Methodology for Analyzing Thematic Co-Authorship Networks

Anthony M. Nwohiri and Andrey A. Pechnikov

***Abstract*—An interesting feature that networks present is the community structure property, under which a graph topology is organized into modules called communities. This paper proposes a co-authorship graph-based methodology for studying thematic communities. The graph is divided into thematic communities to identify their basic characteristics, such as research thematic direction, researcher count, community count and relationship between researchers. This information could be used to make incentives policies that would support pertinent research areas. The proposed methodology was tested using data retrieved from mathematical portal Math-Net.Ru. The findings of the tests indicate that more research in robots and robotic systems, combustion/explosion, and data protection techniques/systems need to be promoted. It was demonstrated that the mathematical models employed were adequate and applicable to different research domains. However, this would need complete and reliable basic bibliographic data on research co-authorship within the relevant discipline over a long enough period.**

***Keywords*—co-authorship; community, graph, modularity, scientometrics.**

I. Introduction

Scientometric analysis is closely associated with quantifying the scientific research personnel and accomplishments to forecast the process of scientific development, improve the quality of existing scientific research, as well as reveal and develop new areas of research [1] [2] [3]. It would be ideal if the findings from such analysis were not only of scientific interest, but also actually used to determine objective criteria for research quality and make decisions on research funding.

The narrow disciplinary focus of scientific research apparently creates opportunities for in-depth understanding of the topic under investigation. However, such a narrow disciplinary tradition inhibits the creation of a comprehensive picture of the system under study, which would have allowed viewing the system through multiple disciplinary lenses and making connections across different knowledge formations [4]. Major global problems cannot be tackled effectively without interdisciplinary collaboration [5] [6].

The general regularities of scientific advancement involve the concurrent growth in differentiation of scientific disciplines and the strengthening of integration linkages between various fields of scientific knowledge [7] [8]. The study of research collaborations can be used to evaluate patterns in the evolution of various scientific directions, identify research centers, research schools and important personalities. Moreover, researchers are faced with data mining challenges -- sifting large volumes of data for correlations, patterns, and trends. In addition, the use of data visualization techniques reduces the cost of identifying associations and, subsequently, of presenting analysis results [9] [10].

Methods and approaches to solving such problems largely depend on the scale of research. Global-scale tasks involve building global map of science[1]. The map includes over 130 million scholarly publications from all over the world, algorithmically organized into more than 116,000 research clusters. An earlier and smaller work [11] used data from the Institute for Scientific Information, the founder of Web of Science, which is now part of Thomson Reuters. The study was based on citation data retrieved from over 6,000 journals organized into 172 categories. Interestingly, the map is still maintained on the website of one of the authors[2].

The main limitation when it comes to evaluating the scientific activity of Russian organizations is that a significant part of scientific results is published in Russian journals, many of which are not indexed by Scopus and Web of Science databases and are not cited by foreign (non-Russian) authors.

Russia's own large-scale project "Map of Russian Science"[3], which was widely launched by the Russian Ministry of Education and Science in 2013, was an unsuccessful adventure. The stated objective, which was "*...to obtain a clear picture of the 'landscape' of Russian science, its priorities in various fields and directions*"[4], remained unachieved.

Large-scale projects are usually created using a single method that may be applied across various scientific fields. This is not always convenient for analyzing distinct scientific disciplines that get lost in visual representations, potential data gaps and high citation barriers in major projects. It was recommended by Milman and Zhurkovich [12] to modify the scientometrics toolkit and analyze

[1] Map of Science. URL: https://sciencemap.eto.tech (accessed: 13th February, 2023).
[2] No title. URL: https://www.leydesdorff.net/map06 (accessed: 15th February, 2023).
[3] Map of Russian Science. URL: https://ru.wikipedia.org/wiki/Карта_российской_науки (accessed: 14th February, 2023).
[4] Poisk Newspaper. URL: https://poisknews.ru/magazine/9328/?ysclid=lozni10nj2479160561 (accessed: 15th February, 2023).

specialized information flows. One of the authors used chemical technology and applied chemistry in a separate paper he wrote 30 years ago to illustrate this [13].

Numerous studies with a bibliographic focus have examined scientific collaboration across a range of disciplines and tackled different facets of this problem. Using data for the last 60 years, Seltzer and Daniel [14] showed that the likelihood of co-authoring a paper in general economics journals is higher than in economic history journals. Chebukov et al. [15] used bibliomaniac analysis, social network analysis, and graph theory to investigate patterns in scientific collaboration in the field of mathematics and related fields. Johal et al. [16] considered the so-called "political co-authorship", where the formation of a group of co-authors is, for example, aimed at attaining publication in a highly regarded medical journal. Basarab et al. [17] investigated the case of co-authorship for one local unit of a higher education institution and made recommendations on how to intensify joint research.

More traditional works in computer science [18], economics [19] and physics [20] have been devoted to the study of the properties of co-authorship graphs.

This paper investigates collaborations between mathematicians – communication (through co-authorship) between representatives of various mathematical disciplines. Coordinated efforts are deployed towards finding joint solutions to certain problems. Co-authors form the so-called co-authorship networks [21]. Such networks tend to break into communities (groups, modules, clusters) with more connections internally than externally [22]. Suppose that authors belonging to the same community are conducting research in an area of mathematics that is common to them (albeit in a rather broad sense). Then, such communities and the relationship between them would reflect the main features and internal characteristics of the areas of mathematics and relationship between these areas.

Our goal is to use co-authorship to build and analyse the network of relationships among Russian mathematicians. For this purpose, the following tasks are addressed:

- Building a co-authorship graph of Russian mathematicians based on data retrieved from mathematical portal Math-Net.Ru[5];
- Dividing the above graph into thematic communities and analysing them to identify main research areas.
-

The study's theoretical goal is to develop methods for analysing how scientists from various domains that are part of a single scientific discipline communicate with one another. The proposed approach was tested for the mathematics discipline and was shown to have a great potential for development. From a practical point of view, some of the findings can be directly used for making policies, such as those aimed at advancing trending or promising mathematical research.

II. Methodology for Investigating Thematic Communities of Scientific Co-Authorship

A co-authorship graph, whose vertices represent authors of publications and whose edges indicate the relationships between co-authors, serves as the mathematical model for the object of our research. Newman [23] defines co-authorship relationship as follows: “*...two scientists are considered connected if they have co-authored a paper together*”, since “*...most people who have written a paper together will know one another quite well*”, but “*... there are many scientists who know one another to some degree but have never collaborated on any work*”. So, this definition shows that a relationship between two researchers does not depend on the *number* of articles they have co-authored; rather, such a relationship exists if the two authors have co-authored at least one paper that may or may not have other co-authors.

Co-authorship graphs are one of the frequently studied constructs of scientific collaboration, perhaps because of the inherent obviousness of defining the "acquaintance" of researchers, as documented by the very articles that the researchers themselves publish. Co-authorship graphs can be built using a variety of data sources, such as the "Authors"[6] section of the scientific electronic library eLIBRARY.RU, where required data can be manually retrieved. Although the procedure for building a co-authorship graph is well-defined, it depends significantly on initial data [15]. For instance, in an article by Bredikhin et al. [21], such a procedure is presented for the case of data from the REPEC database of economic journals.

The next section in this paper describes the process we used to build a co-authorship graph for mathematical journals. The standard notation for the co-authorship graph is $G=G(V,E)$, where V is the set of vertices representing authors, and E is the set of edges connecting pairs of authors who have co-authored at least one article. The graph is an undirected graph, without loops and multiple edges.

An interesting feature that real networks present is the clustering or community structure property, under which a graph topology is organized into modules commonly called communities or clusters [24]. These communities are subgraphs that have more connections internally than externally [23] [25] [26]. Figure 1 shows a typical illustration of an undirected graph with three communities.

[5] Pan-Russian mathematical portal. URL: https://www.mathnet.ru (accessed: 15th February, 2023).

[6] Authors search. URL: https://elibrary.ru/authors.asp (accessed: 14th February, 2023).

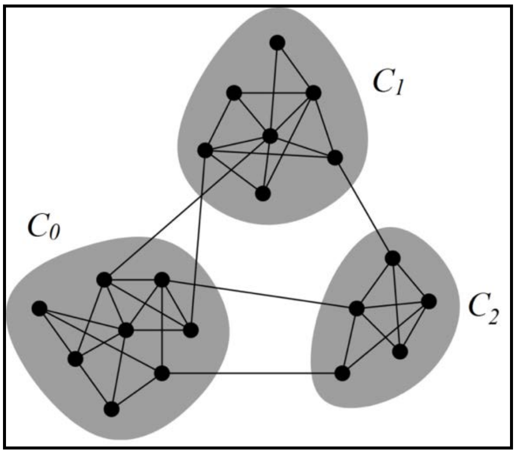

Fig. 1 A small network with community structure of the type considered in this paper. In this case there are three communities, shaded in grey, which have dense internal links but between which there is only a lower density of external links.

Let k be the number of communities into which graph $G=G(V,E)$ is divided, and let V_0, V_1, ..., V_{k-1} be the subsets of vertices making up the communities. No pair of subsets intersects, i.e. $V_i \cap V_j$ (i,j=0,...,k-1; i≠j), and their union gives a set V: $V_0 \cup V_1 \cup \ldots \cup V_{k-1} = V$.

For undirected graphs, density-based clustering represents a theoretically sound way of formalizing this problem. The term "modularity" is used to measure the strength (quality) of a particular division of a network into modules (communities) in the sense that each community has dense internal links but only a lesser density of external links (see Fig. 1).

One of the formal definitions of modularity measure Q, which is used here, is given by Newman and Girvan [27]. This quantity measures the fraction of the edges in the network that connect vertices of the same type (i.e., within-community edges) minus the expected value of the same quantity in a network with the same community divisions but random connections between the vertices. With this definition, $Q \in [-1,1]$ and a division is considered good if $Q>0.7$.

Open-source network analysis and visualization software package Gephi[7] was used to divide graph $G(V,E)$ into communities. In Gephi, we implemented the possibility of varying the community count and evaluating the quality of the resulting division using Q. This made it possible for us to run several experiments to determine the best division and/or division with a given community count.

One of the main assumptions made here is that authors belonging to a specific community primarily focus on some common theme that defines that community's scientific field. So, it becomes necessary to identify the theme (scientific direction, scientific discipline) of a group of scientists based on their collaborative works. Such a broad formulation alone generates a problem that requires separate formulation and research. The approach taken to tackle such a problem largely depends on the scope of the study. Obviously, different classifiers of scientific fields and techniques for ascertaining whether researchers are associated with a particular field will be employed in each specific case.

The nomenclature of scientific specialties will be used to categorize mathematical communities, while community themes will be determined based on the nomenclature of the most eminent researchers belonging to those communities.

Following the above steps, $G(V,E)$ is divided into k communities -- C_0, C_1, ..., C_{k-1}. Each community C_i is associated with a triple (V_i, E_i, t_i), where V_i is a set of vertices for community C_i, E_i is a set of edges connecting V_i, and t_i is the theme of C_i. The basic characteristics of thematic communities, such as the number of researchers in each community, the number of links between researchers, the number of links between communities, etc., may therefore be calculated by analyzing the results. Some communities may also have the same theme, in which case generalized data, such as the total number of scientists in each aggregated scientific field, would be available.

Comprehensive interpretation of findings and conclusions depend significantly on the scale of the problem and the scientific disciplines being investigated. This we did in relation to analysis of co-authorship among Russian mathematicians.

III. Initial Data and Co-Authorship Graph of Mathematicians

Over the past 15 years, information system *Math-Net.Ru* has been accumulating data heterogeneously. So, we selected a time interval of 2000 to 2020 for the study in order to date (ascribe a date in the past) articles published between 2000 and 2020; 56 leading Russian journals fully indexed in *Math-Net.Ru* were selected from among 140 journals. Some of the selected journals are *Algebra and Analysis*, *Mathematical Notes*, *Siberian Mathematical Journal*, and others. Detailed information about these 56 journals is given in Znamenskaya et al. [28] (2022, p. 194).

From 2000 to 2020, these journals published almost 62,000 articles written by about 33,000 authors. Znamenskaya et al. [28] showed that 54% of articles in these journals are single-author papers, 30% are written by two authors, 10.5% by three authors, and 5% by four to five people. Teams of five or more authors account for less than 1% of the papers and have little or no influence on the characteristics of the co-authorship graph.

A sample of all authors who had co-authored some works during the specified period in any of the 56 journals was created from the *Math-Net.Ru* database, and this sample was used to form a set of vertices V of the graph. Researchers in *Math-Net.Ru* have integer codes assigned to them as vertex labels. The use of integer codes eliminated some errors associated with author identification. Set of edges E was formed based on a sample from all possible pairs of co-authors according to the following rule: for any pair $i,j \in V$, edge $(i,j) \in E$ exists if and only if there is at least a paper where authors i and j are co-authors.

Thus, an initial version of the co-authorship graph $G(V,E)$ containing more than 26,000 vertices and 36,000 edges was constructed. Its principal characteristic is that it

[7] Gephi. The Open Graph visualization Platform. URL: https://gephi.org (accessed: 13th February, 2023).

decomposes into more than 4,000 connected components. The maximum component has 11,860 vertices and 20,858 edges, while the second maximum contains just 100 vertices, and the minimum component has 3).

The subgraph with the maximum component was selected. Next, vertices of degree 1 (leaf vertices) were removed via several consecutive iterations. Removal of such vertices would not affect graph connectivity subsequently but would make a meaningful analysis of communities much easier. As a result, a co-authorship graph denoted as $G_{co\text{-}auth}$ was built. It had one connected component with 8,166 vertices and 17,164 edges. Additional study was conducted on this resulting graph. Graph diameter d (number of edges in the shortest path connecting the two most distant vertices) was 26, and the average path length in the graph was 9.7.

IV. Building and Exploring Thematic Communities

Visualization software package *Gephi* was used to divide Gco-auth into communities. Maximum modularity measure Q = 0.933 was achieved when the graph was divided into 74 communities. The entire list of communities and their key attributes are displayed in Table 1. Rows in the table are ordered in ascending order of specialty codes (assigned by the Russian Higher Attestation Commission (VAK). VAK is a national government agency in Russia that oversees awarding of advanced academic degrees. Rows with the same specialty code are ordered in descending order of vertex count in a community. For clarity, communities of the same type are grouped in a single block.

The affiliations of "leaders" are also known, but are not listed in Table 1, because many scientists have several affiliated organizations in the Math-Net.Ru database, making it difficult to determine which is the primary one. A summary of “leaders” in terms of status, geographic location and affiliation is presented in Figures 2a, 2b and 2c.

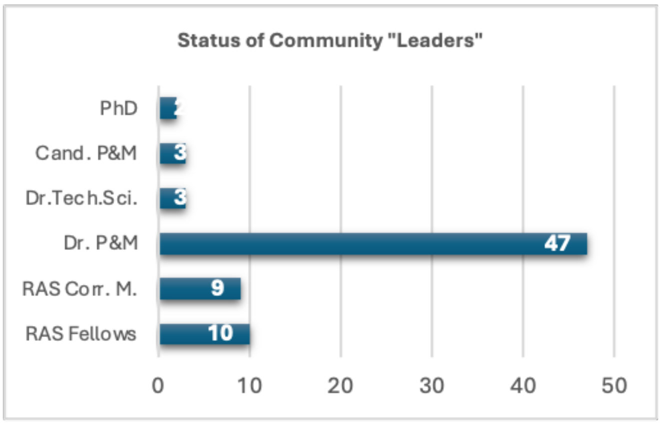

Fig. 2(a) Status of Community "Leaders"
Abbreviations: ***PhD,*** Doctor of Philosophy; ***Cand. P&M,*** Candidate of Physical & Mathematical Sciences; ***Dr.Tech.Sci.,*** Doctor of Technical Sciences; ***Dr. P&M,*** Doctor of Physical & Mathematical Sciences; ***Corr. M. RAS,*** Corresponding member of the Russian Academy of Sciences; ***RAS Fellow,*** Fellow of the Russian Academy of Sciences.

Fig. 2(b) Geographic location of "Leaders" by city

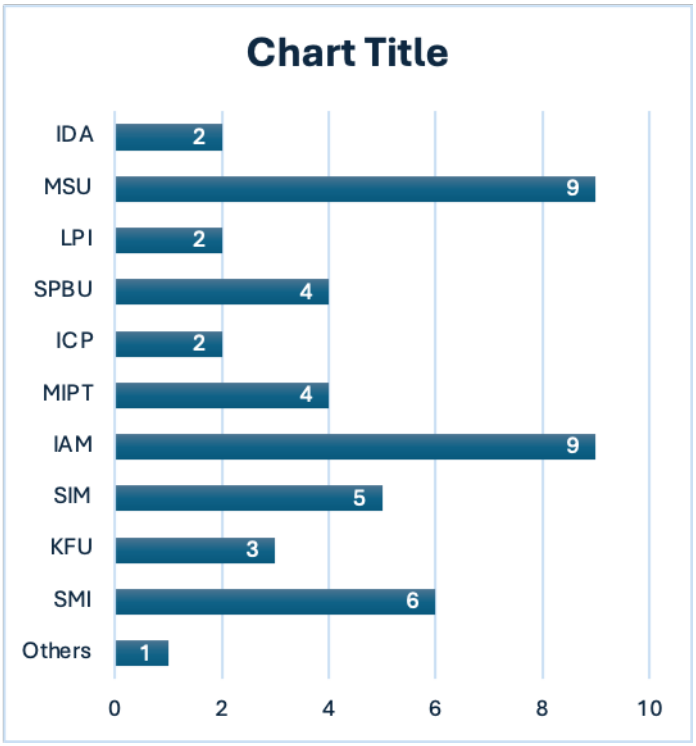

Fig. 2(c) Affiliate research institutions of "leaders"
Abbreviations: **IDA,** Institute of Design Automation; **MSU,** Moscow State University; **LPI,** Lebedev Physical Institute; **SPBU,** St. Petersburg State University; **ICP,** Institute of Control Problems; **MIPT,** Moscow Institute of Physics and Technology; **IAM,** Institute of Applied Mathematics; **SIM,** Sobolev Institute of Mathematics; **KFU,** Kazan Federal University; **SMI,** Steklov Mathematical Institute; **Others,** Other universities and institutes.

Table 1 Thematic communities of the co-authorship graph

t_i	i	$\|V_i\|$	$\|E_i\|$	***Leader***	***Status***	***Location***
01.01.01	19	143	258	Semenov E. M.	Dr. P&M ≈ PhD	Voronezh
01.01.01	20	141	239	Kusrayev A. G.	Dr. P&M ≈ PhD	Vladikavkaz
01.01.01	66	138	294	Aptekarev A. I.	Corr. M. RAS	Moscow
01.01.01	4	112	192	Konyagin S.V.	RAS Fellow	Moscow
01.01.01	39	101	171	Tikhomirov V. M.	Dr. P&M ≈ PhD	Moscow
01.01.01	58	38	64	Smolyanov O. G.	Dr. P&M ≈ PhD	Moscow
01.01.01	64	19	31	Bogachev V. I.	Corr. M. RAS	Moscow
01.01.02	70	201	389	Sadovnichy V. A.	RAS Fellow	Moscow
01.01.02	1	153	252	Tarkhanov N. N.	Dr. P&M ≈ PhD	Krasnoyarsk
01.01.02	32	135	273	Gonchenko S. V.	Dr. P&M ≈ PhD	NizhnyNovgorod
01.01.02	15	117	225	Ushakov V. N.	Corr. M. RAS	Yekaterinburg
01.01.02	28	45	94	Pochinka O. V.	Dr. P&M ≈ PhD	Moscow
01.01.02	65	42	79	Tonkov E. L.	Dr. P&M ≈ PhD	Izhevsk
01.01.02	57	35	60	Broer Henk W	PhD	Netherlands
01.01.02	42	18	28	Konyaev Y.A.	Cand. P&M ≈ PhD	Moscow
01.01.03	16	236	465	Shabat A. B.	Dr. P&M ≈ PhD	Moscow Oblast
01.01.03	52	205	449	Orlov Y. N.	Dr. P&M ≈ PhD	Moscow
01.01.03	13	86	145	Slavyanov S. Y.	Dr. P&M ≈ PhD	St. Petersburg
01.01.03	48	80	152	Kudryashov N. A.	Dr. P&M ≈ PhD	Moscow
01.01.03	6	78	160	Dobrokhotov S. Y.	Dr. P&M ≈ PhD	Moscow
01.01.03	22	77	165	Kulish P. P.	Dr. P&M ≈ PhD	St. Petersburg
01.01.03	11	55	100	Arefieva I. Y.	Corr. M. RAS	Moscow
01.01.03	49	48	84	Tyutin I. V.	Dr. P&M ≈ PhD	Moscow
01.01.03	10	42	88	Kostomarov D. P.	RAS Fellow	Moscow
01.01.04	68	203	369	Bukhshtaber V. M.	Dr. P&M ≈ PhD	Moscow
01.01.04	5	128	272	Golubyatnikov V.P	Dr. P&M ≈ PhD	Novosibirsk
01.01.04	9	109	194	Repovsh D.D.	PhD	Slovenia
01.01.04	25	38	80	Mikesh J.	Dr. P&M ≈ PhD	Czech Republic
01.01.05	17	223	401	Shiryaev A.N.	RAS Fellow	Moscow
01.01.05	30	135	221	Vatutin V. A.	Dr. P&M ≈ PhD	Moscow
01.01.06	29	264	565	Chubarikov V. N.	Dr. P&M ≈ PhD	Moscow
01.01.06	43	229	496	Mazurov V. D.	Corr. M. RAS	Novosibirsk
01.01.06	23	121	233	Kirichenko V.V.	Dr. P&M ≈ PhD	Kyiv
01.01.06	33	116	192	Vostokov S.V.	Dr. P&M ≈ PhD	St. Petersburg
01.01.06	54	96	210	Goncharov S.S.	RAS Fellow	Novosibirsk
01.01.06	2	74	126	Guterman A.E.	Dr. P&M ≈ PhD	Moscow
01.01.06	21	55	92	Abyzov A.N.	Cand. P&M ≈ PhD	Kazan
01.01.07	26	187	369	Yevtushenko Y. G.	RAS Fellow	Moscow
01.01.07	31	179	337	Vasin V. V.	Corr. M. RAS	Yekaterinburg
01.01.07	40	162	339	Vabischevich P. N.	Dr. P&M ≈ PhD	Moscow
01.01.07	50	103	179	Ilyin V.P.	Dr. P&M ≈ PhD	Novosibirsk
01.01.07	67	102	171	Malozemov V. N.	Dr. P&M ≈ PhD	St. Petersburg
01.01.07	55	76	165	Belotserkovsky O.	RAS Fellow	Moscow
01.01.07	27	68	115	Zorkaltsev V. I.	Dr.Tech.Sci.	Irkutsk
01.01.07	63	68	129	Ostapenko V. V.	Dr. P&M ≈ PhD	Novosibirsk
01.01.07	72	65	127	Shaidurov V. V.	Corr. M. RAS	Krasnoyarsk
01.01.07	45	26	56	Kholodov A. S.	RAS Fellow	Moscow
01.01.09	53	151	280	Bure V.M.	Dr.Tech.Sci.	St. Petersburg

01.01.09	69	131	252	Gimadi E.H.	Dr. P&M ≈ PhD	Novosibirsk
01.01.09	7	62	121	Borodin O. V.	Dr. P&M ≈ PhD	Novosibirsk
01.02.01	3	234	517	Borisov A.V.	Dr. P&M ≈ PhD	Moscow
01.02.01	24	89	225	Ovchinnikov M.	Dr. P&M ≈ PhD	Moscow
01.02.04	41	130	281	Radchenko V. P.	Dr. P&M ≈ PhD	Samara
01.02.04	18	95	207	Paimushin V. N.	Dr. P&M ≈ PhD	Kazan
01.02.04	0	115	183	Nazarov S. A.	Dr. P&M ≈ PhD	St. Petersburg
01.02.05	35	139	284	Kovalev V. L.	Dr. P&M ≈ PhD	Moscow
01.02.05	47	66	119	Kosterin A.V.	Dr. P&M ≈ PhD	Kazan
01.02.05	73	25	68	Kozelkov A. S.	Dr. P&M ≈ PhD	Sarov
01.03.01	37	135	353	Sazonov V. V.	Dr. P&M ≈ PhD	Moscow Oblast
01.04.02	12	183	339	Pavlov M. V.	Cand. P&M ≈ PhD	Moscow
01.04.02	71	91	150	Gerdzhikov V. S.	Dr. P&M ≈ PhD	Bulgaria
01.04.02	14	50	93	Manko V. I.	Dr. P&M ≈ PhD	Moscow Oblast
01.04.17	56	36	80	Smirnov N. N.	Dr. P&M ≈ PhD	Moscow
05.02.05	36	47	97	Pavlovskiy V. E.	Dr. P&M ≈ PhD	Moscow
05.13.10	51	124	264	Mandel A. S.	Dr.Tech.Sci.	Moscow
05.13.10	38	96	187	Novikov D. A.	RAS Fellow	Moscow
05.13.16	34	160	371	Galanin M. P.	Dr. P&M ≈ PhD	Moscow
05.13.16	59	72	160	Petrov I. B.	Corr. M. RAS	Moscow Oblast
05.13.18	61	305	761	Chetverushkin B.	RAS Fellow	Moscow
05.13.18	62	151	300	Kozubskaya T. K.	Dr. P&M ≈ PhD	Moscow
05.13.18	60	128	324	Galaktionov V. A.	Dr. P&M ≈ PhD	Moscow
05.13.18	8	90	183	Gasnikov A. V.	Dr. P&M ≈ PhD	Moscow Oblast
05.13.18	46	68	195	Markov M. B.	Corr. M. RAS	Moscow
05.13.19	44	21	50	Smyshlyaev S. V.	Dr. P&M ≈ PhD	Moscow

Abbreviations: ***ti,*** community type (specialty code assigned by the Higher Attestation Commission (VAK)); ***i,*** community number in the graph division (number automatically assigned to a community by Gephi); ***|Vi|,*** number of vertices within community Ci; ***|Ei|,*** number of edges between vertices in Ci; ***Leader,*** researcher who has the largest number of co-authors among others researchers in Ci (hereinafter we will use the not-quite-precise term "leader" in quotes); ***Status,*** academic status/academic degree/title of the "leader"; ***Location,*** city where the main affiliation of the "leader" as indicated in the Math-Net.Ru database is located; ***Dr. P&M,*** Doctor of Physical & Mathematical Sciences; ***Dr.Tech.Sci.,*** Doctor of Technical Sciences; ***Cand. P&M,*** Candidate of Physical & Mathematical Sciences; ***RAS Fellow,*** Fellow of the Russian Academy of Sciences; ***Corr. M. RAS,*** Corresponding member of the Russian Academy of Sciences.

Note (with respect to Table 1): A *Candidate of Science* is the first of the two science doctorate degrees awarded in Russia and in some Commonwealth of Independent States; A *Doctor of Science* is the second doctorate degree, and it is higher than a *Candidate of Science* degree; A *Candidate of Science* degree and a *Doctor of Science* degree are all more or less equivalent to a *PhD* degree in the West. They have similar scientific requirements; The designation "*Corresponding Member"* is not an academic degree or academic title, but a particular kind of membership in the Russian Academy of Sciences. Same with the designation *"Fellow*". Membership of the Russian Academy of Sciences is considered very prestigious; *Fellows* and *Corresponding Members* of the Russian Academy of Sciences mostly hold the academic degree of *Doctor of Science*. As shown in Fig. 2c, Moscow State University, the Institute of Applied Mathematics and Steklov Mathematical Institute are Russia's centres of mathematical science. The three are all located in Moscow, which is evidenced by Fig. 2b, where Moscow accounted for almost a half of the "Leaders" count.

The following approach was used in determining the type of each of the 74 communities. Among all the authors that make up a particular community, the first three with the largest co-author count were selected. Next, based on data from *Math-Net.Ru*, expert judgement was then used to determine the community type that matches the VAK nomenclature[8] (codes) of specialties for these selected authors. The point here is, specialty code is not a compulsory attribute in the database, and then we had to investigate additional information, such as the codes for journals where these authors have published some papers. The reason for the use of the old nomenclature is that for many authors in *Math-Net.Ru*, their information has not been updated since after 2021.

[8] Order No. 1027 of the Russian Ministry of Science and Higher Education dated October 23, 2017, "On Approval of the Nomenclature of Scientific Specialties for Academic Degrees". URL: https://docs.cntd.ru/document/542610966 (accessed: 17th February, 2023).

For example, for community C_{38} (row i=38 in Table 1), these are Academician D.A. Novikov[9] (19 co-authors, VAK specialty code 05.13.10), Dr. V.N. Burkov[10] (16, specialty code of his 1975 doctoral dissertation is 05.00.00), and Dr. V.A. Ivashchenko (14). According to information available on the personal pages of these scientists in *Math-Net.Ru*, their community is designated as "Control in Social and Economic Systems", which corresponds to code 05.13.10. A fragment of graph $G_{co\text{-}auth}$, consisting of two communities, highlighted (where possible) with different shades of dark colour, is shown in Figure 3. Inscriptions in the upper part of the figure indicate the community type for C_{44} and C_{29}, while vertices representing "leaders" are drawn larger and labelled with the names (last name first) of these people.

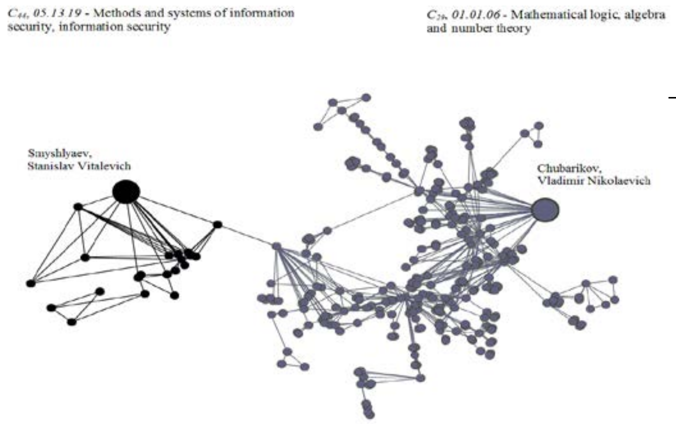

Fig. 3 A fragment of $G_{co\text{-}auth}$.

Table 2 Scientific specialties of communities and their summary characteristics

Code	**Name of scientific specialty**	**Comm. count**	**Scientist count**	**Average scientist count per community**
01.01.01	Real, complex & functional analysis	7	692	98.9
01.01.02	Differential equations, dynamical systems & optimal control	8	746	93.3
01.01.03	Mathematical physics	9	907	100.8
01.01.04	Geometry & topology	4	478	119.5
01.01.05	Probability theory & mathematical statistics	2	358	179.0
01.01.06	Mathematical logic, algebra & number theory	7	955	136.4
01.01.07	Computational mathematics	10	1036	103.6
01.01.09	Discrete mathematics & mathematical cybernetics	3	344	114.7
01.02.01	Theoretical mechanics	2	323	161.5
01.02.04	Mechanics of deformable solids	3	340	113.3
01.02.05	Fluid, gas & plasma mechanics	3	230	76.7
01.03.01	Astrometry & celestial mechanics	1	135	135.0
01.04.02	Theoretical physics	3	324	108.0
01.04.17	Chemical physics, combustion & explosion, physics of extreme states of matter	1	36	36.0
05.02.05	Robots, mechatronics & robotic systems	1	47	47.0
05.13.10	Management in social & economic systems	2	220	110.0
05.13.16	Application of computer technology, mathematical modelling & mathematical methods in scientific research	2	232	116.0
05.13.18	Mathematical modelling, numerical methods & software system	5	742	148.4
05.13.19	Methods & systems of information protection, information security	1	21	21.0

V. Discussion

The numerical features of communities, such as vertex count and edge count, vary significantly, as Table 1 illustrates. Moreover, the average co-author count (average degree of vertices) in communities ranges from 3.1 to 5.7. Some characteristics not listed in the table also vary considerably. For example, diameter d_i varies from 5 to 16, and the average path length apl_i from 2.4 to 6.2.

These findings align with the widely held beliefs regarding the geography of science in Russia[11]. They also subtly validate the suitability of our model.

A rigorous analysis of the communities yielded quite interesting results. For example, in community C_{42} (01.01.02, *Differential Equations*, *Dynamical Systems and Optimal Control*) "led" by a *Candidate of Science*, there are 18 researchers, 12 of whom work at the Moscow Power Engineering Institute (including the "leader"), 2 at RUDN University, 3 at universities in Uzbekistan and 1 at a university in Tajikistan. Based on publications by members of this community, a rather narrow research field – "Regular and singular perturbation theory" – was designated as the theme of this community.

Community C_{57} (of the same type as C_{42}), whose "leader" is Prof. Dr. Henk W. Broer from the Netherlands, has 35 members from foreign countries (Brazil, UK, Italy, Canada, etc.), that are experts in the subject of chaotic dynamics. All joint papers written by the community members have been published in English-language journal, *Regular and Chaotic Dynamics*[12].

To list all our findings from investigation of all communities would take a lot of space. Furthermore, that would be outside the purview of this paper. A quick intermediate result that needs further exploration is that

[9] Dmitry Alexandrovich Novikov. URL: https://www.mathnet.ru/php/person.phtml?option_lang=rus&personid=33032 (accessed: 24th February, 2023).
[10] Vladimir Nikolaevich Burkov. URL: https://www.mathnet.ru/php/person.phtml?option_lang=rus&personid=51029 (accessed: 24th February, 2023).

[11] Geography. URL: https://geographyofrussia.com/karty-nauka-v-rossii (accessed: 3rd December, 2023).
[12] Regular and Chaotic Dynamics. URL: http://rcd.ics.org.ru (accessed: 3rd December, 2023).

small communities with "leaders" who have a low status in Russian mathematics (perhaps for the time being) focus on important but very narrow research areas.

Table 2 provides further significant insights on the primary research areas. With the grouping of 74 communities into 19 scientific specialties, we were able to identify six research areas involving both the largest number of communities and the maximum member count. The six are: “*Computational Mathematics*”, “*Mathematical Physics*”, “*Differential Equations, Dynamical Systems and Optimal Control*”, “*Mathematical Logic, Algebra and Number Theory*”, “*Real, Complex and Functional Analysis*”, “*Mathematical Modelling, Numerical Methods and Software Systems*”.

The total number of researchers working in these mathematics fields accounts for over 60% of the investigated author set that was used to build co-authorship graph $G_{co\text{-}auth}$. In terms of membership size and community count, three research areas are conspicuously different from others. They are: “*Robots, Mechatronics and Robotic Systems*”, “*Chemical Physics, Combustion and Explosion, Physics of Extreme States of Matter*”, and “*Methods and Systems of Information Protection, Information Security*”. The other 10 areas occupy an intermediate position.

VI. Conclusion

The paper proposed a methodology for investigating thematic communities based on the co-authorship graph of researchers working within the same scientific discipline. Breaking down the graph into thematic communities made it possible to reveal the basic characteristics of communities, such as community type (thematic research field), number of scientists in each community and number of links between them.

Consequently, a closer examination of the resulting communities enabled us to:

a. build an inventory of the main research areas within a particular scientific discipline,
b. carefully examine communities that have clear deviations from the average characteristics, and offer insightful justifications for such deviations,
c. create a list of aggregated research areas that include communities of the same type and acquire generalized data, such as total community count and scientist count in each aggregated area.

The proposed research methodology was tested using *Math-Net.Ru* data, and the findings are reported. The main results include a list of six research areas in mathematics that were most actively developing in Russia between 2000 and 2020, as well as the three fields with the fewest number of members and communities.

From a practical point of view, the findings can be immediately applied for making incentives policies that would promote research in those identified areas—robots and robotic systems, combustion and explosion, data protection methods and systems—which are of absolute relevance and practical value today. It might be suggested, for example, that the Russian Science Foundation (https://www.rscf.ru/en/), whose primary objective is to support basic research and development of leading research teams in various scientific fields, hold targeted competitions in these areas (e.g., within the framework of the young researcher competition projects).

Apart from mathematics, our method is immediately applicable to any other science discipline. However, this would require having available collaboration data for that particular discipline. Moreover, a large amount of such data, collected over a long period of time, as accurate as possible, would be required.

VII. Acknowledgment

The authors would like to thank the Math-Net.Ru team for the data provided and for many years of productive collaboration.

References

[1] S. L. Yang, Q. L. Yuan, and J. H. Dong, Are Scientometrics, Informetrics, and Bibliometrics Different?, Data Science and Informetrics, vol. 1, pp. 50 –72, 2020. https://doi.org/10.4236/dsi.2020.11003

[2] J. Qiu, R. Zhao, S. Yang, and K. Dong, Informetrics: Theory, Methods and Applications, first ed., Springer, Singapore, 2017.

[3] P. Kalachikhin, Combined Methods for Forecasting Scientific Achievements, Scientific and Technical Information Processing, vol. 48, no. 4., pp. 231–238, 2021. https://doi.org/10.3103/S014768822104002X

[4] A. F. Repko, R. Szostak, and Buchberger M. P., Introduction to Interdisciplinary Studies, third ed., SAGE Publications, 2019, 448 p.

[5] H. O. Witteman, and J. E. Stahl, Facilitating interdisciplinary collaboration to tackle complex problems in health care: Report from an exploratory workshop, Health Systems, vol. 2, no. 3, pp. 162–170, 2013. https://doi.org/10.1057/hs.2013.3

[6] M. Ullah, A. Shahid, I. Din, M. Roman, M. Assam, M. Fayaz, Y. Ghadi, and H. Aljuaid, Analyzing Interdisciplinary Research Using Co-Authorship Networks, Complexity, vol. 2022, no. 2524491, 13 pp, 2022. https://doi.org/10.1155/2022/2524491

[7] S.A. Lebedev, The main models of development of scientific knowledge, Herald of the Russian Academy of Sciences, vol. 84, pp. 201-207, 2014. https://doi.org/10.1134/S1019331614030095.

[8] R. Hazra, M. Singh, P. Goyal, B. Adhikari, and A. Mukherjee, Modeling interdisciplinary interactions among physics, mathematics and computer science, Journal of Physics: Complexity, vol. 4, no. 4, 2023. https://doi.org/10.1088/2632-072X/ad0017

[9] Q. Li, Overview of Data Visualization. Embodying Data: Chinese Aesthetics, Interactive Visualization and Gaming Technologies, pp. 17–47, 2020. https://doi.org/10.1007/978-981-15-5069-0_2

[10] K. Eberhard, The effects of visualization on judgment and decision-making: a systematic literature review, Management Review Quarterly, vol. 73, pp. 167–214, 2023. https://doi.org/10.1007/s11301-021-00235-8

[11] L. Leydesdorff and I. Rafols, A global map of science based on the ISI subject categories, Journal of the American Society for Information Science and Technology, vol. 60, no. 2, pp. 348–362, 2009.

[12] B. Milman and I. Zhurkovich, Analytics and bioanalytics on the maps of science, Analytics, vol. 2, pp. 34–41, 2013 (In Russian).

[13] B. L. Milman and Y.A. Gavrilova, Analysis of citation and co-citation in chemical engineering, Scientometrics, vol. 27, pp. 53–74, 1993.

[14] A. J. Seltzer and S. H. Daniel, Co-authorship in economic history and economics: Are we any different?, Explorations in Economic History, vol. 69, pp. 102–109, 2018.

[15] D. E. Chebukov, A. M. Nwohiri, A. A. Pechnikov, and E.A. Znamenskaya, Analysis of Co-Authorship Pattern in Mathematics-Related Fields, Journal of Harbin Engineering University, vol. 44, no. 12, pp. 851–861, 2023.

[16] J. Johal, M. Loukas, R. J. Oskouian and R. S. Tubbs, Political co-authorships in medical science journals, Clinical Anatomy, vol. 30, no. 6, pp. 831–834, 2017. https://doi.org/10.1002/ca.22932.

[17] M. A. Basarab, E. V. Glinskaya, I. P. Ivanov, A. V. Kolesnikov and V. I. Kuzovlev, Study into the Structure of the Scientific Coathorship Graph Using Social Network Analysis, Cybersecurity Issues, vol. 1, no. 19, pp. 31–36, 2017. (In Russian., abstract in English).

[18] M. A. Brandão and M. M. Moro, The strength of co-authorship ties through different topological properties, Journal of the Brazilian Computer Society, vol. 23, no. 5, 2017. https://doi.org/10.1186/s13173-017-0055-x

[19] O. Gerasimova and I. Makarov, Higher School of Economics Co-Authorship Network Study, 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, May 01-03, 2019, pp. 1–4, https://doi.org/10.1109/CAIS.2019.8769556.

[20] O. I. Ivanov, A. M. Kovalenko, A. V. Kolobov, V. V. Koroleva, A. V. Leonidov, and E. E. Serebryannikova, Topology of the Co-Authorship Graph in the Field of Physics in Russia, Bulletin of the Lebedev Physics Institute, vol. 47, pp. 223–227, 2020. https://doi.org/10.3103/S1068335620080060

[21] S. V. Bredikhin, V. M. Lyapunov, N. G. Scherbakova, "The structure and parameters of the unweighted co-authorship network based on DB REPEC data," Problems of Informatics, vol. 3, no. 52, pp. 56–67, 2021, (In Russian). https://doi.org/0.24412/2073-0667-2021-3-56-67.

[22] A. M. Jaramillo, H.T.P. Williams, N. Perra, R. Menezes, The structure of segregation in co-authorship networks and its impact on scientific production, EPJ Data Science, vol. 12, no. 47, 2023. https://doi.org/10.1140/epjds/s13688-023-00411-8

[23] M.E.J. Newman, The structure of scientific collaboration networks, Proceedings of the National Academy of Sciences of the USA, vol. 98, no. 2, pp. 404–409, 2001.

[24] F. D. Malliaros and M. Vazirgiannis, Clustering and community detection in directed networks: A survey, Physics Reports, vol. 533, no. 4, pp. 95–142, 2013. https://doi.org/10.1016/j.physrep.2013.08.002.

[25] P. Bedi and C. Sharma, Community detection in social networks, Wiley interdisciplinary reviews: Data mining and knowledge discovery, vol. 6, no. 3, pp. 115–135, 2016. https://doi.org/10.1002/widm.1178

[26] J. Zhang, X. Yang, X. Hu and T. Li, Author Cooperation Network in Biology and Chemistry Literature during 2014-2018: Construction and Structural Characteristics, Information, vol. 10, no. 7, 236, 2019. https://doi.org/10.3390/info10070236

[27] M. E. Newman and M. Girvan, Finding and evaluating community structure in networks," Physical Review E, vol. 69, no. 2, 2004. https://link.aps.org/doi/10.1103/PhysRevE.69.026113

[28] E. A. Znamenskaya, A. A. Pechnikov, and D. E. Chebukov, Analysis of co-authorship in mathematical journals of Math-Net.Ru, Proceedings of the 24th Conference on Scientific Services & Internet (SSI-2022), Russia, September 19-22, 2022, pp. 190–202 (In Russian but with English abstract). https://doi.org/10.20948/abrau-2022-5.

Anthony Nwohiri is a senior lecturer at the University of Lagos, Nigeria. He obtained his BSc and MSc degrees in Information Technology, St. Petersburg State University, Russia, in 2008 and 2010 respectively. He received his Ph.D. degree in Computer Science from St. Petersburg State University / Petrozavodsk State University in 2014. Dr. Nwohiri's research interests are webometrics, web mining, network analysis, algorithms, data mining, and data science. He is a member of the Computer Professionals (Registration Council of) Nigeria (CPN), Nigeria Computer Society (NCS), Association for Information Science and Technology (ASIS&T) and Association for Computing Machinery (ACM). Email: anwohiri@unilag.edu.ng

Andrey Pechnikov received his Ph.D. degree in Physics and Mathematics in 1981 from the then Leningrad State University (now St. Petersburg State University), and Doctor of Technical Science (Professor) in 2011 from Petrozavodsk State University. Currently, he is a Leading Researcher in the Laboratory for Mathematical Cybernetics at the Institute of Applied Mathematical Research (Karelian Research Centre, Russian Academy of Sciences). Prof. Pechnikov has authored above 100 publications, and his research interests include mathematical modeling, discrete mathematics, graph theory, webometrics, and scientometrics. Email: pechnikov@krc.karelia.ru