

Eugenia Smyrnova-Trybulska, Nataliia Morze, Olena Kuzminska

Mapping and Visualization of a Research Network: Case Study

Edukacja - Technika - Informatyka nr 1(23), 327-332

2018

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach
dozwolonego użytku.



EUGENIA SMYRNOVA-TRYBULSKA¹, NATALIJA MORZE²,
OLENA KUZMINSKA³

Mapping and Visualization of a Research Network: Case Study

¹ Doktor habilitowany, University of Silesia in Katowice, Faculty of Ethnology and Sciences of Education in Cieszyn, Head of Department of Humanistic Education and Auxiliary Sciences of Pedagogy, Poland

² Profesor doktor habilitowany, Borys Grinchenko Kyiv University, Corresponding Member of National Academy of Pedagogical Sciences of Ukraine, Vice-rector on Informational Technologies, Ukraine

³ PhD, University of Life and Environmental Sciences of Ukraine, Department of Information and Distance Technology, Ukraine

Abstract

The article focuses on trends and methods as well as tools used for mapping and visualization of scientific domains as a research methodology which is attracting more and more interest from scientific information and science studies professionals; several factors which have contributed to such developments are also discussed. Science mapping or bibliometric mapping is a spatial representation of how disciplines, fields, specialties, and individual documents or authors are related to one another (Small, 1999). Different approaches have been developed to extract networks using the selected units of analysis (authors, documents, journals, terms, etc.). Furthermore, the authors looks at the most popular computer programs used for the processing of bibliographic and bibliometric data. In addition, plans and activities are presented concerning a comprehensive analysis of IRNet project (www.irnet.us.edu.pl) networking activities and numerous publications, using special ICT tools and study methods, first of all Pajek and presenting a several examples.

Keywords: networking, bibliographic and bibliometric data, mapping, visualization, applications

Introduction

Mapping and visualization of scientific domains as a research methodology is gaining more and more popularity among scientific information and science studies professionals. At least two factors have contributed to that. The first is the digitization of resources, scientific literature and other sources of knowledge, and hence, growth and development of bibliographic databases available online, such as ISI Web of Knowledge, Scopus, Google Scholar. The other is a non-linear increase in computing power, use of supercomputers and distributed technology (Chen et al., 2010; Zhao, Strotmann, 2008; Osińska, 2010). Conception of the mapping and visualization of scientific domains as a research methodology presented on the Figure 1.

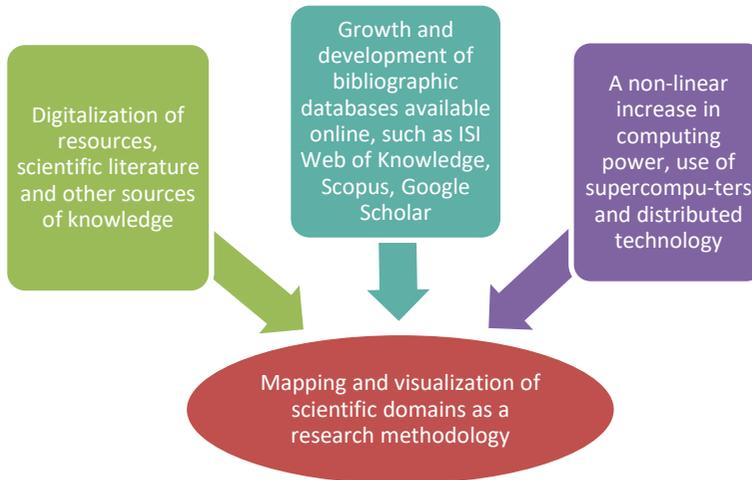


Figure 1. Mapping and visualization of scientific domains as a research methodology

Mapping and Visualization of IRNet project Networking Activities

Within the framework of the IRNet project (www.irnet.us.edu.pl) research is being conducted in several WPs which are separate, yet simultaneously connected through interrelated stages, which roughly address our several research questions. We created a bibliographic database containing all the published articles, books from 2014 to date by all IRNet researchers (178 publications).

This database also includes information regarding the number of authors, authors' affiliation, country, year, research areas, keywords, the source of publication (e.g., journal, book, etc.) and certain other data. Consequently, this dataset enables us to assess the main trends in IRNet scientific publications production. The time frame of the analysis is the last 4 years of IRNet existence, in which we have been able to trace its knowledge production and dissemination.

Based on the dynamics of international co-authorships, we will be able to map and trace international collaboration patterns and thus infer IRNet geographical influence scope, i.e. its international influence (Research Question).

By means of additional research utilizing information available in the Institute for Scientific Information (ISI), namely in the Science Citation Index (SCI), we assess the geographical pattern of the citations of IRNet scientific activities, publications, production. This enables us to evaluate to what extent IRNet project scientific production has been cited at the European and world level.

Methodology of research

The main indicators of research effectiveness – bibliometric indicators – is a powerful information tool to support the development of science. So today, all

over the world, they constitute obligatory elements in the reporting of scientific institutions, universities, research groups and individual scientists.

In this case, we consider an information analyst as a tool for monitoring scientific information flow within the framework of the international project IRNet, the influence of the publications of its members in the international scientific information space, as well as predictive and advisory activities to improve representativeness of publications in international scientometric databases, the approximation of the level of scientific publications of individual scientists to world standards.

The articles became the theoretical basis of research, they show the current processes in Biblio – and scientometrics. The article's metadata of participants of the project became the practical base and they were placed in the scientometric database BDWeb of Science and other international journals. Practical research methods were prepared using the programs: Bibexel (metadata analyses), and program Pajek: the graph editing and visualization of the graph structure of co-authorship – of the method of Kamada-Kawai (Kamada, Kawai, 1988).

For the study of science as a process of scientific communication, which is relevant to this study, one can use the tools of network analysis. As a tool for analysis and visualization of large scale networks we can recommend Pajek (Pajek, 1996). However, for complex types of analysis, one must have competencies helping one to work with specialized products and to develop programming skills.

To carry out a network analysis of participants of the project we will build a network of co-authorship for research activity and presentation of results in scientific publications.

Analyses of data for the purpose of network analysis were performed using Bibexel, starting with obtaining initial data for this program. These data (up to end of 2016) were assembled using built-in tools of the scientometric database Web of Science. Data for each participant in the project IRNet are files with the extension .txt, which were combined using the program Bibexel: in the new window Type new/file name here, enter name which needs to have the merged file, select all the files you want to merge, and run Files -> Append all selected files to another.

With this program, the files were created for visualization in the Pajek program: the file edit Analiz.net (Mapping -> Create net-file for Pajek), and the file number of publications of each participant of the project.vec (Mapping -> Create vec file). Because not all publications of the project participants belonged to the metric databases of the Web of Science database, editing of data for network analysis was performed manually in Notepad.

The procedures in Pajek's main window are organized in accordance with the types of data objects that they use as input. Partitions, Vectors, Permutations can be used to store properties of vertices. Objects of Cluster type define subsets of vertices according to certain characteristics.

For building a graphics editing program Pajek (<http://mrvar.fdv.unilj.si/pajek/>) the following objects were used:

1. Networks are the main objects (vertices and lines), and the default extension is: .net. In this case, the created file is Analiz.net of the 83 vertices – authors' articles and links between them, reflecting the co-authorship in publications.

2. Vectors – file Analiz.vec, contains 83 entries, where for each vertex (author) its quantitative characteristic (real number) – the number of publications is indicated.

3. Partitions – file Analiz.clu, where for each vertex the class is defined to which it belongs. In this case, there are 10 classes, in line with the countries' mission project participants:

To build a network visualization of co-authorship in the program Pajek, it is necessary to download the file Analiz.net in the Networks field. Vectors – Analiz.vec Partitions – file Analiz.clu, and select Draw -> Network + First Partition+ First vector.

The result (Figure 2) obtained is the visualization of the connectivity graph of the authors in the framework of the project.

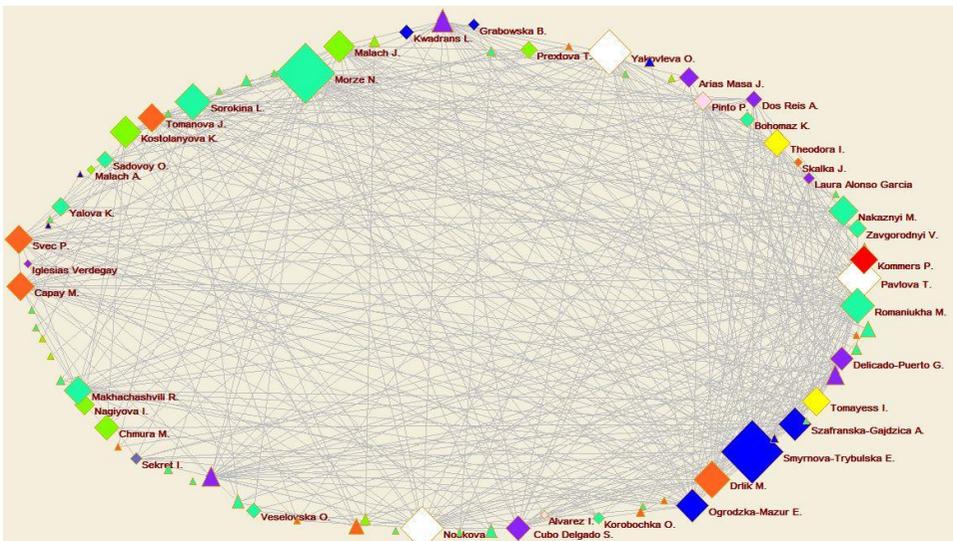


Figure 2. Example of network's visualization of co-authorship

Source: own work.

As for building the visualization of networks linking scientists with special software and settings, project participants' "labeled" rectangles (diamond) and the corresponding vertices have the signatures – the names of the scientists, not the participants – triangles (Triangle); shapes-marks are of different colors depending on the country, which represent the authors, and the size that is determined by the number of each author's publications. As can be seen in the Figure the scientists have different publication activity, some who are not the participants of the project have a greater number of publications than some of the participants. The latter can be considered as a basis for widening the circle of participants or an invitation of active scientists to other projects.

The distribution of the contribution of the project participants by the number of posts is not uniform: the center is represented by more active members. For each network node the author's name and the number of publications are added (which also affects the size of the geometric area). The color is also allocated to the country offices participating in the project.

It should be noted that the visualization, basically, is a brief synthesis of the results of the analysis that allows us to understand the context of the transition to the data that underpin it. The credibility presented in the form of a visualization of the analysis results will be high, if there is a verification of the results of the analysis, that is, all the visualization elements retrieved can be traced back to the primary data and methods of their processing.

The results of this visualization can be used to make optimal decisions on the management of information resources with the aim of improving processes of scientific communication and the evaluation publications contribution of each participant. To determine the scientific impact of the project participants in the development of pedagogical science an additional analysis needs to be conducted.

Conclusion

This article presents trends, methods as well as tools used for the mapping and visualization of scientific domains as a research methodology which is gaining more and more popularity among scientific information and science studies professionals; furthermore, several factors contributing to this development are discussed.

In the market there is an extensive offering of various types of software for processing of bibliographic and bibliometric data; software for mapping and visualization as well as interdisciplinary research – open source license as well as commercial license programs.

The author of the article, a coordinator of the international IRNet project, describes certain plans, methodologies and activities relating to comprehensive review of IRNet project networking activities and numerous publications, using special ICT tools and study methods; results of these activities will be presented in subsequent publications.

References

- Chen, Ch., Ibekwe-Sanjuan, F., Hou, J. (2010). The Structure and Dynamics of Co-citation Clusters: A Multiple-perspective Co-citation Analysis. *Journal of the American Society for Information Science and Technology*, 61 (7), 1386–1409.
- Kamada, T., Kawai, S., (1988). An Algorithm for Drawing General Undirected Graphs. *Information Processing Letters*, 31, 7–15.
- Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *zin*, 2 (96), 41–51. Pobrane z: <http://repozytorium.umk.pl/handle/item/463> (1.02.2017).
- Osińska, W. (2016). *Wizualizacja informacji*. Toruń: Wyd. UMK.
- Program Pajek. Pobrane z: <http://mrvar.fdv.unilj.si/pajek/> (1.02.2017).
- Small, H. (1999). Visualizing Science by Citation Mapping. *Journal of the American Society for Information Science*, 50 (9), 799–813.
- Smyrnova-Trybulska, E. (2017). *Mapping and Visualization of Scientific Bibliometric Domains and Research Network Activities*. In: M. Chruby (ed.), *DLSC2017 Conference Proceedings* (p. 301–308). Brno: University of Defence.
- Zhao, D., Strotmann, A. (2008). Information Science during the first Decade of the Web: An Enriched Author Cocitation Analysis. *Journal of the American Society for Information Science and Technology*, 59 (6), 916–937.

Acknowledgement

The research leading to these results has received, within the framework of the IRNet project, funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement No: PIRSES-GA-2013-612536