Information systems. Computer sciences. Issues of information security

Информационные системы. Информатика. Проблемы информационной безопасности

UDC 001.18:004.94:008.2
https://doi.org/10.32362/2500-316X-2023-11-3-17-29

RESEARCH ARTICLE

Dynamics of link formation in networks structured on the basis of predictive terms

Sergey O. Kramarov [1, 2],
Oleg R. Popov [3, @],
Ismail E. Dzhariev [2],
Egor A. Petrov [2]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *Surgut State University, Surgut, 628408 Russia*
[3] *Southern Federal University, Rostov-on-Don, 344006 Russia*
[@] *Corresponding author, e-mail: cs41825@aaanet.ru*

Abstract
Objectives. In order to model and analyze the information conductivity of complex networks having an irregular structure, it is possible to use percolation theory methods known in solid-state physics to quantify how close the given network is to a percolation transition, and thus to form a prediction model. Thus, the object of the study comprises international information networks structured on the basis of dictionaries of model predictive terms thematically related to cutting-edge information technologies.
Methods. An algorithmic approach is applied to establish the sequence of combining the necessary operations for automated processing of textual information by the internal algorithms of specialized databases, software environments and shells providing for their integration during data transmission. This approach comprises the stages of constructing a terminological model of the subject area in the Scopus bibliographic database, then processing texts in natural language with the output of a visual map of the scientific landscape of the subject area in the *VOSviewer* program, and then collecting the extended data of parameters characterizing the dynamics of the formation of links of the scientific terminological network in the *Pajek* software environment.
Results. Visual cluster analysis of the range of 645–3364 terms in the 2004–2021 dynamics of the memory and data storage technologies category, which are integrated into a total of 23 clusters, revealed active cluster formation in the field of the term *quantum memory*. On this basis, allowing qualitative conclusions are drawn concerning the local dynamics of the scientific landscape. The exploratory data analysis carried out in the *STATISTICA* software package indicates the correlation of the behavior of the introduced *MADSTA* keyword integrator with basic terms including periods of extremes, confirming the correctness of the choice of the methodology for detailing the study by year.
Conclusions. A basis is established for the formation of a set of basic parameters required for an extensive computational modeling of a cluster formation in the semantic field of the scientific texts, especially in relation to simulations of the formation of the largest component of the network and percolation transitions.

Keywords: information network, algorithm, database, term, cluster, visualization, mapping, dynamics, network analysis

• **Submitted:** 11.08.2022 • **Revised:** 01.11.2022 • **Accepted:** 02.03.2023

For citation: Kramarov S.O., Popov O.R., Dzhariev I.E., Petrov E.A. Dynamics of link formation in networks structured on the basis of predictive terms. *Russ. Technol. J*. 2023;11(3):17–29. https://doi.org/10.32362/2500-316X-2023-11-3-17-29

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Динамика формирования связей в сетях, структурированных на основе прогностических терминов

С.О. Крамаров [1, 2],
О.Р. Попов [3, @],
И.Э. Джариев [2],
Е.А. Петров [2]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *Сургутский государственный университет, Сургут, 628408 Россия*
[3] *Южный федеральный университет, Ростов-на-Дону, 344006 Россия*
[@] *Автор для переписки, e-mail: cs41825@aaanet.ru*

Резюме

Цели. Для моделирования и анализа информационной проводимости сложных сетей с нерегулярной структурой возможно применение известных в физике твердого тела методов теории перколяции, позволяющих количественно оценить, насколько данная сеть близка к перколяционному переходу, и тем самым сформировать модель прогнозирования. Объектом исследования выступают международные информационные сети, структурированные на основе словарей модельных прогностических терминов, тематически относящихся к перспективным информационным технологиям.

Методы. Применен алгоритмический подход, согласно которому задается последовательность комбинирования необходимых операций по автоматизированной обработке текстовой информации внутренними алгоритмами специализированных баз данных (БД), программных сред и оболочек, предусматривающих их интеграцию при передаче данных. Данный подход, в частности, включает этапы построения терминологической модели предметной области в библиографической БД Scopus, затем обработку текстов на естественном языке с выводом визуальной карты научного ландшафта предметной области в программе *VOSviewer* и далее – сбор расширенных данных параметров, характеризующих динамику формирования связей научной терминологической сети в программной среде *Pajek*.

Результаты. Визуальный кластерный анализ, составляющий в динамике 2004–2021 гг. диапазон 645–3364 термов категории «Технологии памяти и хранения данных», интегрированных суммарно в 23 кластера, выявил активное кластерообразование в области терма «quantum memory» (квантовая память), позволяющее делать качественные выводы о локальной динамике научного ландшафта. Проведенный в программном пакете *STATISTICA* разведочный анализ данных свидетельствует о корреляции поведения введенного интегратора ключевых слов MADSTA с базовыми термами, включая периоды экстремумов, что подтверждает правильность выбора методики детализации исследования по годам.

Выводы. Заложена основа для формирования комплекса базовых параметров, необходимых при обширном вычислительном моделировании кластерообразования в семантическом поле научных текстов, особенно в отношении симуляций формирования наибольшего компонента сети и перколяционных переходов.

Ключевые слова: информационная сеть, алгоритм, база данных, термин, кластер, визуализация, картирование, динамика, сетевой анализ

• **Поступила:** 11.08.2022 • **Доработана:** 01.11.2022 • **Принята к опубликованию:** 02.03.2023

Для цитирования: Крамаров С.О., Попов О.Р., Джариев И.Э., Петров Е.А. Динамика формирования связей в сетях, структурированных на основе прогностических терминов. *Russ. Technol. J.* 2023;11(3):17–29. https://doi.org/10.32362/2500-316X-2023-11-3-17-29

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The study of information dissemination in social networks with a random topology is an urgent task for the optimization of the modern sociotechnical systems, which is confirmed by the undiminished interest in the problems of social network analysis (SNA) [1–3].

A distinct category of complex networks, along with social and biological networks, is comprised by information networks, also called "knowledge networks." An example is the citation networks between scientific publications, whose structure quite accurately reflects the structure of the information stored in its vertices (articles), which defines the terminology "information network."

When applied to an analysis of the relationships between classes of words in the thesaurus, an information network can also be seen in conceptual terms, representing the structure of language, or perhaps even the mental constructs language represents [4].

To model and analyze the information processes occurring in the networks with an irregular structure, it is possible to apply percolation methods [5] known in solid-state physics.

Percolation, which is derived from Latin (*percolare*), means "leakage, seepage." For a long time, this simple probabilistic model was the ideal basic model used in physics to demonstrate phase transitions and critical phenomena. As a mathematical object, it was first considered in the classic work of Broadbent and Hammersley in 1957 [6], in which the term as well as geometric and probabilistic concepts were introduced.

Methods for solving various theoretical and applied problems over the last decades have brought new insights into the mathematical study of percolation [7]. Although the penetration of a fluid inside a porous stone, the spreading of an epidemic or the dissemination of the information in a social network seemingly have nothing in common, all three aspects converge in mathematics to comprise an additive component [7–9].

Regardless of the physical nature and the model of the system, percolation theory in its most general form addresses questions concerning the probability that there is an open path from 0 to infinity (or whether there is an infinite cluster of connected pores or nodes). Thus, the problem is reduced to finding an answer to the question whether such paths exist for a given probability p. The theory is mainly concerned with the existence of such a cluster and its structure with respect to the filling probability p.

To mathematically describe this criticality, it is necessary to define a percolation model. As an example, we selected the simplest model represented by an infinite two-dimensional (or square) grid. The intersection points of the lines are called nodes (graph vertices), while the lines themselves are called links (graph edges). There are two grid models: permeability of links and permeability of nodes.

In the first model, each link is mathematically occupied with probability p or free with probability $1 - p$. The occupied links then connect the nodes into clusters. This model can be used to simulate the process of liquid penetration inside a porous stone and epidemic propagation.

In the node percolation model, we occupy not a link, but each node with probability p, leaving it free with probability $1 - p$. In general, link percolation is considered less general than node percolation due to the possibility of reformulating link percolation as node percolation on another grid, but not *vice versa*.

Percolation theory mainly focuses on the appearance of an infinite cluster with increasing probability p. To characterize this phenomenon, one typically takes the size of a giant cluster S, which is defined as:

$$S = \lim_{N\to\infty} \frac{S_1}{N}, \quad (1)$$

where N is the system size (total number of nodes); S_1 is the number of nodes in the largest cluster.

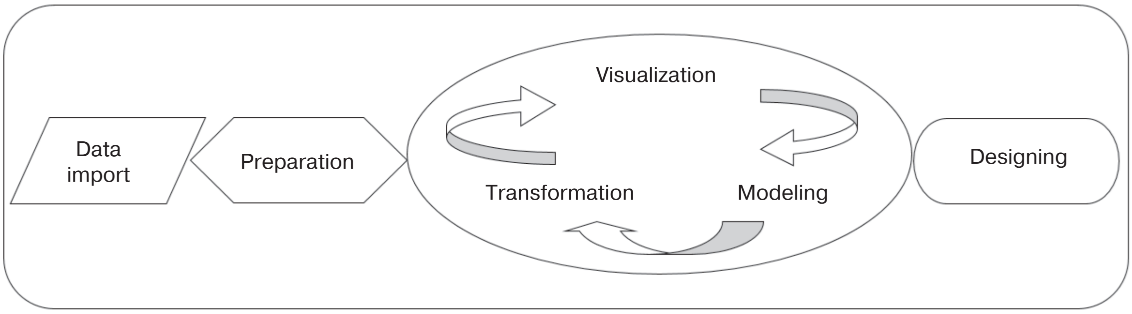


Fig. 1. Science data scientific project implementation algorithm diagram [11]

As the probability p increases, there must exist a critical p_S, called the percolation threshold or critical point, above which a non-zero value of S can be found. This determines the percolation transition of the system with respect to the control parameter p, while S is the corresponding parameter of order.

To describe the features of finite clusters, the distribution of clusters by size can also be used:

$$p_S = \frac{m_S}{\sum\limits_S m_S}, \qquad (2)$$

where m_S is the number of clusters with size S.

Models have now been constructed for three-dimensional and higher grid dimensions or for other percolation conditions.

Extensive computational modeling of social networks, both on classical 2D-regular grids and for higher-dimensional networks, has yielded results confirming the mapping of the phase of transient behavior characterized by the largest component in the proposed non-consensus opinion model with the well-known physical problem of seepage in incompressible fluids [9].

The use of computational modeling methods of random networks with a large number of links has led to analytical solutions, such as the construction of stochastic models describing the dynamics of node state changes and percolation transitions that predict the dynamics of social network behavior [2].

By analogy with social networks, we believe that a properly constructed research methodology will allow us to quantify how close the information network is to the threshold of percolation, and thus form a model for its prediction.

In this regard, it makes sense to apply mapping and network analysis techniques to study the patterns that affect the percolation threshold in information dissemination and clustering in networks structured on the basis of dictionaries of model predictive terms extracted from network resources and thesauruses, as well as scientometric and bibliographic databases.

In the article [10], the priority importance of information and communication technologies (ICT) and the field of information sciences was substantiated under the conditions of convergent tendencies of "joint" interdisciplinary links in the development of complex geoinformation systems. This position determined the main choice of the subject area for the present study.

TOOLS AND METHODS

The basic toolkit of the stages of the scientific project associated with the processing of large amounts of information comprised of science data is shown in Fig. 1. The main mechanism of knowledge generation for the final design lies in a large central block, including data visualization and modeling. Analysis of the obtained knowledge at each step will require their transformation and optimization, including narrowing the range of observations of interest, calculation of a set of summary statistics, average values, etc.

The algorithmic approach seeks to maximize the objectivity of the search and the speed with which it allows a user to delve into a given subject area of research.

The basic steps and tools for implementing the empirical part of the algorithm are considered in [12]. A variant modified to solve current research problems consists of the following iterations:

1) formation of a terminological model of the subject domain based on a matrix defining the system maturity level (SML) of self-organizing intelligent systems;
2) formation of algorithmic queries of the Scopus[1] bibliographic database by a special formula and output the data in .csv format for further processing and analysis in software tools;
3) preprocessing, vectorization, clustering of the obtained terminological database by internal algorithms of *VOSviewer*[2] software environment;
4) primary visual analysis of intra- and inter-cluster interactions by *VOSviewer* program's basic interfaces by years and terms;

[1] https://www.scopus.com/. Accessed March 01, 2022.
[2] https://www.vosviewer.com/. Accessed March 22, 2022.

5) a detailed study of the parameters of scientific terminology network links by clusters in the dynamics of development with the help of algorithms of the *Pajek*[3] software product;
6) data processing and visualizing the dynamic relationships of term network parameters in the *STATISTICA*[4] environment.

There are several ways to achieve this goal. The initial stage of this work is the formation of a terminological model of the subject domain, on the basis of which the database for further research is formed.

The basis for the selection of a given pattern of keywords is the method of calculating the SML matrix of self-organizing intelligent systems (IS) [13], allowing the maturity index of the IS to be quantified. The SML indicator, which is applied at the system level, represents a maturity index from 0 to 1.

In the framework of this study, level 4 of a sociotechnical system maturity is of interest, which has the maturity index range 0.60–0.80, and can be described as follows: "the predicted technologies do not go beyond research and some prototyping, and the requirements of socioeconomic adaptation of the new technologies can be developed by reaching a compromise between communities" [13].

In order to determine the basic structures of the subject domain by the expert method, four categories of promising trends in the development of ICT corresponding to the fourth level of the SML matrix were identified:
1) human-computer interfaces;
2) computing engineering;
3) memory and data storage technologies;
4) electronics and communications.

The present study proposes an approach based on the extraction of stable keyword patterns by analyzing the corpora of the subject-oriented texts, including not only stationary databases of scientometric and bibliometric information, but also using dynamic network thesauruses, one of which is represented by the network encyclopedia Wikipedia[5] [14].

At the same time, the desired level of reliability of the obtained information can be increased by forming a combined algorithm of the term extraction. This algorithm allows a multilevel selection based on an extended expert environment using the internal services of Wikipedia at the first stage, while, at the second stage, scientifically reliable algorithmic ways of providing information from highly authoritative qualitative research databases, such as Scopus and Web of Science[6], are used [15].

[3] http://mrvar.fdv.uni-lj.si/pajek/. Accessed April 03, 2022.
[4] https://www.statistica.com/en/. Accessed June 15, 2022.
[5] https://www.wikipedia.org/. Accessed March 15, 2022.
[6] http://www.webofknowledge.com/. Accessed March 04, 2022.

The extended terminological basis is obtained by processing the initial sets of basic keywords obtained by expert method in the bibliographic database Scopus using a special formula. The extracted useful database for the study falls within the dynamic range of 645–3364 terms, including related publications in which their joint inclusion occurs.

In [16], the main computational methods leading to an automatic mode of knowledge discovery in publications are classified in detail, including distributive semantic modeling. In addition to term-level modeling analysis, analysis at the level of the dissemination topic—so-called thematic modeling allows a deeper understanding of the information dissemination process. In the aspect of information retrieval and filtering, two generative models are widely used: probabilistic latent semantic analysis (PLSA) and, more commonly, latent Dirichlet allocation, which, in turn, is a generalization of PLSA.

Recently, a fundamentally different and more universal approach to thematic models based on network modeling has been proposed: the stochastic block model [17].

However, the use of these methods for analyzing scientific literature is rare [16–18].

Based on the study of programs for visualization and mapping of science and technology as the primary tool for thematic cluster analysis and visualization of the data obtained, the software package *VOSviewer* was selected. This tool is freely available and well integrated with bibliographic databases, including Scopus [19, 20].

VOSviewer's internal algorithms provide vectorization, normalization, term-document matrix construction, bibliometric mapping and initial clustering of text data, dynamically changing in the context of the assigned research tasks.

The maps generated by *VOSviewer* include various items that can include publications, researchers, or terms. *VOSviewer* maps the strength of a reference reflecting the number of publications in which two terms occur together (in the case of a matching occurrence reference).

For each pair of items i and j, *VOSviewer* requires as input the similarity s_{ij} ($s_{ij} > 0$). To determine the similarity between items, the frequency of matching is typically determined using a similarity measure. Different types of similarity measures can be applied: strength of association, Jaccard index, Pearson correlation, cosine measure.

In *VOSviewer*, the similarity s_{ij} is calculated using the association force AS defined in the equation:

$$AS_{ij} = \frac{s_{ij}}{s_i s_j}, \tag{3}$$

where s_i is the similarity of the element of the ith component; s_j is the similarity of the element of the jth component; s_{ij} is the similarity of the pair. All the listed quantities have the dimensionality equal to one.

After calculating the similarity between the elements, a special technique for their mapping is applied [16]. *VOSviewer* determines the location of elements on the map by minimizing the function:

$$V\left(x_1,\ldots,x_n\right)=\sum_{i<j}s_{ij}\left\|x_i-x_j\right\|^2, \quad (4)$$

where x_i, x_j are locations of nodes i and j in two-dimensional space; n is the number of the nodes in the grid; $\left\|x_i-x_j\right\|$ are the Euclidean distances between nodes i and j, provided that:

$$\frac{2}{n(n-1)}\sum_{i<j}s_{ij}\left\|x_i-x_j\right\|=1. \quad (5)$$

Consequently, the idea of *VOSviewer* is to minimize the weighted sum of the squares of distances between all pairs of elements. The squared distance between the pair of elements is weighted as the similarity between the elements.

As a result, a complex associated structure of the network under study is formed, whose nodes and terms are calculated by the weight of different elements according to three basic criteria: degree of nodes, as well as the distance and strength of links between nodes; here, the size of nodes depends on the weight of a particular term.

Particular attention should be paid to the dynamics of the formation of the so-called largest or giant component of the network. A classic example of a discrete probability distribution is the Poisson distribution model of random numbers. Based on this, Erdős and Rényi [21] proposed an extremely simple network model, which they called a random graph. It was shown that a random graph has an important property, which can be called a phase transition to a state when a large fraction of all vertices are connected together into one giant component.

Using the Poisson distribution, the heuristic argument [4] can be used to calculate the expected size of the giant component of random networks. Suppose u is a part of vertices of the network which do not belong to the giant component. Therefore, the probability that a vertex does not belong to the giant component is also equal to the probability that none of the network neighbors of the vertex belongs to the giant component, i.e., u^k, where k is the degree of the vertex.

After applying the averaging procedure, the expression for the probability of the Poisson distribution of degrees p_k in the self-consistency relation for u within a large graph size can be represented as follows:

$$u=\sum_{k=0}^{\infty}p_k u^k=\mathrm{e}^{-z}\sum_{k=0}^{\infty}\frac{(zu)^k}{k!}=\mathrm{e}^{z(u-1)}, \quad (6)$$

where e is the Eulerian number; k is the degree of a vertex; z is the average degree of all vertices N of the network.

The fraction S of the network occupied by the giant component is $S = 1 - u$. By averaging the expression for the random graph model of Erdős and Rényi by the probability of the Poisson distribution of degrees, we obtain the following self-consistency relation within the large size of the graph:

$$S=1-\mathrm{e}^{-zS}. \quad (7)$$

The appearance of the giant component indicates a phase (percolation) transition at the point $z = 1$, where the divergence of the mean size $\langle s\rangle$ of the non-giant components also occurs when studying the behavior of the random graph. If $z < 1$, the only non-negative solution of this equation is $S = 0$, while if $z > 1$, there exists also a nonzero solution which determines the size of the giant component.

The tasks of general network analysis are performed by the *Pajek* software. This software environment based on large network visualization graphs offers many different efficient (sub-quadratic) network analysis algorithms [22].

From a bibliometric point of view, the methods offered by *Pajek* include clustering and main path analysis. The software is used not only to reveal the global structure of knowledge networks, but also to operationalize and measure the stability of the resulting network models [23]. An important feature of *Pajek* is its close connection to *VOSviewer*, which allows direct two-way communication between these network environments, as well as exporting data in formats of widely used other external tools, including programming language R, *Statistical Package for the Social Science*s (*SSPS*) and *Excel* [22].

The most fundamental approaches to the study of knowledge networks are related to basic descriptive statistics of the network structure, such as measuring the number and size of components in the network, as well as calculating various measures of centrality (degree, closeness, betweenness). From the point of view of the tasks set in this paper, these latter characteristics of network interactions are effectively computed in *Pajek*.

The universal integrated system *STATISTICA* was used as a software package for statistical analysis. When properly used, the system saves the user from routine calculations by clearly displaying the results of cluster analysis [24], leaving the specialist to interpret the results and formulate conclusions. At the same time, it provides cues, which are important for the implementation of functions of data analysis, data management and visualization. In terms of implementing iterations of the algorithm of this study, the system is integrated with *Excel*.

RESULTS AND DISCUSSION

The data obtained from the bibliographic database Scopus are taken from the first year of publications, in which the occurrence of each of the keywords in the articles exceeds the lower threshold of 10 articles, and is inclusive up to 2021. The bibliographic data are detailed by year for mapping and primary visual analysis of clusters. In our case, the overall time range of the study covers the period from 1978 to 2021.

As part of the objectives of this research phase, the memory and data storage technologies category was selected as the basis for the selection of the established keywords.

In order to analyze the behavior of the trend, four keywords were selected from the terminology model: *phase-change memory*, *patterned media*, *quantum memory*, *DNA digital data storage*. From the resulting database the different dates of the initial inclusion of the terms follow.

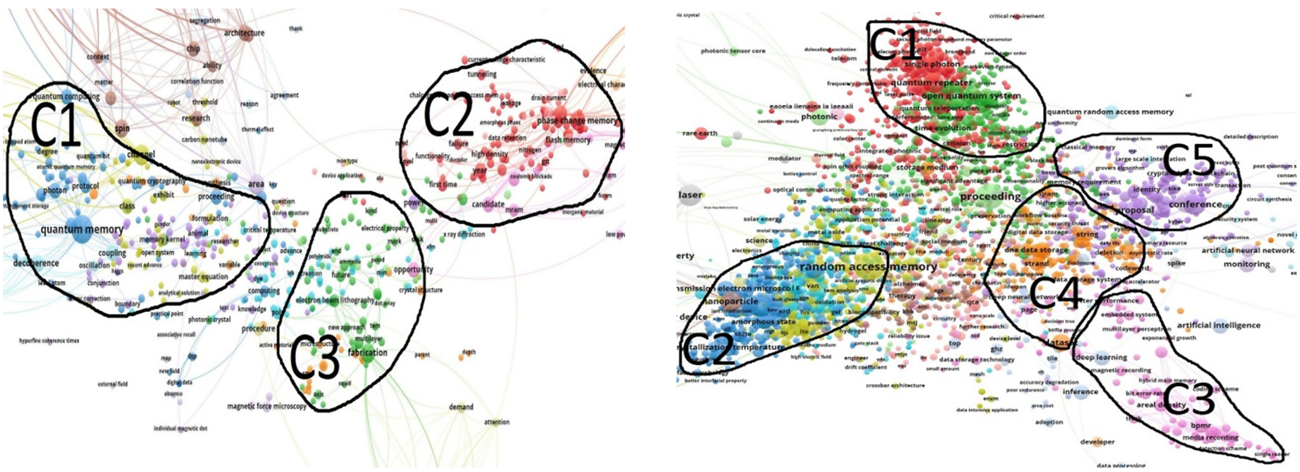
In this regard, the Scopus query formula was optimized. For example, the database of publications for the keyword *quantum memory*, limited to the year 1978, was specified by the formula: TITLE (quantumANDmemory) AND (LIMIT-TO (PUBYEAR, 1978)).

Full bibliographic descriptions of extracted publications were automatically downloaded in .csv format for the further processing and analysis in software tools, including a total of 33 information fields for each record.

It should be noted that in the context of the issues we are considering, text data from the TITLE and ABSTRACT fields of the database of uploaded documents was used. These fields are selected in *VOSviewer* automatically when loading the source document.

In order to understand the clearer structure of the network in the *VOSviewer* environment, an additional parameter, which implements the network integration of the specified keywords, with the name *memory and data storage technologies all* (*MADSTA*), was introduced. The timing of the introduction of this integrator is related to the period of active research phase since inclusion of the fourth keyword *DNA digital data storage*, i.e., from 2004 to 2021.

In the above time interval, the comparative visual analysis of the obtained range of 645–3364 terms integrated into a total of 23 clusters shown in Fig. 2 presents a well-defined clustering dynamic around the terms *quantum memory*, *phase-change memory*. At the same time, the noticeable activation of research by 2021 in the area of the term *DNA digital data storage* pushed the term *patterned media* to the periphery of the scientific landscape. At the same time, the formation of a new cluster associated with the term *post-quantum cryptography* is clearly visualized near the *quantum memory* area. This can be explained by the fact that many cryptographers are now actively developing new algorithms for quantum key search [25] in order to prepare for a future time when quantum computing is likely to present a security threat.



(a) (b)

Fig. 2. *VOSviewer* visualization map showing the dynamics of data clustering for the keyword integrator *MADSTA*, marked clusters in the area of *quantum memory* (C1), *phase-change memory* (C2), *patterned media* (C3), *DNA digital data storage* (C4), *post-quantum cryptography* (C5): (a) the beginning of the period 2004; (b) the end of the period 2021

For the further analysis of clustering, an automated transfer of data is performed from one *VOSviewer* network environment to another using the *Pajek* network calculator [21].

Working in *Pajek* starts with three file extensions characterizing certain types of data: networks, partition, vectors. When data are loaded into the program, the parameters characterizing the dynamic state of network links are investigated. We note that the *Pajek* network calculator function ideally suits it for performing such tasks.

The data processed in *Pajek* are shown in the summary table. Only parameters characterizing the dynamic state of network links are presented here for the term integrator.

The results of exploratory data analysis implemented in the universal integrated package *STATISTICA* with a brief description of the parameters are shown below.

The Total Link Strength parameter characterizes the number of links in a simple network. Another important parameter for the study is the average number of links per node known as the Average Degree. The Average Degree of all nodes in an individual network is a measure of the structural cohesion of the network. Since this index does not depend on the size of the network, it can be compared between networks of different sizes.

In Fig. 3a, the dynamics of Average Degree behavior leads to average values of 20–30 link units per node. This can be explained by the fact that the interest of the scientific community in this area does not decrease steadily. In our case, the presence of maxima in 2007 and 2009 may be a problem area.

The proposed hypothesis assumes the influence of the single keyword *quantum memory* on the other words by the effect of the aftereffect. In Fig. 3b, an increasing trend in the total number of links for *quantum memory* can be clearly seen. Since this indicates that the direction in this period was relevant, the total number of links for the remaining words determines the dynamic behavior of the overall network. The introduction of the parameter *MADSTA* shows a smooth increase in the total number of links in the network, which leads to the analysis of the next dynamic parameter—Density.

Density is responsible for the number of lines in the simple network, expressed as a fraction of the maximum possible number of lines. The parameters for determining the correlation dependence (Density) cover the maximum time period because the keywords were entered in the studied network alternately and at different time intervals.

Table. Dynamics of parameters characterizing the state of terminology network links for the keyword integrator *MADSTA*

Year	Total Link Strength	Average Degree	Density	Degree Centralization	Betweenness Centralization
2004	6140	19.039	0.029	0.139	0.073
2005	15292	29.896	0.029	0.146	0.192
2006	15167	29.827	0.029	0.184	0.092
2007	34415	44.899	0.029	0.211	0.054
2008	22696	27.831	0.017	0.194	0.120
2009	41099	40.833	0.020	0.117	0.034
2010	27976	30.017	0.016	0.236	0.201
2011	30205	22.366	0.008	0.151	0.118
2012	27010	22.813	0.009	0.210	0.240
2013	35111	29.419	0.012	0.172	0.101
2014	32007	29.230	0.013	0.294	0.236
2015	32552	29.232	0.012	0.101	0.075
2016	31683	27.780	0.012	0.110	0.065
2017	35182	27.358	0.011	0.084	0.052
2018	34183	26.726	0.010	0.117	0.070
2019	38765	25.878	0.009	0.148	0.140
2020	43288	27.285	0.009	0.126	0.065
2021	44552	26.488	0.008	0.082	0.072

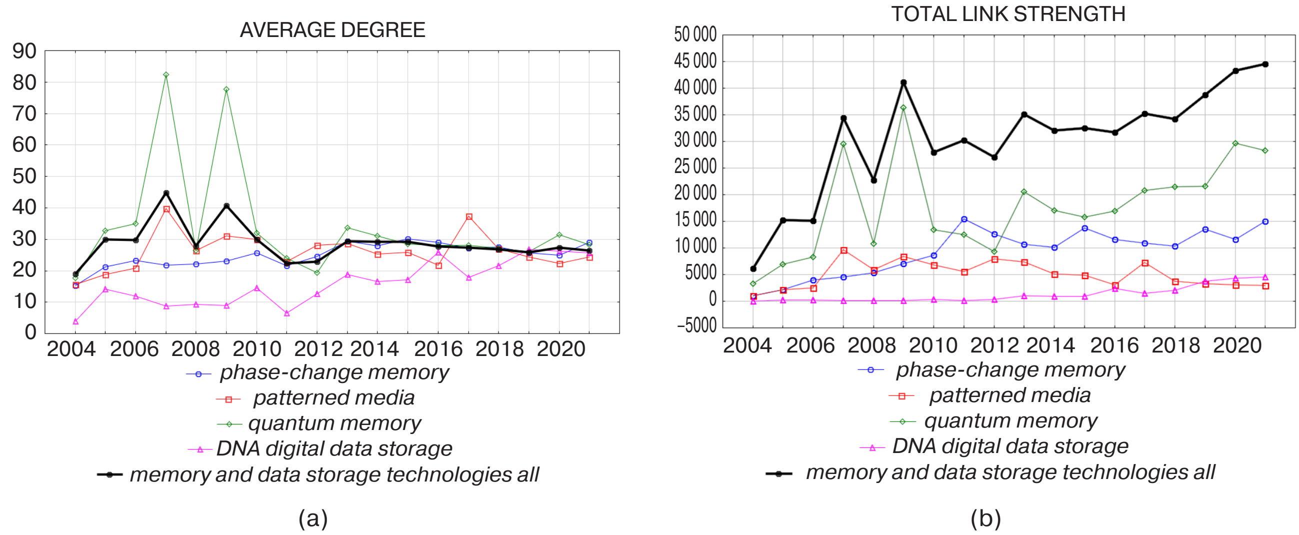


Fig. 3. Dependence of the average degree and the total link strength on time:
(a) dependence of the average degree of all vertices, taking into account the introduced parameter, on time;
(b) dependence of the total number of links in the network on time

Figure 4 shows that the integrative increase in the number of links in the network complicates the interaction structure due to the growth of the maximum possible number of lines. Consequently, the dependencies will tend to return minimum values. In this case, the behavior of the *MADSTA* parameter in the graph reveals that the integration of 4 keywords does not lead to changes in the dynamics of the behavior of the index of network density. The *MADSTA* parameter, which correlates with the curves, confirms the correctness of the choice of methodology for detailing the study by year.

The first parameter Degree Centralization (DC) represents the variation of the degree centrality of the network vertices divided by the maximum degree value that is possible in a network of the same size.

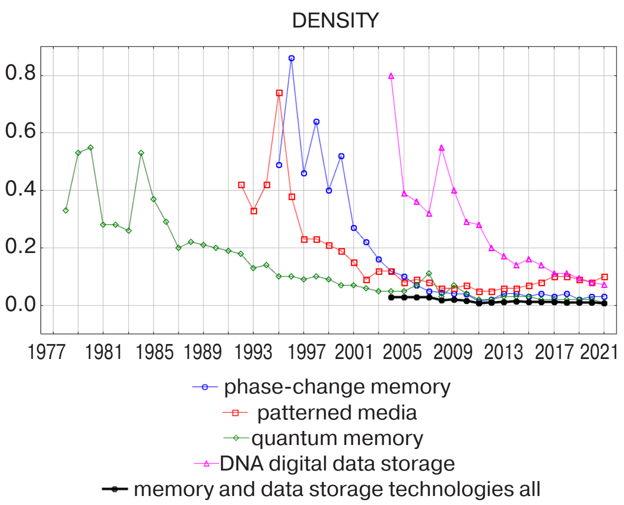


Fig. 4. Dependence of the network density index on time

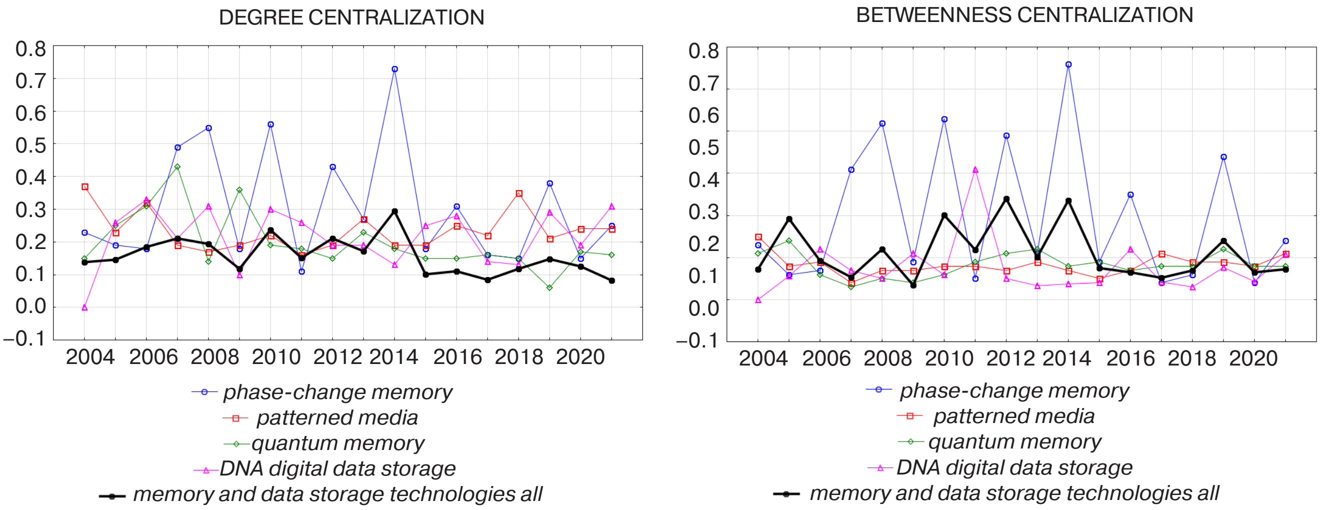
As can be seen in Fig. 5a, oscillations of the DC indicator for the four keywords result in a mutual relationship over the interval from 2004 to 2021. The values of the integrating parameter in this relationship show a weak correlation.

Betweenness Centralization (BC) represents the variation of betweenness scores between network nodes divided by the maximum value of betweenness score possible in a network of the same size. For a single node, betweenness shows the level of its inclusion in combinations of links between other nodes.

In Fig. 5b shows that the oscillation of the BC indicator of keywords in this dependence on the interval from 2004 to 2021 is synchronous, which refers to the same phase of the curve behavior. Graphically, there is a strong correlation between the extremums of keywords and *MADSTA* on a given interval. This can be explained in terms of the emergence of the trend associated with the oscillations of the *phase-change memory* indicator comprising an impulse for other terms, in which the semantic links on the shortest paths between the nodes are more clearly manifested.

Thus, empirical research provides an opportunity to find latent dependencies of dynamic parameters of network interactions on time during the exploratory analysis stage, creating a basis for quantitative assessment of clustering and percolation transitions.

The analysis of the graphs shows that the introduction of the integrating parameter *MADSTA* displays the interaction of the keywords under study. This is confirmed by the close correlation on the graphs of the dependencies of the dynamic parameters of the network (average degree, betweenness score and density) on a given interval. The formation of new clusters (C4, C5) is explained by the increased total number of links.



(a) (b)

Fig. 5. Dependence of DC and BC of the network on time: (a) dependence of the degree of centrality of the network vertices on time; (b) dependence of betweenness centralization index of the network vertices on time

CONCLUSIONS

The algorithmic approach implemented in the work allows for the maximum objectivity of the search and the speed with which it allows a given subject area of research to be analyzed. The level of reliability of the obtained information is increased due to the formation of a combined algorithm for term extraction based on an expanded expert environment and scientifically reliable qualitative research databases.

By implementing algorithmic queries using the Scopus bibliographic database, an extended terminological base is created, which in terms of 2004–2021 dynamics contains the range of 645–3364 terms of the memory and data storage technologies category, providing data output in the .csv format for further processing and analysis in specialized software tools.

A comparative analysis of the data visualization map of the terms performed in the software environment *VOSviewer*, which are integrated into a total of 23 clusters, reveals active clustering associated with the term C5 *post-quantum cryptography* in the area of cluster C1 (*quantum memory*), allowing qualitative conclusions to be drawn about the local dynamics of the scientific landscape.

The result of empirical study of the dynamics of the formation of interactions of terms in the *Pajek* network calculator and subsequent processing of the obtained parameters in the package *STATISTICA* is the construction of time series on the change in the average degree and the total number of links, network density, the degree centralization and the betweenness score of the bibliographic network.

To continue the analysis, it is necessary to supplement the obtained empirical material with data on general network parameters characterizing the number and size of components in the network, the distribution of distances in the network, and the degree of distribution of network fragments, including the presence of a giant component.

The formation of a complete set of basic parameters is necessary in the subsequent mathematical and computational modeling of information networks, during which the dynamics of network clustering and achieving the threshold of percolation over time will be evaluated.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Maltseva D., Batagelj V. Journals publishing social network analysis. *Scientometrics*. 2021;126(4): 3593–3620. https://doi.org/10.1007/s11192-021-03889-z
2. Zhukov D.O., Khvatova T.Yu., Zal'tsman A.D. Modeling of the stochastic dynamics of changes in node states and percolation transitions in social networks with self-organization and memory. *Informatika i ee Primeneniya = Informatics and Applications*. 2021;15(1):102–110 (in Russ.). https://doi.org/10.14357/19922264210114

СПИСОК ЛИТЕРАТУРЫ

1. Maltseva D., Batagelj V. Journals publishing social network analysis. *Scientometrics*. 2021;126(4): 3593–3620. https://doi.org/10.1007/s11192-021-03889-z
2. Жуков Д.О., Хватова Т.Ю., Зальцман А.Д. Моделирование стохастической динамики изменения состояний узлов и перколяционных переходов в социальных сетях с учетом самоорганизации и наличия памяти. *Информатика и ее применения*. 2021;15(1):102–110. https://doi.org/10.14357/19922264210114

3. Chkhartishvili A.G., Gubanov D.A., Novikov D.A. Models of influence in social networks. In: *Social Networks: Models of Information Influence, Control and Confrontation*. *Studies in Systems, Decision and Control*. 2019;189:1–40. https://doi.org/10.1007/978-3-030-05429-8_1
4. Newman M.E.J. The structure and function of complex networks. *SIAM Rev*. 2003;45(2):167–256. https://doi.org/10.1137/S003614450342480
5. Dashko Yu.V., Kramarov S.O., Zhdanov A.V. Sintering of polycristalline ferroelectrics and the percolation problem in stochastically packed networks. *Ferroelectrics*. 1996;186(1):85–88. https://doi.org/10.1080/00150199608218039
6. Broadbent S.R., Hammersley J.M. Percolation processes: I. Crystals and mazes. *Math. Proc. Cambridge Philos. Soc.* 1957;53(3):629–641. https://doi.org/10.1017/S0305004100032680
7. Li M., Liu R.-R., Lu L., Hu M.-B., Xu S., Zhang Y.-C. Percolation on complex networks: Theory and application. *Phys. Rep*. 2021;907:1–68. https://doi.org/10.1016/j.physrep.2020.12.003
8. Brunk N.E., Twarock R. Percolation theory reveals biophysical properties of virus-like particles. *ACS Nano*. 2021;15(8):12988–12995. https://doi.org/10.1021/acsnano.1c01882
9. Shao J., Havlin S., Stanley H.E. Dynamic opinion model and invasion percolation. *Phys. Rev. Lett*. 2009;103(1):018701. https://doi.org/10.1103/PhysRevLett.103.018701
10. Bodrunov S.D. *Noonomika* (*Noonomics*). Moscow: Kul'turnaya revolyutsiya; 2018. 432 p. (in Russ.). ISBN 978-5-6040343-1-6
11. Wickham H., Grolemund G. *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc.; 2016. 492 p.
12. Popov O.R., Kramarov S.O. The study of information dissemination in networks arranged from a set of forecasting terms. *Vestnik kibernetiki = Proceedings in Cybernetics.* 2022;1(45):38–45 (in Russ.). https://doi.org/10.34822/1999-7604-2022-1-38-45
13. Popov O.R. Adaptation of world practices to the problem of long-term technological forecasting of the state of self-organizing intelligent systems. *Intellektual'nye resursy – regional'nomu razvitiyu = Intellectual Resources – Regional Development.* 2021;2:91–98 (in Russ.).
14. Kogai V.N., Pak V.S. Algorithmic model of computer system of keywords extracting from text based on ontology. *Problemy sovremennoi nauki i obrazovaniya = Problems of Modern Science and Education.* 2016;16(58):33–40 (in Russ.).
15. Panin S.B. Modern scientometric systems "WoS" and "Scopus": publishing problems and new guidelines for Russian university science. *Gumanitarnye issledovaniya Tsentral'noi Rossii = Humanities Researches of the Central Russia*. 2019;3:51–65 (in Russ.). https://doi.org/10.24411/2541-9056-2019-11030
16. Thilakaratne M., Falkner K., Atapattu T. A systematic review on literature-based discovery: general overview, methodology, & statistical analysis. *ACM Computing Surveys*. 2019;52(6):Article 129. https://doi.org/10.1145/3365756

3. Chkhartishvili A.G., Gubanov D.A., Novikov D.A. Models of influence in social networks. In: *Social Networks: Models of Information Influence, Control and Confrontation*. *Studies in Systems, Decision and Control*. 2019;189:1–40. https://doi.org/10.1007/978-3-030-05429-8_1
4. Newman M.E.J. The structure and function of complex networks. *SIAM Rev*. 2003;45(2):167–256. https://doi.org/10.1137/S003614450342480
5. Dashko Yu.V., Kramarov S.O., Zhdanov A.V. Sintering of polycristalline ferroelectrics and the percolation problem in stochastically packed networks. *Ferroelectrics*. 1996;186(1):85–88. https://doi.org/10.1080/00150199608218039
6. Broadbent S.R., Hammersley J.M. Percolation processes: I. Crystals and mazes. *Math. Proc. Cambridge Philos. Soc.* 1957;53(3):629–641. https://doi.org/10.1017/S0305004100032680
7. Li M., Liu R.-R., Lu L., Hu M.-B., Xu S., Zhang Y.-C. Percolation on complex networks: Theory and application. *Phys. Rep*. 2021;907:1–68. https://doi.org/10.1016/j.physrep.2020.12.003
8. Brunk N.E., Twarock R. Percolation theory reveals biophysical properties of virus-like particles. *ACS Nano*. 2021;15(8):12988–12995. https://doi.org/10.1021/acsnano.1c01882
9. Shao J., Havlin S., Stanley H.E. Dynamic opinion model and invasion percolation. *Phys. Rev. Lett*. 2009;103(1):018701. https://doi.org/10.1103/PhysRevLett.103.018701
10. Бодрунов С.Д. *Ноономика*. М.: Культурная революция; 2018. 432 с. ISBN 978-5-6040343-1-6
11. Wickham H., Grolemund G. *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc.; 2016. 492 p.
12. Попов О.Р., Крамаров С.О. Исследование распространения информации в сетях, структурированных из набора прогностических терминов. *Вестник кибернетики*. 2022;1(45):38–45. https://doi.org/10.34822/1999-7604-2022-1-38-45
13. Попов О.Р. Адаптация мировых практик к проблеме долгосрочного технологического прогнозирования состояния самоорганизующихся интеллектуальных систем. *Интеллектуальные ресурсы – региональному развитию*. 2021;2:91–98.
14. Когай В.Н., Пак В.С. Алгоритмическая модель компьютерной системы выделения ключевых слов из текста на базе онтологий. *Проблемы современной науки и образования*. 2016;16(58):33–40.
15. Панин С.Б. Современные наукометрические системы «WoS» и «Scopus»: издательские проблемы и новые ориентиры для российской вузовской науки. *Гуманитарные исследования Центральной России*. 2019;3:51–65. https://doi.org/10.24411/2541-9056-2019-11030
16. Thilakaratne M., Falkner K., Atapattu T. A systematic review on literature-based discovery: general overview, methodology, & statistical analysis. *ACM Computing Surveys*. 2019;52(6):Article 129. https://doi.org/10.1145/3365756

17. Gerlach M., Peixoto T.P., Altmann E.G. A network approach to topic models. *Sci. Adv*. 2018;4(7):eaaq1360. https://doi.org/10.1126/sciadv.aaq1360
18. Zelenkov Yu. The topic dynamics in knowledge management research. In: Uden L., Ting I.H., Corchado J. (Eds.). *Knowledge Management in Organizations* (*KMO 2019*)*: Proceedings of the 14th International Conference*. 2019. P. 324–335. https://doi.org/10.1007/978-3-030-21451-7_28
19. Van Eck N.J., Waltman L. Visualizing bibliometric networks. In: Ding Y., Rousseau R., Wolfram D. (Eds.). *Measuring Scholarly Impact: Methods and Practice*. Springer; 2014. P. 284–321. https://doi.org/10.1007/978-3-319-10377-8_13
20. Van Eck N.J., Waltman L. *Accuracy of citation data in Web of Science and Scopus*. arXiv preprint arXiv:1906.07011. 2019. https://doi.org/10.48550/arXiv.1906.07011
21. Erdös P., Rényi A. On random graphs I. *Publ. Math. Debrecen*. 1959;6:290–297. Available from URL: https://studylib.net/doc/25387956/on-random-graphs--1959--by-p.-erdos-and-a.-renyi
22. De Nooy W., Mrvar A., Batagelj V. *Exploratory social network analysis with Pajek: Revised and epanded edition for updated software.* 3rd ed. Series: Structural Analysis in the Social Sciences. Cambridge: Cambridge University Press; 2018. 484 p. https://doi.org/10.1017/9781108565691
23. Doreian P., Batagelj V., Ferligoj A. (Eds.). *Advances in network clustering and blockmodeling.* John Wiley & Sons; 2020. 414 p.
24. Zuanazzi N.R., de Castilhos Ghisi N., Oliveira E.C. Analysis of global trends and gaps for studies about 2, 4-D herbicide toxicity: A scientometric review. *Chemosphere*. 2020;241:125016. https://doi.org/10.1016/j.chemosphere.2019.125016
25. Sigov A.S., Andrianova E.G., Zhukov D.O., Zykov S.V., Tarasov I.E. Quantum informatics: Overview of the main achievements. *Russ. Technol. J.* 2019;7(1):5–37 (in Russ.). https://doi.org/10.32362/2500-316X-2019-7-1-5-37

17. Gerlach M., Peixoto T.P., Altmann E.G. A network approach to topic models. *Sci. Adv*. 2018;4(7):eaaq1360. https://doi.org/10.1126/sciadv.aaq1360
18. Zelenkov Yu. The topic dynamics in knowledge management research. In: Uden L., Ting I.H., Corchado J. (Eds.). *Knowledge Management in Organizations* (*KMO 2019*)*: Proceedings of the 14th International Conference*. 2019. P. 324–335. https://doi.org/10.1007/978-3-030-21451-7_28
19. Van Eck N.J., Waltman L. Visualizing bibliometric networks. In: Ding Y., Rousseau R., Wolfram D. (Eds.). *Measuring Scholarly Impact: Methods and Practice*. Springer; 2014. P. 284–321. https://doi.org/10.1007/978-3-319-10377-8_13
20. Van Eck N.J., Waltman L. *Accuracy of citation data in Web of Science and Scopus*. arXiv preprint arXiv:1906.07011. 2019. https://doi.org/10.48550/arXiv.1906.07011
21. Erdös P., Rényi A. On random graphs I. *Publ. Math. Debrecen*. 1959;6:290–297. URL: https://studylib.net/doc/25387956/on-random-graphs--1959--by-p.-erdos-and-a.-renyi
22. De Nooy W., Mrvar A., Batagelj V. *Exploratory social network analysis with Pajek: Revised and epanded edition for updated software.* 3rd ed. Series: Structural Analysis in the Social Sciences. Cambridge: Cambridge University Press; 2018. 484 p. https://doi.org/10.1017/9781108565691
23. Doreian P., Batagelj V., Ferligoj A. (Eds.). *Advances in network clustering and blockmodeling.* John Wiley & Sons; 2020. 414 p.
24. Zuanazzi N.R., de Castilhos Ghisi N., Oliveira E.C. Analysis of global trends and gaps for studies about 2, 4-D herbicide toxicity: A scientometric review. *Chemosphere*. 2020;241:125016. https://doi.org/10.1016/j.chemosphere.2019.125016
25. Сигов А.С., Андрианова Е.Г., Жуков Д.О., Зыков С.В., Тарасов И.Е. Квантовая информатика: обзор основных достижений. *Russ. Technol. J.* 2019;7(1):5–37. https://doi.org/10.32362/2500-316X-2019-7-1-5-37

About the authors

Sergey O. Kramarov, Dr. Sci. (Phys.-Math.), Professor, Advisor to the President of the University, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia); Chief Researcher, Surgut State University (22, Energetikov ul., Surgut, 628408 Russia). E-mail: maoovo@yandex.ru. Scopus Author ID 56638328000, ResearcherID E-9333-2016, https://orcid.org/0000-0003-3743-6513

Oleg R. Popov, Cand. Sci. (Eng.), Associate Professor, Expert-Analyst of the Temporary Scientific Team of the Department of Technology and Professional and Pedagogical Education, Southern Federal University (105/42, Bolshaya Sadovaya ul., Rostov-on-Don, 344006 Russia). E-mail: cs41825@aaanet.ru. ResearcherID AAT-8018-2021, http://orcid.org/0000-0001-6209-3554

Ismail E. Dzhariev, Junior Researcher, Postgraduate Student, Department of Automated Information Processing and Management Systems of the Polytechnic Institute, Surgut State University (22, Energetikov ul., Surgut, 628408 Russia). E-mail: dzhariev2_ie@edu.surgu.ru. https://orcid.org/0000-0003-4068-1050

Egor A. Petrov, Junior Researcher, Postgraduate Student, Department of Automated Information Processing and Management Systems of the Polytechnic Institute, Surgut State University (22, Energetikov ul., Surgut, 628408 Russia). E-mail: petrov2_ea@edu.surgu.ru. https://orcid.org/0000-0002-4151-197X

Об авторах

Крамаров Сергей Олегович, д.ф.-м.н., профессор, советник президента университета, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78); главный научный сотрудник, БУ ВО «Сургутский государственный университет» (628408, Россия, Сургут, ул. Энергетиков, д. 22). E-mail: maoovo@yandex.ru. Scopus Author ID 56638328000, ResearcherID E-9333-2016, https://orcid.org/0000-0003-3743-6513

Попов Олег Русланович, к.т.н., доцент, эксперт-аналитик временной научной группы кафедры технологии и профессионально-педагогического образования, ФГАОУ ВО «Южный федеральный университет» (344006, Россия, Ростов-на-Дону, ул. Б. Садовая, д. 105/42). E-mail: cs41825@aaanet.ru. ResearcherID AAT-8018-2021, http://orcid.org/0000-0001-6209-3554

Джариев Исмаил Эльшан оглы, младший научный сотрудник, аспирант кафедры автоматизированных систем обработки информации и управления Политехнического института БУ ВО «Сургутский государственный университет» (628408, Россия, Сургут, ул. Энергетиков, д. 22). E-mail: dzhariev2_ie@edu.surgu.ru. https://orcid.org/0000-0003-4068-1050

Петров Егор Аркадьевич, младший научный сотрудник, аспирант кафедры автоматизированных систем обработки информации и управления Политехнического института БУ ВО «Сургутский государственный университет» (628408, Россия, Сургут, ул. Энергетиков, д. 22). E-mail: petrov2_ea@edu.surgu.ru. https://orcid.org/0000-0002-4151-197X

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt