Subject: Library and Information Science

Production of Courseware
-Content for Post Graduate Courses

Paper No : 10 Informetrics and Scientometrics

Module : 17 Webometrics, Cybermetrics and Nettometrics

Development Team

Principal Investigator & Subject Coordinator	Dr. Jagdish Arora, Director INFLIBNET Centre, Gandhinagar
Paper Coordinator	Dr I K Ravichandra Rao Retd Professor, Documentation Research and Training Centre
Content Writer	Dr N S Harinarayana Associate Professor, University of Mysore
Content Reviewer	Dr I K Ravichandra Rao Retd Professor, Documentation Research and Training Centre

Module 17

Webometrics, Cybermetrics and Nettometrics

I.   Objectives

After reading this unit, you will be able to

- Understand the growth and development of webometrics research
- Have a bird-eye view of webometrics research and its applications in LIS domain and other social science domain.
- Familiarize yourself with various techniques available for webometrics research

- Know various data gathering tools available for conducting webometrics research

II.   Module Structure

1.0 Introduction

Library and information science (LIS) domain has a long history of measuring the research output of an organization or individual using quantitative techniques derived from mathematics and other subfields. sociology of science, librametrics, bibliometrics, scientometrics and informetrics are some of the terms used often for measuring the output of the research. The term bibliometrics was first used in 1969 by Pritchard. He defined bibliometrics as "the application of mathematical and statistical methods to books and other media of communication". By 1970 the term bibliometrics has gained considerable attention and mentioned in LIS literature including Library and Information Science Abstract and other LIS literature (Hood & Wilson, 2001).

Scientometrics was first coined by Vassily V. Nalimov & Z. M. Mulchenko in 1969. The term 'scientometrics' is the Russian equivalent of 'naukometriya'. Scientometrics though overlap with the term 'bibliometrics' has wider ambit when compared to bibliometrics. Scientometrics has been defined as "the study of the quantitative aspects of science as a discipline or economic activity. It is part of the sociology of science and has application to science policy-making. It involves quantitative studies of scientific activities, including, among others, publication, and so overlaps bibliometrics to some extent" (TagueSutcliffe, 1992, p. 1).

As the field of bibliometrics and scientometrics expanded, the new term 'informetrics' was introduced to include information science research in the early 1990s. The definition provided by Tague-Sutcliffe (1992, p. 1) defines informetrics as "the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists. Thus it looks at the quantitative aspects of informal or spoken communication, as well as recorded, and of information needs and uses of the disadvantaged, not just the intellectual elite. It can incorporate, utilise, and extend the many studies of the measurement of information that lie outside the boundaries of both bibliometrics and scientometrics. …Two phenomena that have not, in the past, been seen as a part of bibliometrics or scientometrics, but fit comfortably within the scope of informetrics are: definition and measurement of information, and types and characteristics of retrieval performance measures."

The above definitions help in identifying the similarities between the terms such bibilometrics, scientometrics and also distinguish between the terms.

Since 1990s the term "informetrics" has been often used in the world of quantitative analysis of scholarly communications. As the Web gains its momentum and became major source of information and scholarly communication shifting its boundaries from print to online, researchers in scientometrics and informetrics have tried to quantify web pages or sites, because of this, new term "webometrics" was coined in 1997

by Almind & Ingwersen (1997). Today 'Webometrics' has become one of the major areas of research in LIS domain. The subsequent section of this paper discusses the origin of Webometrics, its applications in information science and other areas of social sciences.

2.0 Webometrics

Webometrics is the quantitative study of web related phenomenon (Thelwall, Vaughan & Björneborn, 2005). The growth of web as a major source of information attracted large number of researchers in applying traditional bibliometrics and scientometrics technique to new form of communication media that is World Wide Web.

According to Björneborn & Ingwersen (2004), the definition of webometrics is "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches." Mike Thelwall is one of the pioneer in the field of webometrics and according to him, webometrics is "the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study" (Thelwall, 2009), which emphasizes a small subset of relatively applied methods for use in the wider social sciences. The purpose of giving this alternative definition is to help publicise appropriate methods outside of the information science discipline rather than to replace the original definition within information science.

The introduction of Web Impact Factor (WIF) by Ingewersen (1998) which was similar to that of Web of Science's (WOS) Journal Impact Factor (JIF) for assessing the website helped webometrics to establish as a method to assess the web phenomenon. The WIF measure may be defined as the number of web pages in a web site receiving links from other web sites, divided by the number of web pages published in the site that are accessible to the crawler. As web become major sources of information key areas were identified to apply webometric techniques to the web.

3.0 Key Areas of Webometrics

Thelwall, who has contributed immensely for the growth of webometric research, has identified key areas of webometrics (Thelwall, 2009; 2012). They are:

1. Link Analysis,
2. Web Citation Analysis,
3. Search Engine evaluation, and
4. Web Data analysis or Measuring Web 2.0.

3.1 Link Analysis

Link analysis is the quantitative study of hyperlinks between web pages. Early webometric studies have conducted this type of research extensively. Link analysis has been used in webometrics study for identifying the patterns in academic or scholarly web spaces. Link analysis has direct analogies to traditional bibliometric studies. In traditional bibliometric related studies, the researcher often used to do studies by measuring or quantifying the research contributions of an organization or researchers, in a similar way link analysis can be done by identifying how popular a website is?, which link of web site or page is more popular and why? This kind of a research can be done using link analysis in webometrics. The web crawling tools such as SocSciBot (www.socscibot.wlv.ac.uk) introduced by the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK and other freely available link analysis web crawlers have made the process of crawling web data simpler. Some of the major studies in this field of research have been reviewed by Kousha (2005) in his paper entitled "webometrics and scholarly communications: an overview".

The 'Webometrics Ranking' is one of the best examples for link analysis of academic university websites. Webometrics Ranking of World Universities (www.webometrics.info/en) has been conducted since 2006 to identify the web presence of world universities through link analysis.

3.2 Web Citation Analysis

Another key area of webometrics research is web citation analysis. Web citation analysis is nothing but counting online citations to published items such as journal articles. Earlier, bibliographic citations found in traditional citation indexing databases such as SCI, SSCI, and A&HCI were considered as major source for identifying the impact of research publications. The web provided new methods to overcome the limitations of the bibliographic citation databases, by helping in finding web citations for articles published online largely for open access journals. Research on web citations have found that there was correlation between the citations found in bibliographic citations and citation found on the web for particular research papers (Vaughan and Shaw, 2003). The introduction of Google Scholar has helped webometricians to use this citation database to identify the impact of the research publications easily. Because of the introduction of tools such as Google Scholar and Altemetrics, the research in this area has seen considerable growth in recent time. Two of the research publications that have been mentioned below on this subject throw more light on how this kind of studies can be conducted using webometrics. These studies were done by Kousha & Thelwall (2005, 2007) on Google Scholar and Google Web/URL citations.

3.3 Search Engine evaluation

Another key area of research that has been conducting in webometrics is search engines evaluations. The major webometrics studies have concentrated on the coverage of the web and the accuracy of the reported

results. It was known fact that none of the search engines have indexed whole of the information available on the web. Still 70% of the information available on the web has not been indexed by any of the search engines available today (Metayer, 2010). The reliability and validity of the data obtained through commercial engines was also very contentious issues in webometrics, many studies have raised validity and reliability of data obtained through the search engines (Bar-Ilan, 2004; Kousha, 2005). Many of the search engines such as Alta Vista (which was initially used by many webometricians for link analysis), Yahoo and Bing now have stopped providing advanced link search commands feature which were used extensively in beginning of webometric research (Thelwall, 2012). The development of web crawling programs and availability of API (Application Programming Interface) from different information processing industry have now helped to expand webometrics research to other social science domain.

3.4 Web Data analysis or Measuring Web 2.0

As web evolved over the years, webometrics has expanded its research base by simply measuring academic websites to other domain of research area of social sciences. Availability of web crawler and API (Application programming Interface) has helped in extracting larger cache of data easily on the web. Recent studies done by Thelwall and his team (see, Sugimoto & Thelwall, 2013; Kousha, Thelwall & Abdoli, 2012; Wilkinson & Thelwall, 2002; shema, Bar-Ilan & Thelwall, 2012) have shown that how webometrics can be used in other areas of subject apart from library and information science domain, which is the traditional forte of webometric research.

Webometrics has been using in many other areas of studies apart from library and information science domain. Today Webometrics is not only using for research evaluation but also in other social sciences mainly examining web phenomenon. As the information available on the web increases, webometrics has become an important method of research for studying web data.

4.0 Cybermetrics or Webometrics: is there any difference?

The web, a large repository of information has been considered as one of the important medium for research. The availability of structured and non-structured data provided a fertile ground for using bibliometrics, scientometrics and informetrics. Many researchers have proposed various terms to define new area of research. The terms that were proposed: netmoterics, webometry, internetometrics, webometrics and cybermetrics, web bibliometry and web metrics (Thelwall, Vaughan, & Björneborn, 2005).

Bjorneborn & Ingewersen (2004) have made an effort to differentiate these two terms. They have defined webometrics as "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the WWW drawing on bibliometrics and informatric approaches."

On the other hand, Björneborn & Ingewersen (2004) have defined cybermetrics as "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet, drawing on bibliometrics and informetrics approaches." Cybermetrics encompasses statistical studies of discussion groups, mailing lists and other computer mediated communication on the Internet."

By going through above definitions one can easily distinguish between webometrics and cybermetrics. Cybermetrics includes all aspects of Internet related studies including computer mediated communications and covers quantitative aspects of web including studying web network properties. Figure 1 succinctly depicts the difference informetrics, bibliometrics, scientometrics, cybermetrics and webometrics and overlapping nature of each other.
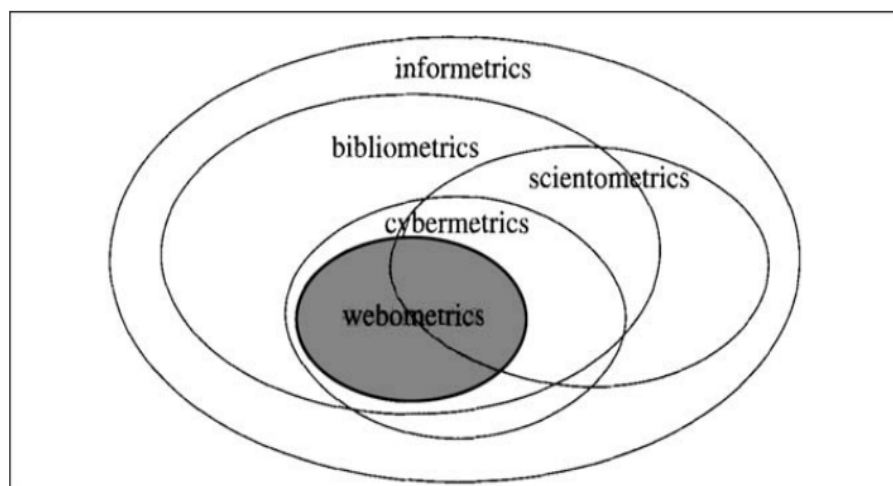

Figure 1: The Sizes of the overlapping between Informatrics, Bibliometrics, Scientometrics, Cybermetrics and Webometrics
(Source: Björneborn & Ingewersen, 2004)

5.0 Network Analysis or Nettometrics

Network analysis has become one of the important areas of research because of availability of huge data sets through Internet. Though network analysis has its origin long back in mathematics, sociology and other field of human interactions, it is in recent times that network analysis gain its momentum because of people interactions through technological means (on the web).

A social network is a social structure made up of individuals (or organizations) called "nodes", which are tied(connected) by one or more specific types of interdependency, such as friendship, kinship, common interest,

financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige (Wikipedia, 2013).

Social network analysis views social relationships in terms of network theory consisting of nodes and ties (also called edges, links, or connections). Nodes are the individual actors within the networks, and ties are the relationships between the actors.
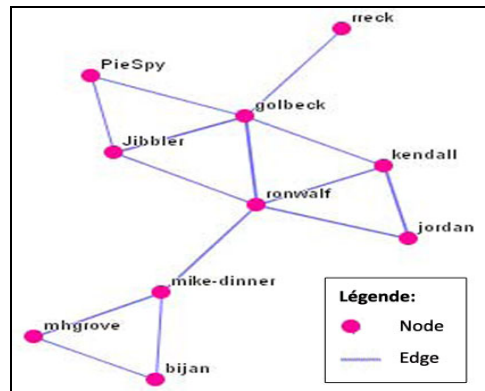

Figure 2: A Simple social network is modeling by a graph with nodes and edges linked together, nodes are the individual actors (e.g., friends in a group) and edges are the relationship between the actors.

Social network analysis has become key technique in sociological research and has been in other field of study as well such as, anthropology, in anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, and sociolinguistics. In information science social network analysis technique is used study the authors' citation pattern, co-citation analysis and co-authorship pattern (Otte, & Rousseau, 2002) .

The study on "Co-authorship networks in the digital library research community" by Xiaoming Liu and others (Liu, Bollen, Nelsson, & Van de Sompel, 2005) is one of the examples for how social network analysis can be used to identifying the co-authorship network pattern among digital library research community. There is a whole lot of opportunity to apply social network analysis to different domain of subject areas. The development of network analysis tools and visualization software has allowed researchers to investigate large networks like Internet and interpret them in an understandable way.

6.0 Summary

This unit introduces the concepts such as webometrics, cybermetrics and nettometrics. The unit discusses the applications of webometrics in various areas of research in library and information science. The difference between webometrics and cybermetrics has been explained. The concept of social networking has also been introduced here.

References

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'Webometrics'. Journal of Documentation, 53(4), 404-426.

Bar-Ilan, J. (2004). Use of search engines in information science research. Annual Review of Information Science and Technology, 38, 231-288.

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. Journal of the American Society for Information Science andTechnology, 55(14), 1216-1227.

Hood, W. H., &Wilson, C. S. (2001). The literature of biblometrics, scientometrics, and informetrics. Scientometrics, 52(2), 291-314.

Ingwersen, P. (1998). The calculation of web impact factors. Journal of Documentation, 54(2) 236-243.

Kousha, K. (2005). Webometrics and scholarly communication: An overview. Quarterly Journal of the National Library of Iran (Faslenameh Ketab), 14(4), 7-16.

Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis, Journal of the American Society for Information Science and Technology, 58 (7), 1055-1065.

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. Scientometrics, 74(2), 273-294.

Kousha, K., Thelwall, M. & Abdoli, M. (2012). The role of online videos in research communication: A content analysis of YouTube videos cited in academic publications. Journal of the American Society for Information Science and Technology, 63(9), 1710–1727.

Liu, Xiaoming., Bollen, J., Nelsson, M. L. , & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. Information Processing and Management, 41(5), 1462 – 1480.

Metayer, E. (2010). Tapping into the invisible web for competitive intelligence. Competia. Retrieved from http://competia.com/tapping-into-the-invisible-web-for-competitive-intelligence/

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information science. Journal of Information Science, 28(6), 441-453.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. PLOS ONE , 7(5). Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0035869

Sugimoto, C. R., & Thelwall, M. (2013). Scholars on soap boxes: Science communication and dissemination via TED videos. Journal of the American Society for Information Science and Technology, 64(4), 663-674.

Tague-Sutcliffe, J. M. (1992). An introduction to informetrics. Information Processing & Management, 28, 1–3.

Thelwall, M. (2009). Introduction to webometrics: Quantitative web research for the social sciences. New York: Morgan & Claypool.

Thelwall, M. (2012). A history of webometrics. Bulletin of the American Society for Information Science and Technology, 38(2). Retrieved from http://mail.asist.org/Bulletin/Aug-12/AugSep12_Thelwall.pdf

Thelwall, M., Vaughan, L., & Björneborn, L (2005). Webometrics. In Cronin, B. (ed.), Annual Review of Information Science and Technology. Medford, NJ: Information Today, Inc., Vol. 39, 81-135

Vaughan, L., & Shaw, D. (2003). Bibliographic and web citations: What is the difference? Journal of the American Society for Information Science and Technology, 54(4), 1313-1324.

Wikipedia, (2013). Social network analysis.Wikipedia. Retrieved from http://en.wikipedia.org/wiki/Social_network_analysis

Wilkinson, D., & Thelwall, M. (2012). Trending Twitter topics in English: An international comparison. Journal of the American Society for Information Science and Technology, 63(8), 1631-1646.