Defining Potential Academic Expert Groups based on Joint Authorship Networks Using Decision Support Tools

© Iryna Balagura[0000-0001-9627-2091], © Sergii Kadenko[0000-0001-7191-5636], © Oleh Andriichuk[0000-0003-2569-2026] and © Ivan Gorbov[0000-0001-6888-0866]

Institute for Information Recording of National Academy of Sciences of Ukraine, Kyiv, Ukraine
balaguraira@sgmail.com

Abstract. We consider a co-authorship network in "Information security" field. The network is constructed using Scopus data for the Ukrainian affiliated authors. We define the key centrality indicators: centrality degree, betweenness centrality and weighted centrality. We have made rankings of authors by centrality indicators and citations and aggregated this data with decision support methods. We demonstrate the methodology and possibility of defining expert groups and academic schools using the scientific database's content. We propose to use decision support methods to define most communicative and cited scientists within co-authorship networks and demonstrate the way of ordinal factorial analysis usage for defining the relative weights of different centrality indicators within complex networks. Empirical results, obtained in the paper, indicate that there are no strong connection between given author's centrality indicators and the number of citations to the author's works. However, some centrality indicators are more influential and significant than others. The approaches suggested in the paper can be applied to detection of central nodes in complex networks in general.

Keywords: Co-authorship Network, Scopus, Centrality, Information Security, Decision Support, Ranking, Ordinal Factorial Analysis.

1 Introduction.

Subsequent paragraphs, however, are indented. Rapid development and general evolution of science, as well as increase of the number of publications of all kinds led to the necessity of complex consideration and organization of a system for statistical analysis of document information stream [1, 2]. Scientometrics, being a conceptually new development stage of analytical processing of documentation and scientific-statistical information, is targeted at resolution of such issues as the most rational selection of effective information, methodology of its evaluation, and efficient ways of its analysis. Necessary conditions of adequate functioning and development indicate extreme relevance of the problem of scientifically grounded, balanced, and efficient state policy in this area [2]. Scientometrics finds practical application in qualitative evaluation of academic publications and definition of development dynamics of both separate academic directions and science in general.

Interaction of academics from different research areas, particularly, within co-authorship networks, is an important and essential part of research process. In academic cooperation studies (in addition to bibliometrics and scientometrics) social network research and expert estimation methods are used [3]. Particularly, one of the most common methods is based on co-authorship network usage, where the nodes represent the authors while edges represent co-authorship links, proportional to the number of joint authorships (publications) [4]. Co-authorship network represents a tool for defining the functional structure of scientific research as a whole, helps us understand and forecast the ways of scientific information dissemination and evolution of academic schools, as well as define the relevance degree of specific research areas. Study of respective networks allows us to define the key publications, research fields, and authorship clusters [5, 6].

Academic co-authorship networks represent an example of complex networks; they can be analyzed using respective quantitative topological indicators, and further interpreted from content viewpoint [7, 8]. For instance, defined co-authorship clusters can represent expert groups and academic schools [9]. The relevance of expert group definition is beyond any doubt, because only professional expert examination can provide thorough and objective estimate of research results, while scientometric indicators in this case will be only the tools of decision-making support [10].

Academic schools, in their turn, are an essential developmental component of scientific cognition and educational processes. In spite of this important key role of an academic school, it is not acknowledged at state level (at least, in Ukraine), as there are no registration mechanisms and ways of legal certification of an academic partnership between supervisors, their students and associates. These factors have a negative imact upon authority, image, and reputation of academic schools [11]. As we can see, definition of academic schools is essential for optimization of joint academic research activity, particularly, for structuring of collegial intellectual creative process, targeted at obtaining and application of conceptually innovative, original knowledge, significant for respective scientific fields.

During search for expert groups and academic schools, certain "important" nodes will be located in the "centre" of respective clusters [12]. Thus, defining potential co-authorship network centre is the necessary condition for definition of respective expert groups and academic schools.

Detection of academic communities is a relevant task while choosing experts for evaluation of scientific research works, solving topical problems in certain areas, and searching for partners to cooperate with. Besides that, in scientometrics it is important to understand the processes that take place during academic collaboration. Academic community structure, intensity of interaction in it, its leaders: these and other aspects led to emergence of a whole new research area – the Science of Team Science (SciTS) [13]. In order to study the key trends of academic cooperation and detect "rich people's clubs" as well as the most highly communicative academics, co-authorship networks are used [14]. Usage of social networks featuring specialists' profiles, such as ResearchGate (https://www.researchgate.net/) and LinkedIn (https://www.linkedin.com/), simplifies the task of looking up specific researchers [15]. Scientific profiles can be found in Google scholar, Scopus, Web of science, and other databases. Besides that, there are resources for unification of information on academics from different databases, such as 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. Academic publication databases represent the most thorough resource to look for academic research groups.

2 Basic centrality indicators of a co-authorship network

Potential expert groups of academics are defined based on centrality indicators of co-authorship networks. In complex network theory there are several types of such coefficients, defined as the level of their centrality in a graph. Some of the concepts were based on complex network theory, while others were derived from sociological research results. There are several basic types of centrality, which are widely used in network analysis: centrality degree, betweenness (mediation) centrality, eigenvector centrality, and others [16–18].

Centrality degree defines the number of other network agents a certain person (agent, individual) is connected to; in co-authorship networks it can be interpreted as the degree of academic interaction. In the simplest case this is the degree of a certain node, which characterizes an author's communicability and can be used to forecast this author's productivity. According to research data, this characteristic does not correlate with average citation level and cannot completely represent all the aspects of authors' communicability [19]. The disadvantage of this indicator for communicative property definition is its inability to take the weights of graph edges (i.e. the number of joint publications of authors) into account.

In [16] the weighted degree of centrality is proposed. It is suggested to calculate centrality in a weighted graph for a specific node as follows:

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

The indicator includes $k_i = \sum_{j=1}^{N} m_{ij}$, i.e. the sum of links to other nodes and $s_i = \sum_{j=1}^{N} \omega_{ij}$, i.e., the sum of weights of respective connections, while α is a coefficient, adjusted for each specific case.

Centrality in the context of mediation (betweenness centrality) defines a node, connecting sub-graphs to each other. In the context of academic cooperation, mediation or betweenness allows us to define authors that connect academic schools:

$$C_B(i) = \sum_{j<k} g_{jk}(i) \tag{2}$$

where $g_{jk}(i)$ is the number of the shortest ways in a graph, which pass through node number $i; i \neq j,k$ [18].

While defining the importance of network nodes, we should consider communicability, significance of adjacent nodes, but we should not lose the information on the general productivity of authors, as the number of co-authors does not directly influence the efficiency of their work [17]. Definition of important nodes is a relevant problem, calling for in-depth study of research subject, as there are many measures, reflecting vertex (node) characteristics of different nature, while the adequacy of their usage is based on their correspondence to the respective experiment purposes.

3 Defining the centers of co-authorship networks for "information security" section of Scopus database

We propose to use decision support methods to define potential academic expert groups and academic research schools in co-authorship networks, and demonstrate the application of "ordinal factorial analysis-based" approach to calculation of relative weights of different centrality indicators of complex networks. For this purpose we use data from Scopus. We consider an example of unification of rankings of centrality and citation measures for researchers in the area of information security. Additionally, the approach allows us to verify the degree of dependence between different centrality measures among themselves and in comparison with citation indicators.

Scopus is one of the largest and most reputable abstract databases in the world. In Ukraine there is a demand for publishing papers cited in the database, particularly, such publications are necessary for obtaining of academic degrees and titles. Scientometric analysis of information security abstracts could show the development of this field in Ukraine in comparison with other countries. Authors of the paper [20] made scientometric research of the "Information security" area in Scopus. Number of published papers by year, countries' ranking by published papers, and other information are presented in [20].

Data from Scopus where gathered by the following keywords: information security, data security, cyber security, network security, cryptography, information assurance, data encryption, computer security. The search query is: (TITLE-ABS-KEY (information AND security) OR TITLE-ABS-KEY (data AND security) OR TITLE-ABS-KEY (security AND of AND data) OR TITLE-ABS-KEY (cybersecurity) OR TITLE-ABS-KEY (network AND security) OR TITLE-ABS-KEY (cryptography) OR TITLE-ABS-KEY (information AND assurance) OR TITLE-ABS-KEY (data AND encryption) OR TITLE-ABS-KEY (computer AND security) AND (LIMIT-TO (AFFILCOUNTRY , "Ukraine")). 1261 documents were obtained by the query for Ukrainian affiliated authors in comparison with 492395 documents by selected keywords for all countries. The number of papers has increased in recent years as a result of legislation change, but it still amounts to just 0.26 % of the word output. VOSviewer and Pajek software were used for analyzing obtained data. Co-word network is shown on Fig.1. Categories of Scopus subject area where defined as Computer Science (832), Engineering (572), Mathematics (247), Physics and Astronomy (154), Energy (120), Social Sciences (110) and others.

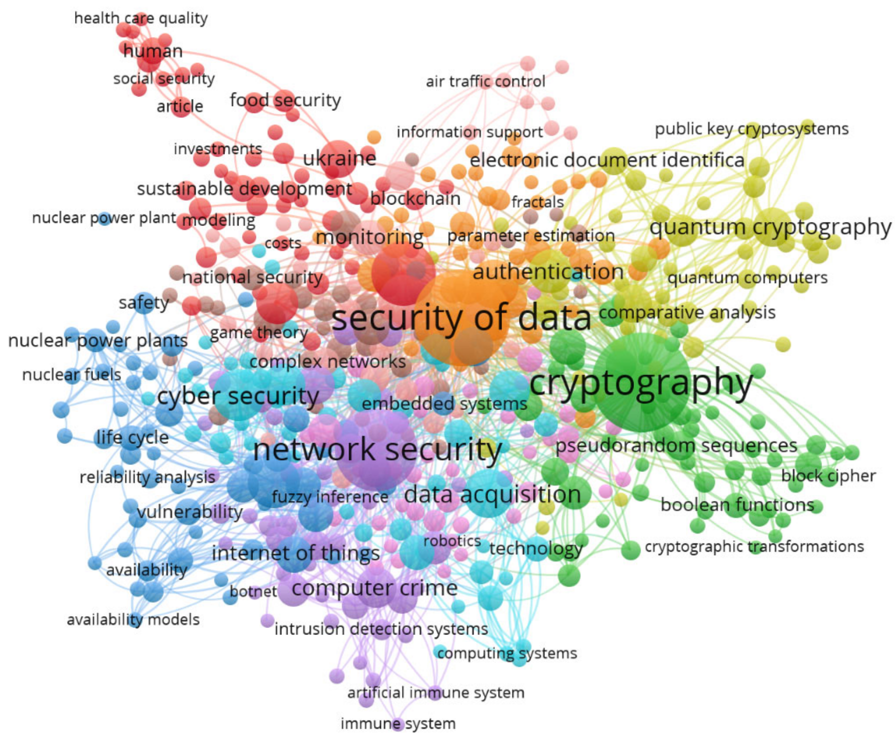

Fig. 1. Co-word network of Ukrainian affiliated authors papers in Information security.

We assume that there are authors' profiles in Scopus and authors have to organize their profiles themselves. Ukrainian rules for translating surnames into English changed several times, so spelling of surnames and names, namesake names may add inaccuracies to the data. On Fig.2 we can see that co-authored network is loosely connected. And on Fig.3 a large network fragement is presented. Existence of academic schools in "Information security" area can be witnessed based on "cliques", which include a productive author with multiple joint publications and a considerable number of "smaller" nodes – "students". Fig. 3 illustrates a fragment of co-authorship network, "built around" Kharchenko V., Sachenko A., Gnatyuk S., and others. The size of a node is proportional to the number of its connections (see Table 1). Concentration of co-authorship links around one or several leaders can indicate the emergence of separate academic schools. Papers of Ukrainian affiliated authors are written in coauthorship with authors from Poland (102), Kazakhstan (48), United States (47), Russian Federation (31), United Kingdom (28), China (25), Germany (25), Slovakia (20), Czech Republic (19), and others.

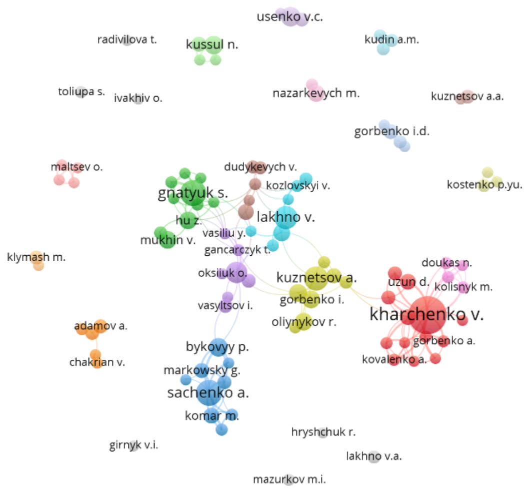

Fig. 2. Co-authorship network of Ukrainian affiliated authors: papers on Information security

Surnames of authors with maximum numbers of connections and joint publications are listed in Table 1. Rankings of authors according to the number of connections and weighted centrality degree are different, because during initial processing of information from the database the weights of connections are defined as proportionally distributed among all authors of each publication. The weighted centrality degree (i.e. the number of papers published in collaboration) reflects the volume of an author's work, while the number of connections characterizes the circle of co-authors of a specific author.

Table 1. Authors of research papers on Information security that have the largest centrality measures.

Degree		Weighted centrality degree		Beetwenness centrality	
Kharchenko V.	17	Kharchenko V.	76	Karpinski M.	0.155710
Gnatyuk S.	10	Sachenko A.	46	Potii O.	0.155269
Karpinski M.	10	Gnatyuk S.	29	Gorbenko Y.	0.151355
Hu Z.	9	Bykovyy P.	25	Kharchenko V.	0.132409
Opirskyy I.	9	Kochan V.	23	Bykovyy P.	0.085545
Sachenko A.	8	Komar M.	23	Gancarczyk T.	0.055159
Vasiliu Y	7	Kuznetsov A.	19	Vasiliu Y.	0.044856
Kochan V.	7	Lakhno V.	18	Opirskyy I.	0.044449
Gorbenko Y.	7	Karpinski M.	17	Akhmetov B.	0.041871
Bykovyy P.	6	Akhmetov B.	16	Gnatyuk S.	0.041140
Akhmetov B.	6	Hu Z.	14	Hu Z.	0.032517
Illiashenko O.	6	Gorbenko Y.	13	Kavun S.	0.031289

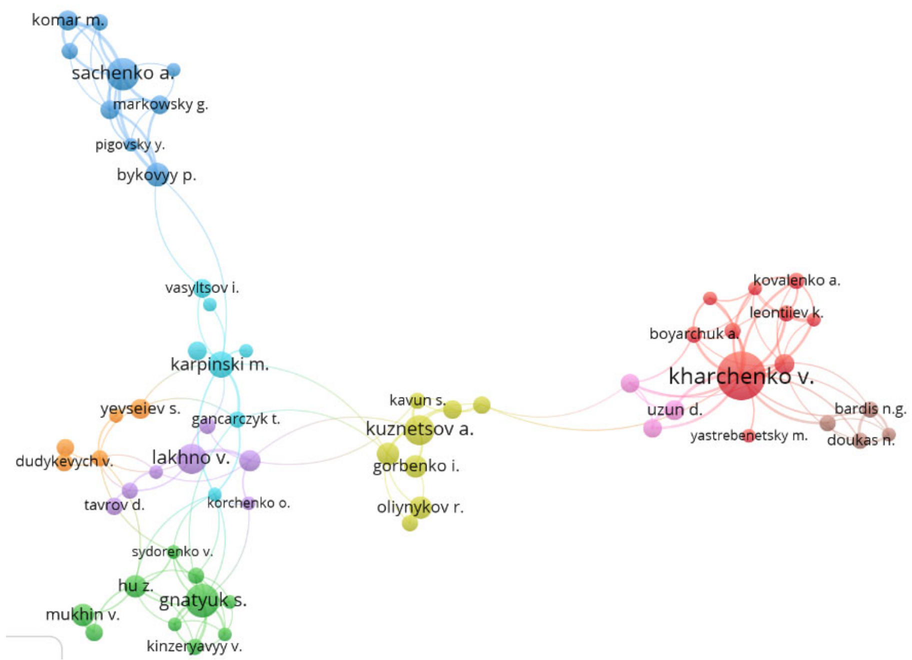

Fig. 3. A large fragment of co-authorship network of Ukrainian affiliated authors: papers on Information security.

Table 1 show that the author ranks changed according to selected centrality indicator. Authors with high centrality indicators are linked with their colleges and form separate groups or academic schools weakly connected to each other. The lists of authors from Table 1, sorted by weighted centrality degrees, betweenness centrality and degree consist of mostly same persons. Each centrality indicator reflects a certain aspect of the authors' communicability.

Also we can define the most productive and cited authors: Table 2 shows authors ranking by number of papers of each author, number of citations, and h-index of selected papers.

We can apply one or several indicators, and combine them with other approaches according to a task. We propose to use decision support methods for integration of several indicators. Decision support methods are currently used for priority setting in different areas, particularly, in weakly structured domains. Calculation of relative criterion weights is an essential component of such processes as strategic planning and resource allocation [21]. Areas of application of these methods range from sustainable development and industry development strategies [21] to information security [22].

Table 2. Authors of research papers on Information security that have the largest number of papers and citations

Number of papers		Number of citations		H-index of papers	
Kharchenko V.	55	Kussul N.	328	Kussul N.	8
Gnatyuk S.	26	Usenko V.	276	Usenko V.	8
Sachenko A.	25	Oliynykov R.	243	Kharchenko V.	8
Kuznetsov A.	22	Kharchenko V.	183	Nazarkevych M.	8
Lakhno V.	21	Kuznetsov A.	162	Kuznetsov A.	7
Gorbenko I.	21	Gorbenko I.	139	Gancarczyk T.	7
Karpinski M.	17	Lakhno V.	114	Lakhno V.	6
Usenko V.	14	Gorbenko Y.	90	Gorbenko I.	6
Gorbenko Y.	13	Nazarkevych M.	89	Mukhin V.	5
Bykovyy P.	13	Sachenko A.	84	Oliynykov R.	5
Kussul N.	13	Gnatyuk S.	69	Gorbenko Y.	5
Oliynykov R.	12	Komar M.	65	Sachenko A.	5

4 Application of decision support methods to defining the central nodes of co-authorship networks

In the context of our current problem, in order to get some information about relative significance of the two criteria, listed in Table 1 (weighted centrality and betweenness), we can try to calculate their weights based on ordinal factorial analysis methods [23-25]. One of the reasons why we are switching to rankings from actual values is because they differ very significantly across different authors (especially, betweenness and citation number, that we are going to use a global criterion below). For instance, betweenness indicator of N.Kussul is below 0.000001, while the same indicator for Y.Gorbenko is 0.151355. In order to apply the approach [23-25], we need some global ranking of alternatives, i.e. authors. To get the ranking, we can ask one or several specialists from the respective field (in our case, information security), to name the leading authors, and rank the authors accordingly. Using these data and data from Table 1, it is possible to calculate the weights of the two criteria. This will allow us to get a rough estimate of their significance, i.e., how important they are for the authors' reputation (reflected in the global ranking based on expert references).

Let us consider a hypothetical example: during an interview, the expert was asked to name up to 10 top specialists in information security area. He placed the authors in the following order: …, Kharchenko, V., …, Gnatyuk, S., …, Sachenko, A., …, Lakhno, V., …, Kussul, N., …, Oliynykov, R., … (see last column of Table 3).

Table 3. A hypothetical ranking of specialists in the area of information security

Authors	Ranking by weighted centrality degree	Ranking by be-tweenness	Hypothetical expert ranking
Kharchenko, V.	1	1	1
Gnatyuk, S.	3	2	2
Sachenko, A.	2	3	3
Lakhno, V.	5	4	5
Kussul, N.	4	6	4
Oliynykov, R.	6	5	6

We use the approach described in [23-25]. For calculation the criterion weights, we should build the respective system of inequalities. The searched weights are the "centers of mass" of the part of the simplex, delimitated by the solution region of the system of inequalities. Strict problem statement and step-by-step solution algorithms can be found in [23-25].

Based on data from Table 3, we can obtain the following criterion weights: weighted centrality (i.e. number of links) – 0.417; betweenness – 0.583. That is, for our hypothetical expert, the number of papers, published jointly with other authors (or number of links to other authors) is a bit less important than an author's role in mediation between academic schools.

We can also use information from Scopus database to obtain objective ranking of authors. Particularly, we can build a ranking based on citation statistics of each author [26]. So, we can define, whether different aspects of an author's centrality actually influence his or her international rating. Let us rank several authors from Table 1 according to the number of citations in Scopus database and define their ratings according to several centrality aspects, such as weighted centrality degree and betweenness centrality (Table 4). It would be problematic to rank authors by ordinary centrality degree (first column of Table 1), because many authors have equal values of this indicator. That is why we focus on the other two centrality measures (second and third columns of Table 1). Let us rank 6 authors, featured in both the second and the third columns of Table 1 by weighted centrality and betweenness, as well as by their respective numbers of citations (see Table 4).

Academics with the largest number of papers in their research fields have large centrality indicators and big number of citations [27]. However, we cannot say there is a direct dependence between these factors. If we try to find the weights of centrality indicators according to the algorithm described in [23, 24], then the solution area turns out to be empty. If we perform a minimal permutation within the ranking (a real expert or rater might agree to perform it) – place author Bykovy in front of Gnatyuk in the global ranking – then the weights are: weighted centrality degree – 0.367; betweenness centrality – 0.633.

Table 4. Ranking of authors according to citations in Scopus and key centrality indicators

Authors	Ranking by number of citations	Ranking by weighted centrality	Ranking by betweenness
Kharchenko V.	1	1	2
Gorbenko Y.	2	6	1
Gnatyuk, S.	3	2	5
Akhmetov B.	4	4	4
Hu Z.	5	5	6
Bykovyy P.	6	3	3

These results (weights) can be interpreted as follows. Empty solution area means that, centrality does not strongly influence the number of citations. That is, the number of references to an author's work only loosely depends on the number of his publication collaborations and on his academic mediation. Moreover, we should keep in mind, that we are only analyzing authors' centrality in one particular topic, i.e. information security, while citations cover all the topics authors publish articles on. If we still compare relative weights of the two centrality factors (although, the dependence is very weak in general), then it turns out that the number of joint publications in Scopus (which is reflected by weighted centrality degree) is less significant (weight equals 0.367) than author's mediation role (weight equals 0.633).

Thus, in addition to separate centrality aspects, we can define some generalized centrality indicator of an author. It can be calculated as arithmetic mean, geometric mean, or weighted sum of normalized ratings of an author according to criteria from Table 1. This approach to aggregation of data on several objects (in our case – authors) is described in [28] and can be applied, because criteria are preferentially independent (which is a necessary and sufficient condition of linear convolution applicability [29])).

5 Conclusions

The search of authors with high centrality levels and their respective co-authors allows detecting of academic schools and expert groups. However, for detailed study we should also use textual (linguistic) analysis of paper abstracts and expert estimates. These approaches would also allow us to outline academic fields more precisely, and assess their development.

We have demonstrated the ranking of publications' authors in the field of "Information security" according to several indicators. All the indicators provide versatile object characteristics. We proposed to create an overall ranking of scientists using whole number of citations per author, weighted centrality and betweeness centrality. We have applied decision support methods for detection of potential expert groups of academics and academic schools with co-authorship networks. The results of final ranking detect simultaneously scientists with wide communication network and high Hirsch-index, who could lead a team or be an reputable expert.

Approaches developed for elicitation of co-authorship networks and elaborated in the paper can and should be used for detection of important nodes in other complex networks.

References

1. Russell, J., Cohn, R.: Scientometrics. Book on Demand, Berlin (2012).
2. Penkova, O.: Scientometrical and bibliometrical research in library and bibliographic theory and practice. Thesis abstract (in Russian). Krasnodar state university of culture and arts, Krasnodar (2002).
3. Alireza, A., Liaquat, H., Leydersdorff, L.: Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. Journal of Informetrics, 6(3), 403-412 (2012).
4. Lande, D., Balagura, I.: Scientometric research of co-authorship networks by "Ukrainika naukova" database (in Ukrainian). Data recording, storage, and processing, 14(4), 41-51 (2012).
5. Ding, Y.: Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. Journal of Informetrics, 5(1), 187-203 (2011).
6. Yan, E., Ding, Y., Milojevic, S., Sugimoto, C.: Topics in dynamic research communities: An exploratory study for the field of information retrieval. Journal of Informetrics, 6(1), 140-153 (2012).
7. Newman, M.: The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. USA, 98, 404–409 (2001).
8. Lande, D., Snarsky, A., Bezsudnov, I.: Internetica: Navigation in complex networks: models and algorithms (in Russian). Librocom (Editorial URSS), Moscow (2009).
9. Lande, D., Balagura, I.: Co-authorship networks, according to "Ukrainika naukova" database (in Russian). In: Applied linguistics and linguistic technologies: Megaling-2012: Collection of research works. NAS of Ukraine, language and information foundation, ed. Shirokov, V., pp. 155-163, ULIF, Kyiv (2013).
10. Tchebotaryov, P. Scientometrics: how to cure and not to mutilate with it? (in Russian). Management of large systems, 44, 14-31 (2013).
11. Kozubtsov, I.: Substantiation of choice of principle for building of a scientific picture of the world of knowledge (in Ukrainian). Scientific-theoretical and socio-political almanac "Grani" ("Edges"), 99(7), 45-48 (2013).
12. Balagura, I., Lande, D., Gorbov, I.: Studying parameters of node importance in networks of co-authors (in Ukrainian). Data recording, storage, and processing, 15(1), 45-52 (2013).
13. Sciabolazza, V.L., Vacca, R., Kennelly Okraku, T. McCarty, C.: Detecting and analyzing research communities in longitudinal scientific networks. PLOS one, August 10, 2017, https://doi.org/10.1371/journal.pone.0182516 (2017).
14. Balagura, I.: Defining expert groups through analysis of reference database «Ukrainika naukova (in Ukrainian). Data recording, storage, and processing, 17(3), 75-82 (2015).
15. 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, 51, 325-335 (2015).
16. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks, 32(3), 245-251 (2010).
17. Balagura, I., Lande, D., Gorbov, I.: Characteristics of networks of co-authors in medical sciences (in Russian). Clinical informatics and telemedicine, 9(10), 141-144 (2013).

18. Lande, D., Balagura, I.: Studying co-authorship networks in legal science according to "Ukrainika naukova" database. Legal informatics, 36(4), 50-57 (2012).
19. Liao, Ch., Yen, H.: Quantifying the degree of research collaboration: A comparative study of collaborative measures. Journal of Informetrics, 6(1), 27-33 (2012).
20. Somayeh Parvin, Farahnaz Sadoughi, Arman Karimi, Masoud Mohammadi & Farzaneh Aminpour. Information Securityfrom a Scientometric Perspective.Webology, 16 (1), 196-209 (2019).
21. Tsyganok, V., Kadenko, S., Andriichuk, O., Roik, P.: Usage of multicriteria decision-making support arsenal for strategic planning in environmental protection sphere. Journal of Multi-criteria Decision Analysis, 24(5-6), 227-238 (2017).
22. Kadenko, S.: Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area. In CEUR Workshop Proceedings. Vol-1813, pp. 8-14 (2016).
23. Kadenko, S.: Defining the relative weights of alternative estimation criteria based on clear and fuzzy rankings. Journal of Automation and Information Sciences, 45(2), 41-49 (2013).
24. Kadenko, S.: Personnel-related decision making using ordinal expert estimates. In: OR52 Keynotes and Extended Abstracts – 52nd Conference of the Operational Research Society, pp. 177-183, Royal Holloway University, London (2010).
25. Kadenko, S.: Determination of parameters of criteria of "tree" type hierarchy on the basis of ordinal estimates. Journal of Automation and Information Sciences, 40(8), 7-15 (2008).
26. Scopus database, http://www.scopus.com, last accessed 2019/11/15.
27. Balagura, I.: Researching the reference database "Ukrainika naukova" as to detection of connection between authors' citations and their academic cooperation (in Ukrainian). In: Proceedings of the annual summary conference "Data recording, storage and processing", pp. 210-214. Institute for information recording of the NAS of Ukraine, Kyiv (2013).
28. Totsenko, V.: Decision support methods and systems: algorithmic aspect (in Russian). Naukova dumka, Kyiv (2002).
29. Keeney, R., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. 2nd edn. Cambridge University Press, New York (1993).