CHAPTER: 1

INTRODUCTION

1.0 Introduction

1.1 Citation Analysis: A Basic concept

1.2 Applications of Citation Analysis
CHAPTER: 1

INTRODUCTION

1.0 Introductory

The words 'Citation' and 'reference' are commonly used interchangeably, but in actuality there is a subtle difference. Webster's unabridged defines the two terms as follows:

Citation - Act of citing passage from a book in its own words, or from another person; also, the passage or words cited; quotations; .... (1).

Reference - A specific direction of the attention; a sign or direction referring a reader to another passage or book (2).

Prince has suggested that these terms should not be used interchangeably. He proposed and adopts:

...the convention that if Paper R contains a bibliographic footnote using and describing Paper C, the R contains a reference to C, and C has a citation from R. The number of references a paper has is measured by the number of items in its bibliography as endnotes and footnotes, etc. while the number of citations a paper has is found by looking it up in some sort of citation index and seeing how many other papers mention it (3).
The origin of citation practice or footnote cannot be easily determined. But the practice seems to have been well established in scientific writing even when the early periodicals started about three centuries ago. Derek de Solla Price has found out that the earliest name of a footnote was 'scholia', which means 'relating to scholarship' (4). This is an indication that the practice of footnoting was considered to be a scholarly practice. Scientific tradition requires that when a reputable scientist or technologist publishes an article, he should refer to earlier articles which relate to his theme. These references are supposed to identify those earlier researchers whose concepts, methods, apparatus, etc., inspired the authors in developing their own articles. Melvin Weinstock has listed 15 reasons why authors of scientific papers are likely to cite earlier publications (5). These are:

(1) Paying homage to pioneers;
(2) Giving credit for related work;
(3) Identifying methodology, equipment, etc.;
(4) Providing background reading;
(5) Correcting one's own work;
(6) Correcting the work of others;
(7) Criticizing previous work;
(8) Substantiating claims;
(9) Alerting researchers to forthcoming work;
(10) Providing leads to poorly disseminated, poorly indexed, or uncited work;
(11) Authenticating data and classes of fact-physical constants, etc.;
(12) Identifying the original publications in which an idea or concept was discussed;
(13) Identifying the original publication describing an eponymic concept or term as, e.g. Hodgkin's Disease, Pareto's Law, Friedel-Crafts Reaction;
(14) Disclaiming work or ideas of others; and
(15) Disputing priority claims of others.

In view of this established 'reference tradition' bibliographic references may be considered as 'unobtrusive measures' of document use. It is seen that authors use the citations as a sort of window dressing to give a scholarly look to their paper. Again, some authors may be in the habit of citing references of other authors who belong to the same school of thought or group. In some cases, it is like mutual backs cratching while in others it is sometime excessive-self-citation. The dominant reasons for citing references is that they provide identification of those earlier researchers whose concepts, methods, and apparatus, etc. have inspired and influenced the authors, it also indicates the probability of a subject relationship between two. It is also possible that some citations are included to ensure that a few
very recent papers are in the bibliography to forestall reviewer's attacks on the up-to-dateness or otherwise of the paper (6).

According to Norman Storer, there are various norms in the social system of science. One of them is organised scepticism, i.e. the members of the scientific community should respond to the contribution of others, both positively and negatively according to one's own evaluation and contribution in question. Again a viable social system must have a provision of destructive incentive or reward both psychologically meaningful and institutionally appropriate. In both the above cases, the practice of citation analysis plays an important part (7). In fact sociologists have come out with their own interesting reasons for practice of citations. It is argued that norms of citation behaviour are deeply embedded in research and scholarly tradition.

Whatever the reasons for citing an article it can be assumed that there is some relationship between the two papers. Its nature will obviously vary greatly. As Martyn says, 'a citation is not a unit, but an event' and is only quantifiable in terms of its frequency of occurrence (8).
1.1 Citation Analysis: A Basic Concept

Bibliometric analysis has now become a well recognised and dependable technique for research. It aims at providing quantitative analysis of the phenomenon going with documents, their organisation, use and services in libraries, information centres and system. It offers to the information workers a type of statistical technique for the study of characteristics and attributed of literature and that of communication media.

The term 'Bibliometrics' was first used by Pritchard to describe "all studies which seek to quantify the processes of written communication ... the application of mathematical methods to books and other media of communication" (9). Fairthorne defined bibliometrics as "the quantitative treatment of the properties of recorded discourse and behaviour pertaining to it" (10).

Before the use of the term 'bibliometrics', it was known as 'statistical bibliography' which is said to have been first used by E. Wyndham Haulme in 1922 (11). It was next used by Gosnell (12) in 1944 and by Raisig (13) in 1962. Within the last forty years bibliographical statistics have been collected and explained in several fields of science for these main purposes: to demonstrate historical movement, 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 (14).

Bibliometric studies fall into two broad categories - those describing the characteristics or features of literature; and those examining the relationship that exist between the component of literature.

Literature descriptions, according to Nicholas and Ritchie (15) concern with the following characteristics or features of documents both published and unpublished:

(1) Bodies responsible for the production and transmission of the information;
(2) Form of transmission;
(3) Medium of communication;
(4) Nature of information conveyed - subject and language characteristics;
(5) Timing and frequency with which information is conveyed;
(6) Amount of information conveyed; and
(7) Geographical origin.

The second category of bibliometric studies aims at establishing relationship that exists inside the literature. One area of bibliometrics is "Citation Analysis" which is the analysis of the citations or
references which form part of the scholarly apparatus of primary communication. Broadus defines the 'true' citation analysis as one which "deals with works cited as having actually been used in preparation of, or having otherwise contributed to, the source paper (16). A distinction is often made between references (made by articles or monographs) and citations (made to articles or monographs). The number of references a paper has is measured by the number of items in its bibliography as endnotes and footnotes, etc., while the number of citations a paper has is found by looking it use in some sort of citation index and seeing how many other papers mention it.

Broadus uses a somewhat different definition when he describes a citation study's methodology: a source publication or group of publications is searched for bibliographic citations. These references, generally footnotes are copied and then analyzed in various ways. The typical study provides four different breakdowns of the total citations; the number of percentage in

(a) each subject field;
(b) each major language;
(c) each form of publication (such as book, periodical, thesis, etc.); and
(d) each age category by five- or ten year intervals.
The viabilities of 'Citation Analysis' are based on two assumptions, firstly the author cites all or atleast the most important sources used in the preparation of work and, secondly, all the sources listed are indeed used (17). Authors tend to cite the documents which are easily accessible to them that involves with geographical and language limitations. It is quite obvious that citations need not be concise. An unfavourable opinion prevails that sometimes authors cite the publications slightly used or perhaps not at all. Studies indicate that authors, who write in English language, for almost all fields, tend to cite the articles published in the same language and the articles in other languages are scarcely cited. Crohin says, "authors seem to be seldom in their scientific, political and personal goals and not to describe their intellectual ancestry" (18). As a consequence, authors prefer superfluous self-citations regardless as how far relevant they are. Usually more citations are received by leading and eminent people. It is because they are well known to everyone.

As Broadus states: Of concern here is the extent to which people become interested in publications because of biographical references in other works. A highly incidence of such use would mean that

(1) the validity of citations tends to be substan-
tiated and
that citation counts have some direct value in predicting use of library materials (16).

Hodges attempts to evaluate the importance of citations to scholars with the following contention:

Access through references is important, and potentially so fruitful for the advancement of knowledge that (a) more fields than are now served by a citation index should be so served, and (b) scholars would benefit from making heavier use of existing citation indexes than they now make.

Concerning the much-debted citation patterns of writers, her study shows: Scientists and humanists both believe in attributing work they use to its originators, in giving readers what they need to check statements, and in relating their work in the same and contiguous areas (19).

Broadus, as a conclusion in his review of citation studies, comments: It may be concluded on the basis of the evidence so far presented that citations are treated seriously by scholars; that they actually are used as leads to specific library materials; and here, it follows that they are predictors of demand. References made in journals or books widely circulated among scholars thus would appear to be the better predictors (16).
There are some basic concepts or techniques in the area of citation analysis, which are frequently used for variety of studies. These are

(a) Direct citation
(b) Bibliographical coupling
(c) Co-citation

(a) **Direct Citation**: If an author is citing any reference in his scientific paper, there is a strong possibility that it may be related to it. This concept of direct citation has been used for variety of researchers to establish relationships among documents of different types. The main credit for developing this concept goes to two researchers Dr. Garfield and Solla Price although the latter also provided a conceptual framework in understanding it.

(b) **Bibliographical Coupling**: It is postulated that any two scientific papers bear a meaningful relationship to each other (they are coupled) either by virtue of their joint descent from the third one, i.e. two papers are said to be coupled if they both cite the same reference (or a test paper), or a group of papers may be forwarded when each paper in the group has at least one reference in common to every member of the group. The strength of the coupling is determined by the number of citations they have in common. This is the basis for Kessler's (19) work on bibliographic coupling. He
suggests that if any two papers have a reference in common then they are bibliographically coupled. A group of papers $G_A$ are related if each member of the group has at least one coupling unit to a given test paper $Po$. The strength of the coupling relationships is measured by the number of coupling units between each paper and the test paper $Po$. Kessler goes on to suggest that bibliographic coupling could be used to show who should read which papers. As papers are published their strength of coupling with other papers should be used to show who must be sent this new paper. This idea would be extremely useful in overcoming language barriers, and obviates any need for subject analysis or translation before the dissemination process. Bibliographic coupling will obviously lik most successfully those papers that provide historical references. Review papers will obviously couple strongly, as will those papers that merely repeat work already done - now termed 'me too literature. Work that is novel or unique will have a low coupling strength, and will thus escape many networks (20).

(c) Co-citation: This concept is based on the philosophy that if two citations are cited together in the latter literature, there is a probability of a relationship among them. The more the frequency of occurrence together, the stronger the chances of relationship
between the two. So, one can say that it is the measurement of the degree of relationship between papers as observed by practising scientists or specialists. Also, because of its dependence on the practising scientists, these co-citation patterns may fluctuate as the interests and intellectual patterns of the field change with time. This concept was for the first time suggested by Small (21) and later on developed by Small in collaboration with Griffith (22). The emergence of co-citations technique has provided clues to the understanding of mechanism of speciality development and mechanism of science.

1.2 Applications of Citation Analysis

The analysis of bibliographical citations is fairly a common investigative technique used both by librarians and information workers on the one hand and science policy researchers on the other hand, for variety of applications. These applications range from both practical studies such as deciding the appropriate collection for a library, appropriate retention period for library holding, etc. to a more theoretical studies of information flow, communication and diffusion of ideas, etc. For the sake of simplicity, we can study these applications of citation analysis in the following points:
(1) Rank list of journals;  
(2) Journal clustering;  
(3) Measure of scholarliness;  
(4) Obsolescence;  
(5) History of speciality;  
(6) Mapping of science;  
(7) Communication patterns;  
(8) Evaluation of efficiency of a scientist;  
(9) Inter-disciplinarity in research areas.

(1) Rank list of Journals: One of the earlier type of studies carried is generally known as 'Preparation of Rank list of Journals'. This involves counting of citations journal-wise and arranging them in the decreasing order. These type of studies are still carried out but only the focus is shifting from one field of knowledge to another. Though these studies, we can get an overall picture of disciplinary structure of science but not of its specialities, because the journals are too broad a unit of analysis to reveal the fine structure of science and its specialities. Nevertheless these studies are also of considerable interest to epistemologists of science.

(2) Journal Clustering: Having analysed an individual author's citation habits, one can isolate those journals that contain papers that are cited most often. A picture can be built up of the key journals in the
centre of any subject area, and the fringe areas. This process of clustering journals together by their cross citation patterns should produce groups more specific than the discipline level. Narin's \((23,24)\) work in this area has produced interesting results. He used 'Journal Citation Index' as raw data. 'Journal Citation Index' is a break down from 'Science Citation Index' of each journal and the journals that have contained articles citing it. From this he produced a matrix of each journal and all the other journals it has cited. A stage further produced a rank listing of the journal cited most often, then second most often, and so on, of each individual journal in turn. For every journal title there will thus be a ranked list of all other journal titles cited.

Cross-disciplinary models demonstrated flow of references from one discipline to another and revealed an overall relational sequence with a small number of journals as the linkage points. Hence one can isolate clusters and groups of journals for a particular discipline. One can see the most cited journals, and the family links between them. Key journal titles will head the listing of all other journals, and maps can be built up of the journal title citation patterns. It is important to note here that the main thing here is not the evolution of a
hierarchical list of most cited journals. It is also possible to study the reasons why this happened. Relating thus to the other factor in the development of science, it may be possible to determine preconditions necessary for the evolution of inter-disciplinary approach. With further refining of tools, we may be able to predict which two fields would get related in future.

(3) Measure of Scholarliness: Windsor and Windsor(25) have proposed an interesting measure of the scholarliness of a subject field. According to them the proportion of papers produced with no references given at all should be the measure of scholarliness. They found that almost one-third of papers in 'Library and Information Science Abstracts' had no references at all, while half of them had four or less. When compared with pharmacy, information science was found to be less scholarly. They also suggest the rather simplistic view that individual articles could be rejected by searchers purely on the absence of any references.

(4) Obsolescence: Citation analysis can also provide some evidence as to the use of various types of library material. This can help the librarian to re-allocate his funds accordingly. Another problem that is normally faced in a growing library is the pressure of 'space required'. Some material in libraries becomes out of date as time progresses. In this context, a librarian
has to frame a policy with regard to the regular depletion of material not much used. In fact, obsolescence studies can provide some solution to this problem. There seems to be four major reasons for this (26):

(i) the information is still valid, but is now incorporated into latter work,
(ii) the information is valid, but is now superseded by latter work,
(iii) the information is valid, but is now in a field of declining interest,
(iv) the information is no longer valid.

These factors, as stressed by Line, can of course be reversed and he suggests three other possibilities:

(i) the information is considered invalid but becomes recognised as valid,
(ii) the information is valid, but inadequate theory of technology delays its exploitation,
(iii) the information is valid, and in a field of unceasing or renewed interest.

Thus published material will change in terms of utility over time, and we must remember that obsolescence, although a general trend, can suddenly be reversed by fashion or technological change. In order to assess obsolescence of a paper we need some measure of its utility. The two major measures of utility are
(i) Citations or references in subsequent literature
(ii) Usage in a library

Assuming that we accept one or the other of our measures the results obtainable are interesting. Papers have a high utility during a short period of time after publication, and then their utility diminishes rapidly as time progresses. Obviously the speed at which the papers obsolescence will vary between subject fields. In order to compare the speed of decay in differing subjects the 'half-life' is used as a measure. A half-life refers to the time taken for half the currently published information to obsolescence. An alternative working definition is the time during which a half of the currently active literature in an area was published. Although the half-lives from an individual study appear to be accurate figures, comparison between various studies suggest a range of possible figures rather than an absolute one. Much of the variation will be due to the problems of citation and usage counting. However, we do have a range of half-lives which can give reasonable indications of the sort of time scale involved for each subject area. These time scales are essential in planning library holdings policy and dissemination techniques. It is important to use them only as indicators, and to realise that each library is unique and
living. Decay and growth is continuous while to study it we are freezing for a moment of time this continuum.

(5) History of Speciality: The citation analysis can be used as an effective tool for creating accurate historical description of scientific fields or scientific problems within a scientific field. It is generally observed that in the initial stage of development of any speciality, only a few scientists from a certain group are likely to participate. Consequently the past literature available with researchers or scientists for citations will be less and to a large extent, to group's own work. So, initially self-citation will be very large and at the same time half-life and average number of citation per paper will be very small. With the passage of time, more scientists will be attracted towards the speciality, more group of scientists will participate and as a result, citation to inter-group will increase; more past literature will be available for citation. This will lead in turn to decrease in self-citation and at the same time increase in half-life and average number of citations per article. During the process of development of a speciality, a stage is always achieved, after which half-life refuses to increase and becomes constant. This stage is called saturation stage in the life of speciality, when it is fully matured and established in terms of its growth.
To sum up it can be said that through citation analysis, it is possible to identify the initial germinating source or a particular speciality and through net-work of its development over a period of time, we can trace the course followed by speciality. If this analysis is coupled with impact analysis one could even work out the pre-conditions necessary for a speciality to obtain an independent status.

(6) Mapping of Science: The emergence of co-citation techniques has provided clues to the understanding of mechanism of speciality development and mechanism of science. In contrast to other measures, co-citation is the measure of the degree of relationship between papers as observed by practising scientists. Unlike others, a number of procedures can be used here to 'normalise' this measure of relationship (that is, to eliminate the effects of sheer size), and to convert what is a measure of proximity or closeness to a measure of 'distance'. The latter step enables one to talk about the 'space' occupied by science and its characteristics (for example, its dimensionality and leading directly to the notion of mapping in science). The idea of 'mapping' science is, by identifying key papers and events through citation analysis. The basic unit of analysis in mapping is the highly cited document. The fact that some documents have been highly cited within a specified time-period confers upon them
a special status as providing important 'ideas' in their respective areas and specialities. It should be possible to identify the corresponding cognitive contents to each highly cited item either by examining the citing context or by querying the cited author. If the analysis of the 'key' papers is compared with the development of new ideas we may be able to trace the origin of scientific revolution. The map of scientific papers would help us in clearly identifying the papers belonging to 'normal science' and those that create scientific revolution as well as their backward origin, i.e. the source that have led to these papers.

(7) Communication Patterns: One of the fundamental problems in scientific communication concerns the way in which the scientists interact with their colleagues in the production and dissemination of research results. One of the methods of investigation of such mutual influence between scientists usually linked in a network would be from the citation analysis of papers they write. The word network here in fact connotes both interconnections between scientists and between the papers they write. This usage, however, corresponds to a particular picture of scientific communication. Each Scientist is envisaged as a node from whom lines of communication run, linking him more or less strongly with his peers, with the number of such lines varying from scientist to scientist. A major impetus for
studies of models of this type was provided by Price, who argued for the existence of 'Invisible Colleges'. The term invisible college has come into use pertaining to a particular core of people who by continual professional intercommunication, exchanging of working papers, data, etc. constitute an information processing machine within a discipline. Within a sufficiently small field of activity, such an invisible college can come close to satisfying the bulk of the current information requirements of the group. People within an invisible college have a strong tendency to cite each other's work. Each invisible college has its neighbouring closely related other invisible colleges, and citation tracing can be used to show such patterns of relationship. The various techniques of citation analysis like direct citation, bibliographical coupling, and co-citation, can be used to study the communication patterns in science. A study of the network of linkages and spatial location of clusters (invisible colleges) co-citing each other, one can work out the probability of a paper of one cluster being cited by the other cluster.

(8) Evaluation of Efficiency of a Scientist: There has been a number of attempts to evaluate the productivity of a scientist on the basis of citations received by his papers or work. One of the first studies supporting the above evidence was done by a psychologist who asked
a panel of experts to test the psychologists who made the most significant contribution in the field. The measures which correlated most highly with that of panel was the number of journal citations to the scientist's work (27). The evaluation of the quality of the paper, its impact through citation analysis is a handy tool and sufficiently reliable.

To sum up it can be stated that citation analysis as a quantitative technique to study the development of science is a very useful one. This will help in giving proper allocation of the funds to various subjects in different libraries. The principle of citation is that we are required to be noticed by our contemporaries, only if we have contributed to knowledge for others to notice us. Such writings as are not referred to by others are in extreme instances, as good as garbage. On the same analogy, libraries should stock such documents as are consulted. The Trueswell's Weeding Technique advocates the weeding of low-circulation and low-use documents from libraries. On the same principle, the test of citation can be applied to ordering periodical subscriptions, so that such periodical publications need not be ordered by libraries. It is undoubtedly true that substantial proportion of periodicals are ordered for record only. That the libraries build for the future is a ready answer given to sceptics. The need for pruning of periodical lists is bound to arise.
in the context of cost escalation. Who is reading what and when is a most relevant question to be asked?

Citation is one technique of finding out about the use of documents in libraries. The results derived from citation technique may be cross-checked by applying other techniques. Since citation analysis has not been tested in the Indian context and there are meagre data about its use in linguistic research, it needs to be tested on our situation.
Bibliographical References


2. Ibid., p. 2092.


5. Weinstock, Melvin. Citation indexes. In Encyclopedia of library and information science, V. 5, p. 16-40.


