Russian Technological Journal.
2023;11(3):17–29

ISSN 2782-3210 (Print)
ISSN 2500-316X (Online)

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

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.

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.

ВВЕДЕНИЕ

Изучение распространения информации в сетях социальных связей, имеющих случайную топологию, является актуальной задачей для оптимизации современных социотехнических систем, что подтверждается незатухающим интересом к проблематике анализа социальных сетей (SNA – social network analysis) [1–3].

Отдельную категорию сложных сетей, наряду с социальными и биологическими, представляют информационные сети, называемые также «сетями знаний». Примером являются сети ссылок цитирования между научными публикациями, структура которых достаточно точно отражает структуру информации, хранящейся в ее вершинах – статьях, что и определяет терминологию «информационная сеть».

Применительно к анализу отношений между классами слов в тезаурусе информационную сеть можно также рассматривать как концептуальную, представляющую структуру языка или, возможно, даже ментальные конструкции, используемые для его представления [4].

Для моделирования и анализа информационных процессов, протекающих в сетях с нерегулярной структурой, возможно применение известных в физике твердого тела методов теории перколяции [5], которая способна ответить на важные вопросы.

Перколяция (лат. *percolare*) в переводе с латинского означает протекание, просачивание. Долгое время эта простая вероятностная модель являлась в физике базовой идеальной моделью для демонстрации фазовых переходов и критических явлений. В виде математического объекта она впервые была рассмотрена в классической работе Бродбента и Хаммерсли в 1957 г. [6], в которой были введены название, а также геометрические и вероятностные понятия.

Методы решения различных теоретических и прикладных задач в течение последних десятилетий привнесли новое понимание в математическое исследование просачивания [7]. Проникновение жидкости внутрь пористого камня, распространение эпидемий или информации в социальной сети, на первый взгляд, не имеют ничего общего, но оказывается, что все три аспекта математическим образом сходятся в аддитивную компоненту [7–9].

В самом общем виде, независимо от физической природы и модели системы, теория перколяции отвечает на вопрос, какова вероятность того, что существует открытый путь из 0 до бесконечности (или существует ли бесконечный кластер связанных между собой пор или узлов). Таким образом, проблема сводится к поиску ответа на вопрос, существуют ли такие пути для данной вероятности p. В основном теория касается существования такого кластера и его структуры по отношению к вероятности заполнения p.

Для математического описания этой критичности следует определить модель перколяции. В качестве примера выберем наиболее простую модель на бесконечной двумерной (или квадратной) решетке. Точки пересечения линий называются узлами (вершинами графа), а сами линии − связями (ребрами графа). Существуют две решеточные модели: проницаемости связей и проницаемости узлов.

В первой модели математически каждая связь занята с вероятностью p или свободна с вероятностью $1 - p$. Затем занятые связи соединяют узлы в кластеры. Эту модель можно использовать для моделирования процесса проникновения жидкости внутрь пористого камня и распространения эпидемий.

В модели перколяции узлов мы занимаем не связь, а каждый узел с вероятностью p, оставляя его свободным с вероятностью $1 - p$. В целом перколяция связей считается менее общей, чем перколяция узлов, из-за того, что перколяцию связей можно переформулировать как перколяцию узлов на другой решетке, но не наоборот.

Теория перколяции в основном фокусируется на появлении бесконечного кластера с увеличением вероятности p. Для характеристики этого явления часто принимают размер гигантского скопления S, который определяется как:

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

где N – размер системы (общее количество узлов); S_1 – количество узлов в самом большом кластере.

С увеличением вероятности p должна существовать критическая p_S, называемая порогом перколяции

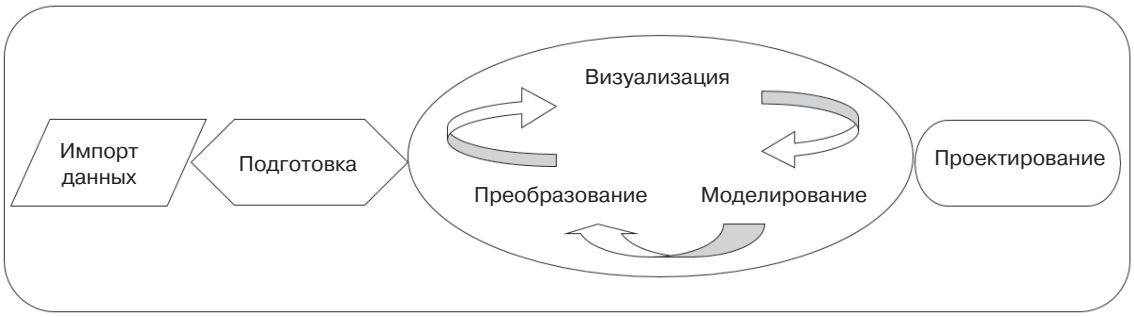


Рис. 1. Схема алгоритма реализации научного проекта science data [11]

или критической точкой, выше которой можно найти ненулевое значение *S*. Это определяет перколяционный переход системы по отношению к управляющему параметру *p*, а *S* является соответствующим параметром порядка.

Чтобы описать особенности конечных кластеров, также используется распределение кластеров по размерам:

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

где m_S – количество кластеров размера *S*.

В настоящее время построены модели для трехмерных и более высоких измерений решетки или для других условий перколяции.

При обширном вычислительном моделировании социальных сетей как на классических 2D-регулярных решетках, так и для сетей более высокой размерности, получены результаты, подтверждающие отображение фазы переходного поведения, характеризующейся наибольшим компонентом в предлагаемой модели динамического мнения (non-consensus opinion (NCO) модель), с известной физической проблемой просачивания в несжимаемых жидкостях [9].

Использование методов вычислительного моделирования случайных сетей с большим числом связей привело к аналитическим решениям, например, к построению стохастических моделей описания динамики изменения состояния узлов и перколяционных переходов, прогнозирующих динамику поведения социальных сетей [2].

По аналогии с социальными сетями полагаем, что правильно построенная методика исследования позволит количественно оценить, насколько информационная сеть близка к порогу перколяции, и тем самым сформировать модель ее прогнозирования.

В связи с этим имеет смысл применить методы картирования и сетевого анализа для изучения закономерностей, влияющих на порог перколяции при распространении и кластеризации информации в сетях, структурированных на основе словарей модельных прогностических терминов, извлеченных из сетевых ресурсов, тезаурусов, наукометрических и библиографических баз данных.

В работе [10] в условиях конвергентных тенденций «стыковых» междисциплинарных связей при развитии сложных геоинформационных систем, обосновано приоритетное значение информационно-коммуникационных технологий (ИКТ) и сферы информационных наук. Это положение предопределило основной выбор предметной области для данного исследования.

ИНСТРУМЕНТ И МЕТОДЫ

Базовый инструментарий этапов выполнения научного проекта, связанного с обработкой больших объемов информации, данных – science data, обозначен на рис. 1. Основной механизм генерации знаний для итогового проектирования кроется в большом центральном блоке, включающем визуализацию и моделирование данных. Анализ полученных знаний на каждом шаге будет требовать их преобразование и оптимизацию, включая в себя сужение круга наблюдений, представляющих интерес, вычисление набора сводных статистических данных, средних значений и т.п.

Алгоритмический подход преследует цель обеспечить максимальную объективность поиска и скорость, с которой он позволяет углубиться в заданную предметную область исследований.

Базовые шаги и инструментарий реализации эмпирической части алгоритма рассмотрены в [12]. Вариант, модифицированный для решения текущих исследовательских задач, состоит из следующих итераций:

1) формирование терминологической модели предметной области на основе матрицы, определяющей уровень зрелости (system maturity level, SML) самоорганизующихся интеллектуальных систем;
2) формирование алгоритмических запросов библиографической базы данных (БД) Scopus[1] по специальной формуле и вывод данных

[1] https://www.scopus.com/. Дата обращения 01.03.2022. / Accessed March 01, 2022.

в формате .csv для дальнейшей обработки и анализа в программных инструментах;
3) препроцессинг, векторизация, кластеризация полученной терминологической БД внутренними алгоритмами программной среды *VOSviewer*[2];
4) первичный визуальный анализ внутри- и межкластерных взаимодействий базовыми интерфейсами программы *VOSviewer* по годам и термам;
5) детальное исследование параметров связей научной терминологической сети по кластерам в динамике развития с помощью алгоритмов программного продукта *Pajek*[3];
6) обработка данных и вывод наглядных динамических зависимостей параметров связей терминологической сети в среде *STATISTICA*[4].

Существует несколько способов достижения поставленной цели. Начальным этапом данной работы является формирование терминологической модели предметной области, на основе которой сформирована база данных для дальнейшего исследования.

Основой для выбора заданного шаблона ключевых слов выбрана методика расчета SML-матрицы самоорганизующихся интеллектуальных систем (ИС) [13], позволяющая количественно оценить показатель зрелости ИС. Показатель SML применяется на системном уровне и представляет собой индекс зрелости от 0 до 1.

В рамках данного исследования интерес представляет 4 уровень зрелости социотехнической системы (sociotechnical system), диапазон индекса зрелости 0.60–0.80, описание «прогнозируемые технологии не выходят за рамки исследований и создания некоторых прототипов, а требования социально-экономической адаптации новых технологий могут быть разработаны за счет достижения компромисса между сообществами» [13].

В целях определения базовых структур предметной области экспертным методом выделены четыре категории перспективных направлений развития ИКТ, соответствующих четвертому уровню SML-матрицы:
1) человеко-машинные интерфейсы (human-computer interfaces all, HCIA);
2) инженерия вычислений (computing engineering all, CEA);
3) технологии памяти и хранения данных (memory and data storage technologies all, MADSTA);
4) электроника и коммуникации (electronics and communications all, ECA).

В настоящем исследовании предлагается подход, основанный на выделении устойчивых шаблонов ключевых слов путем анализа корпусов предметно-ориентированных текстов, включая не только стационарные БД наукометрической и библиометрической информации, но и используя достаточно динамичные сетевые тезаурусы, одним из которых является сетевая энциклопедия Википедия[5] [14].

При этом желаемый уровень достоверности получаемой информации может быть повышен за счет формирования комбинированного алгоритма извлечения терминов. Данный алгоритм допускает многоуровневый выбор на основе расширенной экспертной среды с использованием внутренних сервисов Википедии на первом этапе, и научно достоверных алгоритмических способов предоставления информации из высокоавторитетных качественных исследовательских БД, например, Scopus и Web of Science[6] [15] – на втором.

Расширенная терминологическая основа получена путем обработки полученных экспертным методом исходных наборов базовых ключевых слов в библиографической БД Scopus по специальной формуле. Извлеченная полезная база данных для исследования составляет в динамике диапазон 645–3364 терминов, в т.ч. связанных публикациями, в которых имеется их совместное вхождение.

В [16] подробно классифицируются основные вычислительные методы, приводящие к автоматическому режиму обнаружения знаний в публикациях, включая дистрибутивное семантическое моделирование. Помимо анализа моделирования на уровне термов существует анализ на уровне темы распространения – так называемое тематическое моделирование, позволяющее глубже понять процесс распространение информации. В аспекте поиска и фильтрации информации достаточно широкое применение получили две генеративные модели: вероятностный скрытый семантический анализ (probabilistic latent semantic analysis, PLSA) и, более распространенная, скрытое распределение Дирихле (latent Dirichlet allocation), являющаяся, в свою очередь, обобщением PLSA.

В последнее время предложен принципиально иной и более универсальный подход к тематическим моделям, основанный на сетевом моделировании – стохастической блочной модели (stochastic block model) [17].

Однако использование данных методов для анализа научной литературы встречается редко [16–18].

На основе изучения программ для визуализации и картирования науки и технологий в качестве первичного инструмента для тематического кластерного анализа и визуализации полученных данных выбран программный комплекс *VOSviewer*. Данный инструмент находится в свободном доступе и хорошо

[2] https://www.vosviewer.com/. Дата обращения 22.03.2022. / Accessed March 22, 2022.

[3] http://mrvar.fdv.uni-lj.si/pajek/. Дата обращения 03.04.2022. / Accessed April 03, 2022.

[4] https://www.statistica.com/en/. Дата обращения 15.06.2022. / Accessed June 15, 2022.

[5] https://www.wikipedia.org/. Дата обращения 15.03.2022. / Accessed March 15, 2022.

[6] http://www.webofknowledge.com/. Дата обращения 04.03.2022. / Accessed March 04, 2022.

интегрирован с библиографическими БД, включая Scopus [19, 20].

Внутренние алгоритмы *VOSviewer* обеспечивают векторизацию, нормализацию, построение термдокументной матрицы, библиометрическое картирование и первоначальную кластеризацию текстовых данных, динамически изменяемых в контексте поставленных исследованиями задач.

Карты, созданные *VOSviewer*, включают элементы, которыми могут быть публикации, исследователи или термины. *VOSviewer* картирует силу ссылки, отражающую количество публикаций, в которых два термина встречаются вместе (в случае совпадения ссылки вхождения).

Для каждой пары элементов *i* и *j VOSviewer* требует в качестве входных данных сходство s_{ij} ($s_{ij} > 0$). Чтобы определить сходство между предметами, обычно определяют частоту совпадения, преобразованную с использованием меры сходства. Могут быть применены разные типы мер подобия: сила ассоциации, индекс Жаккара, корреляция Пирсона, косинусная мера.

В *VOSviewer* сходство s_{ij} рассчитывается с использованием силы ассоциации *AS*, определенной в уравнении:

$$AS_{ij} = \frac{s_{ij}}{s_i s_j}, \qquad (3)$$

где s_i – сходство элемента *i*-й компоненты; s_j – сходство элемента *j*-й компоненты; s_{ij} – сходство пары. Все перечисленные величины имеют размерность, равную единице.

После расчета сходства между элементами применяется специальная техника их картирования [16]. *VOSviewer* определяет местонахождение элементов на карте, минимизируя функцию:

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

где x_i, x_j – местоположение узлов *i* и *j* в двумерном пространстве; *n* – количество узлов в сети; $\left\| x_i - x_j \right\|$ – евклидовы расстояния между узлами *i* и *j*, при условии:

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

Следовательно, идея *VOSviewer* состоит в том, чтобы минимизировать взвешенную сумму квадратов расстояний между всеми парами элементов. Квадрат расстояния между парой элементов взвешивается как сходство между элементами.

В результате формируется сложная ассоциированная структура исследуемой сети с узлами, термами, рассчитанными по весу различных элементов по трем базовым критериям: степени узлов, расстоянию и прочности связей между узлами, причем размер узлов зависит от веса определенного терма.

Особо следует обратить внимание на динамику формирования, так называемого самого большого или «гигантского компонента» сети. Классическим примером дискретного распределения вероятностей является модель распределения случайных чисел Пуассона. Основываясь на этом, Эрдёш и Реньи [21] предложили предельно простую модель сети, названную ими «случайным графом». Было показано, что случайный граф обладает важным свойством, которое можно назвать фазовым переходом к состоянию, когда обширная доля всех вершин соединена вместе в один гигантский компонент.

Эвристический аргумент [4] позволяет, используя пуассоновское распределение, рассчитать ожидаемый размер гигантского компонента случайных сетей. Пусть *u* – часть вершин сети, не принадлежащих гигантскому компоненту. Вероятность того, что вершина не принадлежит гигантскому компоненту, также равна вероятности того, что ни один из сетевых соседей вершины не принадлежит гигантскому компоненту, т.е. u^k, где *k* – степень вершины.

После применения процедуры усреднения выражение по вероятности пуассоновского распределения степеней p_k в соотношении самосогласования для *u* в пределах большого размера графа выглядит так:

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

где e – число Эйлера, *k* – степень вершины, *z* – средняя степень всех вершин *N* сети.

Доля *S* сети, занятая гигантской компонентой, равна $S = 1 - u$. Усредняя выражение для модели случайного графа Эрдёша и Реньи по вероятности пуассоновского распределения степеней, получаем следующее соотношение самосогласования в пределах большого размера графа:

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

Появление гигантского компонента говорит о фазовом (перколяционном) переходе в точке $z = 1$, в которой также происходит расходимость среднего размера $\langle s \rangle$ негигантских компонент при исследовании поведения случайного графа. При $z < 1$ единственным неотрицательным решением данного уравнения является $S = 0$, а при $z > 1$ существует и ненулевое решение, которое определяет размер гигантского компонента.

Задачи общего сетевого анализа выполняет программный продукт *Pajek*. Это программная среда с множеством различных эффективных (субквадратичных) алгоритмов сетевого анализа, основанная на графах визуализации больших сетей [22].

С библиометрической точки зрения методы, предлагаемые *Pajek*, включают кластеризацию и анализ основных связей. Программный продукт используется не только для выявления глобальной структуры сетей знаний, но и позволяет операционализировать и измерять стабильность полученных сетевых моделей [23]. Важным свойством *Pajek* служит тесная связь с *VOSviewer*, позволяющая напрямую двусторонне взаимодействовать данным сетевым средам, а также экспорт данных в форматах широко распространенных других внешних инструментов, включая язык программирования R, *Statistical Package for the Social Sciences* (*SPSS*) и *Excel* [22].

Наиболее фундаментальные подходы к изучению сетей знаний связаны с базовой описательной статистикой сетевой структуры, такой как измерение количества и размера компонентов в сети, а также вычисление различных показателей центральности (степень, близость, промежуточность). С точки зрения поставленных в данной работе задач указанные последние характеристики сетевых взаимодействий эффективно вычисляются в *Pajek*.

В качестве программного пакета для статистического анализа использована универсальная интегрированная система *STATISTICA*. Правильное применение системы позволяет избавить пользователя от рутинных вычислений, наглядно отображает результаты кластерного анализа [24], оставляя специалисту интерпретацию результатов и формулировку выводов. При этом она снабжена подсказками, что немаловажно для реализации функций анализа данных, их управления и визуализации. В плане реализации итераций алгоритма данного исследования система интегрирована с *Excel*.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

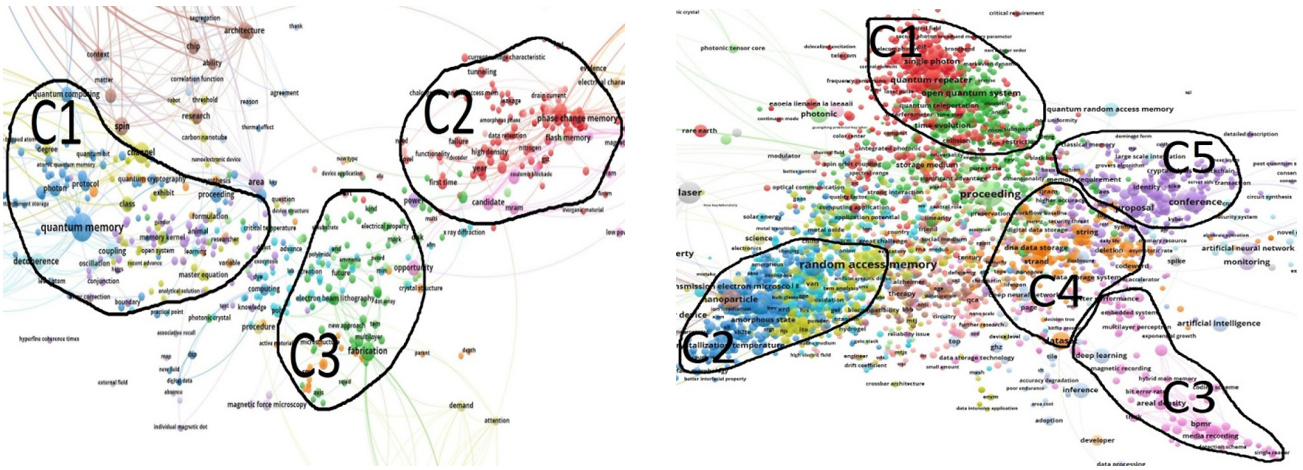
Данные, полученные из библиографической БД Scopus, взяты от первого года публикаций, в котором вхождение каждого из ключевых слов в статьях превышает нижний порог, равный 10 статьям, и до 2021 г. включительно. Библиографические данные детализированы по годам для картирования и первичного визуального анализа кластеров. В нашем случае общий временной диапазон исследования охватывает период с 1978 по 2021 гг.

В рамках задач данного этапа исследования в качестве базы для выбора установленных ключевых слов была выбрана категория «Технологии памяти и хранения данных».

С целью анализа поведения тренда из терминологической модели выбраны 4 ключевых слова: «phase-change memory», «patterned media», «quantum memory», «DNA digital data storage». Из полученной базы данных следуют разные даты начального вхождения терминов.

В этой связи оптимизирована формула запроса в Scopus. Например, база публикаций для ключевого слова «quantum memory», ограниченная 1978 г., задавалась формулой: TITLE (quantumANDmemory) AND (LIMIT-TO (PUBYEAR, 1978)).

Для извлеченных публикаций были автоматически выгружены их полные библиографические описания в формате .csv для дальнейшей обработки и анализа в программных инструментах, включающие по каждой записи в целом 33 информационных поля.



(а) (б)

Рис. 2. Карта визуализации *VOSviewer*, показывающая динамику кластеризации данных для интегратора ключевых слов MADSTA. Обозначение кластеров в области термов: «quantum memory» (C1), «phase-change memory» (C2), «patterned media» (C3), «DNA digital data storage» (C4), «post-quantum cryptography» (C5): (а) начало периода 2004 г.; (б) завершение периода 2021 г.

Необходимо отметить, что в контексте рассматриваемых вопросов работа проводилась с текстовыми данными из полей «название», «аннотация» базы выгруженных документов. Данные области выбираются в *VOSviewer* автоматически при загрузке исходного документа.

Чтобы понять более четкую структуру сети, в среде *VOSviewer* дополнительно введен параметр, который реализует сетевую интеграцию указанных ключевых слов, с названием MADSTA. Время введения данного интегратора связано с периодом активной фазы исследований с момента вхождения 4-го ключевого слова «DNA digital data storage», т.е. с 2004 до 2021 гг.

В указанном временном интервале сравнительный визуальный анализ полученного диапазона 645–3364 термов, интегрированных суммарно в 23 кластера, показывает на рис. 2 хорошо выраженную динамику кластеризации вокруг термов «quantum memory», «phase-change memory». При этом заметна активизация исследований к 2021 г. в области терма «DNA digital data storage», которая вытеснила терм «patterned media» на периферию научного ландшафта. В то же время возле области «quantum memory» явно визуализируется образование нового кластера, связанного с термом «post-quantum cryptography». Это можно объяснить тем, что многие криптографы сейчас активно разрабатывают новые алгоритмы поиска квантовых ключей [25], чтобы подготовиться к тому времени, когда квантовые вычисления станут угрозой безопасности.

Для дальнейшего анализа кластеризации производится автоматизированная передача данных из одной сетевой среды *VOSviewer* в другую – сетевой калькулятор *Pajek* [21].

Начало работы в *Pajek* осуществляется с использованием трех расширений файлов, характеризующих определенные типы данных: networks, partition, vectors. При загрузке данных в программу проводится исследование параметров, характеризующих динамическое состояние связей сети. Отметим, что *Pajek* идеально подходит для выполнения таких задач, поскольку является сетевым вычислителем.

Данные, обработанные в *Pajek*, приведены в сводной таблице. Здесь представлены параметры, характеризующие динамическое состояние связей сети, только для интегратора терминов.

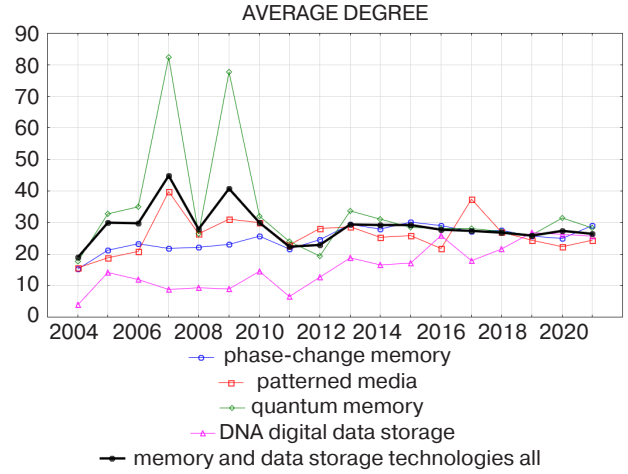
Таблица. Динамика параметров, характеризующих состояние связей терминологической сети, для интегратора ключевых слов MADSTA

Год	Общее число связей (Total Link Strength)	Средняя степень (Average Degree)	Плотность (Density)	Центральность (Degree Centralization)	Промежуточность (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

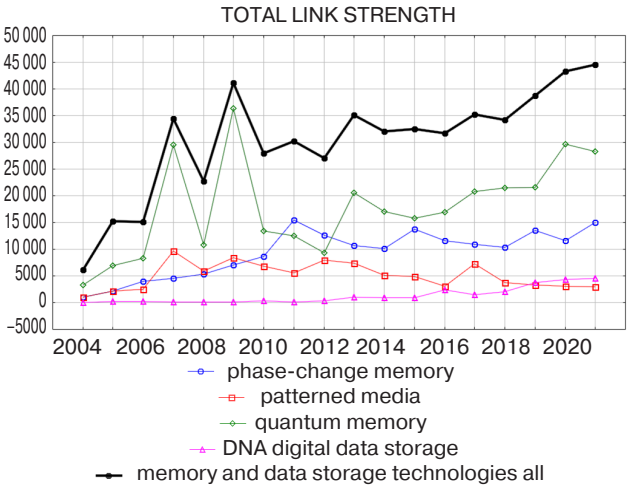
Результаты разведочного анализа данных, реализованного в универсальном интегрированном пакете *STATISTICA*, с кратким описанием параметров приведены ниже.

Параметр Total Link Strength (общее число связей в сети) характеризует количество линий в простой сети. Также одним из важных для исследования параметров является среднее число связей на один узел – Average Degree. Средняя степень всех вершин отдельной сети – показатель структурной сплоченности сети. Данный показатель не зависит от размера сети, поэтому среднюю степень можно сравнивать между сетями разного размера.

На рис. 3а динамика поведения Average Degree приводит к средним значениям 20–30 единиц связей на один узел. Это объясняется тем, что интерес научного сообщества к данному направлению стабильно не угасает. В нашем случае наличие максимумов в 2007 и 2009 гг. может оказаться проблемной областью.



(а)



(б)

Рис. 3. Зависимость средней степени и общего числа связей от времени: (а) зависимость средней степени всех вершин с учетом введенного параметра от времени; (б) зависимость общего числа связей в сети от времени

Выдвигаемая гипотеза предполагает влияние одного ключевого слова – «quantum memory» на остальные слова по эффекту последействия. На рис. 3б отчетливо виден тренд увеличения общего числа связей для «quantum memory». Это свидетельствует о том, что направление в данном периоде было релевантным, следовательно, общее количество связей для остальных слов определяет динамическое поведение общей сети. Введение параметра MADSTA демонстрирует плавное увеличение общего числа связей в сети, что приводит к анализу следующего динамического параметра – Density.

Density (плотность) отвечает за количество линий в простой сети, выраженное как доля от максимально возможного количества линий. Параметры для определения корреляционной зависимости (Density) охватывают максимальный период времени, т.к. ключевые слова входили в исследуемую сеть поочередно, в разные временные промежутки.

На рис. 4 видно, что интегративное увеличение количества связей в сети усложняет структуру взаимодействия вследствие роста максимально возможного количества линий. Следовательно, зависимости будут стремиться к минимальным значениям. При этом поведение параметра MADSTA в графике позволяет выявить, что объединение 4 ключевых слов не влечет за собой изменений в динамике поведения показателя плотности сети. Параметр MADSTA, коррелируя с кривыми, подтверждает правильность выбора методики детализации исследования по годам.

Первый параметр Degree Centralization (DC) представляет вариацию степеней центральности вершин сети, деленное на максимальное значение степени, которое возможно в сети того же размера.

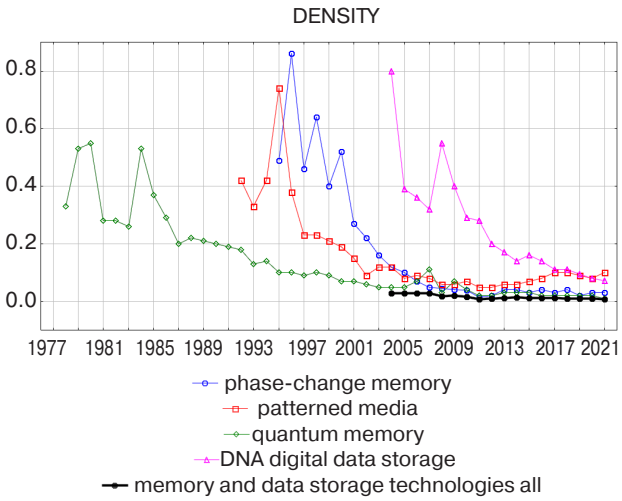


Рис. 4. Зависимость показателя плотности сети от времени

Рис. 5а сообщает, что осцилляции показателя DC по четырем ключевым словам приводят к образованию взаимной связи на интервале от 2004 до

2021 г. Значения интегрирующего параметра в этой зависимости демонстрируют слабую корреляцию.

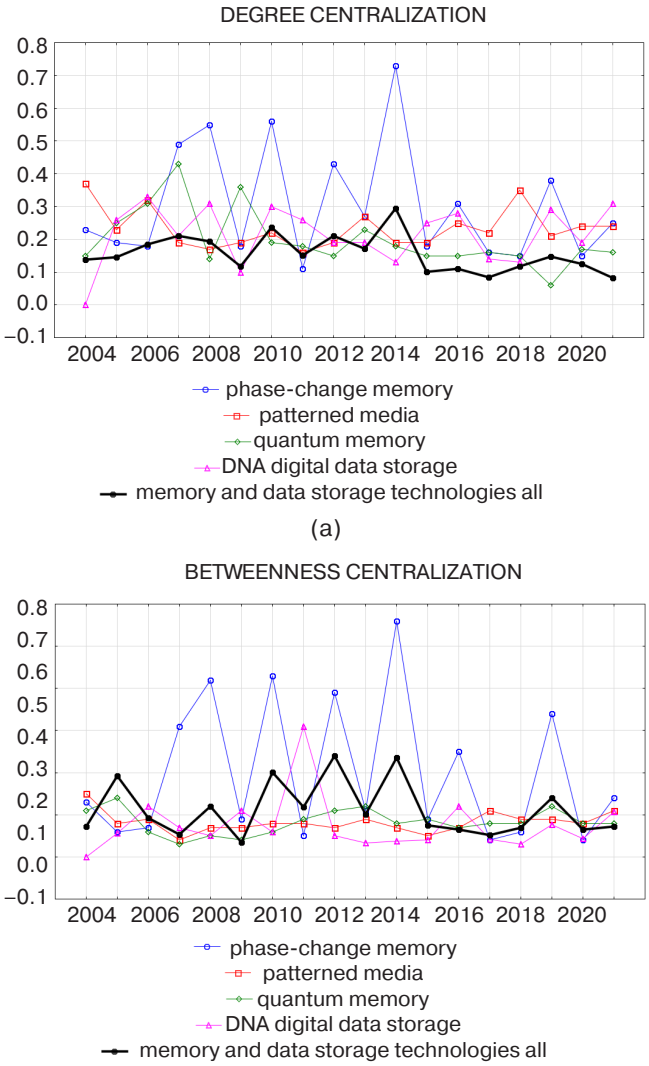


Рис. 5. Зависимость DC и BC сети от времени: (а) зависимость степеней центральности вершин сети от времени; (б) зависимость показателя промежуточности между вершинами сети от времени

Betweenness Centralization (BC) – это вариация показателей промежуточности между вершинами сети, деленное на максимальное значение показателя промежуточности, возможное в сети одинакового размера. Для отдельного узла промежуточность отображает уровень его включенности в комбинации связей между другими узлами.

На рис. 5б видно, что осцилляция показателя BC ключевых слов в данной зависимости на интервале от 2004 до 2021 г. синхронна, что означает одинаковую фазу поведения кривых. Графически на заданном интервале наблюдается сильная корреляция экстремумов phase-change memory и MADSTA. Это объясняется тем, что появление тренда, связанного с осцилляциями показателя phase-change memory, является импульсом для других терминов, при котором семантические связи на кратчайших путях между узлами проявляются более выраженно.

Таким образом, использование эмпирического исследования предоставляет возможность найти на этапе разведочного анализа скрытые зависимости динамических параметров сетевых взаимодействий от времени, создавая основу для количественной оценки кластеризации и перколяционных переходов.

Анализ графиков показывает, что введение интегрирующего параметра MADSTA отображает взаимодействие исследуемых ключевых слов. Подтверждением этому служит тесная корреляция на графиках зависимостей динамических параметров сети (средней степени, показателя промежуточности и плотности) на заданном интервале. Общее количество связей возрастает, что объясняет образование новых кластеров (C4, C5).

ЗАКЛЮЧЕНИЕ

Алгоритмический подход, реализованный в работе, позволяет обеспечить максимальную объективность поиска и скорость, с которой он позволяет углубиться в заданную предметную область исследований. Уровень достоверности получаемой информации повышен за счет формирования комбинированного алгоритма извлечения терминов на основе расширенной экспертной среды и научно достоверных качественных исследовательских БД.

Реализация алгоритмических запросов библиографической БД Scopus позволяет создать расширенную терминологическую базу, составляющую в динамике 2004–2021 гг. диапазон 645–3364 термов категории «Технологии памяти и хранения данных», и обеспечить вывод данных в формате .csv для дальнейшей обработки и анализа в специализированных программных инструментах.

Сравнительный анализ карты визуализации данных термов, выполненных в программной среде *VOSviewer*, интегрированных суммарно в 23 кластера, выявил в области кластера C1 («quantum memory») активное кластерообразование, связанное с термом C5 «post-quantum cryptography», позволяющее делать качественные выводы о локальной динамике научного ландшафта.

Результатом эмпирического исследования динамики формирования взаимодействий термов в сетевом калькуляторе *Pajek* и последующей обработки полученных параметров в пакете *STATISTICA* является построение временных рядов по изменению средней степени и общего числа связей, плотности сети, степеней центральности и показателя промежуточности библиографической сети.

Для продолжения анализа следует дополнить полученный эмпирический материал данными общих

сетевых параметров, характеризующих количество и размеры компонентов в сети, распределение расстояний в сети и степень распределения сетевых фрагментов, включая наличие гигантского компонента.

Формирование полного комплекса базовых параметров необходимо при последующем математическом и вычислительном моделировании информационных сетей, в ходе которых будет проведена оценка динамики кластеризации сети и достижения порога перколяции с течением времени.

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

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

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. 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

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
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. 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. 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 expanded 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

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
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 expanded 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

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

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