has received in a particular period. Citation analysis was not easily undertaken until the launch of science citation index, social science citation index and art and humanities citation index. Garfield’s seminal work on citation analysis, although 25 years old, provides an authoritative historical and conceptual understanding of the topic and its many application (Garfield, 1979). Citation analysis has many critics, and some have focused their criticism on the idea that citation analysis could be used as an input into the RAE – Research Assessment Exercise process (Norris and Oppenheim, 2003). The prime value of citation analysis data to the LIS profession was always in subject retrieval, based on simple but elegant principles: if an article cites an earlier article it is quite possible that the later article has a subject relationship with that earlier article. Citation analysis is that it is a machine driven system of some value. But that it should always be used in subordinate role, as an adjunct to information retrieval processes driven by human agency. This is also how we should intuitively expect bibliometrics to be used in other application, such as research assessment, in the government’s vision of the new “metrics”, research income measures and then citation analysis were to be the main measures and then citation analysis were to be the main measure of research quality (Joint, 2008).

Citation analysis is a method of determining the characteristics of a field or subject by careful examination of the literature of that area. It is based on the principle that “the actual use of the material is indicative of its relevance to current research”. There are two directions that a citation analysis can proceed (i) to the theoretical, which is useful for understanding the development of the literatures as they impact bibliographic and budgetary matter or (ii) to the specific which allows professionals in a particular field to gain knowledge about the details of their field.
(Lowe, 2003). Citation analysis only rarely use thorough investigation of the evaluated science as their starting points, citation analysis in general and the bibliometric indicator. Journal Impact Factor (JIF) in particular play a dominant role in science for the perception of quality and prestige, even finds that citation analysis are considered so important that some journals try to manipulates their way to a higher JIF (Frandsen, 2005). Citation analysis is an effective but some what neglected technique of collection development and evaluation. The method of citation analysis consists of counting and ranking the number of times documents are cited in bibliographies, footnotes, and/or indexing tool (Baker and Lancaster, 1991). Citation analysis has been done through nationally collected citation statistics and proposals have been made to use these analysis for library collection development and evaluation (Bensman, 1985; Broadus, 1985). Data from citation analysis should not be used alone to determine what journals are held by the library. The method has limitations and the best results for collection development should be obtained by using convergence of data from all available sources (Sylvia, 1998).

Citation analysis as a tool for evaluation, citation frequency and impact factor could be helpful in determining the optimum makeup of special and general collections. This can serve as a guide to determines the back files, binding and retention schedules of science journals (Garfield, 1978). Citation analysis of students project and theses help in the evaluation of uses of reference. Collections in university libraries (Okiy, 2003). Citation analysis permits the researcher to comprehend current developments in subject fields, the properties of bibliographical usage, and research trends (Chen, 2004). Owing to the phenomenal growth of literature and rise in subscription costs, lack of space and shrinkage in budget allocation, the library managers have been constrained to determine the collection development and maintenance policies with the optimum use of money and space available and also to meet the current and anticipated needs of the library users (King, 1976). Found the growth of scholarly journals to be 2.6% per year, and there was a steady increase in subscription price on an average of 10-13% every year (Subramanyam, 1975). Citation analysis is the one of the areas of bibliometrics, which deals with the study of relationships of citing and cited contribution. Garfield given fifteen reasons for citing any document, there are
some limitations of citation analysis: (i) a large amount of data is essential to derive solid conclusion, (ii) the collection, processing and analyzing of the raw data is carried out manually only which may act as a barrier regarding to accuracy of the result, (iii) fanciful titles in citations created serious problem in identifying the subject matter, (iv) every reference made at on time has been counted as one citation, if the same reference has been made again using “ibid” or “op.cit” it has been counted again, (v) all occasional papers have been considered as unpublished papers because of their nature (Satish and Kabir, 2001). Citation analysis can be used to study the nature and characteristics of literature cited in the different discipline (Balasubramanian and Bhaskar, 1984). The development of citation analysis has been marked by invention of new techniques and measures, the exploitation of new tools, and the study of different units of analysis. These trends have led to a rapid growth in both the number and types of studies using citation analysis (Smith, 1981). The easiest technique to use is a citation count, determining how many citations have been received by a given document or set of documents over a period of time from a particular set of citing documents (Pinski and Francis, 1976). Two techniques have been devised to identify documents likely to be closely related: bibliographic coupling and co-citation analysis (Kessler, 1963).

2.5.1 Objectives of citation analysis

The objectives of the citation analysis:

(i) To know the distribution of citations in different bibliographic forms,
(ii) To examine the authorship pattern,
(iii) To identify the core journal,
(iv) To know the geographical scattering of citations,
(v) To examine the subject wise break up of citation,
(vi) To find age of cited documents,
(vii) Distribution of monographic citations in local and foreign publication,
(viii) A comparison of all measure described above in the journal,
(ix) Assess researchers, access to information as reflected from citation analysis,
Find out the extent to which the available information resources meet the research needs revealed by the study.

Study the distribution by place of publication,

Identify the language wise distribution,

Compile the rank list of the cited author,

Deduce the citing half-life of the literature cited in 120 and


2.6 Scientometrics

When refer to scientific production and communication it is usually called scientometrics, this field was immigrated in 1960 by pioneering work of ‘Derck de Solla Price’, ‘Mauric Goldsmith’ and ‘Eugengent,. The common form sources of data for analysis is the SCI and SSCI both the institute of the scientific communication (SCI – Science Citation Index), (SSCI-Social Science Citation Index). Correlation between citation impact and per review to make evaluation studies more reliable, it is well known that the evaluation of a person by Scientometrics methods need more caution than bibliometric analysis of countries, institutions or groups (Wormell, 1998). Scientometrics is the newly emerging field that investigates quantitative aspects of science, it is the quantitative arm of the science of science, scientific communication, studies and science policy studies (Conference Report from Library Hi Tech News, 2001). Many types of scientometrics data can be presented as transaction matrix. In all cases the matrix consist of a set of items assigned to each row and column with each cell containing the level of transaction between the row and column items (Kretschmer, 1994).

Scientometrics was suggested by Nalimov and Mulchenko (1969) in their book “Scientometrics : the investigation of science as development of information process”, 1969. According to them scientometrics is a complex of quantitative (mathematical and statistical) method, which are used to investigate the processes
of science (Nalimov and Mulchenk, 1969). Scientometrics is a formed structural part of science of science methodology. Including, the complex of mathematical and statistical methods, used to analyse the quantitative characteristics of science as an enterprise (Veverience, 1994).

2.7 Informetrics

Informetrics is an umbrella term which covers all form of quantitative research of information, discusses bibliometrics, the concept of internationally, science search data base and science citation index. The application of informetrics methods to the solution of chemical problem (Gasteiger, 2006). “Informetrics studies” signifies the new approach of CIS – Centre for Informetric Studies to the scientific study of information flow : improved bibliometric methods are applied not only to scientometric studies and research evaluations of science and technology (SST) but also to the analysis of their mutual, societal, industrial and other specific relations. CIS has also extended traditional bibliometric analysis to cover non-scholarly communities where information is produced, communicated and used (Wormell, 1998). CIS researcher have studied the interesting idea of utilizing informetrics methods on the World Wide Web (www) and started to lay the basis of an emerging area of “webmetrics”. Recent publication have presented a workable method for general informetrics analysis on the www, accompanied by case studies analysis Danish Norwegian, Swedish and international web sites, and their relative visibility on the net compared with relevant positions in the scientific databases (Almind and Ingwersen, 1997). The field of informetrics is concerned with regularities underlying the use of production of information, in particular studying the quantitative properties of information (Ruthvan, 2004).

Wolfram outlines the main foundations of informetrics : the law or generalization in information process and production such as Lotka’s law, Bradford’s law, and Zipf’s law. As Wolfram noted, these mathematical formalisations are not limited to informetrics. Informetrics incorporates the older field of bibliometrics and the new area of cybermetrics and webometrics (Conference Report from : Library Hi Tech News, 2001). “Informetrics” was suggested by German scientist Blackert and Zygel in 1982 as a newly formed branch of science, using mathematical and technical information on theoretical
level and practical information activities. Informetrics is a scientific subdiscipline not yet formed. The problem of its research object to be formed on the concept “Information” is not solved yet (Voverine, 1994). Informetrics covers a wider area of the society and also vast and varied type of information and knowledge resources. Informetric studies signify a new approach to scientific study of information generation, dissemination and use (James, 2008).

2.8 Form-wise distribution

Literature cited in the journal is published in different form like books, journals, conference proceedings, thesis etc. The information regarding the form was collected from the source data and tabulated to find out the most dominant form of literature. Form wise distribution of citations has been done in order to know the most dominant form in which the information is cited. The study will be helpful for information scientists as well as researchers, to know the most dominant form in which the information was being cited in the concept. Form-wise distribution of citation analysis results that in many study, most of citations are given of book, less than that journal citations are given, then other forms of document are cited, online literature citations are also cited, online literature citations are also given. Its citation rate are increasing day by day (Haridasan, 2007; Doreswamy, 2001). In science study journal citations are more than books, patents, proceedings, report, theses (Barroah and Sharma, 1999; Lokhanda, 2007; Satish and Kabir, 2001; Deshpande, 1997).

2.9 Authorship pattern

The distribution of the citations according to the number of authors, in the particular discipline, that means authorship pattern (Doreswamy, 2001). Single author contribute maximum number of article, large number of articles by single authors means that there are no well-established research groups in the area the subject is a new and emerging one (Chen and Chen, 2005). Single authored and two authored citations are maximum, and more than three authored articles are less in number (Sharif, 2004; Barroah and Sharma, 1999). Analysis of authorship pattern show that single authored study is more than multi authored study (Sahoo, 2001; Kagamuramath, 2001; Kannappanavar, 1991; Afolabi, 1983; Deshpande, 1997).
2.10 Geographical distribution

Bibliometrically, the value of place of publication points to the “nationality” of document. Place of publication is the best guide available to nationality of a document (Singh, 2007). Geographical relationship that can be observed from the data is that journals tend to cite journals from their own region more having controlled for self-citations and geographic origin of the cited journal (Frandsen, 2005). The geographic variables are constructed by determining the geographic location of each journal i.e. their place of publication (Kumar, 1984; Sahoo, 2001). Country-wise distribution of citations means geographical distribution the foreign distribution is more than Indian citation distribution (Balasubramanian and Bhaskar, 1984). Geographical analysis of citations provides information about the countries active in a subject field and their relative contribution (Chikate, 2008).

2.11 Chronological distribution

Citations are the references, included by authors in their article. During the period of study which arranging sequential order. Some citations arranged according to the period of publication of cited journals. If the period of the study is long the journals are divided into period of years, having periodicity the periodicity mostly used five, ten years, rarely twenty years periodicity is used. Analysis of chronological distribution show that older documents are less cited than newer ones (Singh, 2007; Kumar, 1984; Lokhanda, 2007; Pillai, 2007; Sahoo, 2001; Kogamuramath, 2001; Mahapatra and Kaul, 1992; Deshpande, 1997; Balasubramanian and Bhaskar, 1984; James, 2008). The cited references were distributed in different time intervals according to their date of publication. This makes the pattern of citations to reveal from quickly the literature on some subject becomes obsolete (obsolescence) (Haridasan, 2007).

2.12 Language-wise distribution

Language is media for communication authors cites different types of document for writing. Selection of document depend upon many factors like subject matter, year of publication, country of origin, method of work, language and availability of source material.
Most of the literature in science is written in English language. English is at the medium for study and teaching specially at the higher level of education. Mostly research scholar in science preferred English language (Beena, 1997; Deshpande and Rajyalakshmi, 1997; Hasso, 1984). Language distribution study shows the most dominant language in which the literature on the subject is cited in the journal. This information is useful for researchers as well as information scientists. English language is the most dominant language in any literature (Haridasan, 2007).

2.13 Ranking of authors

Rank list of author helps the professionals in many way that is for deciding popularity of author or his work, it help to select reading materials, to decide acquisition policy. It helps to researcher also, which document is popular, it indicates the importance of the author and which author is doing very much work in the particular field. Ranking of author is important to give the direction to the professional and researchers and user. There are some risks in author ranking that is how many citations authors will accrue relative to others, it does not support or provide any interpretation of why one author get more or less citations. The goal of this work is not to promote the use of citations as a means of ranking (and subsequent hiring and promotion decisions). Citation data may at best be one of several input to important decisions, and should definitely not be the decisive one (Feitelson, 2004). Ranking of authors is done to know the eminent personalities in the subject, whose work is used by the authors to refine their idea on the subject or topic (Haridasan, 2007).

2.14 Ranking of journals

Ranking of the journals show the usefulness of the journals and their importance in particular subject. This is impact factor on any field of subject. It identify the core journal Solari and Magri told about ranking of journal, it is used in a wide scientific context as a tool for evaluating (researchers and research work) for obtaining overall synthetic perspective of impact factor values, by analyse the ranking of journals. The list is useful only, if it is assumed that all the journals of the ranking are comparable (Frandsen, 2005). It is useful to establish a list of journals mostly cited by the author (Sharif, 2004). There are two approaches to
rank the journals (i) using citation scores. Journals were ranked by citation counts and those with high citation counts were considered to be of high use and thus ranked accordingly, (ii) using user opinion. Journals ranked according to user opinions and those with high score were ranked at the top (Dulle, 2004). Ranking of the journal is necessary to know the most productive journals used in reference, for further study of the subject, this information is useful for the librarians as well as research scholars (Haridasan, 2007; Barroah, 1999). Ranked list of periodicals on a given subject indicating their productivity will serve as an important tool in the hands of the documentalists, librarians, and information scientists in as much as they can review and revise their acquisition list with the help of it. On the other hand measurement of bibliographical scattering of periodicals on a particular subject will reflect the degree of expansion of periodical coverage required in order to achieve optimum bibliographical control (Bhattacharjee, 1975; Deshpande, 1997).

2.15 Author self citation

Self citation is a common phenomenon in citation behaviour. Self-citation inter links various publications and shows their relationships. The author ranking studies often shows the rate of self-citation (Tagliacozzo, 1977). Self citation confined to habits of concerned authors. Moreover senior person will generally get a lot of his earlier works to cite than the younger group of people this will naturally influence the study of citation analysis, mainly the author ranking, it is suggested that the citation analysis study must pay proper attention to the number of self citations provided in an article (Mahapatra and Kaul, 1992). Self citation refers to the number of times the previous papers published in the same journal; the rate of self citation is lower than other author’s citations (Balasubramanian and Bhaskar, 1984).

2.16 Journal self citation

Journal self citation is an interesting bibliometric indicator that gives in indication about the popularity of the journal among its contributors as well as the reader community. The phenomenon of a journal citing itself is termed as journal self citation. The pattern of journal self citations is part of both the citation identify and the citation image, depending on the perspective from a journal (referencing) or to a journal to which citations can be given. For the practical reasons both the
aspects will be treated here in combination, time perspective of a journal, it is supposed that self citations reach an earlier peak after publication than external citations. By age the rate of citation of journal decreases (Mahapatra and Kaul, 1992).

2.17 Citation Indexing

Reviewing the 40 years history of the ISI citation databases, the key advantage of the citation indexing its capacity to bypass the use of normal linguistic forms such as title, keywords or subject heading has not yet been discovered by a large part of the LIS community. The symbolic role played by the citation in representing the content of papers in an important dimension of information retrieval: in combination of various natural language expressions, citation indexes can greatly improve comprehensive literature searches. Citation indexes enable the searcher to locate subsequent and especially current descendants of particular papers or books. As a tool for navigating the literature, citation indexes can expand the scope of the search by retrieving not only those papers that cited a key work, but also those related to the citing references (Wormell, 1998). The main resources for citation analysis are the citation indexes produced by Thomson scientific (Formerly Institute for Scientific Information). Besides their multidisciplinary nature, citation indexing was the major reason. Why this service had an unique position among bibliographic databases for many years. Thomson Scientific is however, no longer the only service offering citation – enhanced data based on the market. In recent years, several database producers have noticed the potential of citation indexing and manually added cited references to a subset of their records. Thomson scientific citation indexes have some limitations that are of crucial relevance for citation analysis as an assessment tool of research performance. Among these constraint is the limited coverage of the citation indexes. As outlined above, Thomson Scientific processes only a selected set of journals for its citation indexes. While these accessed journals tend to be the highest impact peer reviewed journals, they represent only a fraction of scientific work that is documented. Hence, coverage relates to the extent to which the citation indexes cover the written scholarly literature in the field (Neuhaus and Daniel, 2008). The origin of citation analysis as a wide spread assessment tool of research performance can be traced to the mid 1950s, when Garfield proposed the
grounds breaking concept of citation indexing. With the introduction of the science citation index (SCI), the social sciences citation index (SSCI) and the Arts and Humanities Citation Index (A&HCI) by the Institute for Scientific Information Systematic analysis of the impact and influence of scholarly work as well as of trend in science became available (Garfield, 1979). The library of the Chinese Academy of Science has produced the Chinese science citation database (CSCD) since 1989 (Meng, 1995). Nanjing University in Mainland China and the Hong Kong University of Science and Technology produce the Chinese Social Science Citation Index (CSSCI) (Su and Han, 2001). These citation indexes data bases have a great impact on research. Some researchers apply citation data to evaluate the contribution of research fellow, journals, or institutes. Some apply citation data to analyze the structure of specified research field. Some apply citation data to retrospectively evaluate research trends, and to make projection about future trends (Garfield, 1979).

2.18 Collaboration

Co-authorship index (CAI) defined as the average number of authors per document, inter centre collaboration index (IC), number of centres per document, international collaboration rate (percentage of documents authored by two or more countries) (Bonilla, 2008). Collaboration is a sign of scientific achievement and an indication of teamwork which can take place because there is a recognized theoretical structure about which scientist can easily communicate, lack of collaboration in a discipline a sign of scientific underdevelopement or it is an indication that the subject matter and problem of field are more (Lowe, 2003; Li, 2003). Collaboration research has been found to be the dominant pattern in many subject field. Degree of collaboration is different in different subject (Kannappanavar, 1991).

2.19 Obsolescence

Obsolescence defined by Line and Sandison (1974) as the “decline over time in validity or utility of information”. This concept is obvious interest to information theoreticians who concern themselves with the development career and eventual death or incorporation of particular kind of information. But it is also of interest to practical librarians who administer growing collections in finite spaces. Such librarians look to research on obsolescence to help them decide which
Librarians look to research on obsolescence to help them decide which items to keep and which to store or discard in order to make room for new acquisitions. Ideally for remote storage or discarding, research on obsolescence would culminate in simple mathematical formula which could be applied with equal success to any and all libraries. Obsolescence research has produced many mathematical formulas, but unfortunately they have been neither simple nor universally applicable. The best researchers are the one who have admitted that obsolescence is a far more complicated and more hypothetical concept (Gapen and Milner, 1981). Half life is measure of the growth of discipline. In general, within the same subject field, one would expect that a journal that cites primarily newly published material is better than one that cites older source. The citing half life (median citation age) shows how far back in time one must go to account for the age one half of the bibliographic references published in a journal in a particular year (Haridasan, 2007). The term “obsolescence” occurs frequently in the literature of librarianship and information science. In numerous papers we are told how most published literature becomes obsolete within a measurable time, and that an item receives half the use it will ever receive (half life) in a few years. “Obsolescence” is however very rarely defined, and its validity, interest and practical value are often assumed rather than explained (Meadows, 2005; Ijari and Kannappanavar, 1989). Obsolescence study of literature is one of the main area of informetrics and it became an important characteristics of scientific and technological literature. The focus of the most of the obsolescence studies are at the individual document level and may extend up to the journal, sub disciplines or discipline level. The growth of literature and their obsolescence are usually treated together because they represent the initial and the final stages of the information cycle. Both are more time dependent than most other areas of informetrics, where time is not treated as a variable, but more as a boundary defining which data are collected. Growth studies investigate regularities in the creation of literatures or documents over time, which is equated with the growth of information. Buckland explained obsolescence as the relative decreases in use of material as it ages. It is the process by materials become no longer useful or reliable. The term obsolescence is commonly used notions with negative connotation, and is described as the decline of usefulness over time. It can also be described as temporal selective in the use of articles (Pillai, 2007).
2.20 Conclusion

The review of literature shows that vast amount of literature was available in evaluation of collection. There was no uniform pattern in standards, guideline regarding collection of development.

Citation analysis is the major technique of bibliometrics which is useful to evaluate collected data or library collection, it is useful for librarian to collection development and research scientists for their study.

Considering published literature present study has used quantitative evaluation of economics collection.
References


Hambarde, G.K. (1989). Citation analysis of Ph.D. Theses submitted to Chemistry Department of Marathwada University (Unpublished Project Work). Marathwada University, Aurangabad.


Khandare, R.B. (2010). Citation analysis of Ph.D. Theses in History submitted to Dr. Babasaheb Ambedkar Marathwada University, Auragabad. (Unpublished Dissertation). Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.


Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. *Communication among Scientists and Engineers*.


