CHAPTER I

1. Introduction

Science denotes the body of verified and valid knowledge and the systematic generation of new knowledge. It may be viewed as a systematic, objective, deliberate and controlled search for insights and accurate knowledge about a phenomenon. Science is a process which evolves new pieces of information. Scientific information generated in a discipline is an index of the scientific research carried out in that area. Scientific knowledge yields social benefits when made freely available to all potential users. Scientific research is incomplete unless the findings of the scientists are communicated to others in the form of scholarly communication.

Scholarly communication is exceedingly important to the understanding of the genesis and evolution of disciplines. Within and outside the academy, disciplines have emerged in diverse fields of cultural and intellectual endeavors. Since the advent of organized academic activities, disciplines have become spheres of individual and collective activity in science, technology, the humanities, and the social sciences. Societies, specialized research institutes, university departments, and conferences have emerged attesting to the continuous
social formation of these activities. Often, publications include both monographs and journals, yet increasingly journals constitute the majority of published scholarship. Journals emerge as vehicles of communication between scholars and scientists who, through formal and informal acculturation, accept disciplinary consensus as expressed in the journals’ pages. This implicit acceptance of the major tenets of a given discipline permits communities of like-minded scholars and researchers to emerge.

Scientific communication is the only link between scientists in various parts of the globe. The recorded form of the scientific communication is scientific literature in that discipline. An objective analysis of scientific literature output in a discipline is considered a necessary prerequisite for an adequate and clear understanding of the nature and structure of scientific research carried out in that area. Research in a specific subject field results in literature output which is latent with a number of bibliometric indicators.

Bibliometrics is a branch of Library and Information Science. Bibliometric analysis counts up on bibliographic information on literature output. The periodical article is one of the attributes to research and development in various disciplines. Without any exemption the information scene in every discipline has been suffering from a major barrier known as “information explosion”, which is taken to be a synonym to knowledge explosion.
1.1 Knowledge Explosion

The development of new knowledge has in the last few decades far outstripped the rate of development than all of the preceding centuries. This trend is marked by the term `Knowledge Explosion' otherwise known as `Information Explosion'. This has its impact on the information storage and retrieval system in the process of scientific communication. The situation is immeasurably complicated by the fact that the book is no longer the primary medium of scientific communication. The conventional form of document has been gradually replaced mostly by scientific journals.

1.2 The Scientific Journal

In the beginning was the word, says the Bible. The printed word is the basic unit for Education and communication. The conventional Book is an evidence of human cultural history of social knowledge. The printed word and paper have been associated with human history for a longer period than any other concept. The development of printing technology could reduce the time-lag between writing and publishing only to some extent and an alternative to the book was felt as a must by the scientific community. `Necessity is the mother of invention' is an old adage. An appropriate solution was found and that was the Journal.
The scientific journal made its maiden appearance during the early 17\textsuperscript{th} century in Great Britain and France simultaneously. The dawn of the modern journal started with publishing of the `Philosophical Transactions of the Royal Society, London’ and `Journal de Scavans’ from France contemporarily during the year 1605. Since then periodicals multiplied in quantum of publications with research articles as their fundamental unit of information.

The refereeing procedure and bibliography of citations appended to each research article at its end have come to stand as the in-built mechanism as well as quality measuring technique of research output. It took centuries for concretising an idea that the scientific output can be analysed for meaningful results. Feedback from the concerned scientific community to any research output provided an evaluative process leading to the establishment of truths in every field. Library and Information Scientists came to the scene studying the quantum and quality of literature output based on the bibliographic apparatus. Establishment of standards for the compilation of a bibliography paved the way. Each component field included in a record of the compiled bibliography when counted and analysed provided fruitful inferences.

It was Bradford who shaped the analytical study into a subject though a few precursors had done some work including the famous Science Historian J.D.Bernal. There had been continuous analysis of the
contributions of scientific literature output and their use in the regeneration process. Milestones had been created by Bradford, Zipf and Lotka and Bibliometrics originated.

According to S.R. Ranganathan, literary warrant is one of the main attributes for the birth of a new discipline. Unaware of this fact, they were working towards the formation of a new branch in library and information science called Bibliometrics. Bibliometrics covered under the umbrella region library and information science, is one of the component fields of study forming part of the agglomeration 'Social Sciences'. When compared to various other disciplines, Library and Information Science is one which has been christened and baptized as a subject of study in recent decades though it has got a history of its own equal to that of the Human cultural history.

1.3 Literary Warrant and Disciplines

Communication is culture. Library is one of the forms of mass media communication and as such is considered to be a vital component of culture. Library and information science is one of the very few subjects such as Education, Philosophy, Epistemology and Cybernetics that deals with the universe of knowledge as the core subject of study. It is acclaimed to be a science of sciences. Today, literature output involved in the communication of science has been at wuthering heights.
“The sum total of information in a field of study is known as its literature.” The literature in every branch of science is growing rapidly and now accounts for several millions publications in each field. Proliferation of literature from each and every country, in every subject and in various languages leading to problems of information organisation and retrieval is marked by the term ‘information explosion’ posing challenges to the measures of bibliographic control. The situation is reflected in the enormity of surrogates covered in the published secondary sources such as abstracts, indexes and reviews in various subjects and those secondary sources include various kinds of bibliographies.

1.4 From Bibliography To Bibliometrics And Scientometrics

Bibliography, as a generic term, ‘is an exhaustive list of reading materials contributed by an author, or on a particular person or in a micro subject’. It has come a long way in the history of libraries and knowledge organisation. Today, published and priced bibliographies have taken printed, electronic and online forms which provide a bibliographic control mechanism to library professionals as well as scholars. Bibliographic information apparatus is a wonderful tool which, when analysed, reveals a spectrum of results on the behaviour of literature. Bibliography is not a librarian's tool exclusively. Information on scholarly publications in the form of bibliography is necessarily found appended at the end of each article and that gives the clue to a scholar for the furtherance of his/her
information journey. Bibliography becomes a scholar's instrument unfolding the treasure of literature to his community amidst the chaos and confusion arising out of information explosion.

The impact of Information explosion has been great and a few decades ago, it was widely felt among the librarians that the profession was at crossroads not knowing the direction to proceed to find solutions to the barriers to scientific communication. Any effort to provide a bibliography to a client became futile due to the exorbitant cost and widely scattered nature of information. Manually prepared bibliography provided the mechanism, to some extent, to have a control over the literature output with a smaller amount of effectiveness.

The advent of computers into libraries changed the scene. It did not take much for the librarian to realise that the solution was very much with the electronic publications rather than the printed hardcopies of reading materials. And also the bibliographic details found in the citations give a leading clue for the study of the personality in the subject field and problems of identification also. Librarians began to look into the bibliographic compilations for analysis in order to have an effective bibliographic control. It started with citations and later encompassed bibliographies in large measures.
Citation analysis retains the identity of individual article in the context of its citations included in its bibliography. While citation analysis is restricted to each article and the bibliography of citations appended to it, the analysis of bibliographic records is known as Bibliometrics.

Bibliometric studies reveal a wide range of results leading to inferences regarding even the future Nobel laureates. Such studies dealing in depth of bibliographic apparatus led to the evolution of a field called "Bibliometrics".

1.5 Bibliometrics: Scope

Bibliometrics is the technical name for a range of analytical methods using information-quo-information found organised in bibliographic description of reading materials such as books, periodical articles, reports, patents, software, designs, prototypes, and blueprints to develop descriptive statistics, multidimensional analyses, and graphical representations of the output of science. Bibliometrics is often used to

- Clarify and assist in the analysis and formulation of science policy by highlighting the networks of players or subjects that make up scientific research;
- Providing strategic analysis of the relative position of research performers;
- Sketching profiles of the activities and performance of individual centers;
Graphically presenting studies of strategic or innovative subjects.

- Publication counts, involving counting of scientific publications published by a researcher or a research group;
- Citation counts, involving identifying the number of times a specific article is cited in other scientific journal publications; these measures are considered to address questions of quality, influence, and the transfer of knowledge;
- Co-citation analysis, identifying pairs or groups of articles that are cited together in other articles or publications; from these pairs or groups of articles a 'cognitive structure' can be derived, providing information on the direction and flow of scientific thought;
- Co-word analysis, involving assigning keywords to a paper or article by a professional reader; papers which have the same keywords and sets of words are linked to each other via a clustering technique;
- Scientific mapping, involving developing a visual model or 'map' of the realm of scientific fields representing the structure of literature output of particular scientific fields."

Bibliometric studies fall mainly into two broad groups -- those describing the characteristics or features of literature (descriptive studies), and those examining the relationships formed between components of literature (behavioural studies, sometimes referred to as citations studies, but not restricted to them).

The descriptive studies are probably the most easily understood, and can be thought of as resembling in many ways a population census. Both have to provide comprehensive and accurate data about...
changing population in one case of people and in other of literature. Literature descriptions are assembled from the bibliographic study those that provide data on the condition or character of the literature as a whole. There is, therefore, an important distinction between bibliographic and bibliometric data.

1.6 Empirical Laws of Bibliometrics

The three fundamental laws which laid the formation of bibliometrics are

a. Bradford's law of scattering of scientific papers
b. Zipf's law of word occurrence
c. Lotka's inverse square law of scientific productivity

1.7 Bradford's Law of Scattering

Bradford\(^1\) revealed a pattern of how literature in a subject is distributed in journals. “If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several other groups of zones as the nucleus.”

\(^1\) Bradford, S C. Sources of Information of Specific Subjects. Engineering. Vol.137. 1934. 85-86
Bradford's formula.

\[ R(n) = \frac{N \log n}{s} \quad (1 \leq n \leq N) \]

Where \( R(n) \) = Cumulative total of items contributed by the sources of rank 1 to n.

\( N \) = the total number of contributing sources.

\( S \) = constant characteristic of the literature. We then can say that

\[ R(N) = \frac{N \log N}{S} \] is the total number of items contributed by N sources.

In other words Bradford\(^2\) describes that "If all the scientific journals relevant to a given subject are assigned ranks according to the number of articles they carry on the subject, are assigned ranks according to the number of the articles and the journals are grouped into m zones each containing the same number of articles, with the first zone containing the highest ranking journals and the last zone containing the lowest ranking journals, then the number of journals in the succeeding zones form a geometric series with a common ratio \( P_m \) depends only on m.

The statistical regularity pointed out by Bradford's law provides an objective means of determining zones of relative richness or value to a

\[^2\text{Bradford, 1934.} \]
given kind of literature. This has implications to the various processes in a library. For example, a library can safely stock the journals which belong to the core or nuclear zone. It is advisable to extend the purchase list to the next zones till the budget limits permit. If at all the library budget is elastic, a point will be reached at which it would be desirable to obtain copies of articles in the journals on demand rather than subscribing to the journal.

Lancaster\textsuperscript{3} provides an excellent hypothetical example of applying Bradford's law in periodical collection building, while discussing the principle of diminishing returns. Brookes\textsuperscript{4} is of the view that if the total expenditure on periodical provision is limited to the fraction ‘f’ of the sum needed to cover the subject completely, the buying of periodicals may be supplemented by the buying of photocopies of the relatively few relevant papers published in the peripheral periodicals. While preparing bibliographies we are faced with the problem of coverage, the journals that are to be scanned etc. size of a bibliography and the periodicals that should necessarily be included in the list of items to be covered. On the application side of this Bradford's distribution can be fruitfully used to

\begin{flushright}
\end{flushright}
estimate the total law, the studies of Goffman and Morris and I.K. Ravichandra Rao\textsuperscript{5} are significant.

Goffman and Morris\textsuperscript{6} found that the pattern of journal usage in the Allen Memorial Library follows a Bradford distribution. Ravichandra Rao, through his analysis of circulation data collected from six Canadian University Libraries, proved that the rank distribution of transactions follow a Bradford distribution. Bradford's law is very much akin to the Pareto's law relating to wealth distribution and the 80-20 principle used in warehouse management and the Mandelbrot distribution.

Naturally this law is applied to study not only the scattering of publications, but also in other spheres of activity also. A study conducted by Garg and Lalitha Sharma\textsuperscript{7} of R & D indicators in Indian Industry using Bradford's law bears testimony to this fact. By analysing the R & D expenditure of 452 in-house R & D units in different sectors of the Indian Industry, they had identified 19 in-house R & D houses as the core, 60 as the medium and the rest as small. As compared to medium and small-level in-house R & D units, there is a heavy concentration of

Performing a Bradford Analysis

1. should start with some subject area of interest (e.g. citation analysis).
2. locate all articles that cover the subject during some time period (e.g., 1990-2000). Using a database with comprehensive coverage like Library Literature, can speed things up.
3. record the number of relevant articles found as a function of the journals names where the articles originated.
4. summarize the data by counting how many journals have published x relevant articles (e.g. One journal PUBLISHED 5 articles; One journal published 4 articles; Two journals published 2 articles each; 12 journals published 1 article each).
5. rank the results in decreasing order with the most prolific journal given rank 1.
6. arbitrarily select a core of j journals. The number of articles found in the core j journals is denoted as a zone. Denote the first grouping (i.e., the core) as zone 0. (e.g., select a core of 3
journals (j=3) which will have 12 articles associated with it (a=12)).

7. divide the remaining data into "Bradford Zones" such that each zone contains a articles (12 articles), dividing the data into Bradford Zones.

Bradford’s law states that there is some constant k such that the zth zone containing articles consists of kj journals where the initial, or core zone, is the 0th zone. K is also known as the "Bradford multiplier" and is determined by the data. The proportion k0:k1:k2:k3:... is another view of the same phenomenon. To proceed further, one has to determine the value of k: consider the following:

Zone 0 contains 3 journals

Zone 1 contains 6 journals

Zone 2 contains 12 journals

Thus these data contain the proportion 3:6:12 = 1:2:4 = 20:21:22

Therefore, k = 2.

Apply the kj formula to the data. Now the number of journals needed to provide multiples of the cores quickly grows to very large numbers. This is a classic example of the scattering effect i.e
Bradford's law of scattering a measure of the rate of obsolescence by distinguishing the usage of the levels of items. Essentially, this is a method of clustering.

For example, 9 journals have 429 articles, the next 59 have 499, and the last 258 have 404. We roughly get three groupings (ranging from 404 to 499) of articles. Three groupings, 9 titles, 9*5 titles, 9*5*5 titles.

1.8 Bibliographic Scattering

Taking a topic or a subject, arbitrarily chosen, identifying characteristics of an article, the scattering or distribution is shown to follow a clustering process of ranked groups, first noted by Bradford. Bradford's law is almost universal.

Quentin L. Burrel, Statistical laboratory, Dept. of Mathematics, University of Manchester in his article "On the growth of Bibliographies with the time: An exercise in Bibliometrics prediction" says,

Bradford sought ways of providing an efficient, service for abstracting and indexing scientific and technical literatures. As is surely well known Bradford's investigation concentrated on the productivity of journals more particularly he sought to identify those
journals which were relevant in the sense of publishing articles in a particular subject area during the period of study. His empirical studies gave rise to the Bradford law of bibliometrics.

In Bradford's content we have a population of academic journals and in compiling a bibliography we wish to identify those journals which produce articles relevant to our field of interest and the number of articles so provided by each. While a no of authors have concentrated on the importance of having as nearly as possible a complete search of the possible sources, an important practical point raised by Bradford is that 'even when the actual producer during a period of years had been ascertained, new sources would certainly appear during a further period'.

1.9 Models of Bradford's Law

Mathematical expressions or models have been suggested that reflect the graphical relationship observed by Bradford. These models can be classified as

1. empirical

2. causal and

3. formal
Empirical models

Empirical models are only justified a `posteriori’ by their ability to give a good fit to sufficient and varied data.

Causal models

The mathematical relationship is usually derived from some known underlying causative principles characterizing the process. Causal models are explanatory in nature, and may be either deterministic or stochastic.

Formal models

Formal models may be used in situations where the underlying process is too complex or dependent on too many factors to be expressed as a simple causal relationship. Some examples of the three categories of models are the following

Egghe\(^8\) (1985) A generation of Leimkuhlers\(^9\) model, it hinges on a mathematical interpretation of Bradford's law.


Burrell: (1980), This is a model of library circulation based on a stochastic process and could be classed as a causal model.

Basu\textsuperscript{10} (1992,1993), A formal model of the most probable unequal distribution of a set of articles in a set of journals, called the Random Hierarchical model.

1.10 Zipf's Law

In the field of 'Language' studies statistics on word frequencies have been compiled for over a century. The first major work is by Dewey - Relative frequency of English words published in 1923. The major effort at statistical analysis of word frequencies was made by zipf's in the psychology of language published in 1935. Striking regularities in the frequencies distributions of words number of the different words $W(k)$ occurring $k$ times were noted in different language, authors and types of literature. In a very elaborate subsequent work Zipf extended the investigation beyond language in economics, geography etc established a universal law of frequency distribution encompassing a wide diversity of statistical data. Zipf also propounded a principle of least effort as a rationale for the rank-frequency distribution.

Zipf had developed and extended an empirical law, governing a relation between the rank of a word and the frequency of its appearance in a long text. If "r" is a rank of a word and 'f' is its frequency then Zipf's law is stated as follows:

\[ rf = c \]

Where 'c' is a constant. Zipf derived his law from a general principle of least effort. Words whose cost of wage is small or whose transmitted demands the least effort are frequently used in a large text.

It states that in a long textual matter, if the words are arranged in their decreasing order of frequency then the rank of any given word of the text will be inversely proportional to the frequency of occurrence of the word.

Mathematically, Zipf's law can be written as

\[ r = \frac{1}{f} \quad \text{or} \quad rf = c; \]

Where 'r' denotes rank word, 'f' stands for frequency of occurrence of the word and 'c' is a constant. Zipf's law can be effectively used in the generation of semi-automatic or automatic indexes useful for an information retrieval system. Its use has increased tremendously with the emergence of natural language indexing of textual matter especially in electronic form. Several studies aimed at finding out the pattern of
frequency distribution of descriptors of a thesaurus and the distribution of indexing terms are available.

A prominent one among them is that of Fedorowicz\textsuperscript{11} Zunde and Slamecka\textsuperscript{12} have developed a function for the optimum distribution of indexing terms by the number of postings. This should make it possible to transmit information with maximum efficiency. Zipf's law provides a measure of the richness in vocabulary of an author. This technique can be used for deciding the correct authorship of disputed works.

For example, if there is difference of opinion as to the correct author of a work, the work predilections of the attributed authors can be analysed either manually or using a computer. Once the frequency of occurrence of favourite words are decided the disputed text can be analysed to see similarity and thereby decide the author conclusively.

The law is also used for identifying words more frequently used in different foreign languages. These words are taught first in the instructional programmes of foreign languages. Emilie C White\textsuperscript{13} observes that the super-imposition of the Bradford distribution over the


linear Zipf distribution, which demonstrates the emergence of more used and popular items may yield a technique to describe the pattern of books used by library patrons. She feels that this law, when applied to circulation data, these formulations can support such policies as shortened loan periods for heavily used books and the identification of a core collection.

For analysis, this can be applied by counting all of the words in a document excluding the stop-words (a, an, the, therefore, etc.) with the most frequent occurrences representing the subject matter of the document. We could also use relative frequency instead of absolute frequency to determine when a new word is entering a vocabulary.

Naranan\textsuperscript{14} has shown some striking similarities to Zipf's Law in any democrative election in which two or more candidate stand, the candidates are ranked in order of decreasing frequency of the votes they individually individually attract. In this example the candidates are the sources and the votes are the items.

1.11 Lotka’s law

Lotka’s\textsuperscript{15} proposition led to a whole gamut of studies on scientific productivity. Such productivity studies have gained momentum in the post-second world war period. This in fact, has culminated in the rise of a new discipline called Scientometrics. It is defined as the `study of the measurement of scientific and technological progress’. It provides an understanding of the structure of scientific activity, the disciplines being researched, the organisations involved, the strength and deficiency in the scientific groups and their communication channels and at different levels of aggregation. It follows a trajectory of econometrics in the use of quantitative data, concepts and models and extensive use of mathematical and statistical techniques of modelling and data analysis.

Scientific productivity studies have been made from different angles. Impact of social change on scientific productivity, relationship of publication output on scientific recognition, identification of elites in different disciplines, occurrence of discoveries in different cultures etc. are some of the approaches made in this line.

An interesting study of scientific productivity made three decades back is that of Yuasa\textsuperscript{16} Yuasa’s was a statistical study of the scientific achievements in various countries that showed the shifting of the world scientific dominance from one country to another. He found out that this dominance shifted from Italy to Britain, then to France, from France to Germany and finally to USA in the 20th Century. Price\textsuperscript{17} who had traced the development of science in Babylon and plotted the growth of big science from little science had observed that Lotka’s law applied equally well to the productivity of scientists in the 17\textsuperscript{th} as well as in the 20\textsuperscript{th} century. This meant that majority of publications emanated from a handful of people. This is very much similar to the already existing Price’s square root law. Similarly, the conclusion of an extensive review of early studies of scientific productivity made by Narain was that scientific talent was highly concentrated in a limited number of individuals.

Generally, Lotka’s law is an inverse square law that for every 10 authors contributing one article, 25 will contribute two; 11 will contribute three, and 6 will contribute four each. There is general decrease in performance among a body of authors following 1:n:n\textsuperscript{2}, this ratio shows that some produce much more than the average which seems agreeably true for all kinds of content creation.


Thus, we can say that Lotka's law may not be constant in value, but in following Inverse Square. Our challenge will then be to find the correct square in different mediums & fields. Other authors like the Spanish scientist A Lara, the Russian, Marschakova the German, D Schmidmayer, the Belorussian V Lazarev, the Lithuanian, O Voveriene to name a few proposed their own solutions, defining the limits of competence of these three-metrics.

1.12 The Logarithmic Law

Analysis of the data form many social contents over the past two years has led to the discovery that the general form of the flr distribution is logarithmic. Two special cases of this logarithmic law were first noted in 1935 one by Zipf's in the statistics of natural language vocabularies and one by Bradford in the distribution of scientific papers relating some specified topic over the contributing serials.

The law states that in a homogeneous activity the sources of the sources of rank 'r' is given by \( g(r) = K \log (a+r) - \log (a+r-1) \)

Where \( K \) is a constant dependent on the sample size. The cumulative activity of the first \( r \) sources is given by \( G(r) = q(r) = K \log (a+r) \)
So, the conformity with the law is most easily shown by plotting the successive values of \( G(r) \) for \( r = 1,2,3... \) in succession against \( \log (a+r) \) - which is easily done if semi-logarithmic graph paper is used. If the cumulative plot is linear then the data conform to the log law. The log-law is simply a mixed Poisson distribution of a regular kind. It can be derived theoretically by considering an infinite population of mixed abilities in a very natural way.

Because social activities are susceptible to social regulation within the group, the pure logarithmic form of the law is relatively rarely found in empirical data. But the variant forms are easily recognized and, of course, are of special interest because they reflect social interactions with the group.

1.13 The Social Log Law and the Individual

The social distribution of judgments is logarithmic. And so, by analogy with the Gaussian model, it is reasonable to consider the hypothesis that the individual distribution is also logarithmic.

A simple model to account for Bradford's law was proposed by Naranan\(^{18}\) in 1970. The most significant first step in the model was to recognize the Bradford's law is equivalent to a simple power law

distribution of articles in journals. Specifically $J(p)$ the number of journals carrying exactly $p$ articles is of the form of $J(p) = Kp$ - ($K$ being constant and $p = 270$) to explain the power law relation. Naranan was inspired by a model proposed by Fermi in 1949 in cosmic ray astrophysics.

Naranan investigated statistical laws governing distribution of number of citations to articles, citation to journals etc. and found them also to be power laws similar to $J(p) = Kp$.

It has been found from the earlier studies that collaboration in research varies from discipline to discipline and for the same discipline from time to time and from one country to another collaborative research has also been found to be influenced by a number of factors such as funding use of large-scale equipments etc. However, the general conclusion of many a study has been that multiple authorship (i.e. collaborative research) is on the increase in science and technology.

Bibliometrics, as any other branch has been growing in quantum and quality over the years and has given birth to new branches such as Informetrics and scientometrics. The quantum of literature output in various disciplines is more from developed countries.

1.14 Scientometrics

The term Scientometrics originated as a Russian term for the application of quantitative methods to the history of science, but its scope
and objectives have widened considerably. Scientometrics is a wide-ranging field with vague boundaries. It is a generic term for a system of knowledge, which endeavors to study the Scientific and Technological system, using a variety of approaches within the areas of Science and Technology Studies (STS).

According to Pouris, "Scientometrics is for science what econometrics is for economics". Both disciplines attempt to study social phenomena with the rigour provided by the scientific method. The founding father of Scientometrics, de Solla Price, had envisioned Scientometrics to become a hard science like econometrics. This vision is becoming true. Scientometrics followed the trajectory of econometrics in the use of quantitative data, concepts and models and extensive use of mathematical and statistical techniques of modeling and data analysis.

In terms of methodology, Scientometrics techniques can be classified into two categories: one-dimensional (or scalar) and two-dimensional (or relational) techniques. One-dimensional techniques are based on direct counts (or occurrences) and graphical representation of specific bibliometric entities (e.g., publications and patents) or particular data-elements in these items such as citations, keywords or addresses.

One-dimensional techniques are used to generate scalar indicators for monitoring the state-of-the-S&T system. Scalar indicators are increasingly being exploited for science policy purposes—both as descriptive and diagnostic tools.

The two-dimensional techniques are based on co-occurrences of specific data-elements such as number of times the keywords, classification codes, citations and addresses are mentioned together. Multi-dimensional statistical techniques—continuous and discontinuous—are being used to represent scientific universes. Discontinuous methods such as classifications are widely used to scale down the level of observations throughout the fractal structure of science and technology networks. Continuous methods, such as multidimensional scaling, allow the representation of the relational (or structural) features of the data in the forms of ‘maps’. Keyword or classifications codes are content indicators of the knowledge conveyed of the documents. Clusters are indicators of the topics or centers of interest contained in the documents. Maps are strategic indicators of the relevant positions of the topics in the knowledge space covered by the documents.

### 1.15 Scientometrics For Decision Makers

Scientometrics can be used for

- Monitoring the health of science
1.16 Health Of Science

A study commissioned by the Royal Society, London, indicates that the concept of ‘Health of science’ is in practice reduced to measuring the performance and identifying the national strengths and weakness on a comparative basis. This essentially involves benchmarking of the S&T system. A group of different countries can be selected as comparators. Here some crucial issues are: Is a country doing more or less research in a particular field or subfield as compared to other as compared to other nations? Is it doing better than others” is it doing more research in a particular field compared to some other field, or it is doing better in one field compared to another?

The strengths and weakness of a country or a region by computing Activity Index (AI) and Attractivity Index (AAI) (or Citability Index) from the counts of publications and citations (or impact factors of journals (after normalisation) in which the articles are published).

\[ AI = \frac{\text{The country’s shared in world’s publications output in a given field}}{\text{The country’s shared in world’s publications output in all science fields}} \]
AAI = The country’s shared in attracted by publications in a given field. The country’s shared in attracted by publications in all science fields

To assess the research performance in a worldwide perspective, it is not sufficient to use indicators based on publication or citation counts. The phase of the development of scientific fields should also take in to account. The development phase can be classified three categories, depending upon the rate of increase in publication activity:

Hot Topics: Significant increase in publication output worldwide

Cold Topics: Significant decrease in publication output worldwide

Stable Topics: No Significant increase or decrease in publication output worldwide

In general, if a country publishes much below the world average on a hot topic, it has two alternative implications:

- The country has failed to pick up new developments, and it would be worthwhile to investigate the reasons for this.
- The country has wisely resisted jumping the bandwagon on a passing fad.

For stable topics, equal or above world average activity is a potential sign of healthy development, while significantly lower activity indicates a weakness. However, a strong increase over a time in a
country’s publications in on a stable topic may indicate that either a new valuable development has been initiated (which should show up in a relatively high impact scores in citation analysis) or that too much effort is committed to the run-of-the-mill research.

For cold topics, a significantly higher activity indicates that a country is putting too much effort on a topic, where scientific payoff is ‘lean’. Occasionally, a higher activity signifies a breakthrough, which should show up in a relatively high citation rates.

The health of science also depends upon transience and renewal of the scientific population.

1.17 Authors Of Publications

The population of authors can be divided into two groups: transients and continuants. Transients are authors who publish only one paper in their lifetime and then disappear from the public arena of scientific research. A certain amount of transience is a natural concomitance of scientific research, but an extreme proportion of transients is unhealthy.

Since the calculation of the precise extent of transience for a large population of authors is quite difficult, an approximate indicator of
transience \( (q_t) \) can be constructed by choosing a thick slice of time, say 5 to 7 years.

\( q_t = \) Percentage of authors publishing only one paper during the time slice.

Renewal of scientific population can be expressed by the ratio of birth and death rates of the population of authors, i.e., by the ratio of new coming authors and those who have finished their scientific career.

Renewal indicator \( q_r = \) birth rate/death rate

The population is stationary if \( q_r = 1 \),

it is renewing if \( q_r > 1 \) and

dies out if \( q_r < 1 \).
References