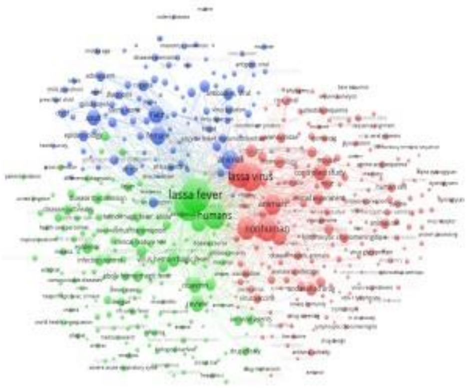
Athens
2019
IFLA WLIC

Submitted on: 03/07/2019

Complementing Bibliometrics with Network Visualization to Support Scientific Spheres

Nirmal Singh
Assistant Librarian
Guru Angad Dev Veterinary and Animal Sciences University
Ludhiana (Punjab)-141004, INDIA
E-mail: nirmal02@yahoo.co.in



Abstract

Science evaluation in both quantitative and qualitative terms is imperative to ascertain the growth and trends in a knowledge domain, strengths and gaps in research, symbiotic collaboration opportunities at national and international level, and impact of research reflected in terms of quality and performance indicators, viz. h-index, impact factor, number of citations, etc. Bibliometrics and its alternatives including scientometrics, altmetrics, webometrics, and librametrics are being widely used as tools for exploring the growth and structure of scholarly communications in different forms and formats encompassing various disciplines. The Library and Information Science Professionals (LISPs) have experience and expertise in bibliographic data handling. Therefore, LISPs can actively execute bibliometric studies. However, the volume and complexity of datasets in science disciplines have grown so enormous that use of special tools has become essential for inferring from the data available in form of various units of analysis. Network visualization software is used as supplemental aid to support bibliometrics in an interactive mode. These software usually have text mining functionality to facilitate the construction and visualization of network maps based on the co-occurrence of key terms extracted from a body of scholarly communication, viz. co-authorship, co-occurrence of institutions, bibliographic coupling and co-citation links. This paper discusses as detailed below the need for scientific evaluation, application of bibliometrics in science evaluation, role of data/network visualization in complimenting bibliometrics and various network visualization software.

Keywords: bibliometrics, collaboration analysis, network visualization, science mapping, scientometrics.

Need for science evaluation

Since last quarter of 20th century, academic and scientific literature has witnessed exponential growth, reflected by the parallel progression in size and complexity of bibliographic databases (Synnestvedt et al., 2005). The publish or perish paradigm, emergence of various formats of information resources, and pressure for securing research funds, etc. are contributing factors. Hence, quantitative and qualitative appraisal is pivotal for understanding the research dynamics to be able to underpin the emerging trends and sudden changes in body of scientific knowledge (Chen, 2006).

The scientific evaluation involves various components, i. e. appraisal of research performance and impact of research, comparison with peers, supporting policy decisions, spotting research trends, etc. The science has become so enormous, complex and specialized that it is impossible to understand the trends for decision making based on one's personal knowledge and acquired experiences (Pendlebury, 2008). Bibliometrics is the cost-efficient alternative to peer review (Bornmann et al., 2018), providing intricate mechanism for evaluating scholarly communications. It is used as a yardstick for understanding the flow of scientific information and planning of science among policy makers, scientists and academicians, and administrators, by processing and predicting various features of interest, including scholarly communication, performance, development, and dynamics (Borner *et al.*, 2010). O'Connor and Voos (1981) elucidated that measurement of bibliographic information is must to resolve many practical problems. The development of *Science Citation Index* by Eugene Garfield gave new impetus to statistical appraisal of scientific literature, making bibliometrics a powerful field of study of science (van Raan, 2004). Of late, this has also been applied for assessment of research productivity of scientists and universities for performance based allocation of funds (Eyre-Walker and Stoletzki, 2013).

Various metrics based on citations and bibliographic data, viz. impact factor, citation index, h-index, g-index, i10index, etc. have been used as performance and quality indicators. Further, this would assist in describing and predicting the frontiers of research in respective knowledge domain.

Role of LIS professionals

Gumpenberger et al. (2012) attribute bibliometrics as an ideal field for contemporary librarians, the experts in data retrieval, filtration and analysis.

Astrom and Hansson (2013) also retrieved the observations of Gumpenberger et al. (2012) and considered university libraries as 'organizational locale' for bibliometric activities because of expertise and experience of librarians in bibliographic data management, their access to data sources and unbiased approach towards the subject(s). Vinyard and Colvin (2018) observed that faculty was interested in taking advantage of the experience of librarians in evaluating journals and impact of their articles.

Bibliometrics

Beginning with a simple statistical method of counting for evaluation and quantification of growth of literature, bibliometrics has flourished to have wider applications. The terms bibliometrics, scientometrics, webometrics, altametrics, etc. represent the quantitative and qualitative appraisal of different forms and formats of scholarly communications.

Alan Pritchard found the term "scientific bibliography" coined by Hulme (1923) to be redundant. He introduced the term 'bibliometrics' in 1969 and defined it as "application of mathematics and statistical methods to books and other forms of written communication"

(Pritchard, 1969). The manifestation of 'books and all forms of written communication' by Pritchard includes different forms and formats in communication process. Potters' observations were also in consonance with those of Pritchard (Potter, 1981).

Reitz (2004) has given an explicit definition of bibliometrics as the "use of mathematical and statistical methods to study and identify patterns in the usage of materials and services within a library or to analyse the historical development of a specific body of literature, especially its authorship, publication, and use". Over the years, scope of bibliometric analysis has broadened to include analyses of books, journals, databases, websites, social networks, etc.

These definitions reveal that bibliometrics involves evaluation of the intricate information processes by analysing its interconnected units.

Review of bibliometrics studies to support science strategies

Several bibliometric studies have been conducted to support science decisions, research policies and collaborative research initiatives. A few relevant are listed here:

During 2015, the European Commission funded Science-Metrics to analyse bibliometric indicators for European policies based on the volume and impact of scientific publications produced from 2000 to 2013 in Europe and rest of the world to support development of European research and innovation policies. (European Commission, 2015).

Similarly, The Research Council of Norway commissioned Science-Metrics to conduct bibliometric study to evaluate scientific performance and national and international collaborations of Norway in science in general and in thematic areas of high relevance to Norway in particular. Potential key partners for Norway by strategic themes for strengthening or reinforcing its collaboration were suggested based on findings of study (Beaudet *et al.*, 2014). Science-Metrics conducted another study for Department of Fisheries and Oceans, Canada to support DFO's implementation of International Science Strategy to target science collaborations with countries having research strengths in its 11 priority research areas. (Picard-Aitken et al., 2009)

The studies by Irawan *et al.* (2016), Gunashekar, et al. (2017) Junquera and Mitre (2007) divulge implications of bibliometrics in evaluating the evolution, structure and complexity of science in different contexts.

Data visualization to complement bibliometrics

Bibliometrics is an acknowledged discipline for study of scientific indicators. However, with exponential growth of academic and scientific communication processes, use of special tools is required to extract interpretable information from large datasets comprising interconnected bibliographical and citation data. Knowledge generated through such tools can be used to infer the organisation of science in more advantageous way to improve it on various scales. The difficulties in processing plethora of scholarly information for interpretation of hidden niceties resulted into emergence of bibliometrics/ network visualization software. In contrast to traditional graphs static in nature, visualizations are interactive, leading to better interpretation of bibliometric results. Visualization software usually combines the trinity of algorithms, bibliographic metrics and graphics to present data in an easy to understand mode. These also allow researchers to explore particular parts of networks using various options and enhance the interactivity and promote involvement of users with data. Sangam and Mogali (2012) referred these "visual tools for abstraction, summarization and presentation of large volumes of data". These leverage the perceptual abilities of humans for cartography of science and technology to get insight into important aspects of cognitive structure of scientific research (Meyer and Glanzel, 2009).

Network Visualization also called Science Mapping provides an overview of the scientific landscapes to make these more manageable and get expedient insights to figure out data informed-decisions by exploring state and development of scientific knowledge and practices, by depicting connections and context of nodes and edges in a richer, realistic and informative way. This compliments bibliometrics by allowing researchers and users to study beyond flat data to draw more symptomatic inferences.

Network visualization tools:

Bibexcel**:** Bibexcel, developed by Olle Persson, Professor at Umea University, Sweden allows bibliometric analysis of data retrieved from Web of Science, Scopus A&HCI, etc., or any textual data having a similar formatting structure. It generates data files for importing to Excel, vis-a-vis. another programme(s) accepting tabbed data records for further analysis. Bibexcel tool-box has number of options to perform different type of bibliometric analysis, and its' output can be further mapped in Pajek, NetDraw and can be processed in SPSS. (Persson et al., 2009).

Pajek**:** Pajek is visualization tool applying proficient algorithms for analyzing large networks containing up to a billion of vertices. It requires execution of several operations in a sequence to get desired results. Extracting sub-networks from a large network, shrinking selected parts of networks, analyzing the weak, strong, and biconnected components, searching for shortest paths, k-neighbors, maximum flow, centralities of vertices and centralizations of networks, clustering coefficients, clustering in networks, generating different types of random networks, etc. are the operations supported by Pajek. Result output of Pajek can be exported and further analysed using R, SPSS, and Excel (http://mrvar.fdv.uni-lj.si/pajek/).

Gephi**:** Gephi can be used to import, visualize, spatialize, filter, manipulate and export different types of networks. It permits the manipulation of structures, shapes and colors to discover patterns and supports large network having upto 1,00,000 nodes and 10,00,000 edges. Software has extendibility feature letting addition of algorithm(s) easily without having programming experience. The layout algorithms can be run in real-time and several algorithms can be run in different workspaces, simultaneously. Its' result visualization can be exported as SVG or PDF file. Gephi architecture supports communication of data sources with existing software, third parties databases or web-services (https://gephi.org/features/)

VOSviewer**:** VOSviewer is Java based programme for creating and visualizing maps based on network data retrieved from Web of Science, Scopus, Dimensions, and Crossref, PubMed and RIS. It has text mining functionality to construct and visualize network maps, viz. co-authorship, co-institutional, bibliographic coupling and co-citation maps. It supports network interoperability with Pajek and provides three types of visualizations, viz. network visualization, overlay visualization, and density visualization, and produces high resolution screenshots. VOSviewer permits fractional and full counting approach for assigning weightage to units of analysis (van Eck and Waltman, 2017).

CiteSpace**:** CiteSpace combines trinity of information visualization methods, bibliometrics, and data mining algorithms to ascertain patterns and trends in given subject(s). Primary source of data inputs for CiteSpace is Web of Science, however, it also provides simple interfaces for obtaining data from PubMed, arXiv, ADS, and NSF Award Abstracts. CitSpace allows adjustment of visual attributes of display and can be used to generate geographic map overlays viewable in Google Earth. Identification of fast-growth topical areas, citation hotspots, automatic labelling of clusters, collaboration networks, structural and temporal analyses of networks, etc. are features of CiteSpace http://cluster.cis.drexel.edu/~cchen/citespace/)

***Science of Science (Sci2) Tool*:** Science of Science (Sci2) Tool is modular toolset designed by researchers at Indiana University, specifically for analysis and visualization of science knowledge domains. It can be used for data visualization in other disciplines too. It allows users to clean, analyze, and visualize a wide variety of data formats including XML, Endnote, Scopus data, NSF files, CSV, etc. Sci2 Tool is built on the Cyberinfrastructure Shell, permitting easy integration and utilization of datasets, algorithms, tools, and computing resources. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro, meso and macro levels (Sci2 Team, 2009).

***CitNetExplorer*:** CitNetExplorer, viz. 'Citation Network Explorer' is Java based software developed on Garfield's work on algorithmic historiography. It can handle large citation networks containing millions of publications and citation relations. The citation networks can be imported to CitNetExplorer directly from Web of Science. Exporting citation network in Pajek file format, zoom and scroll functions to explore visualization, smart labeling algorithm to prevent overlapping, identification of connected components, clusters, core publications, and shortest and longest paths applying various algorithms are the main features of CitNetExplorer (Centre for Science and Technology Studies, 2019).

SciMAT: Science Mapping Analysis software Tool, known as SciMAT is based on longitudinal science mapping approach, allowing users to conduct citation and network analysis using integrated methods, algorithms and measures for science mapping. It allows use of various normalization and similarity measures over data and constructs maps using different algorithms and visualization techniques. It supports network analysis, performance and quality analysis, and temporal analysis. SciMAT supports processing of data in Web of Knowledge and RIS format and generates reports in HTML and LaTex format. (Cobo *et al.*, 2012) https://sci2s.ugr.es/scimat/)

HistCite: HistCite, developed by Eugene Garfield, the inventor of Science Citation Index, is bibliometric analysis and information visualization software to identify significant publications. Though, HistCite is designed to import bibliographies created by searches of Web of Science database, it allows users to enter bibliographies from other sources, manually. It converts bibliographies into diagrams called historiographs.

Publish or Perish: Publish or Perish is a software programme for citations analysis. It allows processing of a variety of data sources. It empowers users to analyze and present the research impact through various metrics, viz. publications and citations, average papers and citations per paper, author and year, h-index and related parameters, average annual increase in the individual h-index, etc. (Harzing, 2019).

Others: In addition to above listed software, various other network visualisation/ science mapping tools are also available, which are not discussed here in detail due to space limitations. These include, Network Workbench, IN-SPIRE, CopalRed, BiblioTool, Science Assessment Integrated Network Toolkit, etc.

Conclusion

Bibliometrics is a time tested, reliable and proven method of science evaluation. Albeit it has been criticised over the period for being a quantitative measure, the assessment of various metrics including impact factor, citations, h-index, g-index, etc. makes it performance and quality evaluation tool too. The network visualization further enhances the significance of bibliometrics in science evaluation by presenting data in dynamic and interactive genre.

Despite variation in features and functions, these software aim at presenting the relations and context of units of analysis based on their appearance in scholarly communications. Several bibliomertic/ network visualization software are available in the open domain and library and information science professionals can use these for betterment of scientific fraternity.

Acknowledgements: Author is highly grateful to Dr. H. S. Banga, University Librarian, GADVASU, Ludhiana and Dr. Trishanjit Kaur, Prof. of Library & Information Science and Dean, College Development Cell, Punjabi University, Patiala for their motivation, support and guidance for writing this article.

References:

- Astrom F. and Hansson J. (2013). How implementation of bibliometric practice affects the role of academic libraries. *Journal of Librarianship and Information Science*, 45(4), 316–322.
- Beaudet A., Campbell D., Cote G., Haustein S., Lefebvre C., Roberge, G. (2014). *Bibliometric Study in Support of Norway's Strategy for International Research Collaboration*, *Report*. Brussels: The Research Council of Norway. Avaibale at: http://www.science-metrix.com/sites/default/files/science-metrix/publications/bibliometric studyinternational.final_.pdf
- Borner K., Huang W., Linnemeier M., Duhlon R. J., Phillips P., Ma N., Zoss A. M., Guo H. and Price M. A. (2010). Rete-netzwerk-red: analyzing and visualizing scholarly networks using the Network Workbench Tool. *Scientometrics*, 83(3), 863–876.
- Bornmann L., Hug S. and Marewski J. N. (2018). Bibliometrics-based heuristics: What is their definition and how can they be studied? Available at: http://arxiv.org/abs/1810.13005.
- Centre for Science and Technology Studies (2019). *CitNetExplorer - Analyzing citation patterns in scientific literature*. Leiden University, The Netherlands. Available at: http://www.citnetexplorer.nl/
- Chen C. (2006). CiteSpace II: Detecting and visualizing emerging trends. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.
- Cobo M. J. *et al.* (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*, 63(8), 1609–1630.
- van Eck N. J. and Waltman L. (2017). VOSviwer manual. Available at: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.6.pdf
- European Commission (2015). *Analysis of bibliometric indicators for European policies, 2000-2013*. Brussels. Available at: https://ec.europa.eu/research/innovation-union/pdf/bibliometric_indicators_for_european_policies.pdf.
- Eyre-Walker A. and Stoletzki N. (2013). The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biology*, 11(10).
- Gumpenberger C., Martin W. and Gorraiz J. (2012). Bibliometric practices and activities at the University of Vienna. *Library Management*, 33(3), 174–183.
- Gunashekar S., Wooding S. and Guthrie S. (2017). How do NIHR peer review panels use bibliometric information to support their decisions? *Scientometrics*, 112(3), 1813–1835.
- Harzing A. W. (2019). *Publish or perish*. Available at: https://harzing.com/resources/

publish-or-perish

- Hulme F. W. (1923). *Statistical bibliography in relation to the growth of modern civilisation*. London: Grafton.
- Irawan D., Priyambodho A., Rachmi C. N., Wibowo D. M. and Fahmi A. K. (2016). Bibliometric study to assist research topic selection: a case from research design on Jakarta's groundwater (part 1). *Research Ideas and Outcomes*, 2, p. e9841. doi: 10.3897/rio.2.e9841.
- Junquera B. and Mitre M. (2007). Value of bibliometric analysis for research policy: A case study of Spanish research into innovation and technology management', *Scientometrics*, 71(3), 443–454.
- Meyer M. and Glanzel W. (2009). Persson's universe of bibliometrics – Has his mapping changed the discipline? In Astrom, F. et al. (Eds.) *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th Birthday*. International Society for Scientometrics and Informetrics, 39–46.
- O'Connor D. and Voos H. (1981). Empirical laws, theory construction and bibliometrics. *Library Trends*, 30(1), 9.
- Pendlebury D. A. (2008). *Using bibliometrics in evaluating research*. Available at: http://wokinfo.com/media/pdf/UsingBibliometricsinEval_WP.pdf.
- Persson O., Danell R. and Wiborg Schneider J. (2009). How to use Bibexcel for various types of bibliometric analysis. ISSI, 9–24. Available at: http://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A232746&dswid=-7102
- Picard-Aitken M., Campbell D. and Cote G. (2009). *Bibliometric study in support of fisheries and oceans Canada's international science strategy*. Montreal. Available at: http://www.science-metrix.com/pdf/SM_DFO_International_Science_Strategy.pdf.
- Potter W. G. (1981). Introduction. *Library Trends*, 30(1), 5.
- Pritchard A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, 25(4), 348–349.
- van Raan A. F. J. (2004). *Measuring Science: Handbook of Quantitative Science and Technology Research*. H. F. Moed, W. Glanzel and U. Schmoch (Eds.). Dordrecht: Springer, 19–50.
- Reitz J. M. (2004). *Dictionary for Library and Information Science*. London: Libraries Unilimited.
- Sangam S. L. and Mogali S. S. (2012). Mapping and visualization softwares tools: a review. *Content Management in Networked Environment*, p. 11.
- Sci2 Team (2009). *Sci2 Tool : A Tool for Science of Science Research and Practice, Indiana University and SciTech Strategies*. Available at: https://sci2.cns.iu.edu
- Synnestvedt M. B., Chen C. and Holmes J. H. (2005). CiteSpace II: Visualization and knowledge discovery in bibliographic databases. *AMIA Annual Symposium proceedings*, 724–8.
- Vinyard M. and Colvin J. B. (2018). How research becomes impact: Librarians helping faculty use scholarly metrics to select journals. *College and Undergraduate Libraries*, 25(2), 187–204.