The detection of actual research topics using co-word networks

Lande D.V., Balagura I.V, , Andrushchenko V.B.

Institute for Information Recording
of National academy of science of Ukraine
Kyiv, Ukraine
balaguraira@gmail.com

Abstract—The purpose of research is to compare co-word and co-author networks from Web of science and show the possibilities of scientific priorities' searching for scientific teams in computer science. The paper proposed a new algorithms for detecting actual research topics, which is calculated on the analysis of co-word and co-author networks. We offer to compare co-words networks of certain team and the main world's tendencies. The algorithm is available to search common topics for international researches of scientific teams and compares co-words networks of certain team and worlds co-word network. The algorithm determines the scientific topics that are relevant to global trends, as well as the scientific clusters that perform them. The examples of calculation of the algorithm for computer science for authors from Belarus have been proposed. The results of the research can be used to determine the scientific topics of research to address the urgent needs of any country, to correct the research of organizations.

Keywords—scientometrics; co-author networks; co-word network; actual research topics;computer science; Belarus.

I. INTRODUCTION

Researchers are always searching for the most actual problems of our society and trying to solve them. There are well-known international science resources, which are used for searching and publication main scientific results . Arrangement of publications by their significance on certain topic is very helpful in such databases. There are numbers of rating systems world-wide which are being used for publications, journals and scientists ranking. Citing indexes, eigen-factor and others are widely used for the scientific work estimation and all of them are based on the amount of a publication citing [1]. Unfortunately there are some papers, that aren't popular and aren't cited. The main point of the popularity is the actuality of topic in the world in the right moment. So, the actual research topic identification is very important for any scientific work. Project "State of Innovation" defines the most current innovative trends and the most innovative companies in the world [2]. The researchers are determined the dynamics of growth in the number of publications in the database Web of Science for assessing the origin of innovation and database Derwent World Patents Index for assessing the innovative capacity and so on. The global leaders in the number of patent research and development are China and South Korea according to the reports. For example, in information technology computing is by far the most active subsector, comprising 83 percent of IT's

overall activity. Development and innovation could be identified with new research areas. The methods are based on the study of citations networks and abstracts of patent databases, citation time and frequency downloads. Another project of Web of science database is "research fronts", that consist the list of 100 "hot" papers (which is 0.1 percents of publications on the same industry and the same period, the most cited papers in a short period of two years) and 44 new "research fronts" (new research areas that marked a significant increase in citations of "hot" papers). There are used 10 research fields and co-citation maps. In the last report Research Fronts 2016 of Clarivate Analytics we could see the most actual themes in mathematics, computer science and engineering. It is mainly focus on hesitant fuzzy sets, configuration design and heat transfer analysis, the Keller-Segel chemotaxis model, partial differential equations, cloud manufacturing, the internet of things, multiple-input multiple-output systems, measurement-device-independent quantum key distribution, Li-ion batteries and bio-inspired algorithms [3]. ǀ

Most actual and new research topics may be recognized using of all described methods. But researches of some scientific teams don't match with actual topics [4]. At this time it is expedient to find main priorities of science development and choose right topics for international cooperation. Research teams already have some scientific heritage and haven't to change all topics.

The purpose of research is to compare co-word and co-author networks from Web of science and show the possibilities of scientific priorities' searching for scientific teams in computer science. We propose to use the methods of co-author and co-word network analyzes on the base of international resources. It will allow to analyze main key-words and topics of Ukrainian researches and to find main points, which is most close to most actual international researches.

Co-authors network is a network structure where nodes are scientists and links are co-authorship, size of nodes and width of lines are depends of network characteristics and common paper numbers. A co-word network are built on occurred pairs of terms and shows their interconnections. According to the algorithm from paper [5] terms will be extracted using frequency characteristic in abstracts. Co-word and co-author networks could be used for identification and description of scientific groups and research topics, the most communicative

Figure 1. Co-author network

researchers and main principles of science communication. For the analysis we have to use main principles and instruments of complex networks, that are described in many works [2-9]. Co-word networks often are used to describe the rsearch topic statistically for identifying themes and trends in a subject domain [10,11].

Co-word networks research teams publications allow to find common "narrowed" line of research with a clearly defined system of concepts (terms); common terminology may differ in detail from the general in a separate sciences; reduce the noise information that facilitates the work of experts in the knowledge that forms the model domain. In [8] co-word analysis was employed to reveal patterns and trends in the research by measuring the association strength of terms representative of relevant publications.

We use bibliographic information from Web of science. As an example the analysis of publications of computer science for Belarus authors are considered. Web of science is the biggest and high quality research databases. Lots of articles dedicated to the scientometric analysis of this resource.

A. Development of the algorithm

We assume that crossing of key-words from scientific team papers with whole co-word network of research area could show topics and authors, which are the most close for international researches and their work could be adjust to higher

level. And we could call them actual topics for the scientific team.

So, proposed algorithm of detection of priority scientific topics consists of next stage:

- Detecting of the area of researching.
- Co-author networks building and calculating of main characteristics, extracting scientific teams and the most communicative authors of country [12, 13].
- Co-word networks building and main words groups, matching with technical dictionaries [14].
- Co-word networks building for all researching area.
- Calculating of $(A \cap B)$ for co-word networks.
- Searching of $(A \cap B)$ in co-word networks of our teams and matching with co-author networks, searching the teams with the crossing of words from the international networks.
- Describing research area of possible cooperation with foreign partners, building international co-author networks, searching for partners.

B. Co-author and co-word network analysis

Web of science is the main scene for the international representation of researches produced by Clarivate Analytics company. About 50000 Belarus papers from the Core Collection since 1975 were indexed by database. Optics,
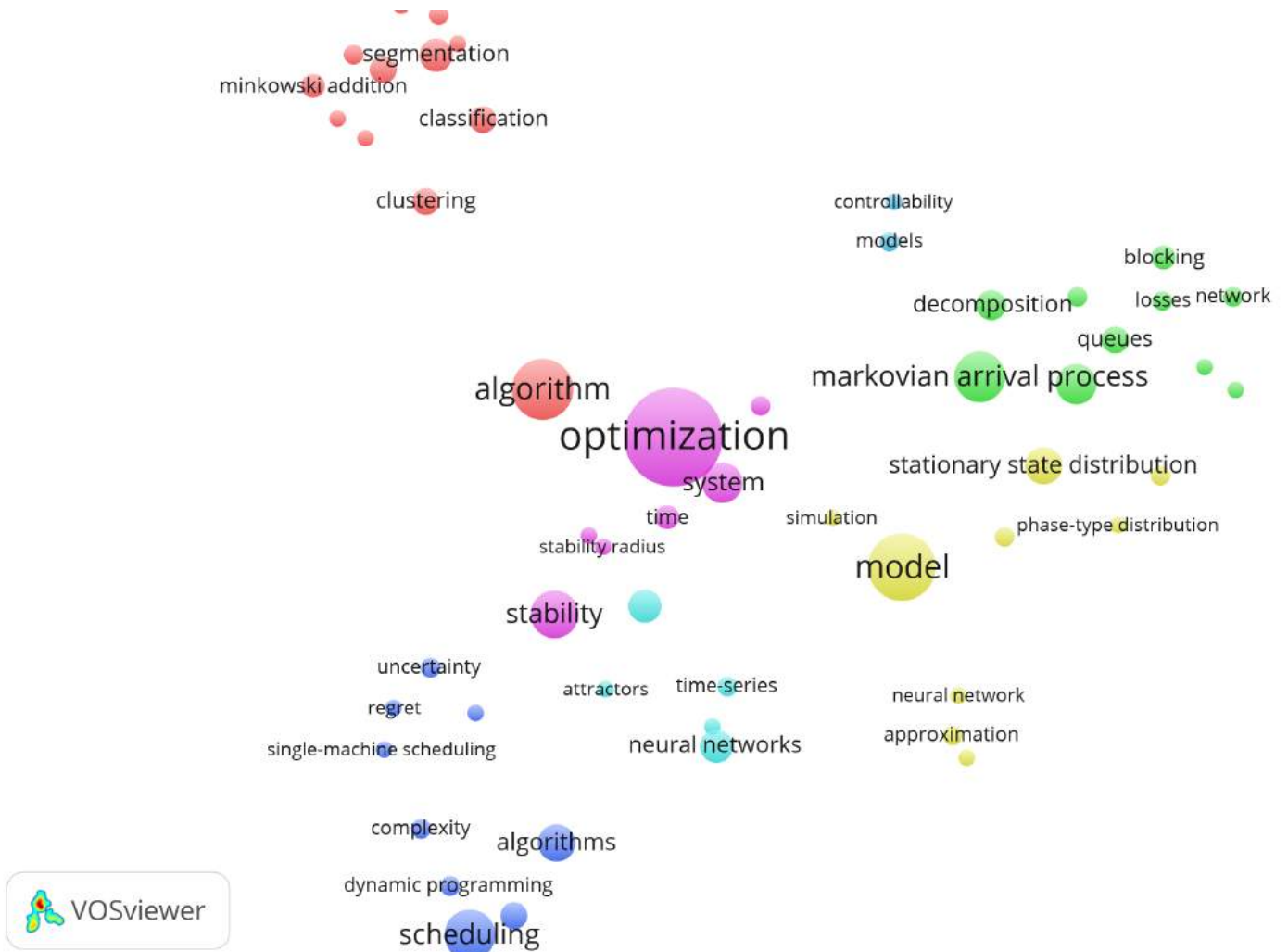
208

Figure 2. Co-word network

applied physics, material science, engineering are main areas of the publications from Belarus in Web of science. Web of science contains about 964 documents of computer science with Belarus scientists in 1997-2017 years. This part was analyzed with VOSviewer and Pajec tools. The co-author, co-word, co-citation networks was build. (Fig. 1, Fig.2).

According to proposed algorithm:

- We selected The computer science in Web of science data base.
- We chose the country Belarus, so have to build co-author network and search for clusters (Fig.1). Input data file was .txt format.
- Co-word network of the country team consist of 2357 key-words and shown on Fig.2 The distance between the nodes is inversely proportional to the number of coincidental occurrences of terms. Main terms according to clusters are optimization, stability, system; algorithm, clustering, segmentation, classification; model, stationary state distribution, neural networks, approximation, simu-

lation; markovian arival process, decomposition, blocking and others.

- Co-word networks building for all researching area. About 20 thousand words were used: complex networks, neural networks, quantum computers, recognition, statistical analysis, big data, simulation and others. Most papers dedicated to interdisciplinary applications.
- We used own software and detect 100 common words for Belarus and world publications. Among them: neural networks, algorithm, optimization, clustering, system, process, programming models and others.
- Last steps could be used by experts for describing actual research topics according to key-words. For more detail description we could use descriptions of main scientific teams from co-author networks and technical classifications. Also the co-word map we could use for the base of ontology for different scientific area.

Previously in [15], the testing of the algorithm was presented on the example of the analysis of the term networks

on scientific publications in physics in the Scopus database of the research team of the Kiev National University Taras Shevchenko. Scopus contains more Ukrainian and Belorussian articles than Web of science. And there are more possibilities for searching common research topics among separate organizations. So for complex analysis of actual topics of the country it will be useful to work with all databases, which are available. The results of scientometric analysis is depends on input data and it is necessary to choose the most full database for detecting future collaboration. For example, if we choose searching for Ukrainian and Belorussian cooperation we could use the biggest national resources and transform terms to one language. Also widely used in Ukraine and Belarus are Google Schoolar, Arxiv, Index Copernicus, РИНЦ and even such resources as Research Gate. But Web of science and Scopus could be used for the results revising as the most authoritative resources for international collaboration.

II. Conclusion

The algorithms, which based on co-word and co-author analysis for detecting actual topics and priorities, are proposed. We offer to compare co-words networks of certain team and the main world's tendencies. The algorithm is for activation of international research and popularity of papers witch is depend on topics. Using of the algorithm could help scientists to get more authority in the world. The results of the research can be used to determine the scientific topics of research to address the urgent needs of any country, to correct the research of organizations. The co-author and co-word networks of research output for computer science from Belarus were built, which shown the possibility for detecting the most popular research topics. Detected key-words are neural networks, algorithm, optimization, clastering, system, process, programing models and others. A further research need to use and compare several databases for increasing number of papers and searching scientific teams.

References

[1] Milojevic S., Radicchi F., Bar-IIan J. " Citation success index – An intuitive pair-wise journal comparison metric ", Journal of Informetrics, vol. 11, issue 1, pp. 223-231, 2017.
[2] " State of Innovation Report 2017 ", Clarivate Analytics, 143 p., 2017.
[3] " Research fronts 2016 ", Institutes of Science and Development, Chinese Academy of Sciences, The National Science Library, Chinese Academy of Sciences, Clarivate Analytics, 100 p., 2016.
[4] Newman M., Strogatz S., Watts D. " Random graphs with arbitrary degree distributions and their applications ", Physical review E, vol. 2, issue 64, pp.116, 2001.
[5] Lande D. , Snarskii A. , Yagunova E. , Pronoza E. "Volskaya Hierarchies of Terms on the Euromaidan Events: Networks and Respondents " Perception 12th International Workshop on Natural Language Processing and Cognitive Science NLPCS 2015 Proceedings, 22-24 September 2015, pp. 127–139, materials, Krakow, Poland., 22-24 September, 2015.
[6] Albert R., Albert-Laszlo Barabasi " Statistical mechanics of complex networks", Reviews of Modern Physics, vol. 74, pp 47-97, 2002.
[7] Bolshakova H. , Vorontsov K. , Efremova N. , Klyshinskiy E., Loukachevitch N., Sapin A. " Automatic treatment texts on natural language and analysis of data: training allowance. ", in Russian, NIU HSE, 269 p. , 2017.
[8] Wiesner A., Morbach J., Marquardt W. " Information integration in chemical process engineering based on semantic technologies ", Computers & Chemical Engineering, vol. 4, issue 35, pp. 692–708, 2011.
[9] Evangelos Pafilis, Sune P. Frankild, Julia Schnetzer, Lucia Fanini, Sarah Faulwetter, Christina Pavloudi, Katerina Vasileiadou, Patrick Leary, Jennifer Hammock, Katja Schulz, Cynthia Sims Parr, Christos Arvanitidis, and Lars Juhl Jensen " ENVIRONMENTS and EOL: identification of Environment Ontology terms in text and the annotation of the Encyclopedia of Life Bioinformatics " Bioinformatics (Oxford), vol. 11 , issue 31, pp. 1872-1874, 2015.
[10] Choi Jaewoo, Kang Woonsun " Themes and Trends in Korean Educational Technology Research: A Social Network Analysis of Keywords " Procedia - Social and Behavioral Sciences , vol. 131, pp.. 171–176, 2014.
[11] Xiuwen Chen, Jianming Chen, Dengsheng Wu, Yongjia Xie, Jing Li " Themes and Trends in Korean Educational Technology Research: A Social Network Analysis of Keywords "Procedia Computer Science, vol. 91, pp.. 547–555, 2016.
[12] Franceschini Fiorenzo, Maisano Domenico, Mastrogiacomo Luca " Empirical analysis and classification of database errors in Scopus and Web of Science " Journal of Informetrics, vol. 10,issue 4, pp.. 933–953, 2016.
[13] Balagura I. V., Lande D. V. " Linguistic studies of research collaboration using the «Ukrainika naukova» abstracts database ", Data recording, storage and processing, vol. 16, issue 3, pp. 45–53, 2014.
[14] Lande D. V., Andrushchenko V. B., Balagura I. V. " Formation of the subject area on the base of wikipedia service ", Open Semantic Technologies for Intelligent Systems, pp. 211–2014, materials, Minsk, February,2017.
[15] Lande D. V., Balagura I. V., Pohorilyi S. D., Dubchak N. A. "Scientometric analysis of co-word networks taken from publications of a scientific team ", Data recording, storage and processing, vol. 19, issue 1, pp. 34–39, 2017.

ОПРЕДЕЛЕНИЕ АКТУАЛЬНЫХ НАУЧНЫХ НАПРАВЛЕНИЙ С ПОМОЩЬЮ СЕТЕЙ ТЕРМИНОВ

Ландэ Д.В., Андрущенко В.Б., Балагура И.В.
Институт проблем регистрации информации
Национальной академии наук Украины
Киев, Украина

Цель исследования - сравнить сети терминов и соавторов, построенные на основе реферативной базы данных Web of science, показать возможности поиска научных приоритетов по информатике. В статье предложен алгоритм поиска актуальных направлений с помощью анализа сетей авторов и терминов. Предлагается сравнить сети терминов научных коллективов и основные мировые тенденции. Алгоритм заключается в поиске общих тем для международных исследований научных коллективов, при этом происходит стравнение сетей терминов коллектива и общей для научного направления. Выполняется определение научных тем актуальных на мировом уровне, а также научные коллективы, которые их выполняют. Рассмотрен пример использования алгоритма для авторов из Беларуси в области информатики. Результаты исследования могут быть использованы для определения перспективных научных тем в стране и отдельных организациях на основе реферативных и научных баз данных.