Popular Scientometric Analysis, Mapping and Visualisation Softwares: An Overview

Ashok Kumar **J Shivarama** **Puttaraj A Choukimath**

Abstract

Measurement of scientific productivity has been regarded as main indicator of ascertaining impact of research over scientific community. To showcase the impact of research using science mapping and visualisation, social network analysis has been developed over the period of time. These methods help researchers to understand the structural, temporal and dynamic development of a discipline. The present paper provides a comprehensive overview of widely used softwares used for scientometric analysis, mapping and visualisation.

Keywords: Scientometrics, Data Analysis, Data Mapping, Visualisation Softwares, Citation Datbases

1. Introduction

At present the rate of publishing of scholarly communication is like flood of scholarly literature which is being published regularly, hence it is exposing the weaknesses of current scientometric analysis based methods for evaluating this scholarly literature. A novel and promising approach to examine and evaluate this big amount of literature by using scientometric analysis softwares like Bibexcel, Pajek, CiteSpace, SAINT, Publish or Perish, Network Workbench, SITKIS, Vantage Point or Excel etc. This paper develops the most comprehensive list of the softwares available to date, assessing the potential value of data analysis by each one. This paper also overview the building and validating metrics drawn from the citation data for Social Network Analysis (SNA), Mapping and Visualisation. These softwares are developed by the experts to help in managing highly specialised databases to organize the large scale data collected in a way which can be frequently updated, and to work with network analysis for mapping and visualisation.

2. Scientometric Analysis, Mapping and Visualisation Softwares

Scientometric softwares can be defined in the light of definition of "Software" as provided in the Online Dictionary of Library and Information Science (ODLIS) as follows "A set of computer based programs, designed and developed to analyse citation based bibliographic data as input to perform the specific tasks i.e. structural analysis of scholarly communication, mapping of scientific research, creation of metrics based social maps, information representation and organisation, visualisation of research, micro level analysis (co-word, co-author, cited references, bibliographic coupling, co-citation) etc. as output.

A scientometric analysis, mapping and visualisation software facilitates its users to draw the maps for visual representation of scientific research based on citation data, to study the structural, temporal and dynamics of a subject discipline. Most of these softwares are based on the modern algorithms, mathematical and statistical methods, graphs theory, network science, etc.

10th International CALIBER-2015
HP University and IIAS, Shimla, Himachal Pradesh, India
March 12-14, 2015

Figure 1: Scientometric Analysis, Mapping and Visualisation Softwares

2.1 Common Features of these Softwares are

- Facilitate structural, temporal and dynamic analysis of a subject discipline.
- Facilitates and supports mapping & visualization of a discipline.
- Able to import input data from the data sources, editing and cleaning of acquired data.
- Helps to execute metrics based evaluation of the data, creation of maps and networks for visualisation.

2.2 Purposes

Scientometric Analysis, Mapping and Visualisation Softwares can be used for following purposes:

1. To study Structural and temporal analysis of information and dynamics of scholarly communication.

2. Mapping of scientific research/subject and building of Metrics based social maps.

3. Facilitates application of modern science analysis, mapping and visualisation techniques and methods.

4. Information representation, organisation and network visualisation.

5. Performs micro level analysis such as:

i. Co-word analysis – keyword based approach of analysis

ii. Co-author analysis – authorship based approach of analysis

iii. Cited references – document based approach of analysis (bibliographic coupling, co-citation analysis, author bibliographic coupling, author co-citation, Journal co-citation and journal bibliographic coupling).

3. Databases for Scientometric Analysis

There are many bibliographic database used widely for doing Scientometric based analysis. Some of these are: Google Scholar, Web of Science, Scopus, Microsoft Academic Research, and PubMed etc. Web of Science, Scopus, PubMed and Google Scholar are majour and popular sources of bibliometric data for doing Scientometric based analysis. Each of these databases has its own advantages and limitations.

3.1 Google Scholar (https://scholar.google.co.in):

Google Scholar is an unpublished bibliographic database offered by Google. It allows researchers to create their Google Scholar page by using their Gmail account having an affiliating address such as academic institution, fields of interest and citations. Through its "cited by" feature, it provides access to abstracts of articles that have cited the article being viewed. Previously this feature was only found in Scopus and Web of Knowledge.

Coverage: Google Scholar currently covers metrics of articles published during the year 2009 and 2013 (both inclusive). The metrics based on the citations from all the articles that were indexed in Google Scholar in June 2014. This also includes citations from

articles that are not themselves covered by Scholar Metrics, Google Scholar includes following items below in order to avoid the misidentification of articles indexed in it:

a) Journal articles from websites which follows guidelines of Google Scholar.

b) Articles published in selected conferences in Computer Science and Electrical Engineering.

c) Articles preprint submitted to the digital repositories i.e. arXiv, SSRN, NBER and RePEC.

Google Scholar doesn't covers following items:

a) Court opinions, patents, books, and dissertations

b) Publications having less than 100 articles published between 2009 and 2013 and

c) Publications didn't receive any citations to the published articles during 2009 and 2013.

Features of Google Scholar:

- It allows users to search for literature available in digital or physical format online or in respective libraries.
- It indexed the scholarly literature available in the form of full-text articles, technical reports, preprints, theses, books, and selected Scholarly Web pages.
- It only provides access to the abstract and citation details of resources required prior subscription.
- The most relevant results for the searched keywords will be listed first, in order of the author's ranking, the number of references that are linked to it and their relevance to other scholarly literature, and the ranking of the publication that the journal appears in.
- It keeps up with the recent developments in subject areas.
- It allows an author to check who is citing publications, creation of author's profile/individual page.
- Google Scholar indexed peer-reviewed online journals, scholarly books and other non-peer reviewed journals.

3.2 Web of Science (http://portal.isiknowledge.com)

The Development of Web of Science (WoS) as world leading citation database is the result of efforts made by the Eugene Garfield of ISI (father of Citation Indexing), when he launched the Science Citation Index (SCI). Web of Science (WoS also known as Web of Knowledge) is an online subscription-based bibliographic citation database maintained by Thomson Reuters. Web of Science is a Multidisciplinary (Science, Social Sciences, Arts, Humanities) database, having unmatched coverage of research data.

Coverage: Data Indexed carefully in WoS has coverage from the year 1900 to the present. One of the largest discovery platform with the most complete records in every subject selected on the basis of their impact. 100+ years of abstracts, Over 90 million records covering 5,300 social science publications in 55 disciplines, 800 million+ cited references, 8.2 million records across 160,000 conference proceedings.

Web of Science consists of the coverage of the following seven online databases

i) Conference Proceedings Citation Index (CPCI): covering more than 160,000 conference titles in the field of Sciences from1990 onwards.

ii) Science Citation Index Expanded (SCI Expended): covering more than 8,500 Science journals from 1900 onwards.

iii) Social Sciences Citation Index (SSCI): covering more than 3,000 Social Science journals from the year 1900 onwards.

iv) Arts & Humanities Citation Index (AHCI): covering more than 1,700 Arts & Humanities journals from 1975. 250 major scientific and social sciences journals are also covered additionally.

v) Index Chemicus (IC): Covering more than 2.6 million records from 1993 onwards.

vi) Current Chemical Reactions (CCR): Indexed more than one million records from 1986 onwards. It also covers INPI archives from 1840 to 1985.

vii) Book Citation Index (BCI): covering more than 60,000 selected books from 2005.

3.3 SCOPUS (http://www.scopus.com)

Scopus is an english language bibliographic citation subscription based database, offered by Elsevier. It has showcased data from 1995 to present. SCOPUS has links to the 55 million records. It is regarded as one of the largest abstract and citation based database of peer-reviewed literature in the form of journals, books and conference proceedings. It provides coverage of the world's research output in the various disciplines such as science, technology, medicine, social sciences, arts and humanities. Scopus database incorporates various important tools to track, analyze and visualize the research. To maintain transparency in the selection of journal to be included in the database, SCOPUS has established an independent international Scopus Content Selection and Advisory Board consisting of subject librarian and scientists. Scopus covers a wider journal range but it is currently limited to recently published articles (articles published after 1995) as compare to the Web of Science. Google Scholar. Scopus also offers authors profiles (containing details such as affiliations, publications and bibliographic data, references, and details of the citations received by the publication). Alerting features of Scopus database allows its registered users to track the changes to a profile and facilitates calculation of author's productivity index (h-index).

3.4 PubMed (www.ncbi.nlm.nih.gov/pubmed): PubMed is popular freely accessible Bibliographic database, offered by United States National Library of Medicine (NLM). PubMed consists of more than 24 million citations for the disciplines of life science and biomedical literature from MEDLINE database. It includes science journals in biomedical and life sciences, and online books. Citations may also include links to full-text, available from PubMed Central and publishers.

Content: PubMed cover only selected journals that comply with PubMed scientific standards. It provides access to the MEDLINE, selected records from Index Medicus, selected records of the journals (from Science published by American Association for the Advancement of Science, BMJ published by British Medical Journal Group, and Annals of Surgery, published by Lippincott Williams & Wilkins).

Medical Subject Headings (MeSH) are assigned to the items before adding to the MEDLINE database and collection of full-text available books and other subsets of the NLM records. Most of the PubMed records also contain links to the full text articles; some of them are often accessible freely from PubMed Central and their local mirrors sites such the UK PubMed Central.

Information related to the journal items indexed in the PubMed database can be retrieved from the NLM Catalog. As on 28 July 2014, PubMed contains more than 24 million records. 500,000 new records

are added by Scopus each year. As on the same date, 13.1 millions of PubMed records are listed alongwith abstracts and 14.2 million articles are with fulltext links (out of which 3.8 million fulltext articles are available accessible freely). PubMed offers simple and advanced search window.

4. Popular Scientometric Analysis, Mapping and Visualisation Softwares: A Brief Profile

4.1 BibExcel (https://bibliometrie.univie.ac.at/bibexcel/)

BibExcel is free software designed by Olle Perrsson, Department of Sociology, Umea University, Umea, Sweden, to assist the user in analysing the bibliographic data, or any data which is available in text format-able form in similar manner. The concept behind this is generation of data files that can be imported to Excel, or any program that takes tabbed data records, for further processing. This toolbox includes various tools, some are visible in the window and some are hidden in the menues. Many of the tools can be used in combination to achieve the greater result. Bibexcel Features:

i. Able to do most types of bibliometric analysis, (i.e. co-citation, bibliographic coupling, mapping and clustering analysis)

ii. Bibexcel allows easy interaction/interoperability cross walk with other software, e.g. Pajek, Excel, SPSS, etc.

iii. BibExcel's strength is higher degree of flexibility in managing data and analysis.

iv. Possible to use other data sources than Web of Science, in fact can deal with data other than bibliographic records.

v. Able import many different types of data.

4.2 CiteSpaceII (http://cluster.cis.drexel.edu/~cchen/citespace)

CiteSpace is developed by the Dr. Chaomei Chen, a Professor of Informatics, College of Computing and Informatics at Drexel University, Philadelphia, USA, for progressive knowledge domain visualization and analyzing trends and patterns in scientific literature. It is free Java based software (runs on Java Runtime) used for structural and temporal analyses of networks derived on publication data such as Collaboration Networks, Authors Co-citation Networks, and Document Co-citation networks.

Data Source: The major data source for CiteSpace input is ISI WoS database, also supports data from PubMed, arXiv, ADS (http://adswww.harvard.edu/), and NSF Award Abstracts.

Features: It performs various functions to facilitates the understanding and interpretation of visualization, network patterns and historical patterns (i.e. to identification of the fastest growing topics, to find out the hotspots of citations in particular subject domain, fragmentation of a publication network into clusters, automatic labelling of clusters with terms from citing articles, geospatial patterns of collaboration based on Google Earth.

4.3 Gephi (http://gephi.github.io)

Gephi is developed by the dedicated team of young engineers and researchers in computer sciences, web mining, network sciences and information visualization. Presently led by Mathieu Bastian, CTO. Gephi is an interactive Open Graph visualisation and exploration platform for analysis of all type of networks, complex systems, dynamic and hierarchical graphs developed by the Young Team of engineers and researchers in the subject area of computer sci-

ence, web mining, network sciences/complex networks and information visualisation.

Features: Gephi is available under GNU (General Public License), runs on Windows, Linux, MacOS X,. Gephi is used to reveal publication patterns and trends. Gephi used 3D render engine to display large graphs in real-time and to speed up the exploration. Gephi is built-on combined functionalities and flexible architecture to: explore, analyze, spatialize, filter, cluster, manipulate, export, all types of networks. Gephi is able to do exploratory data analysis, link analysis, social network analysis, biological network analysis and poster creation, etc.

4.4 HistCite (http://interest.science.thomsonreuters.com/forms/HistCite)

HistCite software package was developed by the Eugene Garfield, Father of Citation Index. It is used for the purpose of bibliometric analysis and information visualization. HistCite operates on Windows computers with Internet Explorer and trial version is available free of cost. HistCite is designed to identify the significant (most cited) papers retrieved in topical searches of the Web of Science. HistCite operates on windows with internet explorer.

Features: Being bibliometric analysis and information visualization software, it performs following core functions:

i. Performs specific application: Converting of bibliographies into diagrams (historiographs).

ii. It uses the bibliographic information (titles, authors, dates, author addresses, references, etc.) that describe published items for performing bibliometric analysis.

iii. It measures and studies various aspects of a specific scholarly knowledge domain.

iv. It analyse the literature published in particular knowledge domain.

v. Used to know the time and geographic region of publication.

vi. Identify prolific countries/ major contributors to particular subject.

vii. Identify the languages used frequently for publications.

viii. To know the important journals in a particular subject.

ix. To identify the key authors, institutions and articles in particular subject area.

x. To measure authors impact on each other.

4.5 Network Workbench (http://nwb.cns.iu.edu/)

Network Workbench (NWB) developed at Indiana University, USA, is another free Java based Network Analysis, Modelling and Visualization Toolkit. It is designed for handling Large-Scale data in the subjects of Biomedical, Social Science and Physics Research. It includes specific features for the execution of bibliometric studies.

Features

i. NWB supports design, evaluation, and operates in a unique distributed, shared resources environment for managing large-scale data for network analysis, modelling, and visualization.

ii. NWB supports network science research.

iii. NWB users can access or upload major network datasets.

iv. Effective algorithms make NWB to be able to perform large scale network analysis.

v. NWB makes its users to generate, run, and validate network models to study the structural and dynamics of a network efficiently.

vi. NWB provides advanced features for visualization and analysis for doing specific networks analysis

vii. Researchers will have access to the validated algorithms, developed in the past in personal time-consuming processes.

4.6. NodeXL (http://nodexl.codeplex.com/)

Created by Marc Smith and his team at Social Media Research Foundation. NodeXL is free open-source software designed for Microsoft Excel 2007 and 2010 which makes it to explore network graphs easily. NodeXL graph gallery gives access to the community generated network graphs. Different fonts for edge, vertex and group labels can be set in NodeXL. NodeXL has in built auto update feature.

Features

i. Flexible Import and Export of graphs in GraphML format, supports softwares like Pajek, UCINet, and matrix formats.

ii. NodeXL is directly connected to personal social networks i.e. Twitter, YouTube, Flickr and email, NodeXL uses several plug-ins to get information of personal networks from Facebook, Exchange, Wikis and WWW hyperlinks, etc.

iii. To reduce clutter in the graph scale zoom option available into areas of interest.

iv. Flexible Layout is possible by using force-directed algorithms used to lay down the graphs, or dragging of vertices using mouse.

v. In setting the colour, shape, size, label, and opacity of individual vertices network appearance can be easily adjusted by filling in worksheet cells.

vi. Using a set of sliders Dynamic filtering instantly hides vertices and edges.To automatically group powerful graph vertex NodeXL analyse their common attributes.

vii. NodeXL has advanced graph metrics as a result it easily calculates degree, betweenness centrality, closeness centrality, eigenvector centrality, PageRank, clustering coefficient, graph density etc.

viii. NodeXL is able to perform set of repeated tasks with a single click.

4.7 Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/ and http://pajek.imfm.si/doku.php)

Pajek is developed by Vladimir Batagelj and Andrej Mrvar (some procedures were contributed by Matjaz Zaversni). Pajek is non-commercial freely available software programme runs on windows. Pajek is able to execute varoius kinds of network analyses and visualizations activities. Bibexcel software may be used to format input data for Pajek.

4.8 Publish or Perish (http://www.harzing.com/pop.htm)

Publish or Perish (PoP) is a free software program that retrieves and analyzes academic citations Google Scholar to measure the research impact. PoP able to calculate the various citation based metrics and indexes. It retrieves citations of publications via Google Scholar and Microsoft Academic Search for metrics based analysis of citations. PoP was developed by Anne-Wil Harzing, Research Professor and Research Development Advisor of the ESCP Europe Busi-

ness School. It is designed to evaluate the impact of research work not covered in the ISI.

Feature

i. Publish or Perish is able to perform following metrics based analysis:

ii. Total number of papers and total number of citations in a particular subject.

iii. Average citations received by per paper, citations received by per author, papers received by per author, and citations received per year.

iv. Calculation of Hirsch's h-index and related parameters.

v. Calculation of Egghe's g-index.

vi. Calculation of the contemporary h-index.

vii. Calculation of three variations of individual h-indices.

viii. Calculation of the average annual increase in the individual h-index.

ix. Calculation of the age-weighted citation rate.

x. Calculation and analysis of the number of authors per paper.

4.9 R-Project (http://www.r-project.org/)

'R' was initially written by Robert Gentleman and Ross Ihaka, department of Statistics, University of Auckland. Project R is a language based environment for statistical computing and graphics formations. R is a GNU project capable of executing a wide variety of statistical computing (linear and nonlinear modelling of data, execution of classical statistical tests, time-series analysis, classification, clustering of data, etc.) and graphical formation techniques are highly extensible.

Features

i. The strongest ability of R is to produce the quality well-designed publication-plots, mathematical symbols and formulae as and when needed.

ii. R is available as open source software and supports a wide variety of UNIX based platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

iii. The R is an integrated software facilities data manipulation, calculation and graphical representation:

- Data handling and storage facility,
- Calculations on arrays for particular matrices R have a set of operators.
- For data analysis, graphical display whether on-screen or in hard R provides a large, coherent, integrated collection of intermediate tools.
- R is developed on a programming language which incorporates conditionals, loops, user-defined repeatable functions and input and output facilities.

4.10 Science Assessment Integrated Network Toolkit (SAINT) (http://www.rathenau.nl/en/themes/theme/project/bibliometric-software-tools/saint.html)

SAINT is a fully integrated software developed by the Rathenau Institute, The Hague, Netherlands, for handling of large scale data for bibliometric and patentometric research and one of the few packages can also be used for conversion of ISI data into relational database (dbm or accdb or sql files). SAINT is available for downloaded from the official. Source code is also available under open source licence to test and improve SAINT. This toolkit is easy to use

for making research easier and more efficient. The issues related to SAINT can be discussed with community members via discussion forum.

The 1.0 ready to use version has:

i. To download Web of Knowledge bibliographic data a parser program.

ii. A program for splitting the sentences into a database of separate words (i.e. useful in title or abstract level analysis).

iii. Tool for transformation of database table or query into matrix format that can be executed by i.e. Pajek for visualization.

4.11 Sitkis (https://sites.google.com/site/sitkisbibliometricanalysis/)

Sitkis is a free Java and MS Access based bibliometric software tool, helping researchers facilitating them during computation process of research analysis and evaluation of scientific information. Sitkis was developed exclusively for bibliometric analysis. It provides tools for extremely streamlined analysis of bibliometric networks. Utilizing Sitkis, one can calculate large amount of data in few minutes data that can take days to analyse.

4.12 VOSviewer (http://www.vosviewer.com/Home)

VOSviewer is a free Java based program, primarily developed by Nees Jan van Eck and Ludo Waltman at Centre for Science and Technology Studies (CWTS), Leiden University used for data analyses and constructing bibliometric networks visualisation. VOSviewer creates network maps of individual publications, co-authorship, co-citation network, keywords co-occurrence networks, etc. For constructing and visualizing co-occurrence networks of important terms extracted from a body of scientific literature. VOSviewer uses its inbuilt text mining function.

4.13 CitNetExplorer (http://www.citnetexplorer.nl/Home)

In addition like VOSviewer CitNetExplorer is also a free Java based software tool developed by Nees Jan van Eck and Ludo Waltman at Centre for Science and Technology Studies (CWTS), Leiden University. It is used for analyzing and visualizing citation networks of scientific publications for improved understanding to study the structure and dynamics of science communication. It supports direct data import from Web of Science for creation of citation networks. Citation networks can be explored interactively.

Features of CitNetExplorer include:

i. Analyse development of a research field over time.

ii. Identifying the literature on a research topic.

iii. Exploring the publication oeuvre of a researcher.

iv. Supporting literature reviewing.

Data Management

(a) Web of Science data import- data can be imported directly from the Web of Science database for network analysis.

(b) Pajek export- networks can be exported in the popular Pajek file format.

(c) Large networks. It supports very large networks, (including millions of publications and tens of millions of citation relations, are supported.

4.14 Loet Leydesdorff (http://www.leydesdorff.net/)

Loet Leydesdorff provides set of free DOS based software to analyse and evaluate bibliometrics data obtained from the data sources such as Scopus, ISI, and Google Scholar for metrics based analyses such as co-authorship (international and institutional collaboration, collaboration networks), co-words analysis, Co-citation and bibliographic analysis and much more. These softwares are also helpful for preparing data for the creation of relational databases and information visualization by other visualisation tools such as Pajek.

4.15 SciMAT (http://sci2s.ugr.es/scimat/description.html)

SciMAT stands for Science Mapping Analysis software Tool, is a Java based open source science mapping software tool, developed at the research group Sci2s at University of Granada, Spain. SciMAT has been supported by the Project of Spanish Ministry of Education and Science. SciMAT was developed by M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, and F. Herrera which incorporates methods, algorithms, and measures for the science mapping. It is free to perform a science mapping and analysis. SciMAT handles bibliographic data in Sqlite 3 format (amendable at any time), to carry out bibliometric analysis based studies. SciMAT uses several algorithms to edit data. Strategic diagrams, Cluster networks, and Evolution areas (are three techniques jointly used by SciMAT for visualisation). SciMAT has following modules:

(a) dedicated module for management of knowledge base,

(b) science mapping and analysis modules, and

(c) results and maps generation module for visualisation.

Features of SciMAT are

i. Loaders for handling ISI Web of Knowledge format and RIS format.

ii. Bibliometric networks: Bibliometric networks are based on co-word, co-citation, and bibliographic coupling.

iii. Pre-processing: It includes de-duplicating, time-slicing, data reduction and network reduction.

iv. Normalization: Normalization association strength, equivalence index, inclusion index, jaccard's index and salton's cosine are used for normalisation.

v. Mapping: For mapping of data simple Centers Algorithm, Single-linkage, Complete-linkage, Average-linkage and Sum-linkage clustering algorithms are used for mapping.

vi. Analysis: network analysis, performance and quality analysis, and temporal analysis can be executed using SciMAT.

vii. Visualization: for result visualisation strategic diagram, cluster network, overlapping map, evolution map can be created.

viii. Report: Outcomes generated in HTML and LaTeX format.

4.16 VantagePoint (https://www.thevantagepoint.com/)

Vantage Point is Commercial powerful textmining tool for discovering knowledge in search results from patent and literature databases. It has visualization capabilities. VantagePoint is a 32-bit program powerful commercial text-mining tool compatible with

Windows Vista, Windows 7, and Windows 8 platforms. VantagePoint is used for discovering knowledge in search results from patent and literature databases. VantagePoint helps to understand and works with search results retrieved from text databases.

Features of VantagePoint can be classified into following five broad categories

i. Importing – importing the data into VantagePoint and mining the raw data for getting more data from it.

ii. Cleaning – data cleaning means transforming the data into a consistent set, combining the things that needed to analyze as a group, and merging and normalizing data received from diverse sources. VantagePoint uses fuzzy matching techniques data cleaning.

iii. Analyzing – analysing gathered data in a variety of ways.

iv. Reporting – reporting prepares to display end results.

v. Automating – automating encodes entire process to make consistently repeatable.

4.17 CoPalRed (http://ec3.ugr.es/copalred/)

CoPalRed is a spceialised computer program developed by Rafael Bailón-Moreno, Department of Chemical Engineering, University of Granada, Spain. CoPalRed depended on the database for gethering information (i.e. Web of Science, Scopus, Medline, etc). This citation data is used for analysis and CoPalRed transforms this into new knowledge explicitly not avaiable and generated before. The common errors ocurred in a bibliographic databses are corrected or cleaned by the CoPalRed prior to any analysis by its self cleaning system. While doing analysis CoPalRed make preliminary filtering of the data recived and excutes three types of fully automated analysis:

i. Structural analysis. It shows the structural network formation of particular knowledge domain.

ii. Strategic analysis. Defining the relative position of actor within the network, by defining its intensity, external relations (centrality) and internal cohesion (density).

iii. Dynamic analysis. CoPalRed, analyzes the transformations actors over time in a network. It identifies approaches, forks, appearances and disappearances of actors in a network.

CoPalRed main modules are:

- Information capture module
- Module debugging information
- Generation module knowledgebase
- Knowledge Management Module.

4.18 IN-SPIRE™ (http://in-spire.pnnl.gov/)

IN-SPIRE™ Visual Document Analysis was developed by Pacific Northwest National Laboratory, US. It is licensed software and latest released is 5.9 version of IN-SPIRE. It has various tools for exploring textual data, tools for handling boolean and topical queries, and tools for time and trend analysis. The suite of tools allows the user to rapidly discover hidden information relationships by reading only pertinent documents. IN-SPIRE has been used to explore technical and patent literature, marketing and business documents, web data, accident and safety reports, newswire feeds and message traffic, and more. It has applications in many areas, including information analysis, strategic planning, and medi-

cal research. The core features of IN-SPIRE are (a) Quickly Creation of meaningful visualizations and (b) Able to explore and understand large textual database without reading every record. IN-SPIRE™ supports multiple types of text file formats (such as IN-SPIRE: supports encodings for ASCII, UTF-8, UTF-16 and also compatible with PDF, MS-Word, MS-Excel, and RTF files, email, XML, HTML, RSS and spreadsheets.

4.19 UCINET (https://sites.google.com/site/ucinetsoftware/home)

UCINET 6 for Windows is a software package for social network analysis, developed by Lin Freeman, Martin Everett and Steve Borgatti, and published by Analytic Technologies, USA (http://www.analytictech.com/). It incorporates NetDraw network visualization tool. UCINET 6 is available for download free of cost 90 day uses only or can be purchased. It allows students to purchase as discounted price. In addition a number of free tools, including Anthropac, NetDraw and KeyPlayer are also available. UCINET 6 is compatible with Windows operating system (i.e. NT, 98, XP, etc). It can be run via BootCamp, VMFusion Ware, Parallels or Wine to run it on Mac or Linux. It supports analysis of large datasets. UCINET 6 is able to analyse maximum network size about 2 million nodes but practically most UCINET procedures are very slow to run networks size larger than about 5000 nodes. However, the situation varies network to network.

5. Conclusion

The scientometric researchers need to know about the various popular data analysis, mapping and visualisation softwares. Each of such softwares has specific characteristics which are different from each other. Many of these softwares run in java run time environment (JRE). Most of these softwares are available either free of cost or under open source licensing. Some software are able to do scientometric analysis, some are able to create maps and networks, while some are specialised in information visualisation. It is very hard to choose any software as best among all. As many of these softwares are designed and developed in a way to capture the input data from the various popular data sources such as: Scopus, Web of Science, Google Scholar, PubMed and many others, now we can analyse the large scale data in lesser time. Many of the softwares have data convergence and compatibility feature which supports interoperability and crosswalk of input and output data. These scientometric analysis, mapping and visualisation softwares are now widely used by the researchers across the globe.

References

1. JASON Priem and BRADLEY M. Hemminger. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social. First Monday, Vol. 15, (7). Available at http://firstmonday.org/ojs/index.php/fm/article/viewArticle/2874/2570 (Accessed on 31/12/2014).
2. ONLINE DICTIONARY FOR LIBRARY AND INFORMATION SCIENCE. Software. Available at http://www.abc-clio.com/ODLIS/odlis_s.aspx (Accessed on 21/01/2015).
3. SANTHANAKARTHIKEYAN1, S. [et. al.]. Scientometrics Study on Web: Tools and Techniques, International Journal of Education Research and technology. Vol. 4 (1), pp.40-45. Available at http://soeagra.com/ijert/ijertmarch2013/7.pdf (Accessed on 19/12/2014).

4. WIKIPEDIA. Google Scholar. Available at http://en.wikipedia.org/wiki/Google_Scholar. (Accessed on 05/01/2015).

5. GOOGLE SCHOLAR. Google Scholar. Available at https://scholar.google.co.in/intl/en/scholar/metrics.html#coverage. (Accessed on 05/01/2015).

6. WIKIPEDIA. Web of science. Available at http://en.wikipedia.org/wiki/Web_of_Science. (Accessed on 16/01/2015).

7. WEB OF SCIENCE. Web of Science. Available at http://wokinfo.com/citationconnection. (Accessed on 19/01/2015).

8. WIKIPEDIA. Scopus. Available at http://en.wikipedia.org/wiki/Scopus. (Accessed on 21/01/2015).

9. ELSVIER. Scopus. Available at http://www.elsevier.com/online-tools/scopus. (Accessed on 16/01/2015).

10. NCBI. Pubmed. http://www.ncbi.nlm.nih.gov/pubmed/. (Accessed on 21/01/2015).

11. WIKIPEDIA. Pubmed. Available at http://en.wikipedia.org/wiki/PubMed. (Accessed on 16/01/2015).

12. UNIVERSITY OF VIENNA. Bibexcel. Available at http://homepage.univie.ac.at/juan.gorraiz/bibexcel/index.html. (Accessed on 29/12/2014).

13. ASTROM, Fredrik. (2009). Celebrating Scholarly Communication Studies Communication Studies: Festschrift for Olle Persson at his 60th Birthday. E-Newsletter of the International Society for Scientometrics and Informetrics. Vol. 5, pp.1-94. Available at http://www.soc.umu.se/digitalAssets/65/65934_ollepersson60.pdf. (Accessed on 29/12/2014).

14. CHEN, Chaomei. CiteSpace: visualizing patterns and trends in scientific literature. Available at http://cluster.cis.drexel.edu/~cchen/citespace/. (Accessed on 12/29/2014).

15. GEPHI. Gephi. Available at http://gephi.github.io/about/. (Accessed on 12/29/2014).

16. GARFIELD, Eugene. Historiograph Compilation HistCite Guide. Available at http://garfield.library.upenn.edu/histcomp/guide.html. (Accessed on 16/01/2015).

17. WIKIPEDIA. Histcite. Available at http://en.wikipedia.org/wiki/Histcite. (Accessed on 21/01/2015).

18. NWB Team. (2006). Network Workbench Tool. Indiana University, Northeastern University, and University of Michigan. Available at http://nwb.slis.indiana.edu. (Accessed on 19/12/2014).

19. SMITH, M. [et.al.]. (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel. Available at http://nodexl.codeplex.com. (Accessed on 15/01/2015).

20. VLADIMIR, Batagelj [et.al.]. Pajek: Program for Large Network Analysis. Available at http://pajek.imfm.si/doku.php?id=pajek. (Accessed on 12/29/2014).

21. HARZING, A.W. (2007). Publish or Perish. Available at http://www.harzing.com/pop.htm. (Accessed on 14/01/2015).

22. R-PROJECT. The R Project for Statistical Computing. Available at http://www.r-project.org. (Accessed on 14/12/2014).

23. Rathenau Institute. The infrastructure of knowledge: Bibliometric software tools Available at http://www.rathenau.nl/en/themes/theme/project/bibliometric-software-tools/saint.html. (Accessed on 17/12/2014).

24. SURULINATHI, M. Software. Available at http://smartlis.webnode.com/software. (Accessed on 14/01/2015).

25. VOSVIEWER. Vosviewer: visualising scientific landscape. Available at http://www.vosviewer.com/Home. (Accessed on 11/01/2015).

26. JAMALI, Hamid R. Scientometric portal. Available at https://sites.google.com/site/hjamali/scientometric-portal. (Accessed on 22/12/2014).

27. CITNETEXPLORER. Analysing citation pattern in scientific literature. Available at http://www.citnetexplorer.nl/Home. (Accessed on 08/01/2015).

28. COBO, M.J. [et.al.] (2012). SciMAT: a new Science Mapping Analysis Software Tool. Journal of the American Society for Information Science and Technology, Vol. 63 (8), pp. 1609-1630.

29. VANTAGEPOINT. VantagePoint: Serious Software for Serious Professionals. Available at https://www.thevantagepoint.com. (Accessed on 14/01/2015).

30. EC3. CopalRed: Spanish version 1.0. Available at http://ec3.ugr.es/copalred. (Accessed on 16/12/2015).

31. WIKIPEDIA. Copalred. Available at http://es.wikipedia.org/wiki/Copalred. (Accessed on 15/01/2015).

32. PNNL. IN-SPIRE™ Visual Document Analysis. Available at http://in-spire.pnnl.gov. (Accessed on 12/01/2015).

33. BORGATTI, S.P., [et.al.]. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies. Available at https://sites.google.com/site/ucinetsoftware/home. (Accessed on 15/01/2015).

About Authors

Mr. Ashok Kumar, Library and Information Officer, School of Planning and Architecture, New Delhi. Email: informashok1984@gmail.com

Dr. J Shivarama, Assistant Professor, Centre for Library and Information Management Studies at Tata Institute of Social Sciences, Mumbai.
Email:j.shivarama@tiss.edu

Mr. Puttaraj A Choukimath, Asst. Librarian (SS), with Sir Dorabji Tata Memorial Library of 'Tata Institute of Social Sciences', Mumbai.
Email: choukimath@tiss.edu