DOI: 10.63527/1607-8829-2025-3-71-91

D. Lande[1,2] (https://orcid.org/0000-0003-3945-1178),
A. Snarskii[1,2] (https://orcid.org/0000-0002-4468-4542),
D. Manko[2] (https://orcid.org/0000-0003-1848-2952),
I. Linchevskyi[1] (https://orcid.org/0000-0002-2896-9580),
V. Fedotov[1] (https://orcid.org/0000-0001-7821-0989)

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute";
[2]Institute for Information Recording, NAS of Ukraine

Corresponding author: A. Snarskii, e-mail: *asnarskii@gmail.com*

Network Analysis of Co-authorship and Research Directions in Thermoelectrics

This paper presents a methodology for constructing an extended co-authorship network in the field of thermoelectrics by integrating direct co-authorship links with thematic proximity derived from researchers' Google Scholar profiles. The network is built through targeted probing using relevant keywords (e.g., "thermoelectrics"), and link weights are determined by both the number of joint publications and shared research interests. This approach enables the identification of not only direct collaborators but also potential interdisciplinary partners, revealing the latent structure of the scientific community. Additionally, a descriptor network is constructed based on the co-occurrence of keywords across author profiles, forming an ontological map of the field. Thematic clusters are identified using the Louvain algorithm, representing core research directions. For the first time in this context, Large Language Models (LLMs) are employed to interpret cluster content by generating meaningful, human-readable labels from lists of descriptors. This allows for fast, objective, and scalable identification of scientific trends without relying on expert annotation. The analysis shows that the extended network exhibits higher density than the traditional co-authorship network, highlighting the significance of thematic connections. Centrality measures (degree and betweenness) help identify key contributors and structural bridges within the field. The proposed approach supports the analysis of scientific communities, detection of research schools, and collaboration forecasting.

Keywords: generalized co-authorship network, subject domain, LLM, scientometric service, network sounding, topic descriptors, thermoelectricity.

1. Introduction

As a result of the development of scientific information systems, new opportunities have emerged for evaluating the level of scientists, scientific schools, and studying the patterns of

Citation: D. Lande, A. Snarskii, D. Manko, I. Linchevskyi, V. Fedotov (2025). Network Analysis of Co-authorship and Research Directions in Thermoelectrics. *Journal of Thermoelectricity*, (3), 71–91. https://doi.org/10.63527/1607-8829-2025-3-71-91

scientific interaction [1]. Currently, the task of selecting expert groups and predicting [2] the collaboration of scientists in various fields, including thermoelectrics, is relevant. By considering the relationship between the common scientific interests of different scientists and their co-authorship, it is possible to form networks that can be used to address this task.

In the field of complex network theory, structures, properties, and evolution of networks, which consist of numerous interacting elements – nodes and edges, are studied. The study of complex networks is a broad field that ranges from the study of protein structures to telecommunication networks and the Internet. Among other things, this theory can help measure scientific potential, identify scientific leaders and schools, as network analysis allows visualizing and analyzing connections between scientists and their works, as well as understanding how these connections shape the structure of the scientific community. In particular, the theory of complex networks can assist in the following aspects:

- measuring scientific potential: Network analysis can show which scientists and scientific groups are the most active and influential in a particular scientific field;
- identifying leaders: co-authorship network analysis can help identify leaders in the scientific community; scientists who are the most influential may have a large number of co-authors and be frequently cited in publications;
- studying scientific schools: network analysis can also help study scientific schools and determine their influence in the scientific community.

Currently, there are numerous scientists working in the field of complex network studies, including the most influential and others. Among them are Albert-László Barabási [3–6], who explores general issues in the theory of complex networks as a science, as well as questions of influence and interaction in complex networks. Duncan Watts [7–9], in his works, also discusses the "new" science of networks, small-world network models, and their properties. Mark Newman [10–12] focuses on the physics of complex networks, their structure, and functionality. Steven Strogatz [13–15] works in the field of complex systems, network dynamics, and their applications in biology and sociology.

Network analysis can be a useful tool in scientometrics, particularly for analyzing the relationships between scientists and their works. For example, networks can be used to analyze co-authorship between scientists and universities. Citation networks can help researchers identify the most cited works and scientists in a particular field, as well as uncover new trends in science and technology.

This article is dedicated to the task of constructing and studying co-authorship networks. Such networks are constacted taking into account the joint publications of various authors. Each node in the network represents an author, and edges between nodes show the presence of joint works. Data on joint publications, which can be collected from various sources such as scientific databases (e.g., Google Scholar, Scopus), monographs, journals, conferences, are used for constructing co-authorship networks.

Such networks can be used for:

- studying the structure of scientific communities, identifying key players, and detecting the most productive authors and scientific groups;
- determining the scientific influence of authors (collaboration indices can be used to determine which authors are the most influential in a particular field);
- searching for potential partners for scientific collaboration;
- analyzing scientific trends, tracking the popularity of different topics in a particular field, and identifying groups of authors with scientific interests and popular topics within these groups;
- studying the dynamics of scientific communities, changes in community structures, and the emergence of new players in certain fields.

Co-authorship networks are related to network models of subject areas, as they are defined by scientific interactions among scientists. Based on the co-authorship networks, it is possible to identify groups of scientists working in a specific field with similar interests. The combination of co-authorship networks and network models of subject areas can help in understanding scientific disciplines and their development, as well as in assessing the scientific productivity of scientists.

Co-authorship networks have been studied by experts in applied mathematics, scientometrics, sociology. Currently, there is a separate direction called Scholarly Data Mining [16], which is dedicated to in-depth analysis of scientific communication, including co-authorship relationships. The work [17] considers problems that arise at various stages of cooperation analysis related to data collection, network boundary establishment, relational data matrix determination, data analysis, and results interpretation. Gautam Ahuja studied the relationship between the structure of co-authorship networks and innovation in organizations in the work [18]. M.E.J. Newman laid the foundation for a modern co-authorship network research in the work [19], and in paper [20], approaches to constructing term networks as an ontological model of a subject domain were studied, specifically proposing new rules for defining syntactic and semantic relations between terms in text, as well as the directions of these relations in undirected and directed networks built on the basis of a thematic text corpus. The article [2] studied issues related to predicting potential scientific collaboration among scientists based on the analysis of scientific publications.

The so-called co-authorship networks have already become a traditional tool for studying the patterns of scientific collaboration, which allow obtaining not only scientometric evaluations, but also identifying experts for solving complex tasks. The largest scientific information services allow researchers to create their profiles containing relevant scientometric information. A significant number of studies are dedicated to the study of co-authorship networks, as well as the Google Scholar scientometric service (http://scholar.google.com/citations), confirming the relevance of this work.

Many modern scientometric services are based on methods of forming citation networks, co-authorship, identifying significant nodes, network structure, and studying relevant document collections. In particular, the work [22] presents a method for evaluating the importance of nodes in the co-authorship network based on an improved PageRank algorithm. The paper also proposes a scheme for evaluating the contribution of each author to the work. The work [23] analyzes the co-authorship network in order to find interdisciplinary scientific communities, while the work [24] investigates the network of topic flows - Topic Flow Network (TFN), which is constructed using information about each author and the article abstract.

The aim of this work is to present a new approach for constructing a network of connections between scientists through targeted sounding of available scientometric services, forming and further investigating a generalized network of scientists' collaboration taking into account the relationships of their co-authorship and meaningful correlations of their research directions.

The term "sounding" refers to the selection of a small sample of content from large networks that cannot be fully scanned for technological reasons [25]. Many modern studies of scientometric networks use mechanisms for probing them, after which conclusions are drawn about the topology of such networks. However, it has been shown in [26] that this approach can be flawed. The images of scientometric networks obtained as a result of monitoring may significantly differ and only partially reflect the properties of such networks. This can happen because the properties of these images depend heavily on the algorithms used for monitoring. Therefore, sounding a network should satisfy the condition of completeness in extracting information about objects of a certain type. In this study, the minimum citation count of authors of scientific publications serves as a constraint for probing. Scientists with citation counts below a certain threshold are not considered. Thus, in fact, a complete scan is carried out for a defined set of descriptors and this parameter of the node set, and accordingly, the network formed by them is considered.

Also, obviously, a collaboration network can become quite large if it is not limited to a specific topic. In our research, the topic is thermoelectricity. Deviating from this topic and studying co-authors who have only an indirect relationship to it complicates the perception of the formed network and leads to the "topic drift" effect. To overcome this effect, topic filtering is used, i.e. descriptors that are assigned to authors in the scientometric network are used to determine their thematic orientation. Thus, the size and topology of the generalized co-authorship network depend on the thematic orientation (formally defined by tags-descriptors assigned to scientists and the boundary value of citation). It should be noted that identifying clusters in such networks can be considered as a basis for further identification of scientific schools, expert groups. In this case, a scientific school refers to a creative team of researchers united by a common field of research and having a recognized leader.

2. Mathematical notations

Let's formally consider the conditions of the problem, namely, let A be the set of authors, A_i be the author with index i, and P_i be the profile of author A_i. Let D denote the set of all existing descriptors. We are interested in the descriptors that are included in the author's profile. For simplicity, we will assume that a profile is a set of descriptors P_i and $d_j \in D$ is a descriptor with index j. Let $\hat{d}_j^i$ denote the presence indicator of descriptor with index j in author with index i:

$$\hat{d}_j^i = \begin{cases} 1, d_j \in P_i, \\ 0, d_j \notin P_i. \end{cases} \tag{1}$$

The vector $\bar{A}_i = \left(\hat{d}_1^i, \hat{d}_2^i, \dots, \hat{d}_{|D|}^i\right)$ is assigned to the author with index i.

We will consider the scalar product of the corresponding vectors as the thematic proximity of the interests of the authors with indices i and k:

$$Sim(A_i, A_k) = (\bar{A}_i, \bar{A}_k). \tag{2}$$

The co-authorship relation between authors with indices i and k is denoted by $Co(A_i, A_k) \in \{0.1\}$.

Accordingly, in these notations, the connection in the generalized network of co-authorship of scientists between authors with indices i and k is equal to

$$Link(A_i, A_k) = Sim(A_i, A_k) + C \cdot Co(A_i, A_k), \tag{3}$$

where C is a constant that is chosen by experts.

The set of all possible values of $Co(A_i, A_k)$ forms a matrix of simple co-authorship. Thus, the matrix corresponding to the network of scientists is a combination of the network of thematic interests and the co-authorship network. As a result, the matrix of the network of scientists is denser.

3. Algorithm

The algorithm for probing the reference network of a scientometric information service and further forming a network of scientists has been adapted to the real co-author network of the service (in this case, Google Scholar) as follows (Figure 1):

Step 1: Descriptors (keywords or tags specified in authors' profiles) are selected as the base for probing (initially, one descriptor is chosen, in our case, the obvious one – “thermoelectricity” Figure 2).

Step 2. Using the tools of the scientometric service, all authors who have assigned themselves the chosen descriptor/descriptors (tags) are selected. As a result of this selection, authors are sorted in descending order based on their number of citations. To construct the network by probing, authors with citation values no less than a predetermined threshold τ (for example, $\tau = 5000$) are considered.

Step 3. The list of descriptors assigned to the authors identified in step 2 is examined. Descriptors that correspond to the primary topic are selected. This process can be carried out by an expert or automatically, for example, in this case, by the presence of the word fragment "thermoel". In particular, in the present case, the pages of the authors under the first descriptor contain descriptors that correspond to the primary topic, such as thermoelectric, thermoelectrics.

Step 4. For each of the authors identified in Step 2, their co-authors with citation values not less than the threshold τ are considered. Only those scientists whose descriptors are close to the primary topic of thermoelectricity are considered as nodes in the network. These authors are also considered as nodes of the future network of scientists. The corresponding descriptors are also considered, among which such descriptors as thermoelectric, thermoelectrics, thermoelectric materials, topological thermoelectrics, photo-thermoelectric coupling, thermoelectric properties, thermoelectric materials, thermoelectric generator, thermoelectric devices, ionic thermoelectric, etc. were found.

Step 5. For all selected descriptors, authors who have assigned these descriptors to themselves are chosen. If the list of authors with citation values greater than τ for all selected descriptors is exhausted, the process is terminated. Otherwise, the algorithm returns to Step 2.

Obviously, the presented algorithm converges, since the number of scientists covered by the scientometric service is limited.

The weight value of connections between author nodes in the network corresponds to the number of common descriptors assigned to them. Additionally, if there is co-authorship between authors, a certain constant is added to the weight of the corresponding connection, see formula (3).

Thus, the presented algorithm allows to form a matrix of a generalized co-authorship. It is also possible to consider a simple co-authorship matrix by not taking into account the first term in formula (3), that is, by not considering the links between descriptors. Such networks have been studied in many works and are of certain interest. The properties of such a network will be considered below.

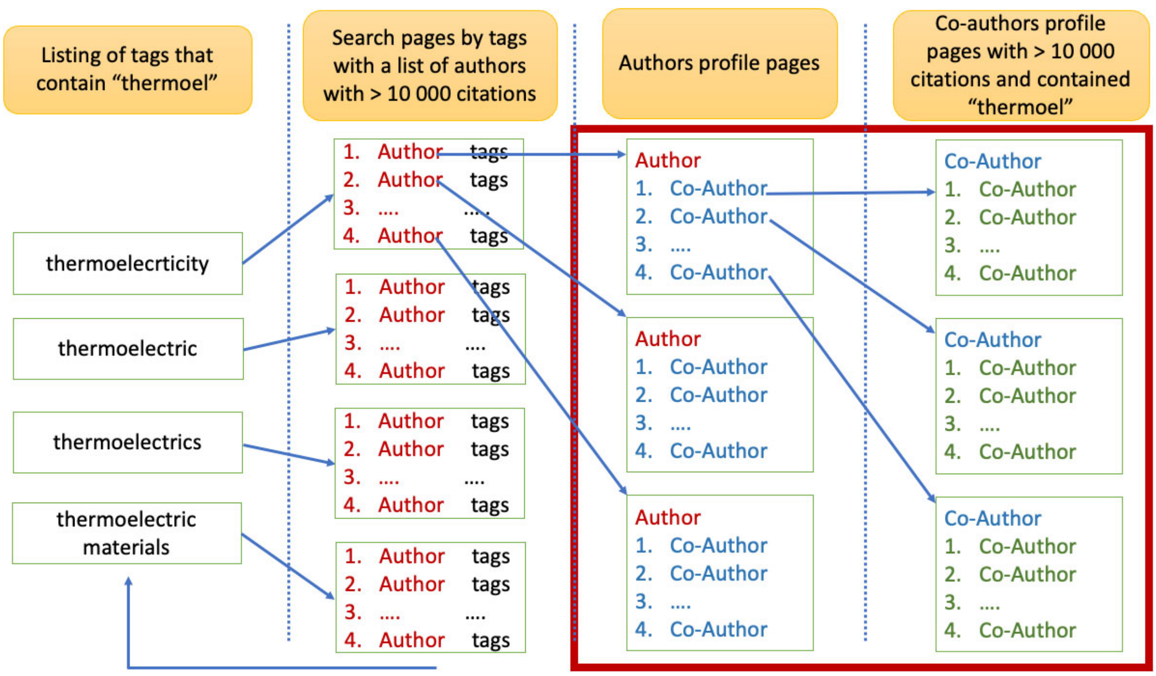


Fig. 1. Advanced service sensing algorithm Google Scholar Citations

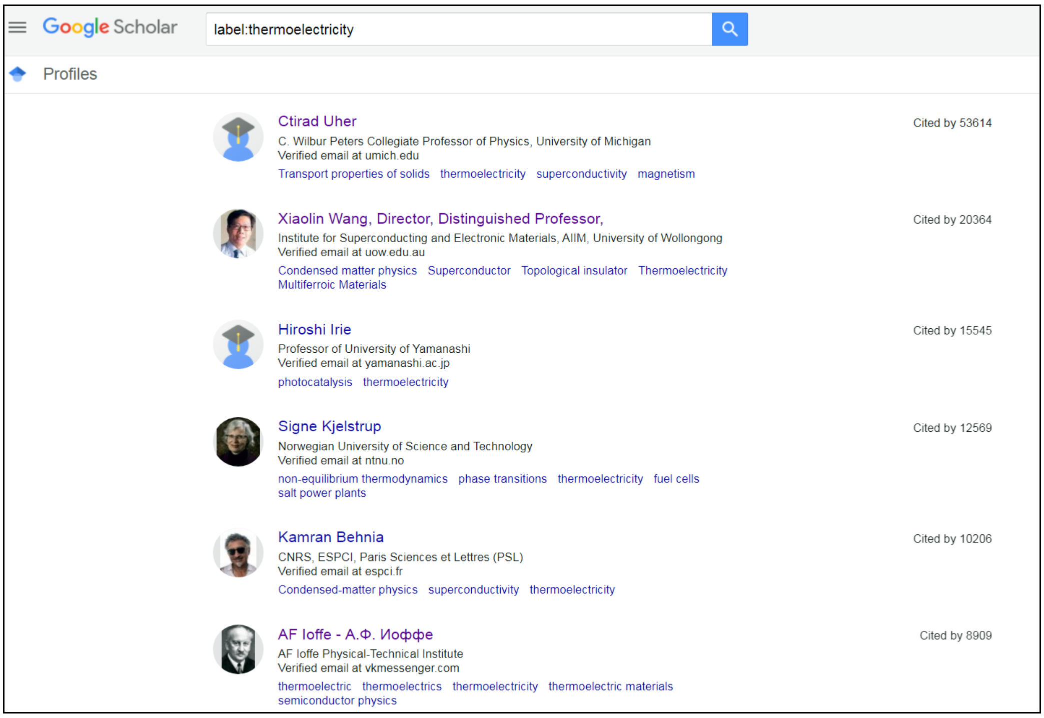


Fig. 2. Fragment of the search results for the descriptor "thermoelectricity"

4. A network of descriptors as a domain model

As an intermediate result, a kind of dual auxiliary network that is formed during the construction of a co-authorship network can be viewed as a network where the nodes are descriptors. The links in such a network can be determined, for example, as follows. Let be A_j the set of descriptors for author i. If $\hat{d}_j^i$ is an indicator of the presence of descriptor j for author i, then the weight of the link between descriptors with indices j and k is $Link(A_j, A_k)$ equal

$$Link(A_i, A_k) = \begin{cases} 1, \exists i\colon\ \hat{d}_j^i \in A_i, \hat{d}_k^i \in A_i, j \neq k, \\ \qquad 0, otherwise. \end{cases} \tag{4}$$

This network can also have weighted edges $Link(A_j, A_k)$ if we consider them as the number of shared occurrences of descriptors with indices j and k across different authors' sets of descriptors.

It should be noted that descriptors not directly related to thermoelectrics can also be included in this formed network as nodes. If such descriptors have low weight, they can be ignored, but if they have significant weight, their deep connections with the primary topic should be investigated as they may define new directions within this topic.

Clearly, this network can be clustered using cluster analysis methods, where individual clusters correspond to different subtopics within the primary topic. Thus, the formed network can be considered as a model of the subject area.

Figure 3 shows the central fragment of the formed network of descriptors corresponding to the topic of thermoelectrics. Table 1 lists the top 30 descriptors related to thermoelectrics with the highest degrees in the given network. The degree distribution of nodes in such a network follows a power law, and its graph in a logarithmic scale is presented in Figure 4 (where the abscissa corresponds to the descriptor numbers sorted by frequency and the ordinate corresponds to the degrees of the corresponding nodes in the network).

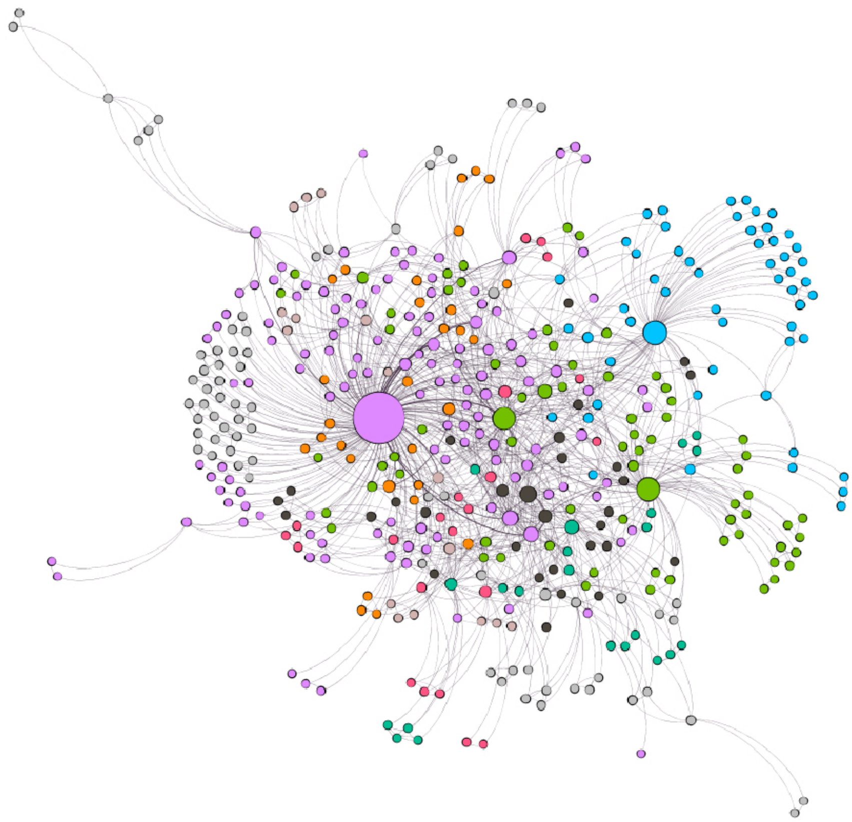


Fig. 3. A fragment of the descriptor network

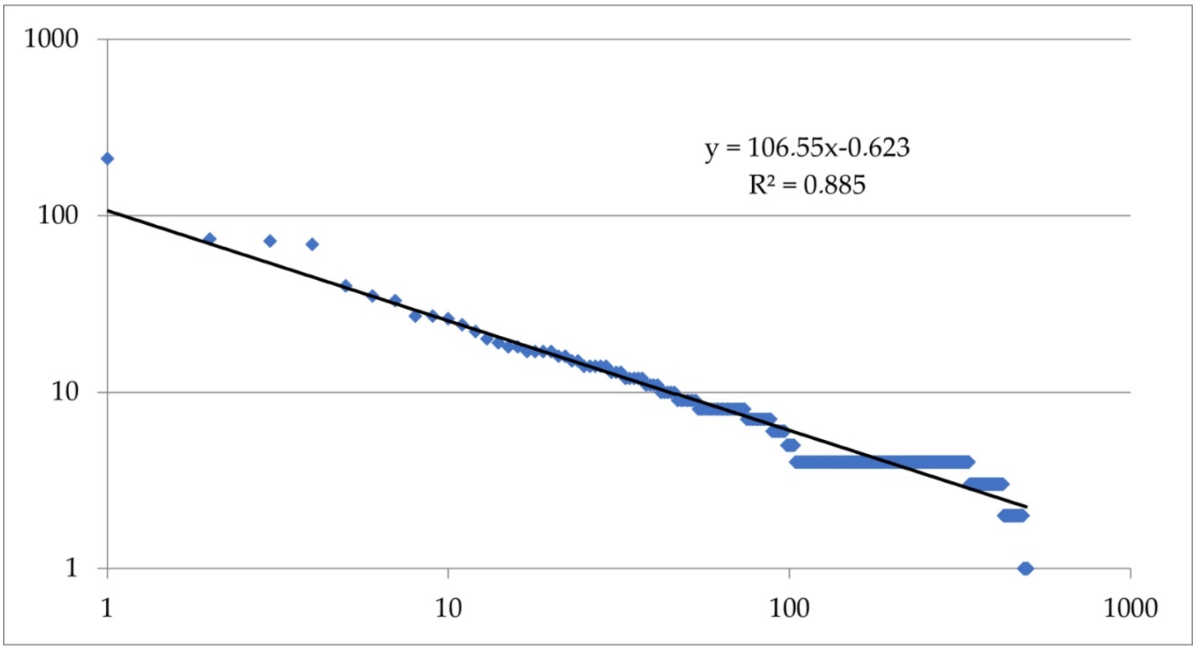


Fig. 4. Ranking distribution of degrees of descriptor network nodes

Table 1

Descriptors with the largest degrees

Rank	Descriptor	Node Degree
1	thermoelectrics	211
2	thermoelectric	74
3	thermoelectric materials	72
4	materials science	69
5	thermoelectricity	40
6	condensed matter physics	35
7	magnetism	33
8	energy materials	27
9	heat transfer	27
10	nanomaterials	26
11	superconductivity	24
12	energy storage	22
13	photovoltaics	20
14	nanotechnology	19
15	thin films	18
16	organic electronics	18
17	optoelectronics	17
18	batteries	17
19	computational materials science	17
20	solid state chemistry	17
21	solar cells	16
22	graphene	16
23	quantum materials	15
24	chemistry	15
25	spintronics	14
26	material science	14
27	materials physics	14
28	catalysis	14
29	nanowires	14
30	fuel cells	13

The network was clustered using the modularity algorithm (Louvain algorithm [26]), which is based on maximizing modularity – a measure that determines how densely connected the nodes are within a cluster compared to a random network.

The formal definition of the modularity measure is given by the formula:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta\left(c_i, c_j\right), \tag{5}$$

where A_{ij} – is an element of the graph's adjacency matrix (1 if there is an edge between vertices *i* and *j*, otherwise 0); k_i, k_j – are the degrees of vertices *i* and *j* (the number of edges incident to each vertex); *m* – is the total number of edges in the graph; c_i, c_j – are the communities to which vertices *i* and *j* belong; $\delta\left(c_i, c_j\right)$ – is a function that equals 1 if $c_i = c_j$, and 0 otherwise; γ – is the resolution parameter (gamma), which controls the weight of the random connection distribution. Typically $\gamma > 0$.

In our case, the value $\gamma = 1$ was used, which allowed the identification of several large clusters within the term network. The analysis of these clusters was performed using a large language model (LLM) [27] through the following prompt:

A cluster analysis of concepts in the field of thermoelectrics has been performed. Below is the list of concepts belonging to one of the clusters. How would you name this cluster? Here are the concepts from it: thermoelectric materials, condensed matter physics, energy storage, photovoltaics, nanotechnology, ...

Performing similar prompts allowed assigning the following main research directions to the identified modularity classes (clusters), which align with the authors' expert assessments:

1. Materials Science and Fundamental Research in Thermoelectrics: Integration of Nanotechnology, Energy, and Advanced Materials
2. Nanomaterials and Innovative Technologies: Photocatalysis, Energy Applications, and Crystal Growth
3. Energy Materials and Computational Materials Science: Condensed Matter Theory, Nanomaterials, and Machine Learning
4. Electrochemical Systems and Novel Functional Materials: Batteries, Catalysts, and Energy Technologies
5. Thermal Processes and Nanotechnology: Heat Transfer, Thermoelectricity, and Quantum Materials
6. Thermoelectrics and Nanotechnology: Energy Materials, Photonics, and Innovative Applications
7. Materials Science and Fundamental Research: Thermoelectrics, Magnetism, and Advanced Nanotechnology
8. Organic Electronics and Nanotechnology: Solar Energy, Computational Modeling, and Applications.

5. A simple co-authorship network

The co-authorship network is of significant interest as it allows for the identification of scientific groups and schools. Figure 5 shows a fragment of the network of scientists in the field of thermoelectrics, constructed using the algorithm with a citation threshold of $\tau = 5000$ and a set of descriptors. As we can see, the network of scientists, which contains 278 nodes, has high connectivity, 1 connected component, and clearly defined clusters, which were identified by modularity classes in the Gephi software (Gephi.org) [27] environment.

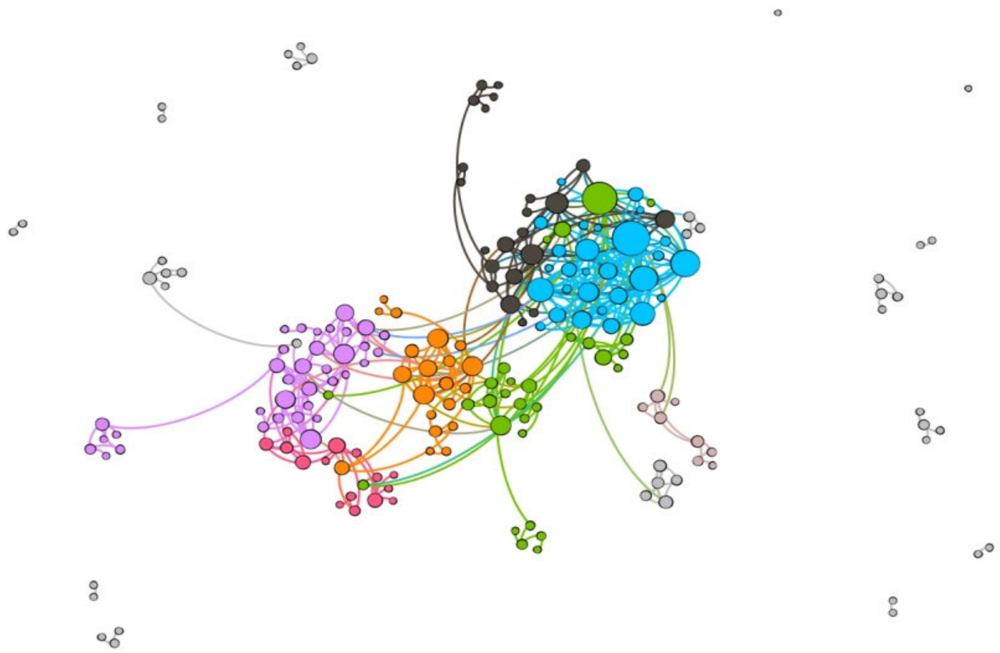


Fig. 5. Fragment of a simple co-authorship network

As can be seen from the figure, authors are grouped into several dense subgraphs, each of which is close to a complete graph. Connections between groups are sparse. In addition to a few large groups, there are a large number of very small groups. Note that the presence of a single node (author) not connected to others does not necessarily mean that the author has no co-authors – it may be that their co-authors have citation counts below the threshold value we set (5000), so they are not included in the network. Including such co-authors would introduce new connections, but the division of the network into main groups (schools) would remain unchanged. Finally, it should be noted that the division into groups was based on the modularity criterion – a measure of the structure of networks that is used to divide them into groups (modules, communities, cliques) with much denser connections within them than between individual groups.

The co-authorship network allows for the analysis of individual nodes (authors), including finding their ranked distribution. Below, Table 2 presents the first 30 authors in the co-authorship network, ranked by degree (the number of links corresponding to the author node) and by betweenness, a characteristic of node centrality determined by the number of paths between any two nodes that pass through the given node relative to all possible paths. As seen in the table, the list of the first 30 authors ranked by degree and betweenness only partially coincide. This is because the former is influenced by the number of publications (considered in the entire network), while the latter considers paths such as "the co-author of my co-author" (indirect links between authors). Figure 6 shows the ranked distribution of nodes in the simple co-authorship network, ranked by degree. As can be seen, this distribution follows a logarithmic function.

Table 2

The author ranks of a simple co-authorship network

Name	**Node Degree**	**Name**	**Node Betweenness**
Kanatzidis	22	G. Jeffrey Snyder	0.1128
G. Jeffrey Snyder	20	Kanatzidis	0.0660
Gangjian Tan (谭刚健)	16	Weishu liu	0.0548
Ctirad Uher	15	David J. Singh	0.0479
Li-Dong Zhao	14	Li-Dong Zhao	0.0374
Jiaqing HE	13	Wang Heng	0.0363
Chris Wolverton	12	Gang Chen	0.0363
Terry Tritt	11	Keivan Esfarjani	0.0359
Xianli Su (苏贤礼)	11	Ronggui Yang 杨荣贵	0.0357
Zhifeng Ren	10	Li Shi	0.0354
David J. Singh	10	Zachary M. Gibbs	0.0351
Zachary M. Gibbs	10	Olivier Delaire	0.0347
Gang Chen	10	Daryoosh Vashaee	0.0325
Brian Sales	10	Mildred S. Dresselhaus	0.0317
Andrew F. May	10	Zhifeng Ren	0.0289
Timothy Hogan	10	Andrew F. May	0.0265
Jian He	9	Ali Shakouri	0.0257
Xinfeng Tang (唐新峰)	9	Terry Tritt	0.0250
Kanishka Biswas	9	S. Joseph Poon	0.0230
Keivan Esfarjani	9	Marisol Martin-Gonzalez	0.0218
Olivier Delaire	8	Georgy Samsonidze	0.0216
Vinayak Dravid	8	Eric Toberer	0.0207
Di Wu	8	A Majumdar	0.0194
Jie Ma	8	Chris Wolverton	0.0191
Zihang Liu 刘紫航	8	Gangjian Tan (谭刚健)	0.0187
A Majumdar	7	Jie Ma	0.0181
Wang Heng	7	Boris Kozinsky	0.0175
Jihui Yang	7	anke weidenkaff	0.0174
Shanyu Wang (王善禹)	7	Maria Ibáñez	0.0174
David N Seidman	7	David R G Mitchell	0.0174

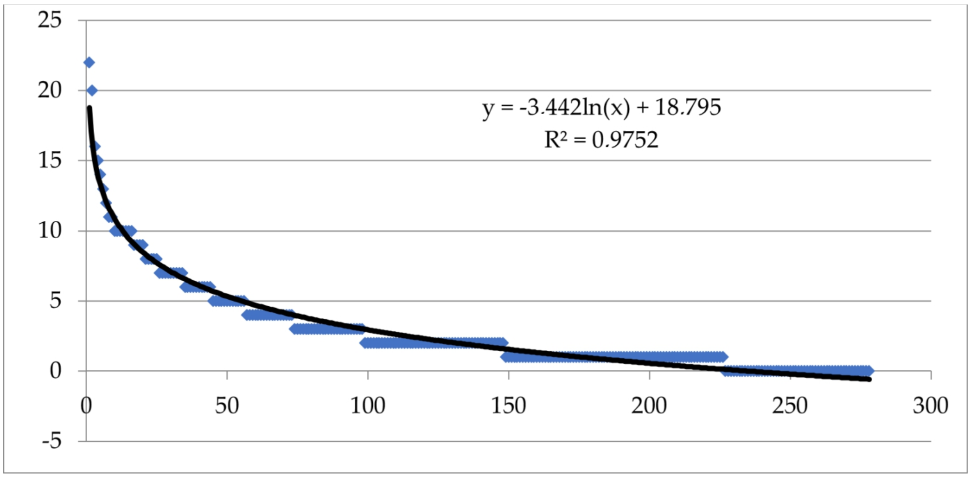


Fig. 6. The rank distribution of authors of a simple co-authorship network by node level

6. Generalized co-authorship network

Figure 7 shows a network formed according to the algorithm described above based on criteria of co-authorship and thematic proximity for the same set of authors and connections calculated according to the expression (3).

As can be seen, clusters in this network are not as clearly defined, but there is an evident "nucleus zone" – a practically fully connected subgraph, as well as numerous "leaves" – the appearance of which is explained by the second component of the formula mentioned above. Obviously, this network is much denser compared to the simple co-authorship network, but the modularity-based clustering method allows for dividing it into separate groups (which are visually less obvious).

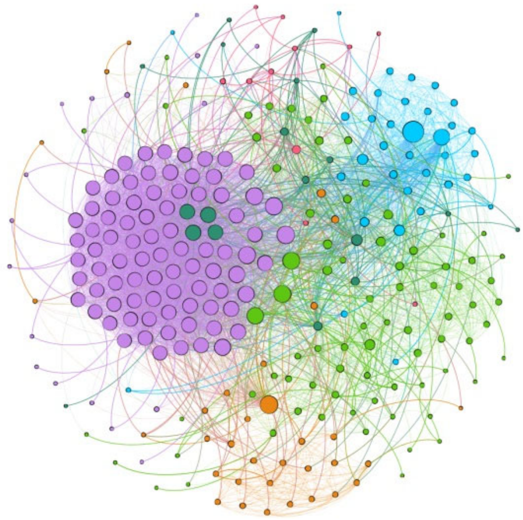


Fig. 6. Generalized co-authorship network

At the same time, the rank distribution of authors' degrees also changes qualitatively. The graph (Figure 7) clearly shows the division of authors into parts – with large ranks (from 1 to approximately 100, Figure 8) and smaller ones (from approximately 100, Figure 9). The plotted trend lines show that these two parts have a fundamentally different kind of trend. For ranks at the beginning of the scale, the distribution is power-law:

$$S(A \leq 97) \approx 137 \cdot A^{-0.1} \tag{6}$$

and for larger values of ranks

$$S(A \geq 98) \approx -45 \cdot \ln A + 251 \tag{7}$$

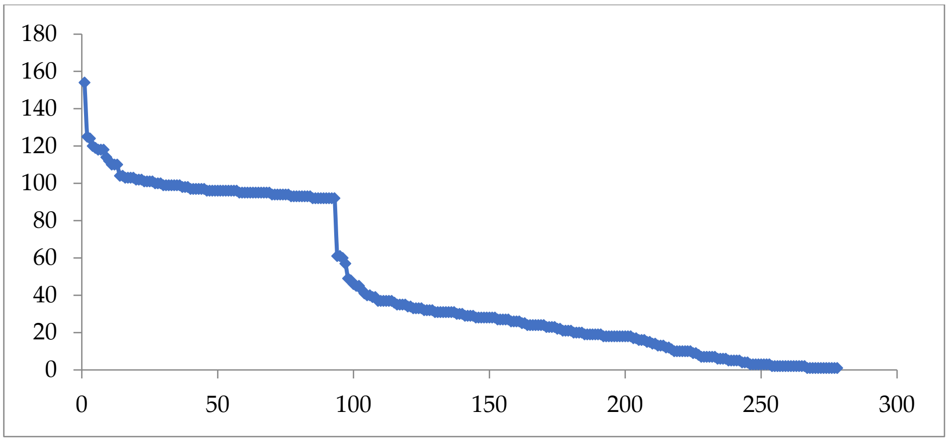


Fig. 7. Ranked distribution of authors by degrees for all authors

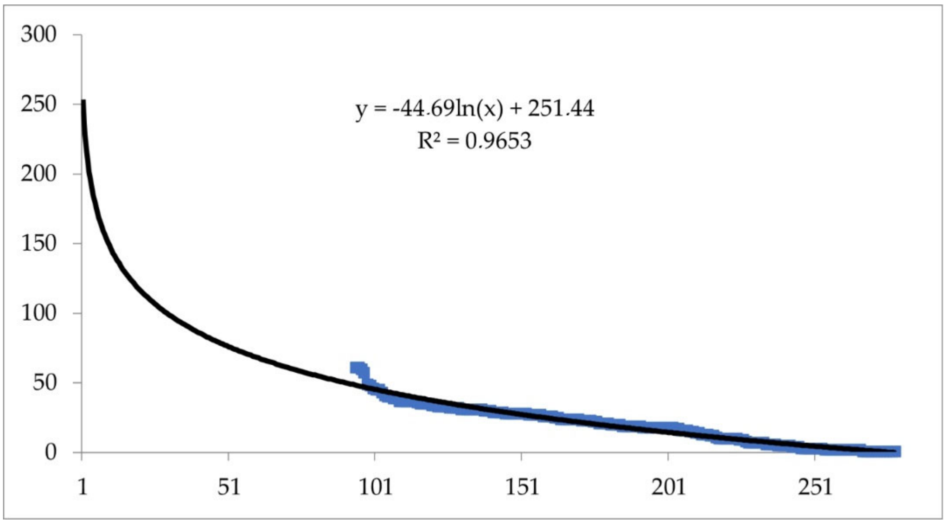


Fig. 8. Ranked distribution of authors by degrees for authors with rank from 1 to 97

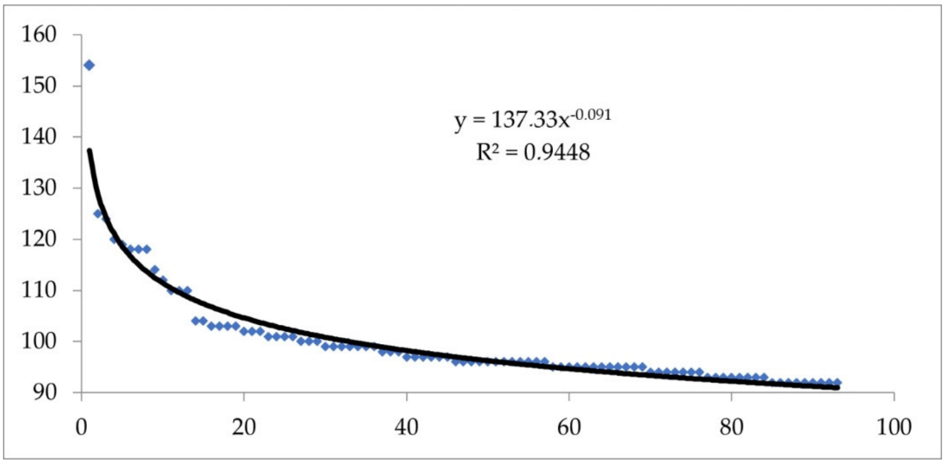


Fig. 9. Ranked distribution of authors by degrees for authors from rank 98 to the end

The qualitatively obtained division of the rank distribution of degrees may be related to the fact that we are dealing with the overlap of two networks – the co-authorship network and the network of correlations between descriptors. As shown earlier, the simple co-authorship network also had a logarithmic rank distribution.

Table 3 shows the list of authors with the highest ranking according to Google's Hirsch index (h-index).

Table 3

List of authors ranked by the Hirsch index (h-index)

Name	**Hirsh-Index (Google)**
Gerbrand Ceder	159
Lidong Chen	107
Brian Sales	97
Jiaqing HE	82
Kornelius Nielsch	74
Shuo Chen （陈硕）	73
Xinfeng Tang (唐新峰)	71
Natalio Mingo	69
Terry Tritt	67
Andreu Cabot	62
Robert A Taylor	61
Armin Feldhoff	59
Kanishka Biswas	58
Eric Toberer	57
Jian He	55
Weishu liu	55
Georgy Samsonidze	51
Subhendra D Mahanti	49
Gerhard Jakob	48
Anders E.C. Palmqvist	47
Choongho Yu	46
Devashi Adroja	46
Nuo YANG	45
Andrew F. May	43
Jorge O. Sofo	42
Gangjian Tan (谭刚健)	40
Wang Heng	38
Graeme R. Blake	37
Matthieu Verstraete	35
Michitaka Ohtaki	34

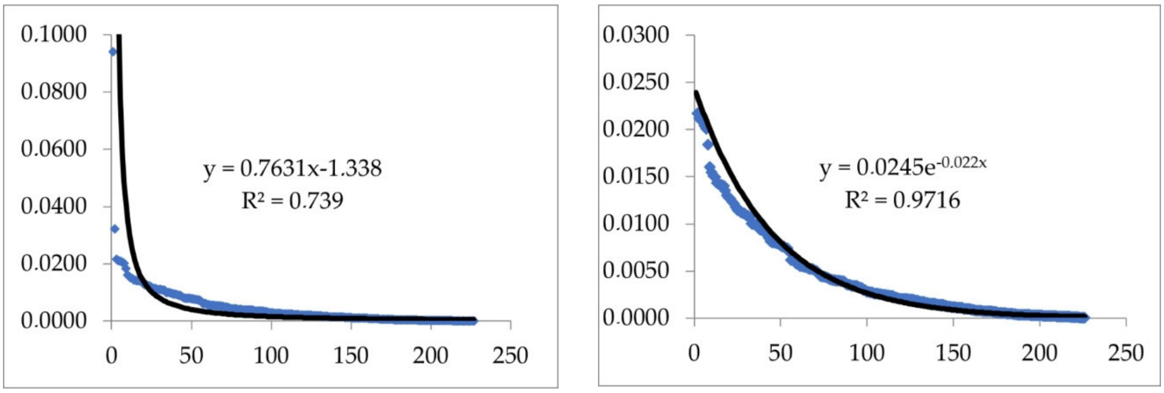
Table 4 lists the first 30 authors of the generalized co-authorship network, ranked by author centrality levels. As can be seen from the table, the list of the first 30 authors ranked by node degree and betweenness only partially match.

Table 4

Lists of the 30 most ranked authors in the generalized network of co-authorship

Name	Node Degree	Name	Node Betweenness
A.F. Ioffe	154	A.F. Ioffe	0.094
Andreu Cabot	125	Michitaka Ohtaki	0.0321
Matthieu Verstraete	124	Brian Sales	0.0217
Weon Ho Shin	120	Terry Tritt	0.0212
Michitaka Ohtaki	119	Kanishka Biswas	0.0209
Andrew F. May	118	Weon Ho Shin	0.0208
Georgy Samsonidze	118	Andreu Cabot	0.0204
Jie Ma	118	Jorge O. Sofo	0.0201
Brian Sales	114	Matthieu Verstraete	0.0184
Jian He	112	Kornelius Nielsch	0.016
Natalio Mingo	110	Andrew F. May	0.0155
Jorge O. Sofo	110	Kamran Behnia	0.0151
Gerhard Jakob	110	Anders E.C. Palmqvist	0.0149
Armin Feldhoff	104	Rachel Segalman	0.0143
Graeme R. Blake	104	Gerhard Jakob	0.0142
Jiaqing HE	103	Armin Feldhoff	0.0142
Kanishka Biswas	103	Boris Kozinsky	0.0141
Boris Kozinsky	103	Marisol Martin-Gonzalez	0.0141
Subhendra D Mahanti	103	Jiaqing HE	0.0135
Shuo Chen （陈硕）	102	G. Jeffrey Snyder	0.013
Robert A Taylor	102	Jian He	0.0127
Anders E.C. Palmqvist	102	Nuo YANG	0.0126
Eric Toberer	101	Devashi Adroja	0.0123
Di Wu	101	Georgy Samsonidze	0.0122
Choongho Yu	101	Rama Venkatasubramanian	0.01178
Nuo YANG	101	Jie Ma	0.0115
Zachary M. Gibbs	100	Boris Kozinsky	0.0115
Gerbrand Ceder	100	anke weidenkaff	0.0112
Weishu liu	100	Maria Ibáñez	0.0111
Xinfeng Tang (唐新峰)	99	David R G Mitchell	0.011

It is interesting to examine the degree centrality distribution of authors based on their betweenness centrality (Figure 10), which follows an exponential function.



a) *b)*

Fig. 10. Rank distribution of node degrees by Betweenness centrality parameter: (a) All authors – the trend line is exponential with an exponent of –0.023 and with low significance R^2 = 0.74; (b) Without the first author – the distribution is exponential with almost the same exponent of –0.022 but with high significance R^2 = 0.97

As seen in Figure 10a, one of the authors has a significantly higher betweenness value compared to the other authors. This author is one of the founders of the scientific direction and technical applications of thermoelectrics – A.F. Ioffe. The fact that the trend of the distribution (black smooth line in Figures 10 and 11) fits much better to the distribution without the first unique author also points to his special position. It should be noted that it is precisely the betweenness of the network by tags and co-authorship that allows determining the most important position of the author within a given topic. His exceptional position led to an increase in the reliability of our empirical data to 0.97.

7. Conclusions

An approach for forming a network of scientists within the subject area of Thermoelectrics has been proposed and implemented. The algorithm for forming this network is limited by markers of knowledge (descriptors) that are predetermined by scientists as participants of the scientometric service in their profiles.

It should be noted that the proposed model for automatic formation of networks of scientists differs fundamentally from existing ones that rely on direct participation of expert individuals in author selection. In the proposed algorithm, both authorship relationships and content correlation of descriptors assigned to authors are used to construct the generalized co-authorship network. Thus, the network scanning program uses knowledge embedded by the authors themselves, which significantly expands the expert environment.

As an intermediate result, the main contemporary research directions in the field of thermoelectrics were identified using an LLM, based on the content of the descriptors and the language model built for this subject area.

It should be noted that the lists of scientists corresponding to the largest nodes of the two networks mentioned are different. In addition, it can be seen that the co-authorship indices of scientists corresponding to the largest nodes, ranked by Hirsch index, do not coincide and are more appropriate for the subject area under study.

Thus, the proposed network has several important advantages for analysis:

- small diameter of the graph and average path length, which can lead to the formation of expert groups of scientists who are not direct co-authors;
- limited number of clusters of scientists that correspond clearly to descriptors, i.e. topics.

In the end, considering not only one criterion of co-authorship increases the variability of solutions and allows regulating the balance between clustering and thematic similarity. In addition to the proposed network, a related network can be considered, whose nodes are descriptors and links are defined by the number of authors assigned to corresponding pairs of descriptors. Such a network can be viewed as a model of the primary subject area. The research results provide a scientific basis for automating and accelerating the process of selecting competent experts to address various issues in the field of thermoelectrics. Although the model was applied within the framework of the Google Scholar service, but the proposed approach can also be used for other scientometric services as well.

The research was supported by a grant from the National Research Foundation of Ukraine (project registration number 2023.04/0087).

Authors' information

Dmytro Lande – Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Information Security.
Andrii Snarskii – Doctor of Physical and Mathematical Sciences, Professor at the Department of General Physics.
Dmytro Manko – Ph.D., Senior researcher in computer engineering.
Igor Linchevskyi – Doctor of Physical and Mathematical Sciences, Professor at the Department of General Physics.
Viacheslav Fedotov – Senior lecturer at the Department of General Physics.

References

1. Ortega J.L. (2015). How is an academic social site populated? A demographic study of Google Scholar Citations population. *Scientometrics*, *104*(1), 1–18. https://doi.org/10.1007/s11192-015-1593-7
2. Lande D., Fu M., Guo W., Balagura I., Gorbov I., & Yang H. (2020). Link prediction of scientific collaboration networks based on information retrieval. *World Wide Web*, *23*(4), 2239–2257. https://doi.org/10.1007/s11280-019-00768-9
3. Barabási A.-L. (216 C.E.). *Network science* (1st edition). Cambridge University Press.
4. Meng X., Varol O. and Barabási A.L. (2024). Hidden citations obscure true impact in science. PNAS nexus, 3(5), 155. https://doi.org/10.1093/pnasnexus/pgae155

5. Liu Y.-Y., Slotine J.-J., & Barabási A.-L. (2013). Observability of complex systems. *Proceedings of the National Academy of Sciences*, *110*(7), 2460–2465. https://doi.org/10.1073/pnas.1215508110
6. Barzel B., & Barabási A.-L. (2013). Universality in network dynamics. *Nature Physics*, *9*(10), 673–681. https://doi.org/10.1038/nphys2741
7. Watts D.J. (1999a). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, *105*(2), 493–527. https://doi.org/10.1086/210318
8. Watts D.J. (2004). The "New" Science of Networks. *Annual Review of Sociology*, *30*(1), 243–270. https://doi.org/10.1146/annurev.soc.30.020404.104342
9. Almaatouq A., Alsobay M., Yin M. and Watts D.J., 2024. The effects of group composition and dynamics on collective performance. *Topics in cognitive science*, 16(2), 302–321. https://doi.org/10.1111/tops.12706
10. Newman M.E.J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, *45*(2), 167–256. https://doi.org/10.1137/S003614450342480
11. Newman M. (2008). The physics of networks. *Physics Today*, *61*(11), 33–38. https://doi.org/10.1063/1.3027989
12. Kirkley A., Cantwell G.T., & Newman M.E.J. (2019). Balance in signed networks. *Physical Review E*, *99*(1), 012320. https://doi.org/10.1103/PhysRevE.99.012320
13. Strogatz S.H. (2001). Exploring complex networks. *Nature*, *410*(6825), 268–276. https://doi.org/10.1038/35065725
14. Newman M.E.J., Watts D.J., & Strogatz S.H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences*, *99*(suppl_1), 2566–2572. https://doi.org/10.1073/pnas.012582999
15. Arbesman S., Strogatz S., & Vitevitch M. (2010). Comparative Analysis of Networks of Phonologically Similar Words in English and Spanish. *Entropy*, *12*(3), 327–337. https://doi.org/10.3390/e12030327
16. Dridi A., Gaber M.M., Azad R.M.A., & Bhogal J. (2021). Scholarly data mining: A systematic review of its applications. *WIREs Data Mining and Knowledge Discovery*, *11*(2). https://doi.org/10.1002/widm.1395
17. De Stefano D., Giordano G., & Vitale M.P. (2011). Issues in the analysis of co-authorship networks. *Quality & Quantity*, *45*(5), 1091–1107. https://doi.org/10.1007/s11135-011-9493-2
18. Ahuja G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly*, *45*(3), 425–455. https://doi.org/10.2307/2667105
19. Newman M.E.J. (2001). From the Cover: The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. https://doi.org/10.1073/pnas.021544898
20. Dmytro Lande, Oleh Dmytrenko, Oksana Radziievska. Determining the Directions of Links in Undirected Networks of Terms. Selected Papers of the XIX International Scientific

and Practical Conference "Information Technologies and Security" (ITS 2019). *CEUR Workshop Proceedings* (ceur-ws.org). – Vol-2577. – pp 132–145. ISSN 1613-0073.

21. Liu J., Li Y., Ruan Z., Fu G., Chen X., Sadiq R., & Deng Y. (2015). A new method to construct co-author networks. *Physica A: Statistical Mechanics and Its Applications*, *419*, 29–39. https://doi.org/10.1016/j.physa.2014.10.006
22. Ullah M., Shahid A., Din I. ud, Roman M., Assam M., Fayaz M., Ghadi Y., & Aljuaid H. (2022). Analyzing Interdisciplinary Research Using Co-Authorship Networks. *Complexity*, *2022*, 1–13. https://doi.org/10.1155/2022/2524491
23. Schäfermeier B., Hirth J., & Hanika T. (2022). Research topic flows in co-authorship networks. *Scientometrics*. https://doi.org/10.1007/s11192-022-04529-w
24. Lande D. (2015). A Domain Model Created on the Basis of Google Scholar Citations. *In CEUR Workshop Proceedings (Ceur-Ws. Org) Vol-1536*, 57–61.
25. Lande D., & Dmytrenko O. (2022). *Research of Topological Properties of Network Reflections Obtained Using Different Algorithms for Scanning Initial Networks* (pp. 347–361). https://doi.org/10.1007/978-3-030-89902-8_26
26. Cai G. (2024, January). Research on the Application of Louvain's Algorithm in Traditional Cultural Media Communication. *In 2024 International Conference on Optimization Computing and Wireless Communication* (ICOCWC) (pp. 1–5). IEEE. DOI: https://doi.org/10.1109/ICOCWC60930.2024.10470776
27. Lande D., Strashnoy L. Advanced Semantic Networking based on the large language models: Monograph. Kyiv: *Engineering*, 2025. 258 p. ISBN 978-617-8180-02-7
28. Cherven K. (2015). *Mastering Gephi Network Visualization*. Packt Publishing.

Submitted: 14.08.2025

Ланде Д.[1,2] (https://orcid.org/0000-0003-3945-1178),
Снарський А.[1,2] (https://orcid.org/0000-0002-4468-4542),
Манко Д.[2] (https://orcid.org/0000-0003-1848-2952),
Лінчевський І.[1] (https://orcid.org/0000-0002-2896-9580),
Федотов В.[1] (https://orcid.org/0000-0001-7821-0989)

[1]Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського";
[2]Інститут проблем реєстрації інформації НАН України

Мережевий аналіз співавторства та напрямки досліджень у термоелектриці

У даній статті представлено методологію побудови розширеної мережі співавторства в галузі термоелектрики шляхом інтеграції прямих зв'язків співавторства з тематичною близькістю, отриманою з профілів дослідників у Google Scholar. Мережа будується шляхом цілеспрямованого зондування з *використанням релевантних ключових слів (наприклад, «термоелектрика»), а ваги зв'язків визначаються як кількістю спільних публікацій, так і спільними дослідницькими інтересами. Такий підхід дозволяє ідентифікувати не лише безпосередніх співробітників, а й потенційних міждисциплінарних партнерів, розкриваючи приховану структуру наукової спільноти. Крім того, мережа дескрипторів будується на основі спільної появи ключових слів у профілях авторів, формуючи онтологічну карту галузі. Тематичні кластери, що представляють основні напрямки досліджень, ідентифікуються за допомогою алгоритму Лувена. Вперше в цьому контексті для інтерпретації вмісту кластера використовуються моделі великих мов (LLM) шляхом створення змістовних, зрозумілих для людини міток зі списків дескрипторів. Це дозволяє швидко, об'єктивно та масштабовано ідентифікувати наукові тенденції без залежності від експертних анотацій. Аналіз показує, що розширена мережа демонструє вищу щільність, ніж традиційна мережа співавторства, що підкреслює важливість тематичних зв'язків. Показники центральності (ступінь та проміжність) допомагають визначити ключових учасників та структурні зв'язки в межах галузі. Запропонований підхід підтримує аналіз наукових спільнот, виявлення дослідницьких шкіл та прогнозування співпраці.*

Ключові слова: узагальнена мережа співавторства, предметна область, LLM, наукометричний сервіс, мережеве зондування, тематичні дескриптори, термоелектрика.

Надійшла до редакції 14.08.2025