

Bibliometric Analysis using Bibliometrix an R Package

Hamid Derviş

Department of Information Management and Records, Kastamonu University, Kastamonu, TURKEY.

ABSTRACT

This study aims to explore the usage of Open-source software in bibliometric analysis. Biblio-metrix an R package for bibliometric and co-citation analysis was used to achieve the research activities. R is an ecosystem software meaning all functions are shared in an open-source environment with the users. We have used Graphene as a subject of research for bibliometric analysis. Graphene is one of the fastest growing research fields in nanotechnology worldwide. A textual query on Web of Science (WoS) Clarivate Analytics using the term “graphene” was performed retrieving 1155 scholarly papers from 2000 to 2017 with having at least one author based in Turkey. Bibliometric results indicate graphene within nanotechnology as a scientific research field is growing steadily. Graphene not only is used in engineering but also can be used in medical technology. Furthermore, this is an ongoing research exploring an Open-source software and its roles in the field of information studies.

Keywords: Bibliometrix, Bibliometric Analysis, R, Graphene, Open-source software.

Correspondence

Hamid Derviş

Department of Information Management and Records, Kastamonu University, Kastamonu, TURKEY.

Email: hderivis@kastamonu.edu.tr

Received: 17-01-2019;

Revised: 15-03-2019;

Accepted: 11-12-2019.

DOI: 10.5530/jscires.8.3.32

INTRODUCTION

Nanotechnology is the study of materials at atomic levels within the 1 to 100 nm range. Naturally, “nano” means “molecular sized” (i.e., at a magnitude of 10^{-9} of a meter). With the advent of microscopic tools, scientists can investigate materials’ properties at nano-scale levels such as higher strength, lighter increased control of light spectrum, lighter weight, which are significant in products manufacturing. Nanotechnology in the mid-1990s was discovered as a research field and it quickly became an important research activity for scientists in a broad scientific field. In the worldwide level nanotechnology has become a research priority. The European Union (EU) has invested heavily in nanotechnology through its Framework Programs (FPs) since 2000. According to the UNESCO’s 2030 science report, developed and developing countries have included nanotechnology as a top research priority for their countries. Moreover, nanotechnology is one of the emerging research fields, which will have the positive impact on both developing and developed countries. For instance, while nanotechnology is used for water purification in emerging countries, it is used for manufacturing better chips for computers in developed countries. At the global level, the total number of publications in nanotechnology-related fields

has increased, due to governmental funding. For example, the Turkish government has also adopted a new approach by becoming the part of the EU’s framework programs and has invested heavily in research and development. The total number of scholarly publications between 1975 and 2015 are some 390 articles (Web of Science) which makes Turkey ranks 18th place in the world.^[1]

Graphene is one of the materials which inherently encompasses nanoscience properties has attracted many researchers from different scientific fields from Materials Science, Biomedical Applications to Technology and Devices. Consequently, the total number of scholarly publications have increased on Graphene, exponentially. Vargas-Quesada^[2] *et al.* mapped the intellectual structure of graphene in United States, China and Europe. The evolution in scientific publications in graphene started steadily in 1988–2003, increased substantially between 2004 and 2009 and reached its peak in 2010–2015. According to authors, it was in 2004 that worldwide publications in graphene picked up its momentum, starting with United States followed by China and EU countries. However, recently not only the top countries publishing scholarly articles in graphene but also Taiwan, Japan, Australia, India, Singapore, Russian Federations, Canada, Hong Kong, Malaysia, Brazil and Iran. Similarly, with the support of administrative council of nanotechnology strategy in South Africa in 2005 resulted in increasing scholarly publications in nanotechnology by Scientific and Industrial Research, University of the Witwatersrand and National Research Foundation.^[3] Although the total number of publications in nanotechnology is low in

Copyright

© The Author(s). 2019 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

South Africa compared to other members of OECD countries, authors stated that it could be increased by direct investment of governmental agencies and private sectors.

Since its discovery, graphene is one of hottest research field due to its variety of application in many different industries. Graphene with its unique molecular structure has a range usage from material science to condensed-matter physics.^[4] Moreover, Nobel Prize in Physics was awarded to scientists who made a breakthrough scientifically by extracting graphene from graphite.^[5] Investing in graphene is promising because scientists are continually experimenting innovative techniques using graphene. For example, according to Massachusetts Institute of Technology (MIT) review graphene can be used in semiconductor's industry instead of silicon, or it can be used for water filtrations, among others.^[6,7] Subsequently, the total number of scientific articles in graphene has started to grow, rapidly.

Quantitative Approach

Although there are several software tools which can be used by scientometricians to analyze or visualize the bibliometric data; selecting a tool by bibliometrician, it depends on what type of analysis is required. One can use Bibliometrix for analyzing and mapping of bibliographic data at the same time. Bibliometrix can be reviewed, changed and improved by the knowledge workers since it is an open source software written in R-packages. It is written in R, an open source language which is a large community of developers and users and, to date, consists of over 16,000 software packages. This means that bibliometrix can be used as a peace of a larger and general data analysis workflow.

Bibliometric methods are used to assess the productivity of scientific outputs quantitatively. Bibliometrics is defined as "the application of mathematical and statistical methods to books and other media of communication".^[8] Derek de Solla Price^[9] paved the way for scientific visualization in his seminal work entitled "Networks of Scientific Papers" by utilizing the bibliographic data of journals Eugene Garfield who pioneered the Science Citation Index was instrumental in establishing the scientometrics as a scientific field.

Although scientometricians use citation analysis to forecast diffusion of knowledge in scientific fields, citation analysis solely does not depict the actual domain of expertise in scientific areas. Bibliometric methods alone do not expose the social structure of the invisible college among scientists.^[10] Social network analysis was used by scientists to study people from different scientific fields such as anthropologists, psychologists, sociologists and recently physicists and mathematicians to measure the communication or diffusion of knowledge in groups, organizations or even countries. Information scientists utilize bibliometrics and Social Network Analysis^[11]

(SNA) methods to assess growth in science. While the former deals mainly with the effects of scientific productivity using citation analysis, the latter primarily focuses on the pattern of relationships among scientists. The network composed of co-authorship among scientists is an accurate indication of their cooperation in research activity. The network structure of nanotechnology research output of Turkey: a co-authorship and co-word analysis study. According to Wellman and Berkowitz,^[12] SNA is a paradigm. Theoretically, it is a premise based on a structured survey of human relations. Gestalt theory was instrumental in shaping SNA by sociologists in the early 1920s. Jacob Moreno and Kurt Lewin were the first scientists using SNA in the social sciences. Lewin,^[13] who worked on group behavior, argued that a person's attitude or behaviour is influenced by his/her position in the social group. Also, they integrated mathematical formulas from graph theory into SNA. Moreno^[14] used network analysis to show social structures among schoolchildren. Moreover, Milgram^[15] proved that no matter how complicated the network structure is, it takes a maximum of six steps from one node (person) in a social network structure for a message to be passed along to another node. According to Stanley Wassermann and Katherine Faust,^[16] the existence of relational information among network members is instrumental in defining and shaping of the network structure.

Bibliometricians apply mathematical and statistical methods to quantify the scholarly communications. Bibliographic data are processed through a workflow: study design, data collection, data analysis, data visualization and interpretation. Aria and Cuccurullo,^[17] stated that bibliometric analysis is a cumbersome activity, which contains many producers. However, there are automated software tools that are used by information scientists or practitioners.^[18] By extracting descriptive and network data within bibliographic literature, one can perform citation analysis. Citation analyses are used to reveal the scientific growth in a specific field at three levels: micro, macro and meso. The conventional method used in citation analysis is bibliographic coupling, co-citation, co-author and co-word. We can say there is bibliographic coupling^[19] between two documents if two records cite the same paper whereas co-citation^[20] measures the most citing reports. Co-author citation depicts the total number of co-occurrences authors' oeuvres in the network structure.^[21] On the other hand, co-word^[22] analysis maps the cognitive structure of the network in a period based on co-occurrence of the words in the abstract, title or keywords in the articles. Co-word citation induces the temporal development in a scientific field in which conceptual structures are formed by using textual discourse.

There are several software tools designed for bibliometric analysis by scholars. For example, BibExcel^[23] which is used to create co-citation networks by extracting a unit of study stored

within bibliographic data. Although BibExcel does not have graphing functionality, co-citation networks can be drawn in Pajek, Gephi, or VOSviewer, to name three. Pajek and Gephi both calculate degree centralities of co-citation networks. CiteSpace^[24] is a freeware Java application for visualizing and detecting trends within scholarly literature. Information scientists for analyzing and mapping bibliographic data from WoS and Scopus online databases mostly use VOSviewer.^[25] Both software creates overlays maps. Compared to most free software (e.g. CiteSpace and VOSviewer), bibliometrix does not focus only on the data visualization but also on the correctness and statistical completeness of the results. For example, VOSviewer only allows you to view the networks but not to analyze the collection according to the different levels of analysis (as proposed in biblioshiny: source impact, source dynamics, document analysis, word analysis, etc.).

Several studies have revealed the importance and, the role of R and its packages in vast scientific fields. One example, Li and Yan^[26] studied and mapped the usage of the R and its packages within the articles in Public Library of Science (PLoS). Bibliometrix is an R statistical package for analysing and visualizing the bibliographic data from WoS and Scopus databases. It is written in R language, which operates under GNU operating system. R is distributed and archived by CRAN network project (<https://cran.r-project.org/>). Aria and Cuccurullo stated that R an open source software with its rich statistical capabilities is an excellent choice for scientific computing. Moreover, R is an open-source environment ecosystem meaning it encompasses statistical algorithms, mathematical functionality and visualization capabilities, which makes it a good candidate for bibliometric analysis. R operates under Windows and Linux operating system environment, with a graphical user interface (RStudio), which makes it user-friendly for the novice or expert users.

Furthermore, bibliometrix proposes a different approach to analyze conceptual structure using Factorial Analysis (FA). FA is a well-known approach in Text Mining domain but it is still little used in science mapping. Bibliometrix covers the whole workflow while the other software only implements a part of it.

METHODS AND DATA

We conducted a topical query and downloaded 1155 bibliographic literature from the online WoS database using the textual term “graphene,” in 2000–2017, with at least an address of Turkey. The term “graphene*” (* = wildcard plural) exclusively was searched in title, abstract, author keywords

(<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>)

(<https://gephi.org>)

(<https://www.r-project.org/>)

and keywords plus in this explanatory article. We choose two citations indexes: Science Citation Index Expanded (SCI-EXPANDED) and Conference Proceedings Citation Index-Science (CPCI-S) from WoS core collections for inquiry. While we downloaded 1155 records, the number of files might change since by the time the more articles are published. All document types: articles, proceeding papers, meeting abstract, editorial material, review and correction. Were selected First, we installed the latest version 1.442 of RStudio in Windows 10. Second, we established the bibliometrix package within R environment to analyze and map the bibliographic data, if it has not been install yet. Then, we used bibliometrix’s functions to create descriptive and co-citation network, respectively. The functions readfile and convert2df embedded in bibliometrix are used. While readfile function load and convert text data to UTF-8 format, convert2df function extract and creates a data frame, which corresponds to the unit of analysis within the exported file from WoS. Eventually, the function biblioAnalysis generates a descriptive data from bibliographic data. The results can be drawn by the generic function (plot) in R. This article aims to conduct a bibliometric and co-citation analysis to answer the following research questions:

- Which authors are the most productive authors in graphene?
- What is the annual scientific publication growth in graphene?
- Which countries do collaborate regarding graphene with Turkey?
- Which journal do scientists mostly publish their articles?
- Who are the most cited scientists?
- What are the conceptual structures of the field graphene?

FINDING AND DISCUSSION

4519 authors wrote a total number of 1052 of articles. It is quite a high number of authors. Fundings have motivated more scientists to participate in publishing scientific articles (This is an assumption). Collaboration is the key amongst authors whereby 19 authors have published solely. The annual percentage growth rate for scientific production is about 47.047, which indicate a steady growth (see Table 1 and Figure 1).

Table 2 and Figure 2 describes and depicts the most prolific authors, respectively. Professor Cıracı has been directing research institutes at Ihsan Doğramacı Bilkent University at Ankara for past 15 years in Turkey. He is followed by Atar, Yola, Kocabas, Shahin, Yakuphanoglu, Metin, Eren, Sen and Balcı. All scientists mentioned above have been instrumental in the diffusion of nanotechnology, especially graphene in

Table 1: Main Information about data.

Articles	1052
Sources (Journals, Books, etc.)	354
Average citation per article	17.66
Authors appearances	4519
Authors of single-authored articles	19
Articles per author	0.463
Authors per Article	2.16
Co-Authors per Articles	4.3
Collaboration Index	2.26

Table 2: The most productive authors.

Articles	1052
Sources (Journals, Books, etc.)	354
Average citation per article	17.66
Authors appearances	4519
Authors of single-authored articles	19
Articles per author	0.463
Authors per Article	2.16
Co-Authors per Articles	4.3
Collaboration Index	2.26

Table 3: The most published articles in journals.

Number of Journals	Number of articles
Physical Review B	69
RSC Advances	39
International Journal of Hydrogen Energy	34
Applied Physics Letters	26
Applied Surface Science	21
Journal of Alloy and compounds	20
Electroanalysis	19
Journal of Physical Chemistry C	18
Electrochimica Acta	14
Composites part B-engineering	13

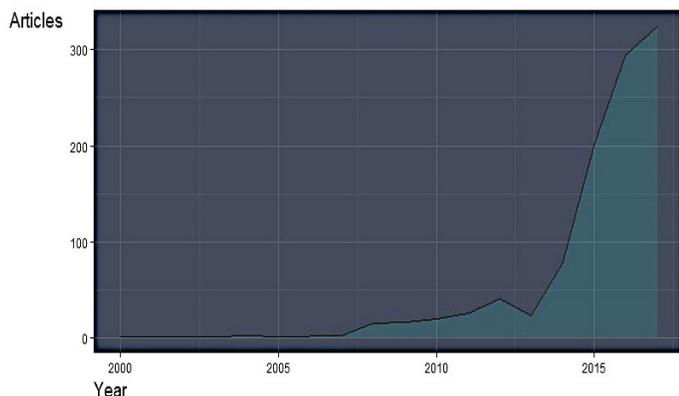


Figure 1: Annual scientific production.

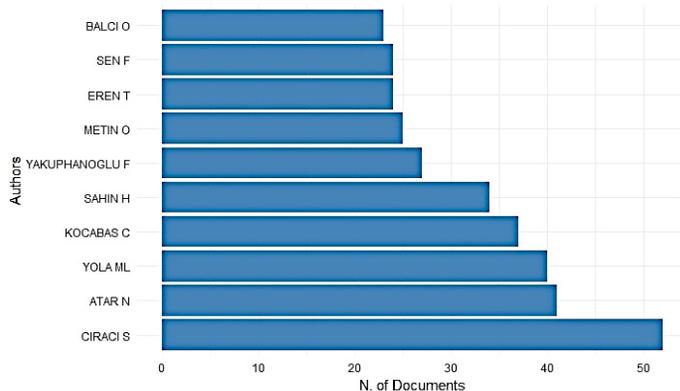


Figure 2: The most productive authors in graphene generated by bibliometrix.

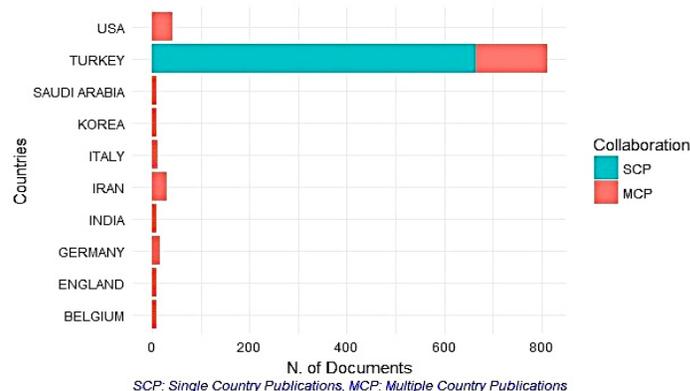


Figure 3: The most productive papers based on collaborative publications.

Turkey. Furthermore, mostly papers written by above authors have the highest citations. Precisely, the first top manuscript has received 172 citations per year (Not shown here). Most productive countries regarding article frequencies are Turkey, the USA followed by Iran (See Figure 3).

Turkish scholars have published a total of 69 articles in the Journal of Physical Review and 39 papers in the RCA advances, which is an online-only peer-reviewed scientific journal in Chemistry, respectively (see Table 3).

Co-citation network structure reveals the most cited people are Novoselov and Geim to whom the Nobel Prize in Physics was awarded. The network structure consists of clusters in which each color represent a component (see Figure 4). Figure 5 shows the temporal structure of the words from 2000 to 2017 depicted research development in the area of graphene. Three clusters are consisting the terms: graphite, transistors, water, adsorption, film and nanosheets and so on. Graphene has many applications in real world. Its unique structure has created an abundance of materials for us to use.

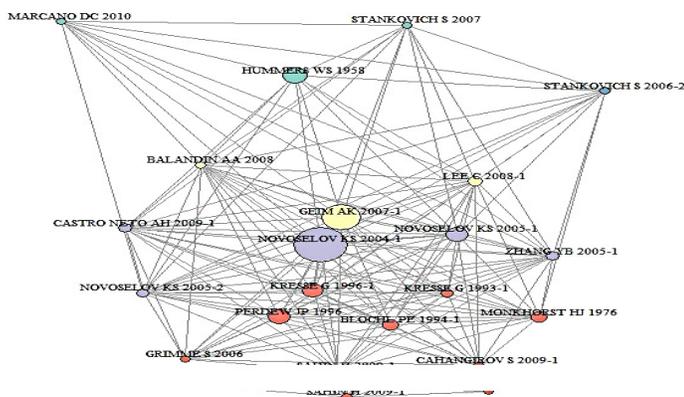


Figure 4: Co-citation network of scientists in graphene.

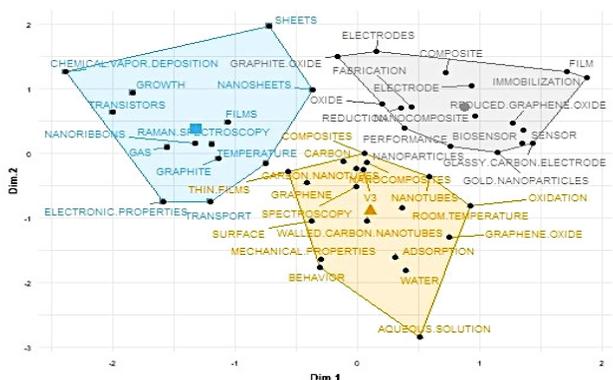


Figure 5: The conceptual structure of graphene.

CONCLUSION

Bibliometrix is a package for bibliometric analysis written in R. R is an ecosystem software meaning it operates in an integrated environment consist of open-libraries, the open-algorithm and open-graphical software. To explore the bibliometrix package, we studied the longitudinal development of graphene between 2000 and 2017 using data acquired from WoS. Graphene is essential because of the Nobel Prize in Physics was awarded to the scientists who extracted graphene from graphite. Graphene not only is used in engineering science but also it is used in biotechnology. There are many software tools for bibliometric analysis. However, using them is cumbersome for novice users. One of the advantages of Open-source tools is that their codes can be studied and modified in a constantly changing scientific field such as the bibliometric analysis. Therefore, further applications of Open-source software is inevitable in bibliometric analysis. Furthermore, this is an ongoing research exploring an Open-source software and its roles in the field of information studies.

ACKNOWLEDGEMENT

Author has not received any financial assistance for this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

WoS: Web of Science; SNA: Social Network Analysis.

REFERENCES

1. Tonta Y. TÜBİTAK Türkiye adresli uluslararası bilimsel yayınları teşvik (UBYT) Programının Değerlendirilmesi. Tübitak UlakbİM. 2017. Available from: http://ulakbim.tubitak.gov.tr/sites/images/Ulakbim/tonta_ubyt.pdf.
2. Vargus-Quesada B, Chinchilla-Rodríguez Z, Rodríguez N. Identification and visualization of the intellectual structure in graphene research. *Front Res Metr Anal.* 2017;2:7. Available from: <https://www.frontiersin.org/articles/10.3389/frma.2017.00007/full>
3. Makhoba X, Pouris A. Bibliometric analysis of the development of nanoscience research in South Africa. *S Afr J Sci.* 2017;113(11-2):1-9.
4. Luv PH, Wang GF, Wan Y, Liu J, Liu Q, Fei-chang MA. Bibliometric trend analysis on global graphene research. *Scientometr.* 2011;88(2):399-419.
5. Nobelprize. The 2010 Nobel Prize in physics. 2010. Available from https://www.nobelprize.org/nobel_prizes/physics/laureates/2010/press.html.
6. Bourzac K. When Will Graphene Be Commonly Used?. *MIT Technology Review.* 2015. Available from: <https://www.technologyreview.com/s/536451/>.
7. Condliffe J. The world's narrowest fluid channel could transform filtration of water and gases. *MIT Technology Review.* 2017. Available from: <https://www.technologyreview.com/lists/innovators-under-35/2017/inventor/radha-boya/>
8. Pritchard A. Statistical Bibliography or Bibliometrics?. *J Doc.* 1969;25(4):348-9.
9. DePrice DJS. Networks of scientific papers. *Science.* 1965;49:510-5.
10. Crane D. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*; Chicago, IL: University of Chicago Press. 1972.
11. Darvish H, Tonta Y. The Network Structure of Nanotechnology Research Output of Turkey: A Co-Authorship And Co-Word Analysis Study. *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Istanbul: Boğaziçi University Turkey.* 2015;15:732-43s.
12. Wellman B, Berkowitz SD. *Social Structures: A Network Approach.* 2nd ed. Cambridge: University of Cambridge. 1997.
13. Lewin K. *Field Theory in Social Science.* New York: Harper. 1951.
14. Moreno J. Who shall Survive?. *Sociometr.* 1934;6:206-13.
15. Milgram S. The Small World Problem. *Psychology Today.* 1967;2(1):60-7.
16. Wasserman S, Faust K. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press. 1994.
17. Aria M, Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis. *J of Informetr.* 2017;11(4):959-75.
18. Guler AT, Waaijer CJ, Palmblad M. Scientific workflows for bibliometrics. *Scientometr.* 2016;107(2):385-98.
19. Kessler MM. Bibliographic coupling between scientific papers. *J Am Soc Inf Sci.* 1963;14(1):10-25.
20. Small H. Co-citation in the scientific literature: A new measure of the relationship between two documents. *J Am Soc Inf Sci.* 1973;24(4):265-9.
21. White HD, McCain KW. Visualizing a Discipline: An Author Co-citation Analysis of Information Science 1972-1995. *J Am Soc Inf Sci.* 1998;49(4):327-55.
22. Calon M, Courtial JP, Turner WA, Bauin S. From translations to problematic networks: An introduction to co-word analysis. *So Sci Inf.* 1983;22(2):191-235.
23. Persson O, Danell JR, Wiborg S. How to use Bibexcel for various types of bibliometric analysis. In *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday.* International Society for Scientometrics and Informetrics. 2009;5-S:5-24.
24. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci.* 2006;57(3):35977.
25. Eck NJV, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics.* 2010;84(2):523-38.
26. Li K, Yan E. Co-mention network of R packages: Scientific impact and clustering structure. *J of Inform.* 2018;12(1):87-100.