Detection of Expert Groups for Scientific Expertise

Iryna Balagura[1][0000-0001-9627-2091], Valentyna Andrushchenko[2]

and Ivan Gorbov[1][0000-0001-6888-0866]

[1] Institute for Information Recording of National Academy of Sciences of Ukraine, Kyiv, Ukraine
[2] National Information Center for Ukraine-EU S&T cooperation, Kyiv, Ukraine
balaguraira@gmail.com

Abstract. Today there is no unified model for selecting experts or specialists on formal grounds, therefore this project proposes an approach to solving the problem of searching for experts based on networks of co-authors. The method of research of scientific collectives is proposed, which allows to find groups of scientists whose activities correspond to the topics on the basis of the network of terms according to the given terms of a certain domain. The work of the algorithm is realized on the basis of information from the Web of science, a comparison of the work of the algorithm with the system Aminer. The proposed method allows searching for more flexible requests. It is advisable to use the authors' databases while searching for scientific groups to exclude identical names and obtain detailed information.

Keywords: Co-author Networks, Co-word Networks, Experts, Scientific Teams, Scientific Database.

1 Introduction

There is no unified model for formal searching of experts or specialists at present time. Therefore, we propose the method of searching of experts using co-authors networks. The actual task is the selection of competent experts for the involvement of scientists in solving of important public problems. Additionally, Ukrainian Government passed the Law of Ukraine "On Scientific and Scientific and Technical Activity" of December 26, 2015, which initiated changes in science in Ukraine, so the qualitative expertise is also needed for developing of the science [1].The urgent for search and formation of the expert groups are also associated with the Ukraine participation in the program "Horizon 2020" and taking the place of Ukrainian scientists as experts in competitive projects [2]. The defining of actual research directions where Ukraine could represent main scientific results to the world community is one of the state priorities. Formation of expert teams is necessary for this task. The proposed method can act as an essential tool for objective decision-making in the creation of such groups. Using the proposed method could help at objective search of international experts to assess challenges and reform of individual sectors in the state based on the

respective scientific papers. Fast and skilled searching of qualified expert groups will assist to the effective implementation of the Law of Ukraine "On scientific and technical expertise" and the Law of Ukraine "On innovation activity." [3].

Another actual task is searching experts for scientific collaborations. On the one hand there are multiple ways to find the partners for forming the multidisciplinary collaboration, and on the other hand, all these instruments are boiled down to generalized ways to find the collaborators via the theme of research and short information on the desired project and describing the work opportunities in research institutions. There are such opportunities like Enterprise Europe Network cooperation opportunities database

(http://een.ec.europa.eu/tools/services/SearchCenter/Search/ProfileSimpleSearch). CORDIS partner search platform (https://cordis.europa.eu/home_en.html). Participants Portal Partner Search (https://ec.europa.eu/research/ participants/portal/desktop/en/organisations/partner_search.html).

Enterprise Europe Network, cooperation opportunities database, provides an opportunity for the search by using the search line and precise options. CORDIS partner search platform implies the usage of the text phrases to find the needed information on the results of supported projects. The new instrument of the Horizon – 2020 participants' portal gives new opportunities for the grantees. It presumes to fill in several options, including the keywords. There are other ways such as using a social network for researchers (ORCID and Research Gate) and through the groups and proposals of business- and employment-oriented service like LinkedIn.

However, there can be applied other attempts, which can allow more precise approaches to find correlations with current research and to widen the frames of the work and its enlargement. Scientometrics – the measures of the research work through investigating the publication activity of the author, research team or research institution. The information on publications can represent the basis for acquiring new arrays of information and provide its further procession to receive new data which can appear new instruments for decision making. Citation indicators are used for describing the nature and main statistics of research. Co-author networks are showed individuals cooperation and could be used for predict next work.

The main idea of the study is to develop the method of qualified and formal detection of expert groups for scientific and technical expertise according to theirs skills and publications.

2 Existing approaches to the search for expert teams

Expert searches can be found on many Ukrainian and foreign scientific, cultural and other websites. It is a method of questioning, and it is conducted in an online or offline form [4]. For candidates, there are specific requirements for experience, skills and job positions for examining one or another field. This expert search method is currently the most widespread in the world. However, for not all types of work, it is possible to announce a long open competition of tasks, and not in each case, the specialist will

pay attention to the announcement. Therefore, the search for experts through questionnaires and contests is not the most effective.

The expert searches are also possible using scientific databases, which contains scientists profiles with information about the publications, citations and other scientometric data. Scientific profiles can be found in Google Scholar, Scopus, Web of science and other databases. There are also resources to combine information about scientists from different databases, including ORCID (https://orcid.org/), "Bibliometrics of Ukrainian Science" (http://www.nbuviap.gov.ua/bpnu/), "Scientists of Ukraine" (http://irbis-nbuv.gov.ua), Aminer (https://aminer.org/) and others. ORCID is a worldwide service that provides a union of publications and identifications of persons of the author. "Bibliometry of Ukrainian-science", "Scientists of Ukraine" created by the VI Vernadsky National Library of Ukraine in the first case on the base of scientific profiles from world scientific and technical resources, in the second case - information on abstracts of papers. AMiner is a database of scientists from artificial intelligence containing 130614292 scholars created in China. AMiner provides detailed information about the author: distribution of publications according to topics, citation, the h-index, g-index, collaboration network, place of work, key skills and ranking among others in the scientific areas. Aminer conducts the sorting, for example, the country, the language of publications, gender, authorship resources, h-indexes and detailed sections, as well as a defined concept [6]. The main disadvantage of such sites is the search restriction according to the identified thematic sections and the absence of a full picture of scientific directions and teams.

In the paper [7] a comparative analysis of the methods of the automatic search of experts was conducted. The expert search models are divided into probabilistic (computations for authors or topics), ranking and network (PageRank, HITS). The choice of method depends on the purpose of the expert search and data source. It is the input data that most determine the result, so the first stage of choosing a database of publications is the most important.

3 The search method of expert teams

It is reasonable for seeking of experts for solving governmental tasks to select out of the research teams that were formed within certain scientific schools and choose specialists from different groups. The sources of experts could are big scientific databases containing information on scientific activities in Ukraine. We propose to use the analysis of co-authors networks for the definition of the scientific groups using algorithms based on modularity [8].

The research deals with the investigation of co-authors and co-words networks appliance for scientometric analysis of abstract databases for describing of the main scientific areas structure. We propose to use the methods of co-author and co-word network analyzes on the base of abstract databases. Co-author network is a network structure where nodes are scientists and links are co-authorships, size of nodes and width of lines are depends of network characteristics and common paper numbers.

Co-word networks could be built on occurred pairs of terms and shows their interconnections. According to the algorithm from paper [9] 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 researchers and main principles of science communication. For the analysis we have to use main principles and instruments of complex networks, that are decrypted in many works [9-16].

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.

We define the co-authorship network G = (V, E) as an object given by a pair of sets (V, E), where V is the set of nodes (authors), E × V × V is the set of edges (co-authorship links) . Then the network of terms can be represented in the form T = (C, L), where C is the set of terms written by the authors of V, and L is the set of relationships between terms. By connecting GUT we get a heterogeneous network H containing vertices V and C, the links E, L and M - the connection between the vertices V and C (the relation the author used the terms in the publications). It is proposed to explore the network of collaborators gradually and to select scientific groups and leaders of scientific directions according to the subject, followed by a network of terms to determine the map of the subject area. The combination of networks of co-authors and network of terms makes it possible to identify scientific groups that are most precisely included in the given problem.

On the first stage there are defined the field and the scientific concepts, by which analyses are provided and then the review file is downloaded and filtered. The result of the first and second stage is data about authors filtered by the several descriptors and connections between them, i.e. the matrix appropriate of the network.

The next stage is forming of the co-authorship network according to the chosen subject field, and the main characteristics and also the calculation of additional parameters. As the result, the main characteristics of scholar's cooperation, scientific collaborations and the most communicative researchers in frames of the definite scientific concepts are defined.

Networks are divided into the clusters using the modularity measure. The modularity of a node is a value that evaluates the density of bonds in a coherent component in comparison with the bundles between the components. In general terms, modularity can be defined as:

$$Q = \sum_{i=1}^{N} \left(e_{ii} - a_i \right)$$

where e_{ij} – the element of the matrix adjacency graph, equal to the ratio of the number of edges, which combine two societies i and j, to the total number of edges in the network, a_i – the ratio of the number of edges connecting the vertices in the community, to the total number of edges:

$$a_i = \sum_{j=1}^{N} e_{ij}$$

The high modularity of the network indicates a strong connection in the clusters and the weak link of the network itself [15].

The fourth stage is dedicated to the selection of full text publications of the most communicative scholars and organizing of the text package to extract core terms (words and word combinations) per scientific concepts. The visualization of terms networks and the concept integrally, calculation of the main parameters are performed. And generalization of results, description of the core characteristics, field trends are also provided on this stage. Then we have to add two networks and form heterogeneous network which consists of co-word and co-authored networks. The father analysis of the union network could identify teams with the most relevant to the concept researches and forecast possible cooperation. Also we propose to rank scientists with centrality measures that allow detect the most collaborative scientists. The identification of important nodes in the network is an urgent task and requires a detailed study of the subject of research, since there are many coefficients that provide versatile characterization of the vertices, and the feasibility of their application is determined only with the purpose of the experiments. The degree centrality, which in fact is an indicator of the number of articles in collaboration, reflects the volume of work of the author, and the number of author ties characterize the circle of his co-authors. In fig. 1 level of centrality of nodes corresponds to their value. Authors with a high level of centrality index are linked in separate groups - scientific groups and not related.

The degree centrality estimate the authors by their communicability and can be used to predict the authors productivity. Centrality degree in weighted graph is:

$$C_D^{\omega\alpha}(i) = k_i^{(1-\alpha)} s_i^{\alpha},$$

where k_i – the sum of edges of between nodes:

$$k_i = \sum_{j=1}^{N} m_{ij}$$

s_i – the sum of weights of the edges, α - coefficient which choosing depends on the case [16]:

$$s_i = \sum_{j=1}^{N} \omega_{ij}$$

Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes:

$$C_B(i) = \sum_{j<k} g_{jk}(i), \quad i \neq j, k.$$

where $g_{jk}(i)$ – number of shortest paths that crossing vertex i [16].

In the sense of scientific collaboration betweenness centrality provides the ability to identify authors which are linking scholarly teams.

So, we propose search method of experts, which contains the following steps:

1. The definition of the field in which the search and related concepts are conducted.
2. The filtration of data from a scientific database.
3. The formation of a co-author network, the main characteristics of the network are calculated, and the network is split into groups based on modularity.
4. The formation of a network of terms for each team of collaborators.
5. The formation of separate heterogeneous networks, which combine co-authors and terms according to the clusters formed in 3.
6. The identification of teams with the most relevant concepts.
7. A ranking of scholars within individual teams is carried out according to the centrality measures.
8. The expert list is formed accordingly.

4 Experimental Results and Discussion

We choose the Web of science database as a source of data. Web of science (Clarivate analytics)consists of 8,700 carefully selected journals (Core Collection) and the citation rates of scientific literature. The main citation indicator of this platform is the Impact Factor (Index of Influence) of the scientific publication. The Web of Science (WoS) includes Science Citation Index Expanded (from the natural sciences), Social Sciences Citation Index (from Social Sciences), Arts and Humanities Citation Index (on Arts and Humanities), Emerging Index (new editions) and other. Science Citation Index Expanded covers 6650 journals of 150 disciplines (astronomy, chemistry, biology, biochemistry, mathematics, physics, medicine, science of materials, pharmacology, etc.). Search depth since 1975. Social Sciences Citation Index - covers 1950 journals of 50 disciplines (anthropology, history, jurisprudence, linguistics, philosophy, politics, psychology, sociology, etc.). Search depth since 1975. Arts and Humanities Citation Index - A Bachelor of Arts and Humanities. It covers 1160 journals in the fields of art, folklore, history, linguistics, archeology, literature, music, philosophy, poetry, religion, theater, radio and television. Search depth since 1975. The web of science contains resecherID, which allows you to identify the author, to identify the main scientific metrics and to rank among experts by citing, collaborating, and publishing activity.

The Web of Science was used as a database for the implementation of the algorithm and the team of the National aviation university as a sample. In the abstract database Web of Science 631 publications of the authors of the NAU were found. The largest number of publications is devoted to aviation and technical sciences, in the humanities and the arts the university is represented by two publications, which al-

most did not appear on the network of terms. The co-authors and a co-word networks was created using the software VOSviewer [6], as shown in fig. 1. The nodes represent the authors, and the edges show the co-authorship of these authors in articles. The larger the node, the more articles published by the author. The network of scientific cooperation assigns nodes to several clusters that describe scientific groups [7]. The clusters are marked with different colors.

The relationship between the terms in publications is shown in Fig. 2, where each node represents a term. The size of the node corresponds to the frequency with which this term was used. Depending on the clusters by a frequency of simultaneous use the terms are displayed in different colors. The same color is indicated by terms that are often used together in one publication. The three most common terms of the network are signals processing, signal radar, aircraft and others. The combination of co-author networks and co-word networks makes it possible to identify the teams that work the most on a given topic. At the next stage, it is possible to rank and allocate leaders of scientific directions. The databases of scientists are used for person identification after receipt on the basis of co-authorship and terms. The databases give an opportunity to test the experience of authors, to separate authors with similar names

Fig.1. Co-author network using Web of science papers of the university team

The network was divided into groups based on the calculation of the modularity of nodes. Modularity of a node is a value that evaluates the density of bonds in a coherent component in comparison with the bundles between the components. The presence of scientific groups that may have signs of scientific schools in the university can be traced clearly to the feature of the majority of the click, which in its essence contains a powerful author with a large number of articles in co-authorship and a significant number of small nodes - the authors-students. In fig. 1 shows a fragment of a network of collaborators' networks, which contains data on author Yanovsky, his co-authors and co-authors of his co-authors.

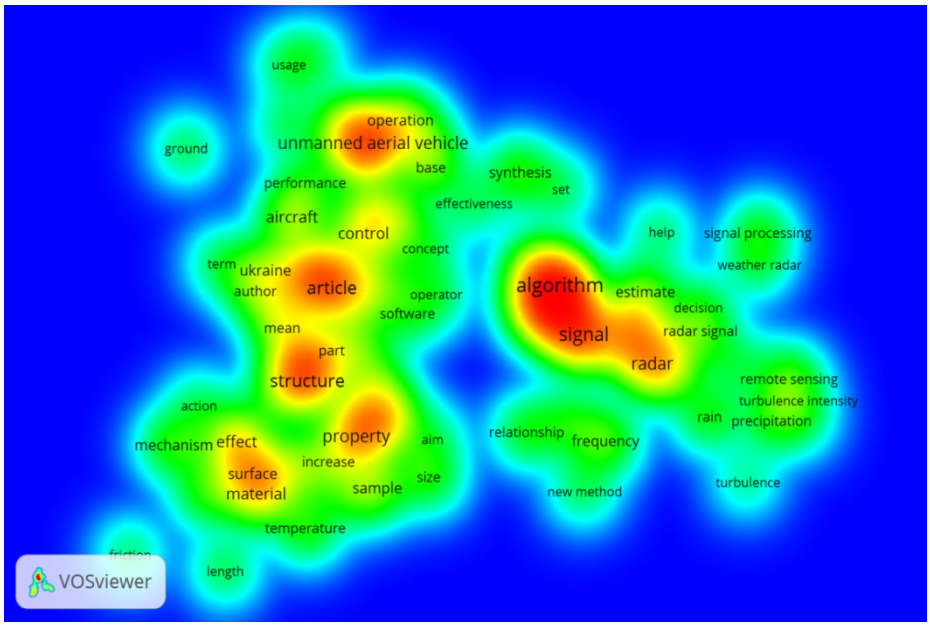

Fig. 2. Co-word network using Web of science papers of the university team

For comparing of searching experts method, we use the concept link prediction in the database Aminer. The Aminer database has found 66 profiles for the concept. In each author's profile there is a percentage of papers of specific topics, but there is no possibility to show the presence of individual concepts.

5 Conclusions

The method for detection of scientific teams is proposed, which allows finding groups of scientists whose activities correspond to the topics by the network of terms according to the given terms of a certain domain. It is shown that the proposed method allows searching for more flexible requests in comparison with existing expert databases. It is advisable to use the databases of scientists while searching for scientific groups to exclude identical names and obtain detailed information about the expert.

The possibility the method usage was shown with examples from Web of science database. The better results could be obtained with adding several international databases.

The aim of the algorithm is to correct researches in Ukraine for activation of international research and popularity of papers which is depend of the topics. Using of the algorithm could help scientists to get more authority in the world.

The method is a part of complex estimation of interdisciplinary and searching priorities and scientific cooperation.

The future research will consist of studying of co-author networks according to concepts with high interdisciplinary degree and developing of methods for searching cooperation partners with data from Web of science, Scopus, Google Scholar. Analyzing of scientific databases allow to detect existing researches and to forecast the possibilities of next cooperation of apart scientists or scientific teams.

The publication contains the results of research conducted under the grant of the President of Ukraine under the competitive project F75 / 173-2018 of the State Fund for Fundamental Research.

References

1. Ukrainian laws Homepage, http://nucpi.nas.gov.ua/news/item/67-zakon-ukraini-pro-vischu-osvitu.html
2. Horizon 2020 Homepage, https://h2020.com.ua/
3. Ukrainian laws Homepage, http://zakon.rada.gov.ua/laws/show/51/95-%D0%B2%D1%80
4. Udding M.N., Duong T.H., Oh K.J., Jung J.-G., Jo G.-S.: Experts search and rank with social network: an ontology –based approach, International Journal of Software Engineering and Knowledge EngineeringVol. 23, No. 01, pp. 31-50 (2013)
5. C. Wei, W. Lin, H. Chen, W. An, and W. Yeh. Finding experts in online forums for enhancing knowledge sharing and accessibility, Computers in Human Behavior, vol.51, pp.325-335 (2015)
6. Robin Brochier,AdrienGuille, Benjamin Rothan, JulienVelciImpactoftheQuerySetontheEvaluationofExpertFindingSystems, CoRR, abs/1806.10813 (2018)
7. Lin, S., Hong, W., Wang, D., & Li, T. A survey on expert finding techniques. Journal of Intelligent Information Systems, 49(2), 255–279.(2017)
8. Lande D.V., Andrushchenko V.B., Balagura I.V.:FormationoftheSubjectAreaontheBaseofWikipediaService,OpenSemantic Technologies forIntelligentSystems : conference, Minsk. – pp. 211-214. (2017)
9. Dmitry Lande, AndreySnarskii, ElenaYagunovaThe Use Of Horizontal Visibility Graphs To Identify The Words That Define The Information Structure Of The TextCEUR Workshop Proceedings. Vol-1108 urn:nbn:de:0074-1108-1. ISSN 1613-0073. Selected Papers of the 15th All-Russian Scientific Conference "Digital libraries: Advanced Methods and Technologies, Digital Collections" Yaroslavl, Russia, October 14-17, 2013. - P. 158-164.(2013)
10. I.V. Gorbov, S.V. Kadenko, I.V. Balagura, D.Yu. Manko, O.V. Andriichuk. Elicitation of possible scientific expert groups at co-authorship network using decision support methods // Data Recording, Storage & Processing , Vol. 15, No 4, P. 77 – 85 (2013).
11. I. V. Balagura, D. V. Lande, I. V. Gorbov. Studying of node importance characteristics in co-author network // Data Recording, Storage & Processing , Vol. 15, No 1, P. 45 – 52 (2013).
12. Lande D., Andrushchenko V., Balagura I.: Data Science in Open-Access Research On-line Resources // Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) - pp. 17-20. Web of Science, Scopus (2018)

13. Lande D.V., Balagura I.V., Andrushchenko V.,B.: The detection of actual research topic using co-word networks, Open Semantic Technologies for Intelligent Systems (OSTIS - 2018): conference, Minsk, - pp. 207-210 (2018)
14. Lande D.V. , Balagura I.V., Dubchak N.A. TheDetectionofActualResearchTopicsinPhysicsforUkrainianScientificGroups // 2017 IEEE InternationalYoungScientistsForumonAppliedPhysicsandEngineering YSF-2017, October 17-20, Lviv, P.64-66 (2017)
15. Zhao S. X., Rousseau R., Ye F.Y. H-Degree as a basic measure in weighted networks, Journal of Informetrics. – V.5, № 4. – P.668–677.(2011)
16. Newman M. E. J. Coauthorship networks and patterns of scientific,Proceedings of the National Academy of Sciences of the United States of America. – V.101,№1. – P.5200-5205. (2004)