Evaluating Distribution and Influence of a Scientific Idea Using Citation Network

Rahila Hasanova[1,2]
Department of Scientific Information and Scientometrics
[1]Institute of Information Technology
[2]Baku Higher Oil School
Baku, Azerbaijan
0000-0002-8220-604X

***Abstract*: The paper examines the issues of evaluating the relevance of a scientific idea, its impact on other fields of science, and its distribution over time and geography using citation network. At the same time, the issues of evaluating the scientific idea and the identification of its real author are considered. Scientometrics analysis such as determining the weight and impact of the scientific idea in the area where it is spread using citation network are carried out.**

***Keywords*: citation network, citation analysis, scientific idea, scientometrics, bibliometrics, impact factor.**

I. INTRODUCTION

Citation analysis is one of the two main areas of bibliometrics (citation analysis and publication analysis) created for the analysis and study of science and scientific relations [1].

Although the term "bibliometrics" is used as a synonym for scientometrics and informetrics, it differs in its scope and essence. Coinciding with the advent of the Internet, Webometrics (or Cybermetrics) uses citation analysis and other principles and methods of bibliometrics to study the structure and characteristics of links on the Internet. For example, the Web Impact Factor indicator has been proposed to evaluate websites in a similar way to the Journal Impact Factor indicator used to evaluate journals [2]. Similarly, the analysis of inbound and outbound links to websites is similar to the "bibliographic citations" used in traditional link analysis [3]. The classic bibliometric laws - Bradford's law, Lotka's law and Zipf's law are also applied on the Internet [4,5].

Citation analysis is a field of scientific numerology that studies citation patterns and is applied to study the impact of research, knowledge flows, and knowledge networks. It is also widely used in knowledge representation and information retrieval.

In recent years, citation analysis has been widely applied to problems such as research, management, and information services, including university ranking, research evaluation, and knowledge domain visualization. This is due both to the availability and accessibility of digital bibliographic data (references, abstracts, etc.) in the online environment, and to the development of computer technology.

II. LITERATURE REVIEW

In general, citation analysis helps in acquiring new knowledge as well as managing and using existing knowledge and is used to: [6-9]:

1. Evaluation of information resources and scientific contributions. The presence of a reference article in the bibliographic list means that it was used by the author when writing the article. Citation is an indicator of the acceptance of an article by the scientific community. However, the number of references to all publications of the journal or to all articles written by a scientist has an impact on the journal, scientist, scientific institution, etc.. This type of citation analysis is called evaluative citation analysis [10].

2. Mapping research areas to explore their intellectual structure. This knowledge management is also intended to assess the characteristics, structure and development of research areas and learned societies. References are similar to the article of the referring author in terms of the method or topic used [11]. Thus, the mapping of literary networks, the analysis of relations in scientific publications based on citations, allows us to study the characteristics and structure of this field of science and society. By comparing periods of time, it is possible to model the history of the development of a field of science and a scientific society. When analyzing a citation network, the field of study is presented by set of objects - documents, authors, journals, etc.. Multivariate Statistical Analysis and Network Visualization Tools are also used to evaluate the relationships between these objects [8].

All these measurements, analyzes and visualizations together provide information about the structure and characteristics of the studied field of science and the scientific community, and also serve to study their development and predict future trends [10].

3 types of reference measurements are used to measure the relationship between two objects:

• reciprocal citation count: the number of citations 2 objects have to each other. This is called cross-citation analysis.

• number of joint links: the number of documents that any 2 objects link to. This is called citation analysis.

• bibliographic citation frequency (BCF): the number of common citations to any 2 entities. This is called bibliographic citation analysis.

Thus, if two documents cite each other frequently, if both documents are cited at the same time, and if both have the same citations, then these documents are similar in methodology and subject matter. [12,13].

3. Monitoring the flow of knowledge and the dissemination of ideas. Links inform the reader about the source of knowledge and ideas [14]. Analyzing how publications were subsequently used, one can trace the spread of ideas across scientific directions and time [15].

For this, a separate article or a set of small articles containing the topic under study, research methodology, and basic theory is determined. Then all references to this set of documents are collected and their distribution by scientific direction and time is studied.

4. Study of users of scientific literature and methods of use. The informational behavior of scientists and engineers is of

great interest to LIS researchers. Data collected through surveys constitute the main source of information for the study [16-19]. Citation analysis improves the results of this type of research by using a different approach to the information behavior of researchers (based on information obtained from citation behavior).

Examining all links made by a group of scientists provides information about their information behavior (what information they use in their scientific research). The frequency with which various types of information (magazine, book, electronic resource, etc.), years, languages, countries and subjects are processed by a group of authors is determined and a group of authors making the same references is compared with this group of authors.

The interaction of scientific directions, areas, journals, organizations or authors is assessed by examining scientific information presented with references [20, 21].

5. Organization, presentation and retrieval of information. Reference links are the only and effective way to find relevant literature, regardless of language or keywords. With this method of organizing information and searching, mobile links were created and access to full-text articles was provided through web technologies and cooperation with publishers, academic databases and libraries. Examples of these technologies and collaborations include CrossRef and OpenURL [22-24].

The reference database is a bibliographic database that contains reference indexes along with authors, keywords, subject headings. Reference indexes organize and present references in such a way that the researcher can easily search through an extensive list of references. However, it is possible to follow articles prior to the current article (by looking at the bibliography at the end of the article), as well as from older articles to newer ones. Users can start with the main article, author, journal, or a combination of both and get the article they wrote, all the publications they cite, and all the publications they cite, which is a very efficient way to search for information. The reference database provides a simple and complete collection for reference analysis. The results of citation analysis can help organize, present, and retrieve information.

For example, citation analysis evaluation helps to find publications of highly influential authors, institutions, countries, and citation network analysis facilitates understanding of the structure of the research area and the relationship between topics, publications and authors [25-30].

The Web of Science citation databases from Calrivate Analytics and Scopus from Elsevier value the use of citation analysis results (impact metrics such as citation count, h-index, and journal impact factor) in information retrieval systems. Search results can be ranked by impact metrics, which in turn allows users to work with more relevant resources.

Due to the dramatic and chaotic increase in the amount of information available on the Internet, research into appropriate ways of organizing and retrieving information is of great importance. Citation analysis provides unique and powerful ways to improve the organization, presentation, and retrieval of information. The best example of this is the web search engine Google, where a similar algorithm is applied to link analysis, which focuses on the information demand of the user and at the same time focuses on quality resources [31].

III. Citation Network Analysis

A citation is an indication of a link between author's article and the articles cited in this article in terms of methodology, topic etc. Multiple links between related documents - reference networks are analyzed to explore the intellectual structure of the research area, knowledge organization and information retrieval. By comparing time intervals, one can learn the history of the development of the field of study and update the knowledge organization system (for example, thesauri). All these are pertain to the problems of analysis of reference networks [10].

A sequence of steps has been proposed to perform a network analysis of links, [32,33]. To do this, a set of objects covering the study area is collected - articles, authors and journals. The level of their interaction is evaluated by the scores obtained through links, and as a result, the structure and characteristics of the study area are extracted through these relationships. Multivariate statistical analysis is used to determine the underlying structure of these relationships. Network visualization tools are used to visually display connections.

Citation network analisys consists of following 6 stages:

1. Definition of the field of study of science: collection of all reliable scientific publications related to this field. The analysis of the reference network depends on the structure and characteristics of a particular scientific field and the scientific community in a certain period of time. Depending on the task of the study, the time period can vary from one year to all years of the existence of that field. Similarly, areas of study and scientific societies can be defined in many ways, from science in general to specific areas [34].

2. Identification of the main sets of objects presented in the given field of study. With the data set collected in the first step, either directly or by selecting the main features, you can perform network feature processing and visualize this data to get a more accurate result. It is interesting that the development of computer technologies makes it possible to carry out network analysis on multidimensional clusters.

3. Measuring relationships between objects in kernel sets. After the main objects are selected, proximity of these pairs of main objects is measured from the referents' point of view. The results are presented through the rows and columns of the matrix. The matrix shows the reference grid of nuclear objects. There are many methods for measuring similarity, the most widely used of which is the Bibliographic Linking Frequency (BCF) method. 2 articles using the same sources in the bibliography are considered bibliographically related, and the number of identical sources is the BCF value for those 2 articles. The higher the BCF value, the closer the relationship between articles [35].

4. Multivariate statistical analysis. The above matrices (eg, citation matrix, co-citation matrix, and BCF matrix) can be analyzed using statistical analysis, network analysis, and

information visualization methods and techniques. The most commonly used statistical analyzes in the analysis of reference networks are factor analysis, cluster analysis, and multivariate scaling (MDS). Cluster analysis divides a set of objects into 2 or more groups (these groups are called clusters), so that the relationship between objects belonging to the same cluster is strong, and the relationship between objects belonging to different clusters is weak. Factor analysis selects a small number of factors based on multiple variables and assigns factor values to the variables. The factor values indicate the degree of association between the observed variables, which helps to classify variables and select those that belong to the category. MDS reduces N-dimensional interactions to smaller sizes (usually 2 or 3), allowing you to visualize and explore relationships between objects.

5. Network analysis and visualization. There are 3 main components in network analysis and visualization: the objects to be analyzed, the relationships between these objects, and display and visualization tools. The first two components correspond to the vertices and rays of the network, and the third corresponds to methods that represent the results of the network in 2- or 3-dimensional space. Network properties such as centrality can be calculated to evaluate the network properties and participating entities.

A large number of software packages have been developed for the analysis and visualization of reference networks (eg CiteSpace, VOSViewer, Sci2) as well as general networks (eg Pajek, UCINet). Some of these software packages are used for reference data.

6. Interpretation and verification of results. By visualizing multivariate statistical analysis such as cluster analysis, factor analysis, and MDS, relationships between objects can be easily discovered. Relationships between multiple objects are first converted into relationships between factors or clusters (which are much smaller in number), as well as relationships between objects and these groups, and then visualized. Based on this visualization, one can begin to interpret the results [32].

IV. Citatİon Networks Analysis and Visualization: VosViewer

Small experiments were carried out using the VosViewer software. From the Scopus database, we tried to explore the field of "journalology", which includes the problems of validity or forgery of journals considered relevant to the modern era. To do this, all articles (64) and references containing this term in the title, abstract or keywords of the article are saved as a CSV file, then a network of links with other keywords is established using the VosViewer program (Fig. 1). Figure 2 shows the citation network of authors related to that field.

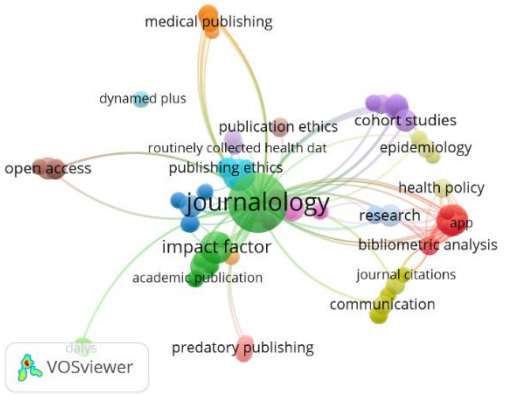

Fig. 1. A network of links in VosViewer to other keywords that match the keyword "journalology" in the Scopus database.

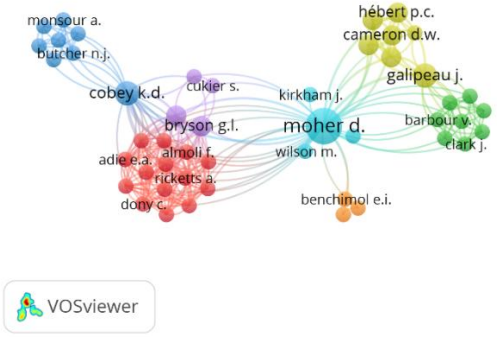

Fig. 2. Reference network of VosViewer authors matching the keyword "journalology" in the Scopus database

In this way, you can evaluate the impact factor of articles, authors, journals, and countries, visualize the link network, and perform multiple analyzes with the link network analysis and visualization tools. Using these tools, you can analyze research areas not only in international academic databases, but also with resources within the country - a single online library that works with national reference indexes. In this regard, the creation of the Azerbaijan Citation Index is of great importance [36,37].

Conclusİon

From studies on citation analysis, it can be concluded that researchers cite more articles that they have cited previously, published in recent years, and popular articles. Also, articles with similar titles are more likely to link to each other.

Thus, referential network analysis tools allow the researcher to evaluate the impact of a scientific idea on other scientific fields and disciplines, to study the interactions between authors, articles, scientific direction and countries. It would be satisfactory if these tools available on the Internet could also be applied to national resources. Thus, it is possible to support the process of scientific awarding by automatically identifying popular scientific fields, authors and articles in the country and conducting various ratings. To do this, first of all, issues such as the creation of a single database of all publications published in the country, then the launch of the Azerbaijan Citation Index and its integration into international academic databases should be resolved.

References

[1]. Fairthorne R. A. Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction. Journal of Documentation, 1969, 25, pp.319-343.

[2]. Ingwersen P. The calculation of Web impact factors. Journal of Documentation, 1998, 54, pp.236-243.

[3]. Rousseau R. Sitations: an exploratory study. Cybermetrics, 1997, 1(1).

[4]. Cui L. Rating health Web sites using the principles of citation analysis: A bibliometric approach. Journal of Medical Internet Research, 1999, 1(1).

[5]. Egghe L. New informetric aspects of the Internet: Some reflections, many problems. Journal of Information Science, 2000, 26, pp.329-335.

[6]. Garfield E. Citation indexing - its theory and application in science, technology, and humanities. New York: John Wiley and Sons, 1979.

[7]. Swanson D. R. Undiscovered public knowledge. Library Quarterly, 1986, 56(2), pp.103-118.

[8]. Small H. Visualizing science by citation mapping. Journal of the American Society for Information Science, 1999, 50, pp.799-813.

[9]. White H. D., Buzydlowski J., Lin X. Co-cited author maps as interfaces to digital libraries: Designing pathfinder networks in the humanities. IEEE International Conference on Information Visualization, July 18–22, 2000, London, England, pp.25-30.

[10]. Borgman C. L., Furner, J. Scholarly communication and bibliometrics. Annual Review of Information Science and Technology, 2002, 36, pp.3-72.

[11]. Cronin B. The citation process: The role and significance of citations in scientific communication. London: Taylor Graham, 1984.

[12]. Borgman C.L. Scholarly communication and bibliometrics. Newbury Park, CA: Sage Publications, Inc. 1990

[13]. White H. D. Author co-citation analysis: Overview and defense. Scholarly communication and bibliometrics, 1990, pp.84-106.

[14]. Rogers E. M., Cottrill C. A. An author co-citation analysis of two research traditions: Technology transfer and the diffusion of innovations. Scholarly communication and bibliometrics, 1990, pp.157-165.

[15]. McCain K. W. Eponymy and obliteration by incorporation: The case of the "Nash Equilibrium." Journal of the American Society for Information Science and Technology, 2011, 62(7), pp.1412-1424.

[16]. Brown C. M., Ortega L. Information-seeking behavior of physical science librarians: Does research inform practice? College and Research Libraries, 2005, 66(3), pp.231-247.

[17]. Brown C. M. The role of Web-based information in the scholarly communication of chemists: Citation and content analyses of American Chemical Society Journals. Journal of the American Society for Information Science and Technology, 2007, 58(13), pp.2055-2065.

[18]. Brown C. M. Communication in the sciences. Annual Review of Information Science and Technology, 2010, 44, pp.287-316.

[19]. Hemminger B. M., Lu D., Vaughan K. T. L., Adams S. J. Information seeking behavior of academic scientists. Journal of the American Society for Information Science and Technology, 2007, 58(14), pp.2205-2225.

[20]. Zitt M., Bassecoulard E. Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. Information Processing and Management, 2006, 42(6), pp.1513-1531.

[21]. Bassecoulard E., Lelu A., Zitt, M. A modular sequence of retrieval procedures to delineate a scientific field: From vocabulary to citations and back. In Proceedings of the 11th International Conference on Scientometrics and Informetrics, Madrid, Spain, June 25-27, 2007, pp.74-84.

[22]. Hitchcock S., Carr, L. Harris S., Hey J. M. N., Hall W. Citation linking: Improving access to online journals. Proceedings of the 2nd ACM International Conference on Digital Libraries, 1997, pp.115-122.

[23]. Hitchcock S., Carr L., Jiao Z., Bergmark D., Hall W., Lagoze C., Harnad S. Developing services for open eprint archives: Globalization, integration and the impact of links, 2000.

[24]. Van de Sompel H., Beit-Arie O. Open linking in the scholarly information environment using the OpenURL framework. D-Lib Magazine, 2001, 7(3).

[25]. Chen H. Visualizing semantic spaces and author cocitation networks in digital libraries. Information Processing and Management, 1999, 35, pp.401-420.

[26]. Chen H., Houston A. L., Sewell R. R., Schatz B. R. Internet browsing and searching: User evaluations of category map and concept space techniques. Journal of the American Society for Information Science, 1998, 49, pp.582-603.

[27]. Chen H., Martinez J., Kichhoff A., Ng. T. D., Schatz B. R. Alleviating search uncertainty through concept associations: Automatic indexing, co-occurrence analysis, and 166 BIBLIOGRAPHY parallel computing. Journal of the American Society for Information Science, 1998, 49, pp.206-216.

[28]. Ding Y., Chowdhury G. G., Foo S., Qian W. Bibliometric information retrieval system (BIRS): A Web search interface utilizing bibliometric research results. Journal of the American Society for Information Science, 2000, 51, pp.1190-1204.

[29]. Lin X., White H.D., Buzydlowski J.W. Real-time author co-citation mapping for online searching. Information Processing and Management, 2003, 39(5), pp.689-706.

[30]. Strotmann A., and Zhao, D). Bibliometric maps for aggregated visual browsing in digital libraries. Proceedings of The SIGIR 2008 Workshop on Aggregated Search, Singapore, July 24, 2008.

[31]. Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 1998, 30(1-7), pp.107-117.

[32]. McCain K. W. Mapping authors in intellectual space: A technical overview. Journal of the American Society for Information Science, 1990, 41, pp.433-443.

[33]. Zhao D. A comparative citation analysis study of Web-based and print journal-based scholarly communication in the XML research field (Doctoral dissertation, Florida State University), 2003.

[34]. Small H. On the shoulders of giants. Bulletin of the American Society for Information Science, 1999, 25(2), pp.23-25.

[35]. Kessler M. M. Bibliographic coupling between scientific papers. American Documentation, 1963, 14, pp.10-25.

[36]. Alguliyev R., Aliguliyev R., Fataliyev T., Hasanova R. Conceptual challenges in developing Azerbaijan citation index. IV International Conference "Problems of Cybernetics and Informatics" (PCI'2012), September 12-14, 2012, pp.50-53.

[37]. Alguliyev R., Aliguliyev R., Fataliyev T., Hasanova R. On necessity of the Azerbaijan citation index (AzCI). IV International Conference "Problems of Cybernetics and Informatics" (PCI'2012), September 12-14, 2012, pp.82-85.