Chapter-4
Scientometrics and Related Aspects

4.1. Introduction
This chapter has discussed in detail about scientometrics and its related aspects. In recent years Library and Information Science has evolved some quantitative measures of study. Bibliometrics, Scientometrics, Webometrics, Informetrics, etc are such fields.

Bibliometrics
Bibliometrics was developed by Alan Pritchard (1969) by naming it as Statistical Bibliography. It is the quantitative measurement of recorded knowledge of different aspects. Bibliometrics covers quantitative analysis of knowledge irrespective of field. It helps in counting and analysing to get to the fact of records.

Scientometrics
It is a set of mathematical and statistical procedures to measure scientific publications and science research. In practice, by using Scientometrics, different metrics studies like publication counting, author ranking, journal ranking, university ranking, citation analysis, etc, are studied to visualise certain inherent characteristics and relationship among different attributes of scholarly publications. Similar scientific fields are Bibliometrics, Informetrics, Webometrics, Virtual ethnography, and Web mining. Scientometrics helps to know the growth pattern, discipline wise growth, total productivity, citation pattern, media of publications, collaboration nature, and many more characteristics and patters of scholarly publication.

Librametrics
This metrics was developed by Dr. S. R. Ranganathan, the father of Library Science of India in the year 1948 [www.netugc.com/librametrics-bibliometrics-scientometrics-informetrics.html]. It was proposed as Librametry by Ranganathan, which was to study Library and Information
Science recorded knowledge and behaviour. This can be defined as the quantitative method used to study Library Science literature and activities. It is the oldest form of metrics study. Librametry was discussed in ASLIB conference in Lamington-Spa. The word ‘Libra’ means ‘library’ and metrics means ‘measurements’. By coining Librametry, Ranganathan meant the application of quantitative methods in library services and management. It helped Dr. Ranganathan for wider library management works such budget allocation for procurement of books and journals, deployment of staff, binding of periodicals, etc.

Webometrics

Webometrics is the study of web and web published literatures along with web based phenomenon. The term Webometrics was coined by Thomas Almind and Peter Ingwersen (1997). Webometrics deals with web construction literatures and web impact factor to assess the quality of web literatures. Webometrics deals with the study of growth of web.

Cybermetrics

It is similar field of metrics study as like as Webometrics i.e. assessing web literatures and web based data. But the specification of Cybermetrics is that it assess all the web related aspects including data like e-mail along with web literatures. This word cyber was first coined by Norbert Weiner and Cybermetrics was used from the shadow of that by William Gibson. (Sen, 2014). Generally it can be defined as the branch of metrics studies where by mathematical and statistical methods are used to measure productivity of web literatures.

Scientometrics as a term was first defined and suggested in 1969 by two Russian scholars, namely, Nalimov and Mulchinko (Sengupta, 1992). Accordingly, scientometrics is a complex of quantitative methods that are used to investigate and study science productivity and different phenomenon, product and derivatives of the process of scientific publications. But scientometrics was introduced and came into prominence with the founding of the journal called "Scientometrics", by T. Braun in 1977, originally published in Hungary and currently from Amsterdam (Netherlands). On the other hand, the term scientometrics is derived from the Russian 'naukometria' and was used mainly in the East and is defined as the study of the measurement of scientific and technological progress (Egghe, 1988).
4.2. History of Scientometrics

Scientometrics is the branch of Bibliometrics which deals with science literatures. Scientometrics can be defined using of bibliometrics techniques in order to study the relationship of different way of science communications. The term ‘Scientometrics’ was first coined by Vassily V. Nalimov in 1960’s.

Since Vassily V. Nalimov coined the term ‘Scientometrics’ in the 1960s. Actually at first, The Russian term ‘naukometriya’ was coined by Nalimov in the year 1969 which was first translated as ‘Science metrics’, and from that the new term Scientometrics was appeared this term has grown in popularity and is used to describe the study of science: growth, structure, interrelationships and productivity. Scientometrics is related to and has overlapping interests with bibliometrics and informetrics. The terms bibliometrics, scientometrics, and informetrics refer to component fields related to the study of the dynamics of a given discipline as reflected in the production of their literature.

Scientometrics has become a very important field of measuring metrics today. Both formal and informal scholarly communications and collaboration are done via this method. Taking cognizance of the fact that in order to get adapted with the contemporary information environment; librarianship today demands for the study of various parameters of the scholarly communications, and scientometrics, giving a fertile ground has become an important field.

4.3. Application of Scientometrics

As discussed above, scientometrics embraces under its umbrella a wide range of issues pertaining to research publications and/or communications in science. To do so, a Scientometrician carries out study both of quantitative and qualitative in nature applying different mathematical and statistical principles. Scientometric studies are pursued with different objectives in different contexts. If a given study is delve in to evaluate the pattern of growth (or decline) of research output on given discipline the output has the propensity to examine the policy in force for scientific research and suggest measures for tuning up the lacunas, if any. Likewise, analysis of different phenomenon in research productivity, publication characteristics tries to explore vital insight into the health of that subject and/or discipline. Study on the
productivity of a given organization is one of the most frequent appearance in the literature of scientometrics. Such studies harbours significant inputs for an organization towards bridging out the gap in research productivity, and thus have vital importance for science policy making. Similarly, study on the various characteristics of scientific productivity of an organization gives an array of scopes to the library mangers as well as the authorities concerned to identify to strength and weakness of library resource adequacy, services rendered, and may more to invigorate the library as a whole. So the application of scientometrics may be clubbed under the following headings:

1. Quantitative evaluation and growth of any basic scientific discipline.
2. Determination of the factors responsible for the increase or decrease in research activity in that area.
3. Assessment, examination and documentation of the organizational and management aspects of science at national, regional and at institutional level.
4. Determination of the productivity of scientists and the mechanics and the pattern of coherence of scientific creativity.
5. Documentation of the impact of research in different parameters, viz. Collaboration, Citation metrics, etc.
6. Measure the strength of library resources and services rendered to the researchers
7. Aid the research community to identify to most influential research outputs
8. Aid the research community to identify and evaluate the chain of various channels of publication of research finding towards reaching out to wide visibility of their publications.
9. Aid the research community to identify the researchers having similar in interest which has an immediate correlation towards establishing the network of collaboration in research.

4.4. Relation of Bibliometrics & Scientometrics

Bibliometrics and Scientometrics are two closely related approaches to measure scientific publications and science in general, respectively. Bibliometrics covers under its umbrella all types of studies on bibliographic data adopting a set of heuristic metrics which helped in
originating Scientometrics. Bibliometrics, as such, are used in quantitative evaluation of different bibliographic data irrespective of discipline and the source data belongs to, but when such a bibliometric study is conducted confining its data sources in the domain of science and applied sciences, it is clubbed under Scientometrics. As such, scientometrics are used only for analysis and evaluation of bibliographic data in the science. Basically it can be said that, Scientometrics is emerged from Bibliometrics and a subset of Bibliometrics.

4.5. Sources of Metric Data

Scopus
Scopus is the largest abstract and citation database of peer-reviewed literature by Elsevier. It was started in the year of 2004. Book Series, Journals and Trade Journals are the only three series covered by Scopus. Scopus also covers author profiles too. It covers more than 69 million of records (as on September, 2018). Scopus is subscription based database.
Scopus covers four different disciplines viz. Life Sciences, Social Sciences, Arts and Humanities, Physical Sciences and Health Sciences. The coverage of Scopus is worldwide.

Google Scholar
Google Scholar is the easiest indexing tool for simple searches as it is similar to the main Google search engine search page. It was started in 2004, November. It covers peer reviewed books, journals, theses, conference papers, dissertations, abstracts, technical reports, etc. Though there is no any formal count of records covered by Google Scholar is found, it is shown that almost 389 million records are covered by it (https://scholar.google.co.in/). Google Scholar doesn’t have the facility of advanced searching like Boolean searching as available in Scopus and Web of Science. But the main positivity of Google Scholar is that it includes informal scholarly communication like presentations and conference papers, as well. This facility is not available in other indexing databases. [https://scholar.google.co.in/ (accessed on 18.11.18)]

ISI Web of Science
The Web of Science (WoS) which was earlier known as Web of Knowledge (WoK) is the citation indexing site which is totally subscription based. It was originaly produced by Institute for Scientific Information (ISI) and currently maintained by Clarivate Analytics. Founded by
Eugene Garfield, one of the originators of many Bibliometric techniques, WoS allows a variety of search options and the ability to follow citations from article to article. WoS covers the indexing of sciences, Social Sciences and Humanities, and have recently expanded to include conference proceedings in addition to journal articles. [https://login.webofknowledge.com/ (accessed on 22.11.18)]

The Major limitation of WoS is that they don’t publish full text of reference article.

4.6. Laws of Metrics Study

Metric research concerns the application of different metric laws developed and applied in conducting different types of studies. Following is a discussion on three widely used fundamental laws which laid the formation of Bibliometrics and Scientometrics:

1. Bradford’s ‘Law of Scattering’
2. Lotka’s Law of ‘Scientific Productivity’ or Inverse Square Law
3. Zipf’s ‘Law of Ward Occurrence’

1. Bradford’s Law of Scattering

Samuel Clementic Bradford formulated this law of bibliometrics in 1934. This law is based on an investigation done by L. Jones in the Science Museum Library in the year 1933. Bradford's law elaborates how literature of a particular subject are scattered in different journals. It helps in determining the number of core journals in any given subject or domain of subject field. i.e., the journal which are containing most relevant articles come first and the less productive in the last then the journals will be grouped into three zones of relevant articles. However the number of journals in each zone will be increasing very rapidly. Journal in a single field can be divided into three parts, each containing the same number of article. Bradfords Law helps to determine the number of core journals in any given field. Bradford’s law states that journals in a single field can be divided into three zones,

- Zone 1: a core journal on the subject, relatively few in numbers that produces approximately one-third of all the articles.
Zone 2: a second zone, containing the same number of articles as the first, but a greater number of journals.

Zone 3: a third zone, containing the same number of articles as the second, but a still greater number of journals.

If the number of journals in the core to the first zone is considered as a constant ‘n’ and to the second zone the relationship is ‘n²’, then, Bradford expressed this relationship as 1: n: n².

Where, ‘1’ is the number of periodicals in the nucleus and ‘n’ is a multiplier.

In order to explain the Bradford’s Law, an empirical dataset of 1245 publications distributed to 215 journals in a given subject taken. The 1245 publications are divided into three zones according to the productivity of the journals in decreasing order. Accordingly, the first zone 7 journals contain 422 items, the second zone 35 journals contain 414 publications and the third zone contains 409 items shared by 168 journals. So for these the first zone is the nucleus zone as it is containing 7 periodicals, followed by 35 periodicals in second zone and 168 in the third zone. Then the zones arranged in geometric progression are:

7:35:173

So, the new series is
7: 7*5: 7*5*5 (approximately)

On substituting 5 = n (known as Bradford’s multiplier)
7: 7n:7n²

i.e. 1:n:n²

Hence, it is proved that Bradford’s Law is valid in this dataset.

2. Lotka’s Law of ‘Scientific Productivity’ or Inverse Square Law

In 1926, Alfred James Lotka, who was a Chemist, Demographer, Ecologist and Mathematician, founded his Inverse Square law correlating contributors of scientific papers to their number of contributions. It provides a fundamental theoretical base for bibliometric studies involving authorships. Lotka’s law describes the frequency of publication by authors in a given field. It states that the number of authors making n contributions is about 1/n² of those making one; and
the proportion of all contributors that make a single contribution is in the region of 60 percent.” This means that out of all the authors in a given field, 60 percent will have just one publication; 15 percent will have two publications \((1/2^2 \times 0.60)\); 7 percent will have three publications \((1/3^2 \times 0.60)\), and so on. Lotka’s law is considered as the most applied law in measuring scientific productivity of authors. In simple way Lotka’s law can be represented as:

\[ x^n y = c \]

where \(x\) stands for the contributions

\(y\) stands for the number of authors, and

\(c\) is the constant.

The law is examined for the following data set:

<table>
<thead>
<tr>
<th>No. of Authors(y)</th>
<th>No. of Papers(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Now, if we put the first row value to the equation \(x^n y = c\),

\[ 1^n 80 = c \]

\(c=80 \ [1^n=1]\)

Now, by putting the second row in equation, we can get the value of \(n\)

\[ x^n y = c \]

\[ 2^n 20=80 \]

\[ \Rightarrow 2^n = 80/20 \]

\[ \Rightarrow 2^n = 4 \]

\[ \Rightarrow n \log 2= \log 4 \]

\[ \Rightarrow n \times (0.30)= 0.60 \]

\[ \Rightarrow n= 0.60/0.30 \]

\[ \Rightarrow n= 2 \]

Thus by putting the values of \(n\) in the other rows the value of \(c\) becomes almost similar. Sen (2010) suggested to work out the exponential value (i.e. \(a\)) for a given set of data, which may
vary based on the volume of dataset. Further statistical test may be carried out for testing of the fitness of the model derived of the Lotaka’s calculated value with the expected value.

3. Zipf’s Law

In 1935, George Kingsley Zipf formulated this law. This law is used to calculate the frequency of words in a text. The law states that in a long text if the words are arranged in decreasing order, then the rank of any given word of the text will be inversely proportional to the frequency of occurrence of the word.

The mathematical equation of Zipf’s Law is as follows:

\[ r \times f = k \]

Where \( r \) = the rank of the word,
\( f \) = the frequency, and
\( k \) = Constant

To apply this law, the words collected from a specific title of article or given text of literary output and then ranked according to their frequency of occurrence. This law states that log of frequency of occurrence are when added to the log of the ranks, the results become almost same for all. Following is an explanation of the Zipf’s Law with an empirical data set.

Suppose, in a title the following words are the most occurring words which are ranked as per their occurrence.

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>853 times</td>
<td>1</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>421 times</td>
<td>2</td>
</tr>
</tbody>
</table>

Log Frequency + log Rank

\[
\log (853) + \log (1) = 2.93 + 0 = 2.93
\]

\[
\log (421) + \log (2) = 2.62 + 0.30 = 2.92
\]
Word : Basal Secretion
Frequency : 311 times
Rank : 3

Log Frequency + log Rank
Log (311) + log (3)
= 2.49 + 0.48 = 2.97

Hence, it is proved that Zipf’s Law is valid.

4.7. Conclusion

Scientometric is a way to measure scientific publications by using different parameters. It helps in quantitative evaluation of data. Scientometrics uses different parameters to quantify and evaluate various inherent characteristics, pattern and relationships among different attributes of the scholarly publications in the domain of science. By adopting a suitably chosen set of metrics one may explore the hidden facts about the productivity of scientific communications which hitherto not evident. The outcome of scientometric studies has wide range of implications right from collection management, resource allocation for collection management, design and development of better indexing techniques to the identification of the trend of scholarly publications in the libraries. Scientometric studies have significant implications in science policy evaluation and drafting of adequate one at different levels viz. organizational, national, etc. It is evident from the literature that newer metrices are being proposed by the scientometricians focusing on certain more in-depth into the behaviour, characteristics and relationship of scholarly publications, which lead to conclude that scientometrics is a fast growing area.