

## МЕТОДЫ И ТЕХНОЛОГИИ ПРИНЯТИЯ РЕШЕНИЙ

УДК 519.711:001

Научная статья

DOI: 10.18287/2223-9537-2025-15-4-578-597



## Методы и модели анализа научной деятельности

© 2025, Д.А. Губанов, Л.Ю. Жилякова✉, О.П. Кузнецов, А.Г. Чхартишвили

Институт проблем управления им. В.А. Трапезникова РАН (ИПУ РАН), Москва, Россия

### Аннотация

Статья посвящена описанию методов анализа научной деятельности и созданных на их основе инструментов для принятия обоснованных решений на всех уровнях управления в научной сфере – от отдельных исследователей до руководителей научных организаций. Проведён обзор отечественных и зарубежных исследований в области наукометрии, выявлены основные тенденции, а также обозначены пробелы, на устранение которых направлен проект ИПУ РАН по созданию информационной системы анализа научной деятельности. Проект включает разработку моделей, методов и алгоритмов анализа научной деятельности учёных, организаций, журналов, конференций в области теории управления на основе данных о публикациях. Эта задача решается с помощью междисциплинарного подхода, включающего сетевой анализ, онтологическое проектирование, методы анализа больших данных и машинного обучения, оптимизации. В основе разработанной системы лежит онтология научного знания в теории управления, основные принципы создания и структура которой описаны в статье. В сочетании с технологическими возможностями этой системы представленная онтология позволяет обеспечить высокий уровень детализации анализа, недоступный в других наукометрических системах (*Web of Science*, *Scopus*, РИНЦ и др.). Показано, что принципы создания и структура разработанной онтологии переносимы на другие области исследований. Рассмотрены современные методы сбора и представления данных, методы сетевого анализа и анализа содержания научных текстов. Представлены основные подходы, методы и модели анализа научной деятельности, рассмотрено применение этих методов в разработанной информационной системе.

**Ключевые слова:** анализ научной деятельности, теория управления, информационная система, классификация, онтология, тематический профиль, тематическое пространство, термин, обработка текстов, сетевой анализ.

**Цитирование:** Губанов Д.А., Жилякова Л.Ю., Кузнецов О.П., Чхартишвили А.Г. Методы и модели анализа научной деятельности. *Онтология проектирования*. 2025. Т.15, №4(58). С.578-597. DOI: 10.18287/2223-9537-2025-15-4-578-597.

**Вклад авторов:** Губанов Д.А. – раздел 3, список источников; Жилякова Л.Ю. – структура статьи, введение, раздел 1, список источников, заключение; Кузнецов О.П. – введение, раздел 2; Чхартишвили А.Г. – раздел 4, редактирование статьи.

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

### Введение

Рост количества научных журналов и увеличивающиеся темпы роста числа научных публикаций затрудняют полный охват информации даже в узких областях науки. Специалисты в одной и той же предметной области не всегда могут узнать друг о друге. Проблемы возникают у редакторов журналов при поиске рецензентов, у организаторов конференций при рассылке приглашений, у лиц, принимающих управленические решения в научной деятельности (НД). Из-за неравномерного представительства журналов из разных дисциплин при разделении на quartiles авторы не всегда могут найти высокорейтинговые журналы для

публикации своих результатов [1]. С развитием информационных и телекоммуникационных технологий всё большие объёмы данных становятся доступными в электронном виде, что открывает большие возможности для автоматизации сбора, структурирования и обработки информации, создания моделей анализа, прогнозирования, планирования и поддержки принятия решений в области управления НД.

Вследствие этого актуально создание систем, автоматизирующих получение информации об объектах НД (публикациях, авторах, научных журналах, организациях, конференциях и др.) на основе анализа публикаций. Такие системы можно разделить на две группы. Хорошо известны базы *Web of Science*, *Scopus*, РИНЦ, *Google Scholar*, *ResearchGate*, *OpenAlex* и др., их главная цель – анализ цитируемости публикаций, на основе которого вычисляются научометрические оценки публикаций и их авторов (индекс Хирша), а также научных журналов (импакт-фактор). Более сложными и менее исследованными являются задачи анализа содержания научных текстов. Систем такого рода гораздо меньше. Можно отметить систему *Semantic Scholar* ([www.semanticscholar.org](http://www.semanticscholar.org)), специализирующуюся на компьютерных науках и медицине, а также разработку *iFORA* (<https://issek.hse.ru/ifora>), которая работает не только с научными публикациями, но и с патентами, бизнес-аналитикой и др.

Несмотря на различия в целях и в методах анализа, указанные системы имеют два обязательных компонента: базу публикаций и средства представления тематического пространства, в котором позиционируются объекты НД. Структура тематического пространства задаётся классификаторами научных областей (УДК [2], MSC2020 [3], PACS [4], OECD [5] и др.).

В настоящей статье дан обзор методов и систем анализа НД; рассмотрена Информационная система анализа научной деятельности (ИСАНД), разработанная в ИПУ РАН [6]. Её основу составляет онтология наук об управлении [7, 8]. На основе классификатора ИСАНД разработан новый тематический класс «Теория управления» в Государственном рубрикаторе научно-технической информации (ГРНТИ) [9], который появится в открытом доступе в следующей редакции ГРНТИ под номером 42.

## 1 Управление научной деятельностью: обзор основных подходов

Для управления наукой требуются комплексные модели НД, интегрирующие методы научометрии, сетевого анализа (СА), теории управления. В качестве объективной оценки работы учёных, журналов, научных коллективов, школ, организаций используются научометрические показатели [10, 11]. Термин «научометрия» в русскоязычных исследованиях был введён в 1969 г. [12] и не потерял своей актуальности спустя десятилетия [13]. В настоящее время исследования по научометрии посвящены периодические научные журналы: *“Scientometrics”* (*Springer*), *“Quantitative Science Studies”* (*MIT Press Direct*), *“Journal of Informetrics”* (*ScienceDirect*), «Управление наукой и научометрия» (РИЭПП), «Библиосфера» (ГПНТБ СО РАН) и др.

Большой потенциал в научометрии и управлении НД имеет применение аппарата сетевых наук. С их помощью становятся возможными исследование возникающих научных связей, явных и неявных сообществ, поиск наиболее влиятельных учёных и научных школ и решение других задач, которые можно сформулировать в рамках сетевой парадигмы. Начало этому направлению положили статьи [14-16]. В обзоре [17] используется рекурсия (исследуется развитие области СА): построена сеть соавторства более 50 тысяч учёных, занимающихся сетевыми науками; проанализирована её топология и динамика; исследованы модели сотрудничества и структурные свойства сети соавторства; определены ведущие авторы, крупнейшие сообщества; проведено сравнение свойств сети с научометрическими показателями.

В настоящее время наукометрия включает ряд методов СА [18]: определение центральности узлов (обзор методов определения центральности см. в [19]) для поиска ключевых авторов и публикаций, влияющих на развитие области исследования [20, 21], кластерный анализ в сетях [22] и его адаптация к наукометрическим сетям [23–25], визуализация связей [26, 27], определение плотности сети и др.

Объёмный наукометрический анализ литературы по теме «научные социальные сети» проведён в [28]. Программная статья [29], посвящённая эволюции сетей научного сотрудничества, процитирована около четырех тысяч раз. Исследованию научных социальных сетей (*LinkedIn*, *ResearchGate*, *Academia.edu*, *Mendeley*) посвящена статья [30]. Научным взаимодействиям в научных сетях посвящено большое количество работ (см., напр., [31–33]). Как правило, в исследованиях рассматриваются отдельные области науки [34, 35], отдельные научные и исследовательские заведения [36, 37]; в [38] проведён анализ гендерного распределения авторов научных публикаций; глобальное гендерное неравенство в науке исследуется в [39]. Исследование научных социальных сетей позволяет представить структуру научного сообщества, выявить влиятельных авторов и организации, проследить возникновение и развитие сотрудничества.

Отдельной областью исследований является анализ индексов цитирования и их влияния на научные сети, который имеет важное значение в оценке научного влияния и создании баз данных (БД) для СА. К основным наукометрическим индексам относятся: для авторов – индекс Хирша ( $h$ -индекс) и его модификации ( $g$ -индекс и  $i$ -индекс); для журналов – показатели, отражающие среднее количество цитируемости недавних статей (к ним можно отнести импакт-фактор и *SiteScore*). В [40] сравниваются различные модификации индекса Хирша на примере данных из области биомедицины; в [41] приведён обзор исследований по цитированию авторов. В [42] приводится сравнение показателей и цитирования с междисциплинарной точки зрения. В [43] исследуется стратификация сетей соавторства, построенных по данным за 50-летний период, на основе  $h$ -индекса. В [44] рассмотрена совместная эволюция сетей соавторства и цитирования, а также их влияние на научные показатели: индекс Хирша автора и импакт-фактор журнала. В [45] предлагается альтернативный индекс, который учитывает семантические связи опубликованных работ в междисциплинарной области знаний, а также цитирование и соавторство учёных, облегчая идентификацию и картирование наиболее релевантных тем и авторов в этой области. Высокую цитируемость имеют обзор литературы по показателям цитируемости [46] и обзор по теории и практике наукометрии [47].

Проблемы, препятствующие развитию наукометрических исследований и научных коммуникаций, и пути их преодоления рассматриваются в [48]. Методика оценки результативности научных организаций представлена в [49], а концептуальная модель системы наукометрического мониторинга результативности НД – в [50]. Задачам управления НД с помощью наукометрического подхода посвящены исследования [51, 52]. Современные подходы в наукометрии изложены в [53]. Международное сотрудничество учёных исследуется в [54, 55], при этом используются новые наукометрические показатели (коцитирование и др.). Анализ количества публикаций и цитирований в различных областях отечественной науки проведён в [56]. Анализу публикаций в информационной системе *Math-Net.Ru* посвящена работа [57], граф цитирования статей российских математиков приведён в [58]. Обзор литературы о влиянии самоцитирования на возникающие искажения в библиометрическом анализе представлен в статье [59]. Сетевая модель коллaborации учёных приведена в [60]. В ИПУ РАН подготовлен специальный выпуск по наукометрии в журнале «Управление большими системами» [61].

Менее исследованными являются задачи анализа содержания научных текстов. В [62] на основе семантического анализа и анализа социальных сетей предложено прогнозирование

успеха научных публикаций. Проведённый на основе набора данных анализ публикаций по химической инженерии за 2010–2012 годы позволил предсказать цитирование через шесть лет после публикации с точностью почти 80%. В [63] исследуются тематики библиометрических исследований посредством сетей цитирования и семантического анализа. Рассмотрено распределение тем научной литературы, в которой используются термины «библиометрия», «наукометрия» и «информетрика». В [64] проводится мета-анализ семантической классификации цитирований на корпусе 60 научных статей в этой области. Исследуются подходы к классификации цитирований на основе их семантического типа. Обзор [65] посвящён кластеризации исследовательских документов на основе семантического анализа и извлечения ключевых слов. Проведён сравнительный анализ алгоритмов извлечения ключевых слов и кластеризации; предложен прототип поисковых систем на основе документов, а также методы семантической классификации исследовательских работ в области компьютерных наук.

Коллективом ФИЦ ИУ РАН разработаны методы семантического анализа научных текстов [66], методы реляционно-ситуационного анализа текстов [67]; интеллектуальная поисковая система *Exactus* (<http://www.exactus.ru/>), на основе которой созданы система выявления заимствований в научных текстах *Exactus Like* [68] (<http://like.exactus.ru/>) и система интеллектуального поиска и анализа научных публикаций *Exactus Expert* [69] (<http://expert.exactus.ru/>). В [70] представлен прогноз путей развития науки, технологий и инноваций. Представлен подход и инструменты интеллектуального анализа текста для построения карт науки особого вида – «Карты науки в квадрате». Работа [71] посвящена семантическому анализу трудов конференции по семантическому анализу.

Разработанный в Московском государственном университете многоязычный энкодер *SciRus-tiny*, использующий семантические векторные представления текстов, предназначен для анализа научных текстов и способен осуществлять поиск близких по тематике публикаций [72]. Открытый бенчмарк *RuSciBench* рекомендован для оценки векторных представлений научных текстов на русском и английском языках из данных библиотеки *eLibrary* [73].

Выявлению plagiarism и нарушения этики публикаций в русскоязычных научных изданиях посвящена статья [74], а выявлению сгенерированных языковыми моделями фрагментов в научных публикациях – работа [75].

## 2 Онтология наук об управлении

Научные тексты перед публикацией получают коды одного или нескольких универсальных классификаторов (УДК [2], MSC2020 [3], PACS [4], OECD [5] и др.), которые строго соответствуют таксономическому принципу: каждый объект классификации относится к одной вершине дерева. Во многих случаях однозначная классификация оправдана, а зачастую и неизбежна: принятая статья попадает в одну определённую рубрику журнала, а доклад включается в одну секцию конференции.

Однако традиционные универсальные классификаторы имеют ряд недостатков и могут не в полной мере отражать соответствие статьи научным тематикам. Их одномерная структура, соответствующая таксономическому принципу, усложняет категоризацию междисциплинарных работ и не позволяет корректно отразить различия в компетенциях специалистов. Поэтому классификатор УДК время от времени подвергается модификациям, в результате которых он оказывается неравномерным по уровням, а, например, в Российском научном фонде требуется указывать в заявках на гранты основной код и два дополнительных. Это означает признание неодномерности позиционирования заявки в тематическом пространстве и позволяет при необходимости провести рецензирование заявки экспертами из разных научных областей. Кроме того, универсальность классификаторов приводит к чрезмерно

крупному членению научных дисциплин и, как следствие, недостаточной детализации разделов. Это хорошо для научных библиотек, которые, как правило, тоже универсальны, но не всегда хорошо для научных журналов, которые специализированы и нуждаются в более глубокой детализации.

Примером научной области, в которой неудобна одномерная классификация, является теория управления. Неудобства проявляются в работе редакций научных журналов по управлению. Для рецензирования статьи, в которой описывается, например, интеллектуальная система, нужен не только специалист по искусственному интеллекту, но и эксперт, владеющий математическими методами, использованными при описании системы, а также компетентный в предполагаемой сфере применения этой системы.

Подход к решению этой проблемы предложен в ИПУ РАН [7]. Тематическое пространство наук об управлении делится на три области: *фундаментальные науки*; *прикладные теории и проблемные области*; *области приложений*. Каждая из этих областей детализируется: среди фундаментальных наук выделяются математика, физика, биология, ...; в проблемных областях – теория автоматического управления, анализ данных, вычислительная техника, ...; в сферах приложений – летательные аппараты, медицина, технологические процессы, финансы и др. Фрагмент *дерева тематического пространства* представлен на рисунке 1. Вершины дерева названы *темами*. Число уровней дерева и ветвлений вершин зависит от требуемой детализации, но три вершины первого уровня, названные *главными*, неизменны.



Рисунок 1 – Фрагмент онтологии тематического пространства  
(прямоугольники – темы и подтемы, листья – термины)

Позиционирование объектов НД в тематическом пространстве многомерно. Как правило, документ (статья, доклад и т.д.) релевантен многим темам, и поэтому он характеризуется вектором релевантностей, который называется *профилем документа*. Этот вектор представляет собой сечение дерева (не обязательно равномерное по уровням); в каждой точке сечения (теме) стоит число из отрезка  $[0, 1]$ ; сумма всех чисел равна 1 (вектор – стохастический). Каждое число – степень релевантности документа данной теме относительно других тем.

Для вычисления профиля документа тематическое дерево снабжается словарём: к каждой висячей вершине-теме (и, быть может, к некоторым промежуточным вершинам-темам) при соединяется множество вершин-терминов, характеризующих данную тему. Тематическое дерево вместе со словарём образует онтологию наук об управлении. В этой онтологии смежные вершины-темы связаны отношением класс-подкласс, а смежные вершина-тема и вершина-термин связаны отношением класс-экземпляр. Введение терминов нарушает древовидность онтологии, т.к. термин может относиться к нескольким темам. Метод вычисления профиля документа в [7] основан на частотности содержащихся в нём терминов.

В программной реализации описанной схемы использована база публикаций работников ИПУ РАН. В [7] приведены некоторые данные пробной эксплуатации разработанной системы на материале публикаций журнала «Автоматика и телемеханика». Набор профилей публикаций работника даёт возможность вычислять профиль работника, характеризующий набор его научных интересов и компетентностей. А наличие профилей работников облегчает решение различных задач по управлению НД, таких как назначение рецензентов и экспертов, подбор команды для научного проекта и т.д.

Составление словаря – это работа, требующая разнообразных компетенций, охватывающих всё тематическое пространство наук об управлении. Для практического использования онтологии удобно, чтобы тематические профили научных объектов строились на фиксированных уровнях дерева тематического пространства.

### 3 Методы и модели

#### 3.1 Сбор данных и извлечение информации

Для анализа НД требуются данные о публикациях и связанных сущностях: метаданные статей, сведения об авторах и их аффилиациях, журналах и конференциях, грантах и т.п. Источники данных – коллекции полнотекстовых документов и открытые библиографические базы. Если доступны только полнотекстовые документы, то, как правило, применяются инструменты извлечения метаданных и ссылок. В частности, библиотека *GROBID* [76] демонстрирует лучшее качество извлечения метаданных и списков литературы среди открытых инструментов. Для специализированных задач могут привлекаться отдельные инструменты (например, *Adobe Extract* для табличных данных). На практике нередко применяют комбинированные конвейеры: несколько инструментов анализа используются последовательно или параллельно, а результаты объединяются и проходят верификацию экспертом.

В частности, в модуле предварительной обработки данных ИСАНД [6] применяется конвейер на основе *GROBID*: модель предварительно дообучена на корпусе русскоязычных текстов по теории управления, что позволяет повысить точность извлечения данных.

Открытые агрегаторы научных данных, такие как *OpenAlex*, предоставляют свободный доступ к информации о публикациях и цитированиях. *OpenAlex* обеспечивает сопоставимое с коммерческими базами *Scopus/ Web of Science* покрытие цитат (на пересечении их корпусов) и индексирует существенно больше журналов открытого доступа (34 тыс. в *OpenAlex*, 6 тыс. в *WoS* и 7 тыс. в *Scopus* на 2024 год) [77]. Качество и полнота метаданных могут различаться по дисциплинам, поэтому целесообразна оценка покрытия БД для конкретной предметной области [77, 78]. В ИСАНД выполняется сбор данных из нескольких источников (в том числе *OpenAlex*), что повышает полноту базы. Каждый источник вносит вклад в общий граф знаний, при этом при загрузке данных в ИСАНД помечается происхождение метаданных.

Ключевая задача при сборе данных – объединение записей об одном объекте (авторе, организации) из разных источников. Внедрение уникальных идентификаторов научных объек-

тов – *ORCID* для авторов, *Research Organization Registry (ROR)* для организаций и др. – упрощает эту задачу. В частности, в *Crossref* поддерживаются *ROR*-идентификаторы фондов, это позволяет унифицировать сведения о грантодателях (см. <https://ror.org/blog/2025-03-05-using-ror-ids-in-place-of-funder-ids>). Для полного автоматического разрешения неоднозначности требуются специальные алгоритмы и эталонные наборы данных. Создаются бенчмарки для сравнения методов идентификации авторов по метрикам точности [79]. В ИСАНД применяется алгоритм объединения записей на основе полученных из разных источников идентификаторов объектов: *ORCID*, *DOI*, *Researcher ID*, *Scopus Author ID*, *ROR* и др.

### 3.2 Структурирование и представление метаданных

Собранные данные о публикациях и связях между ними необходимо представить в удобной для анализа форме. Для этого используются реляционные схемы данных, которые обеспечивают целостность и эффективный доступ к данным, но не являются гибкими (например, при добавлении новых сущностей) и не предназначены для представления сетевых структур (например, для поиска цепочки сотрудничества между учёными).

Графовая модель данных позволяет отразить экосистему науки как сеть объектов. В такой модели любые научные объекты – публикации ( $D$ ), исследователи ( $A$ ), организации ( $O$ ), конференции ( $C$ ), ключевые термины ( $T$ ) и др. – можно представить как вершины графа ( $v \in V$ ), а отношения между ними – как рёбра ( $e = (v_i, v_j) \in E$ ). Можно определить сеть НД как ориентированный граф  $G = (V, E)$ , где  $V$  – объединение множеств разных сущностей, а  $E$  содержит несколько подмножеств  $E_1, E_2, \dots$  типов связей. Например, если  $A$  – множество авторов,  $D$  – множество публикаций, то подграф *соавторства* можно задать как  $G_{auth} = (A, E_{auth})$ , где каждое ребро  $e = (a_i, a_j) \in E_{auth}$  означает совместную публикацию автора  $a_i$  с автором  $a_j$ . Аналогично, *граф цитирования* задаётся как ориентированный граф  $G_{cit} = (D, E_{cit})$ , где дуга  $(d_i \rightarrow d_j) \in E_{cit}$  означает цитирование работы  $d_j$  в работе  $d_i$ . Эти специализированные графы можно анализировать раздельно, однако полное представление даёт интегрированная сеть, включающая все типы объектов и связей. В интегрированном представлении множество вершин  $V = A \cup D \cup O \cup T \cup \dots$ , а каждое ребро помечено типом (например: «автор – статья»; «статья цитирует статью»; «статья содержит термин»; «автор работает в организации»). Этот граф можно воспринимать как многослойную сеть науки: каждый тип связи образует свой слой, но объекты (вершины) одни и те же и могут соединять слои [80]. Для учёта групповых отношений, когда единицами анализа выступают не пары, а группы объектов, применяются *гиперграфы*. Гиперребро может связывать, например, множество авторов  $a_1, a_2, \dots, a_k$ , совместно написавших статью (это единое событие, связывающее всю группу). Гиперграфы усложняют математический аппарат анализа, но способны представить коллективные взаимодействия напрямую, без разбиения их на парные взаимодействия.

Преимущество графового представления – возможность хранить и обрабатывать всю структуру отношений. Добавление новой сущности или типа связи не требует доработки схемы данных – достаточно ввести новый тип узла или ребра. Запросы к графу могут гибко следовать по любым цепочкам связей: например, можно запросить «найти путь от автора А к автору Б через цепочку совместных работ и цитирований длиной не более 4» или «найти всех авторов, статьи которых процитированы в журнале В». Подобные запросы сложно выразить на языке *SQL* в реляционной БД, тогда как графовые базы имеют для этого оптимизированные языки (*Gremlin*, *Cypher*, *SPARQL* и др.) и алгоритмы обхода графов. Сформировались две основные парадигмы для работы с графовыми данными. Первая – семантические сети, основанные на стандартах *RDF/OWL* (*Resource Description Framework/Web Ontology Language*). Семантические хранилища поддерживают язык запросов *SPARQL*, позволяют ис-

пользовать онтологии и выполнять логический вывод. Вторая парадигма – графы свойств [81], в которой узлы и связи могут иметь произвольные свойства (атрибуты), а схема данных часто гибридна. В ИСАНД [6] для хранения и обработки данных о НД используется семантическое хранилище, схема которого задаётся онтологией *OWL*.

### 3.3 Анализ текстов научных публикаций

Анализ содержания научных публикаций позволяет определить тематику работ и измерить их семантическую близость. Публикацию можно представить в виде вектора признаков (профиля), отражающего распределение тем или терминов, разными способами: с помощью классификаторов; тематического моделирования; нейросетевых векторных представлений.

Первый способ – заранее создать *онтологию* или *рубрикатор научных тем* и на их основе классифицировать тексты [82, 83]. В частности, в системе ИСАНД используется разработанный экспертами словарь, покрывающий ключевые понятия (термины) заданной научной области. Это позволяет каждому научному объекту ставить в соответствие стохастический вектор  $\mathbf{p} = (p_1, p_2, \dots, p_n)$  в пространстве из  $n$  предопределённых тематических категорий, где  $p_i \geq 0$  отражает степень принадлежности объекта к теме  $i$ . В ИСАНД профили рассчитываются на основе онтологии наук об управлении. Каждая загруженная в систему публикация автоматически анализируется: производится лемматизация, извлекаются ключевые слова, соотносятся с терминами тезауруса, формируется тематический профиль  $\mathbf{p}$ . Профили хранятся как атрибуты соответствующих узлов (публикаций, авторов, организаций), что позволяет быстро сравнивать объекты по их тематическим векторам. Для измерения близости профилей используется косинусное сходство или метрика на основе  $L_1$ -нормы, например расстояние для двух публикаций  $d_1$  и  $d_2$ :  $d(\mathbf{p}^{(d1)}, \mathbf{p}^{(d2)}) = \frac{1}{2} \sum_i |p_i^{(d1)} - p_i^{(d2)}|$ . Близкие по тематике публикации будут иметь близкое к 1 значение. Это даёт возможность улучшать поиск и рекомендации: в ИСАНД реализованы методы поиска экспертов для рецензирования, поиска тематически близких исследований и т.д.

В противоположность априорному заданию тем *методы тематического моделирования* позволяют автоматически извлекать темы из массива текстов. Наиболее известный метод – *LDA* (*Latent Dirichlet Allocation*) и его варианты [84, 85]. Основная гипотеза *LDA*: существует некоторое (заданное числом  $k$ ) количество скрытых тем, каждая из которых характеризуется статистическим распределением слов  $P(w|z)$ , а каждый документ есть смесь этих тем с определёнными весами  $\theta_d = P(z|d)$ . Параметры модели (распределения  $P(w|z)$  и  $\theta_d$  для всех документов) оцениваются по набору текстов с помощью стохастических методов. В результате работы алгоритма получается набор из  $k$  тем – списков слов с указанием вероятностей  $P(w|z_i)$ . Одновременно для каждого документа получается распределение  $\theta_d = (p(z_1|d), \dots, p(z_k|d))$ , которое является тематическим профилем этого документа для выявленных алгоритмом тем, не привязанных к каким-либо заранее заданным категориям. Тематические модели хорошо подходят для *первичного анализа* корпуса, когда заранее может быть неизвестно, какие тематические кластеры присутствуют. Тематическое моделирование часто применяется для анализа научной литературы, в т.ч. кластеризации публикаций [86] и отслеживания эволюции научных направлений [87].

Можно использовать *векторные представления текстов* на основе предобученных языковых моделей [88]. Эти модели преобразуют текст статьи  $d$  в плотный вектор  $\mathbf{e}_d \in \mathbb{R}^D$  (эмбеддинг), располагая близкие по смыслу публикации рядом в пространстве эмбеддингов (даже если они не имеют общих ключевых слов). Используя эмбеддинги, можно решать разные задачи, в частности проводить тематическую кластеризацию публикаций [89]. Недостатки нейросетевых эмбеддингов и тематического моделирования – отсутствие учёта дрейфа пред-

метной области (поскольку обучающая выборка фиксирована) и сложность интерпретации признаков эмбеддинга.

Для решения конкретных научных задач используются *специальные методы анализа текстов*. В частности, методы извлечения терминов и ключевых слов [90], автоматического реферирования [91], анализа интенции цитирования [92-94], выявления новых научных направлений [95]. Эти методы позволяют составить тезаурус, облегчить анализ больших коллекций работ и оценить влияние научной работы по характеру цитирования. Большие языковые модели активно исследуются, но их применение для анализа НД сопряжено с рисками генерации некорректной информации [96]. Поэтому их следует воспринимать как вспомогательный инструмент для эксперта.

### 3.4 Сетевой анализ

Сетевые модели фокусируются на структуре научного сообщества и коммуникаций в нём. *Сеть соавторства* – неориентированный граф, в котором узлы – авторы, а рёбра соединяют соавторов одной публикации. Анализ такой сети позволяет найти научные группы, научные школы, «мосты» между научными группами, определить силу связей между исследователями и т.п. *Сеть цитирования* – ориентированный граф, в котором узлы – статьи (или другие научные объекты), а ребро  $A \rightarrow B$  означает, что работа  $A$  цитирует работу  $B$ . Граф отражает информационные потоки, позволяет находить наиболее влиятельные публикации и научные направления, строить карты знаний по цитируемости. В зависимости от цели исследования строятся и другие виды научных сетей: граф аффилиации (автор – организация), граф сотрудничества организаций, граф терминов (термин – узел, совместная встречаемость терминов – связь). Эти представления можно рассматривать как проекции единой многослойной сети НД.

Такой подход позволяет выявлять явные и скрытые закономерности, в т.ч. структуру коллaborаций, центры влияния, междисциплинарные связи. Типовые задачи СА: *идентификация ключевых узлов / расчёт влиятельности узлов* – поиск наиболее значимых учёных, работ, организаций по различным показателям [10, 97]; *обнаружение неформальных сообществ* – разбиение сети на плотные кластеры (например, выявление тематических или географических научных групп) [98]; *рекомендация и прогноз связей* – выявление потенциальных связей (для эффективного сотрудничества учёных, для тематического дополнения работ и т.п.) [99]; *анализ динамики* – исследование эволюции сети во времени (например, для оценки роста новых направлений) [100].

В теории сложных сетей разработан широкий спектр мер *центральности* (более 400) (см. [101]), позволяющих ранжировать узлы по их значимости [102]. Мера центральности – это функция  $C$ , которая каждому узлу сети ставит в соответствие действительное число. Простейшая мера  $C_D(v) = \deg(v)$  – степень вершины  $v \in V$  (число связей узла). В контексте науки степень автора в графе соавторства равна числу его соавторов, а степень статьи в графе цитирования – числу её входящих ссылок (цитируемость) либо исходящих (число цитат). Высокая степень означает активное участие объекта в научной коммуникации. Более сложные меры учитывают глобальные свойства. Например, посредничество  $C_B(v)$  показывает, через какие узлы проходят кратчайшие пути в сети: узел с высоким посредничеством служит «мостом» между частями графа. Такой узел может указывать на учёного, соединяющего различные научные сообщества, или работу, цитируемую разными областями, или метод оценки влияния на основе причинно-следственных отношений [103]. Эти показатели позволяют количественно определить наиболее влиятельных учёных, публикации, организации. Библиометрические индексы можно рассматривать как частные случаи мер центральности, однако сетевые показатели дают более полную картину, с учётом всей структуры графа цитирова-

ния. Современные обзоры подтверждают продуктивность сетевого подхода для оценки научного вклада учёных [104].

Важная задача СА – *выявление сообществ в сетях*. Сообществом называют подмножество тесно связанных узлов [98]. Для нахождения сообществ ищется разбиение множества узлов  $f: V \rightarrow \{1, \dots, K\}$  (количество сообществ  $K$  не фиксируется). Для автоматического обнаружения сообществ применяются алгоритмы кластеризации графов, которые обычно оптимизируют модульность – меру  $Q \in [-0,5; 1]$ , сравнивающую плотность внутри кластера с ожидаемой в случайной модели [105]. Методы итерационного укрупнения графа, такие как алгоритм Лувена и его модификации [106], гарантируют получение более связных кластеров. Выявленные в сетях соавторства сообщества помогают обнаружить коллaborации и скрытые социальные структуры (например, научные школы). Кластеры в сети цитирования зачастую соответствуют направлениям исследований.

*Методы прогнозирования связей* [99] позволяют определить, какие связи являются скрытыми или какие связи появятся в будущем (потенциальных соавторов, сотрудничество между группами). В случае со скрытыми связями используется следующая постановка задачи: пусть истинная сеть  $G^* = (V, E^*)$  наблюдаема частично  $G_{obs} = (V, E_{obs})$ , где  $E_{obs} \subseteq E^*$ , необходимо восстановить  $E^* \setminus E_{obs}$ . Для определения потенциальных связей используются меры сходства узлов  $u, v \in V$  в графе [99], в частности, мера *Adamic/Adar*:  $s_{AA}(u, v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log \deg(z)}$ . Актуально направление *анализа динамики сетей* [100]: изучается, как изменяется структура сети со временем  $G(t)$  – например, растёт ли связность научного сообщества, увеличивается или уменьшается фрагментация по группам, как влияют коллaborации на последующую цитируемость работ. Такие исследования находятся на стыке с социологией науки и позволяют делать выводы о тенденциях в организации НД.

### 3.5 Интеграция методов анализа текстов и анализа сетей

Сетевые методы дают в распоряжение исследователей мощный инструментарий для структурного анализа науки [104]. Однако графы не позволяют учитывать смысловое содержание: статьи могут быть связаны через общих авторов или цитирование, но иметь разную тематику. Методы анализа текстов сосредоточены на тематике (на смысле). Их ограничения – отсутствие учёта социальной структуры (исследователи могут быть тематически близки, но в сети находятся далеко друг от друга), значительные затраты на подготовку данных (очистка текста, лемматизация, обучение моделей) и обработку корпусов, особенно полнотекстовых. Предпочтительна интеграция подходов для одновременного учёта того, *кто с кем взаимодействует, и о чём ведутся исследования*. В частности, комбинация признаков социальной сети и семантики текста позволяет предсказывать цитируемость с точностью около 80% [62]. В [107] показано, что качество профилей научных объектов можно улучшить, если учесть связи авторства между объектами. В ИСАНД все данные хранятся в виде графа (что даёт возможность выполнять СА – обход связей, расчёт центральностей и пр.), при этом каждый узел снабжён тематическим профилем. Это позволяет выполнять комбинированные запросы и аналитические сценарии. Например, при поиске рецензента можно отобрать кандидатов по близости профилей и выбрать тех, у кого высокие показатели влияния в сети.

## 4 Основные характеристики разработанной информационной системы

ИСАНД предназначена для обеспечения исследователей, научных коллективов и организаторов науки средствами анализа НД и содержит данные о различных *научных объектах*.

Основой ИСАНД являются массивы публикаций, загружаемые из внешних источников – крупных БД научных материалов. Поскольку одна и та же публикация нередко индексируется в разных системах, в ИСАНД предусмотрена возможность выявления и фиксации дубликатов. То же относится и к другим научным объектам, например авторам или организациям: один и тот же объект может иметь разные представления в разных источниках. Для случаев, когда вероятность совпадения велика (например, у авторов один и тот же уникальный цифровой идентификатор *ORCID*) в системе указывается, что объекты являются дубликатами. Отношение «является дубликатом» считается симметричным и транзитивным.

Особенностью ИСАНД является наличие оригинального *классификатора*, который задаёт структуру *тематического пространства* теории управления. Каждый научный объект в этой системе описывается как вектор в многомерном пространстве, и его можно считать точкой в стандартном симплексе соответствующей размерности. Такой подход обеспечивает корректное сравнение объектов и позволяет применять различные методы анализа.

Структура классификатора имеет три основных уровня. Первый уровень – это *факторы*, которые отражают укрупнённые области исследований, например «Управление в организационных системах». Второй уровень образуют *подфакторы*, детализирующие тематику в рамках факторов, например «Управление в сетевых структурах». Третий уровень представляет собой набор конкретных *терминов*, например, «Социальное влияние».

Для каждого научного объекта в ИСАНД строятся три вида тематических профилей: *базовый профиль* на уровне факторов; *детализированный профиль* на уровне подфакторов; *профиль терминов*. На каждом уровне строится профиль каждой публикации и проводится расчёт профиля для других объектов на основе публикаций, связанных с этими объектами: профиль автора на основе его публикаций (с учётом количества соавторов каждой публикации), профиль журнала на основании опубликованных в нём статей и т.п. С их помощью в ИСАНД осуществляется поиск публикаций, авторов и других научных объектов.

На множестве тематических профилей введена метрика, позволяющая рассчитывать расстояние (от 0 до 1) между научными объектами и находить тематически близкие объекты. Например, исследователь может искать все публикации, находящиеся в тематическом пространстве не далее, чем на расстоянии 0,2 от его собственной работы. В специализированных запросах и сценариях поиска (кого пригласить для рецензирования публикации, кому направить приглашение на конференцию и т.д.) пользователь может самостоятельно задавать радиус тематической окрестности научного объекта (область точек, которые находятся от научного объекта на расстоянии не больше заданного). Алгоритмы расчёта профилей и расстояний между ними описаны в [6].

## Заключение

В статье описана Информационная система ИСАНД в области теории управления и сделана попытка поместить её в более широкий контекст мировой науки о науке – как наукометрии, так и анализа текстов и СА, выявляющих ключевые темы и термины публикаций, взаимосвязи между учёными, организациями и конференциями. Представленный подход к построению онтологии наук об управлении позволяет позиционировать объекты (публикации, их авторов, журналы, конференции и научные организации) в многомерном тематическом пространстве. ИСАНД потенциально представляет собой многофункциональную систему, которая объединяет онтологическую основу, массивы данных и математический аппарат анализа. Информационная система ИСАНД призвана обеспечить исследователей и управленицев инструментами для информационного поиска, изучения НД и принятия обоснованных решений в организации и управлении НД.

## СПИСОК ИСТОЧНИКОВ

- [1] **Kosyakov D., Pislyakov V.** “I'd like to publish in Q1, but there's no Q1 to be found”: Study of journal quartile distributions across subject categories and topics. *Journal of Informetrics*. 2024. Vol.18(1). P.101494. DOI: 10.1016/j.joi.2024.101494.
- [2] ГОСТ 7.90 2007. Система стандартов по информации, библиотечному и издательскому делу. Универсальная десятичная классификация. Структура, правила введения и индексирования: издание официальное.
- [3] Mathematics Subject Classification (MSC2020). <https://mathscinet.ams.org/mathscinet/msc/msc2020.html>.
- [4] Physics and Astronomy Classification Scheme. [https://publishing.aip.org/wp-content/uploads/2019/01/PACS\\_2010\\_Alpha.pdf](https://publishing.aip.org/wp-content/uploads/2019/01/PACS_2010_Alpha.pdf).
- [5] Коды международной классификации (OECD). [https://storage.tusur.ru/files/134958/kody\\_OECD.pdf](https://storage.tusur.ru/files/134958/kody_OECD.pdf).
- [6] **Губанов Д.А., Кузнецов О.П., Курако Е.А. и др.** Информационная система анализа научной деятельности (ИСАНД) в области теории управления. *Проблемы управления*. 2024. №3. С.42-65. DOI: 10.25728/ru.2024.3.4.
- [7] **Кузнецов О.П., Суховеров В.С.** Онтологический подход к оценке тематики научного текста. *Онтология проектирования*, 2016. Т.6, №1. С.55-66. DOI: 10.18287/2223-9537-2016-6-1-55-66.
- [8] **Агаев Р.П., Алекскеров Ф.Т., Алчинов А.И. и др.** Теория управления: словарь системы основных понятий. М.: ЛЕНАНД, 2024. 128 с.
- [9] Государственный рубрикатор научно-технической информации. <https://www.gpntb.ru/grnti.html>.
- [10] **Fortunato S., Bergstrom C.T., Börner K., Evans J.A., Helbing D., Milojević S., ... & Barabási A.L.** Science of science. *Science*. 2018. Vol.359(6379). P.eaa0185. DOI: 10.1126/science.aa0185.
- [11] **Gates A.J., Barabási A.L.** Reproducible science of science at scale: pySciSci. *Quantitative Science Studies*. 2023. Vol.4(3). P.700-710. DOI: 10.1162/qss\_a\_00260.
- [12] **Налимов В.В., Мульченко З.М.** Наукометрия. Изучение развития науки как информационного процесса. М.: Наука, 1969. 192 с.
- [13] **Granovsky Y.V.** Is It Possible to Measure Science? V. V. Nalimov's Research in Scientometrics. *Scientometrics*. 2001. Vol.52. P.127-150. DOI: 10.1023/A:1017991017982.
- [14] **Barabási A.L., Albert R.** Emergence of scaling in random networks. *Science*. 1999. Vol.286(5439). P.509-512. DOI: 10.1126/science.286.5439.509.
- [15] **Watts D.J., Strogatz S.H.** Collective dynamics of 'small-world' networks. *Nature*. 1998. Vol.393(6684). P.440-442. DOI: 10.1038/30918.
- [16] **Girvan M., Newman M.E.J.** Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002. Vol.99(12). P.7821-7826. DOI: 10.1073/pnas.122653799.
- [17] **Molontay R., Nagy M.** Twenty years of network science: A bibliographic and co-authorship network analysis. *Big data and social media analytics: trending applications*. Cham: Springer International Publishing, 2021. P.1-24. DOI: 10.1007/978-3-030-67044-3\_1.
- [18] **Costa L.D.F., Rodrigues F.A., Travieso G., Villas Boas P.R.** Characterization of complex networks: A survey of measurements. *Advances in physics*. 2007. Vol.56(1). P.167-242. DOI: 10.1080/00018730601170527.
- [19] **Saxena A., Iyengar S.** Centrality measures in complex networks: A survey //arXiv preprint arXiv:2011.07190. 2020. DOI: 10.48550/arXiv.2011.07190.
- [20] **Dias A., Ruthes S., Lima L., Campra E., Silva M., Bragança de Sousa M., Porto G.** Network centrality analysis in management and accounting sciences. *RAUSP Management Journal*. 2020. Vol.55. P.207-226. DOI: 10.1108/RAUSP-02-2019-0021.
- [21] **D'Ippoliti C.** “Many-Citedness”: Citations Measure More Than Just Scientific Quality. *Journal of Economic Surveys*. 2021. Vol.35(5). P.1271-1301. DOI: 10.1111/joes.12416.
- [22] **Lancichinetti A., Fortunato S.** Consensus clustering in complex networks. *Scientific reports*. 2012. Vol.2(1). P.336. DOI: 10.1038/srep00336.
- [23] **Ozcan S., Boye D., Arsenyan J., Trott P.** A scientometric exploration of crowdsourcing: Research clusters and applications. *IEEE Transactions on Engineering Management*. 2020. Vol.69(6). P.3023-3037. DOI: 10.1109/TEM.2020.3027973.
- [24] **Lund B., Ma J.** A review of cluster analysis techniques and their uses in library and information science research: *k*-means and *k*-medoids clustering. *Performance Measurement and Metrics*. 2021. Vol.22(3). P.161-173. DOI: 10.1108/PMM-05-2021-0026.
- [25] **Madani F.** ‘Technology Mining’ bibliometrics analysis: applying network analysis and cluster analysis. *Scientometrics*. 2015. Vol.105(1). P.323-335. DOI: 10.1007/s11192-015-1685-4.
- [26] **McLaren C.D., Bruner M.W.** Citation network analysis. *International Review of Sport and Exercise Psychology*. 2022. Vol.15(1). P.179-198. DOI: 10.1080/1750984X.2021.1989705.
- [27] **Yang S., Wang F.** Visualizing information science: Author direct citation analysis in China and around the world. *Journal of Informetrics*. 2015. Vol.9(1). P.208-225. DOI: 10.1016/j.joi.2015.01.001.
- [28] **Busygina T.V., Yuklyaevskaia A.V.** A Scientometric Analysis of the Literature on the Topic “Academic Social Networks”. *Bibliosphere*. 2022. 3. P.101-122. DOI: 10.20913/1815-3186-2022-3-101-122.
- [29] **Barabási A.L. et al.** Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*. 2002. Vol.311(3-4). P.590-614. DOI: 10.1016/S0378-4371(02)00736-7.
- [30] **Van Noorden R.** Online collaboration: Scientists and the social network. *Nature news*. 2014. Vol.512(7513). P.126-129. DOI: 10.1038/512126a.
- [31] **Valizadeh-Haghi S., Shahbodaghi A., Nasibi-Sis H.** ResearchGate social network: Opportunities and challenges. *Journal of Medical Library and Information Science*. 2021. Vol.2. P.1-6. DOI: 10.22037/jmlis.v2i.32545.
- [32] **Ebrahimzadeh S. et al.** Triggers and strategies related to the collaborative information-seeking behaviour of researchers in ResearchGate. *Online Information Review*. 2020. Vol.44(5). P.1077-1096. DOI: 10.1108/OIR-12-2019-0380.

- [33] *Yan W. et al.* How does scholarly use of academic social networking sites differ by academic discipline? A case study using ResearchGate. *Information Processing & Management*. 2021. Vol.58(1). P.102430. DOI: 10.1016/j.ipm.2020.102430.
- [34] *Janavi E., Nadi-Ravandi S., Batooli Z.* Impact of ResearchGate on Increasing Citations and Usage Counts of Hot Papers in Clinical Medicine Indexed in Web of Science. *Webology*. 2020. Vol.17(1). P.130-139. DOI: 10.14704/WEB/V17I1/a212.
- [35] *Dikshit A., Pradhan B., Santosh M.* Artificial neural networks in drought prediction in the 21st century – A scientometric analysis. *Applied Soft Computing*. 2022. Vol.114. P.108080. DOI: 10.1016/j.asoc.2021.108080.
- [36] *Nemati-Anaraki L., Razmgar M., Moradzadeh M.* Scientific impact of Iran University of Medical Sciences researchers in ResearchGate, Google Scholar, and Scopus: an altmetrics study. *Medical Journal of the Islamic Republic of Iran*. 2020. Vol.34. P.142. DOI: 10.47176/mjiri.34.142.
- [37] *Boudry C., Durand-Barthez M.* Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia. edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. *Plos one*. 2020. Vol.15(9). P.e0238583. DOI: 10.1371/journal.pone.0238583.
- [38] *Sánchez-Jiménez R. et al.* Analysis of the distribution of authorship by gender in scientific output: A global perspective. *Journal of Informetrics*. 2024. Vol.18(3). P.101556. DOI: 10.1016/j.joi.2024.101556.
- [39] *Larivière V., Ni C., Gingras Y., Cronin B., Sugimoto C.R.* Bibliometrics: Global gender disparities in science. *Nature*. 2013. Vol.504(7479). P.211-213. DOI: 10.1038/504211a.
- [40] *Bornmann L., Mutz R., Daniel H.D.* Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and technology*. 2008. Vol.59(5). P.830-837. DOI: 10.1002/asi.20806.
- [41] *Bornmann L., Daniel H.D.* What do citation counts measure? A review of studies on citing behavior. *Journal of documentation*. 2008. Vol.64(1). P.45-80. DOI: 10.1108/00220410810844150.
- [42] *Costas R., Zahedi Z., Wouters P.* Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*. 2015. Vol.66(10). P.2003-2019. 10.1002/asi.23309.
- [43] *Jalali Z.S., Introne J., Soundarajan S.* Social stratification in networks: insights from co-authorship networks. *Journal of the Royal Society Interface*. 2023. Vol.20(198). P.20220555. DOI: 10.1098/rsif.2022.0555.
- [44] *Xue H.* Analysis of Effects on Scientific Impact Indicators Based on Coevolution of Coauthorship and Citation Networks. *Information*. 2024. Vol.15(10). P.597. DOI: 10.3390/info15100597.
- [45] *Rivera R.G. et al.* Using scientometrics to mapping Latin American research networks in emerging fields: the field networking index //Scientometrics. 2024. Vol. 129(4). P. 2309-2335. DOI: 10.1007/s11192-024-04970-z.
- [46] *Waltman L.* A review of the literature on citation impact indicators. *Journal of informetrics*. 2016. Vol.10(2). P.365-391. DOI:10.1016/j.joi.2016.02.007.
- [47] *Mingers J., Leydesdorff L.* A review of theory and practice in scientometrics. *European journal of operational research*. 2015. Vol.246(1). P.1-19. DOI: 10.1016/j.ejor.2015.04.002.
- [48] *Гуськов А.Е., Шрайберг Я.Л.* Вызовы для развития научометрических исследований. *Научные и технические библиотеки*. 2023. №2. Р.37-58. DOI: 10.33186/1027-3689-2023-2-37-58.
- [49] *Гуськов А.Е., Селиванова И.В., Косяков Д.В.* Методика оценки результативности научных организаций. *Вестник Российской академии наук*. 2018. Т.88. №5. С.430-443. DOI: 10.7868/S0869587318050092.
- [50] *Гуськов А.Е.* Концептуальная модель системы научометрического мониторинга результативности научной деятельности. *Научно-техническая информация*. Серия 2: Информационные процессы и системы. 2022. № 12. С.14-22. DOI: 10.36535/0548-0027-2022-12-4.
- [51] *Бобров Л.К.* Управление знаниями в зеркале Российского индекса научного цитирования. *Вычислительные технологии*. 2021. №4(26). С.53-72. DOI: 10.25743/ICT.2021.26.4.006.
- [52] *Золотарев Д.В.* Использование результатов фракционного счета научных статей при принятии управленческих решений. *Наука. Инновации. Образование*. 2016. Т.11. №1. С.105-114.
- [53] *Лойко В.И., Луценко А.И., Орлов Е.В.* Современные подходы в научометрии: монография. Краснодар: КубГАУ, 2017. 532 с.
- [54] *Маршакова-Шайкевич И.В.* Научное сотрудничество России со странами ЕС: библиометрический анализ. *Вестник Российской академии наук*. 2010. Т.80. №2. С.124-130.
- [55] *Мохнатчева Ю.В., Цветкова В.А.* Россия в мировом массиве научных публикаций. *Вестник Российской академии наук*. 2019. Т.89. №8. С.820-830. DOI: 10.31857/S0869-5873898820-830.
- [56] *Гохберг Л.М., Сагиева Г.С.* Российская наука: библиометрические индикаторы. *Форсайт*. 2007. №1. С.44-53. DOI: 10.17323/1995-459X.2007.1.44.53.
- [57] *Печников А.А.* Math-Net.Ru как зеркало академического рейтинга журналов RSCI. *Информационное общество*. 2024. №1. С.116-125.
- [58] *Печников А.А.* Исследование тематических сообществ графа соавторства российских математиков. *Научные и технические библиотеки*. 2025. №1. С.33-55. DOI: 10.33186/1027-3689-2025-1-33-55.
- [59] *Писляков В.В.* Самоцитирование и его влияние на оценку научной деятельности: обзор литературы. Часть II. *Научные и технические библиотеки*. 2022. №3. С.85-104. DOI: 10.33186/1027-3689-2022-3-85-104.
- [60] *Мальцева Д.В., Ващенко В.А., Капустина Л.В.* Методология обработки библиографических данных на русском языке для построения сетей коллаборации (на примере базы данных eLibrary). *Социология*: 4М. 2022. №54–55. С.45-78. DOI: 10.19181/4m.2022.31.1-2.2.
- [61] *Новиков Д.А., Губко М.В.* Наукометрия и экспертиза в управлении наукой: предисловие. *Управление большими системами*. Специальный выпуск 44. *Наукометрия и экспертиза в управлении наукой*, 2013. С.8-13.

- [62] **Fronzetti Colladon A., D'Angelo C. A., Gloor P. A.** Predicting the future success of scientific publications through social network and semantic analysis. *Scientometrics*. 2020. Vol.124(1). P.357-377. DOI: 10.1007/s11192-020-03479-5.
- [63] **Mejia C. et al.** Exploring topics in bibliometric research through citation networks and semantic analysis. *Frontiers in Research Metrics and Analytics*. 2021. Vol.6. P.742311. DOI: 10.3389/frma.2021.742311.
- [64] **Kunnath S.N. et al.** A meta-analysis of semantic classification of citations. *Quantitative science studies*. 2021. Vol.2(4). P.1170-1215. DOI: 10.1162/qss\_a\_00159.
- [65] **Nair S.R. et al.** Clustering of research documents-a survey on semantic analysis and keyword extraction. *2021 6th International Conference for Convergence in Technology* (I2CT). IEEE, 2021. P.1-6. DOI: 10.1109/I2CT51068.2021.9418197.
- [66] **Осипов Г.С., Смирнов И.В.** Семантический анализ научных текстов и их больших массивов. *Системы высокой доступности*. 2016. Т.12(1). С.41-44.
- [67] **Ениколовов С.Н., Кузнецова Ю.М., Осипов Г.С., Смирнов И.В., Чудова Н.В.** Метод реляционно-ситуационного анализа текста в психологических исследованиях. *Психология*. Журнал Высшей школы экономики. 2021. Т.18(4). С.748-769. DOI: 10.17323/1813-8918-2021-4-748-769.
- [68] **Sochenkov I., Zubarev D., Tikhomirov I., Smirnov I., Shelmanov A., Suvorov R., Osipov G.** Exactus Like: Plagiarism Detection in Scientific Texts. In: Ferro, N., et al. *Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science*. Springer, Cham. Vol.9626. P.837-840. DOI: 10.1007/978-3-319-30671-1\_76.
- [69] **Osipov G., Smirnov I., Tikhomirov I., Sochenkov I., Shelmanov A.** Exactus expert—search and analytical engine for research and development support. *Novel Applications of Intelligent Systems*. Cham: Springer International Publishing, 2016. P.269-285. DOI: 10.1007/978-3-319-14194-7\_14.
- [70] **Efimenko I.V., Khoroshevsky V.F., Noyons E.C.M.** Anticipating Future Pathways of Science, Technologies, and Innovations:(Map of Science)<sup>2</sup> Approach //Anticipating Future Innovation Pathways Through Large Data Analysis. – Cham: Springer International Publishing, 2016. P.71-96. DOI: 10.1007/978-3-319-39056-7\_5.
- [71] **Хорошевский В.Ф., Ефименко И.В.** Семантическая технология картирования семантических технологий (наукометрический анализ конференций OSTIS). Минск: БГУИР. 2015. С.43-56.
- [72] **Герасименко Н.А., Ватолин А.С., Янина А.О., Воронцов К.В.** SciRus: легкий и мощный мультиязычный энкодер для научных текстов. *Доклады Российской академии наук. Математика, информатика, процессы управления*. 2024. Т.520. №2. С.216-227. DOI: 10.31857/S2686954324700589.
- [73] **Ватолин А.С., Герасименко Н.А., Янина А.О., Воронцов К.В.** RuSciBench: открытый бенчмарк для оценки семантических векторных представлений научных текстов на русском и английском языках. *Доклады Российской академии наук. Математика, информатика, процессы управления*. 2024. Т.520. №2. С.284-294. DOI: 10.31857/S2686954324700644.
- [74] **Chekhoverich Y.V., Khazov A.V.** Analysis of duplicated publications in Russian journals. *Journal of informetrics*. 2022. Vol.16. №1. 101246. DOI: 10.1016/j.joi.2021.101246.
- [75] **Gritsay G. M. et al.** Artificially generated text fragments search in academic documents. *Doklady Mathematics*. Moscow: Pleiades Publishing, 2023. Vol.108. DOI: 10.1134/S1064562423701211.
- [76] **Meuschke N. et al.** A benchmark of pdf information extraction tools using a multi-task and multi-domain evaluation framework for academic documents. *International Conference on Information*. Cham: Springer Nature Switzerland, 2023. P.383-405. DOI: 10.1007/978-3-031-28032-0\_31.
- [77] **Maddi A., Maisonne M., Boukacem-Zeghamouri C.** Geographical and disciplinary coverage of open access journals: Open-Alex, Scopus, and WoS. *PLoS One*. 2025. V.20. №4. DOI: 10.1371/journal.pone.0320347.
- [78] **Culbert J. H. et al.** Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. *Scientometrics*. 2025. Vol.130. №4. P.2475-2492. DOI: 10.1007/s11192-025-05293-3.
- [79] **Subramanian S. et al.** S2and: A benchmark and evaluation system for author name disambiguation. *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 2021. P.170-179. DOI: 10.1109/JCDL52503.2021.00029.
- [80] **Boccaletti S. et al.** The structure and dynamics of multilayer networks. *Physics reports*. 2014. Vol. 544. Issue 1. P.1-122. DOI: 10.1016/j.physrep.2014.07.001.
- [81] **Di Pierro D., Ferilli S., Redavid D.** LPG-based knowledge graphs: A survey, a proposal and current trends. *Information*. 2023. Vol.14. №3. DOI: 10.3390/info14030154.
- [82] **Rao S.X., Egger P.H., Zhang C.** Hierarchical classification of research fields in the" web of science" using deep learning //arXiv preprint arXiv:2302.00390. 2023. DOI: 10.48550/arXiv.2302.00390.
- [83] **Губанов Д.А., Новиков Д.А.** Анализ терминологической структуры теории управления. *Управление большими системами*. 2024. вып.110. С.181-210. DOI: 10.25728/ubs.2024.110.7.
- [84] **Blei D.M., Ng A.Y., Jordan M.I.** Latent Dirichlet allocation. *Journal of machine Learning research*. 2003. Vol.3. P. 993-1022.
- [85] **Blei D.M.** Probabilistic Topic Models. *Communications of the ACM*. 2012. Vol.55, №4. P.77-84. DOI: 10.1145/2133806.2133826.
- [86] **Yau C. K. et al.** Clustering scientific documents with topic modeling. *Scientometrics*. 2014. Vol.100. №3. P.767-786. DOI: 10.1007/s11192-014-1321-8.
- [87] **Blei D.M., Lafferty J.D.** Dynamic Topic Models. *Proceedings of the 23<sup>rd</sup> International Conference on Machine Learning*. 2006. P.113-120. DOI: 10.1145/1143844.1143859.
- [88] **Beltagy I., Lo K., Cohan A.** SciBERT: A pretrained language model for scientific text //arXiv preprint arXiv:1903.10676. 2019. DOI: 10.48550/arXiv.1903.10676.
- [89] **Grootendorst M.** BERTopic: Neural topic modeling with a class-based TF-IDF procedure //arXiv preprint arXiv:2203.05794. 2022. DOI: 10.48550/arXiv.2203.05794.

- [90] **Ebadi A., Auger A., Gauthier Y.** WISDOM: An AI-powered framework for emerging research detection using weak signal analysis and advanced topic modeling //arXiv preprint arXiv:2409.15340. 2024. DOI: 10.48550/arXiv.2409.15340.
- [91] **Rezapour R. et al.** Two-Stage Graph-Augmented Summarization of Scientific Documents. *Proceedings of the 1st Workshop on NLP for Science (NLP4Science)*. 2024. P.36-46. DOI: 10.18653/v1/2024.nlp4science-1.5.
- [92] **Kilicoglu H. et al.** Confirm or refute? A comparative study on citation sentiment classification in clinical research publications. *Journal of biomedical informatics*. 2019. Vol.91. 103123. DOI: 10.1016/j.jbi.2019.103123.
- [93] **Qi R. et al.** Multi-task learning model for citation intent classification in scientific publications. *Scientometrics*. 2023. Vol.128. №12. P.6335-6355. DOI: 10.1007/s11192-023-04858-4.
- [94] **Arnaout H. et al.** In-depth Research Impact Summarization through Fine-Grained Temporal Citation Analysis //arXiv preprint arXiv:2505.14838. 2025. DOI: 10.48550/arXiv.2505.14838.
- [95] **Li X. et al.** Evaluation of unsupervised static topic models' emergence detection ability. *PeerJ Computer Science*. 2025. Vol.11. DOI: 10.7717/peerj-cs.2875.
- [96] **Emsley R.** ChatGPT: these are not hallucinations—they're fabrications and falsifications. *Schizophrenia*. 2023. Vol.9. №1. DOI: 10.1038/s41537-023-00379-4.
- [97] **Vivek N. et al.** Social network analysis as a new tool to measure academic impact of physicians. *Laryngoscope Investigative Otolaryngology*. 2025. Vol.10. №1. DOI: 10.1002/lio2.70060.
- [98] **Fortunato S.** Community Detection in Graphs. *Physics Reports*. 2010. Vol.486, №3-5. P.75-174. DOI: 10.1016/j.physrep.2009.11.002.
- [99] **Lü L., Zhou T.** Link Prediction in Complex Networks: A Survey // *Physica A*. 2011. Vol. 390, № 6. P. 1150-1170. DOI: 10.1016/j.physa.2010.11.027.
- [100] **Holme P., Saramäki J.** Temporal Networks. *Physics Reports*. 2012. Vol.519. P.97-125. DOI: 10.1016/j.physrep.2012.03.001.
- [101] **Chebotarev P.Yu, Gubanov D.A.** How to Choose the Most Appropriate Centrality Measure? A Decision-Tree Approach. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. Piscataway, N.J., United States: IEEE, 2024. DOI: 10.1109/TSMC.2024.3510633.
- [102] **Freeman L.C.** Centrality in Social Networks: Conceptual Clarification. *Social Networks*. 1978. Vol.1. P.215-239. DOI: 10.1016/0378-8733(78)90021-7.
- [103] **Chkhartishvili A.G., Gubanov D.A.** Influence Levels of Users and Meta-Users of a Social Network. *Automation and Remote Control*. 2018. Vol.79, Iss.3. P.545–553. DOI: 10.1134/S0005117918030128.
- [104] **Reia S.M., Silva F.N., de Arruda H.F.** Science of Science: a complex network perspective. *Frontiers in Research Metrics and Analytics*. 2025. Vol.10. DOI: 10.3389/frma.2025.1595966.
- [105] **Miyauchi A., Kawase Y.** Z-score-based modularity for community detection in networks. *PloS one*. 2016. Vol.11. №1. DOI: 10.1371/journal.pone.0147805.
- [106] **Traag V.A., Waltman L., van Eck N.J.** From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*. 2019. Vol.9. 5233. DOI: 10.1038/s41598-019-41695-z.
- [107] **Губанов Д.А., Мельничук В.С.** Построение профилей научных публикаций на основе текстов и связей соавторства (на примере теории управления и ее приложений). *Проблемы управления*. 2025. №1. С.46-52.
- 

## Сведения об авторах



**Губанов Дмитрий Алексеевич**, 1984 г. рождения. Окончил Чувашский государственный университет в 2006 г., д.т.н. (2021). Ведущий научный сотрудник ИПУ РАН. В списке научных трудов 140 работ в области информационного влияния и управления в социальных сетях и анализа сложных сетей. ORCID: 0000-0002-0099-3386; Author ID (РИНЦ): 554667; Researcher ID (WoS): N-6957-2017; Author ID (Scopus): 51261227400. *gubanov@ipu.ru*.



**Жилякова Людмила Юрьевна**, 1970 г. рождения. Окончила Ростовский государственный университет в 1992 г., д.ф.-м..н. (2013). Ведущий научный сотрудник ИПУ РАН. В списке научных трудов более 150 работ в области динамических сетевых моделей, анализа взаимодействия гетерогенных агентов. ORCID: 0000-0002-8987-3777; Author ID (РИНЦ): 175735; Researcher ID (WoS): Q-6865-2016; Author ID (Scopus): 6508231167. *zhilyakova@ipu.ru*.



**Кузнецов Олег Петрович**, 1936 г. рождения. Окончил МГУ им. М.В. Ломоносова, философский факультет (1958), механико-математический факультет (1966). Доктор технических наук, профессор. Главный научный сотрудник ИПУ РАН. Автор более 170 публикаций, в том числе 4 монографий. AuthorID (РИНЦ): 24; Scopus Author ID: 36966033000; WoS ResearcherID J-5275-2018; ORCID 0000-0002-5061-3855. *olpkuz@yandex.ru*.



**Чархишвили Александр Гедеванович**, 1970 г. рождения. Окончил МГУ им. М.В. Ломоносова в 1991 г., д.ф.-м.н. (2006). Главный научный сотрудник ИПУ РАН. В списке научных трудов более 150 работ в области теории управления организационными системами, моделирования

информированности и рефлексии, моделирования информационного управления и противоборства, теории игр, социально-сетевого анализа. ORCID: 0000-0002-2970-1244; Author ID (РИНЦ): 12680; Researcher ID (WoS): M-7933-2013; Author ID (Scopus): 6506434827. *sandro\_ch@mail.ru*.

Поступила в редакцию 17.09.2025, после рецензирования 16.10.2025. Принята к публикации 20.10.2025.



Scientific article

DOI: 10.18287/2223-9537-2025-15-4-578-597

## Methods and models for the analysis of scientific activity

© 2025, D.A. Gubanov, L.Yu. Zhilyakova✉, O.P. Kuznetsov, A.G. Chkhartishvili

V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

### Abstract

This article describes methods for analyzing scientific activity and the tools developed on their basis to support informed decision-making at all levels of management in the scientific domain — from individual researchers to the heads of research organizations. A review of domestic and international studies in the field of scientometrics is provided, highlighting key trends and identifying existing gaps that the Institute of Control Sciences of the Russian Academy of Sciences (ICS RAS) project seeks to address through the creation of an information system for the analysis of scientific activity. The project involves the development of models, methods, and algorithms for analyzing the scientific activity of researchers, organizations, journals, and conferences in the field of control theory, using publication data. This problem is approached through an interdisciplinary framework that integrates network analysis, ontological design, big data processing, machine learning techniques, and optimization methods. The proposed system is built upon an ontology of scientific knowledge in control theory, the main principles of its construction and structure being discussed in this article. Combined with the system's technological capabilities, the presented ontology enables a level of analytical detail that surpasses existing scientometric systems (such as Web of Science, Scopus, Russian Science Citation Index, etc.). It is demonstrated that the principles underlying the development and structure of the proposed ontology can be applied to other research domains. The paper also considers modern approaches to data collection and visualization, as well as methods of network and content analysis of scientific texts. The main concepts, methods, and models for the analysis of scientific activity are summarized, and their implementation within the developed information system is discussed.

**Keywords:** analysis of scientific activity, control theory, information system, classification, ontology, thematic profile, thematic space, term, text processing, network analysis.

**For citation:** Gubanov DA, Zhilyakova LYu, Kuznetsov OP, Chkhartishvili AG. Methods and models for the analysis of scientific activity [In Russian]. *Ontology of designing*. 2025; 15(4): 578-597. DOI: 10.18287/2223-9537-2025-15-4-578-597.

**Authors' contributions:** Gubanov DA – section 3, list of references; Zhilyakova LYu – article structure, introduction, section 1, list of references, conclusion; Kuznetsov OP – introduction, section 2; Chkhartishvili AG – section 4, article editing.

**Conflict of interest:** The authors declare no conflict of interest.

### List of figures

Figure 1 – Fragment of the ontology of the thematic space (rectangles represent topics and subtopics, leaves represent terms)

### References

- [1] Kosyakov D, Pislyakov V. "I'd like to publish in Q1, but there's no Q1 to be found": Study of journal quartile distributions across subject categories and topics. *Journal of Informetrics*. 2024; 18(1): 101494. DOI: 10.1016/j.joi.2024.101494.

- [2] GOST 7.90 2007. System of standards on information, librarianship and publishing. Universal decimal classification. Structure, rules of introduction and indexing: official publication. [In Russian].
- [3] Mathematics Subject Classification (MSC2020). <https://mathscinet.ams.org/mathscinet/msc/msc2020.html>.
- [4] Physics and Astronomy Classification Scheme. [https://publishing.aip.org/wp-content/uploads/2019/01/PACS\\_2010\\_Alpha.pdf](https://publishing.aip.org/wp-content/uploads/2019/01/PACS_2010_Alpha.pdf).
- [5] International Classification Codes (OECD). [https://storage.tusur.ru/files/134958/kody\\_OECD.pdf](https://storage.tusur.ru/files/134958/kody_OECD.pdf).
- [6] **Gubanov DA, Kuznetsov OP, Kurako EA. et al.** ISASA: An Information System for the Analysis of Scientific Activity in the Field of Control Theory and Its Applications [In Russian]. *Control Sciences*, 2024; 3: 35–55. DOI: 10.25728/cs.2024.3.4.
- [7] **Kuznetsov OP, Sukhoverov VS.** Ontological approach to determining the subject matter of scientific text [In Russian]. *Ontology of designing*, 2016; 6(1): 55-66. DOI: 10.18287/2223-9537-2016-6-1-55-66.
- [8] **Agaev RP, Aleskerov FT, Alchinov AI. et al.** Control Theory: Dictionary of the System of Basic Concepts [In Russian] Moscow: LENAND, 2024. 128 p.
- [9] State index of scientific and technical information [In Russian]. <https://www.gpntb.ru/grnti.html>.
- [10] **Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, ... & Barabási AL.** Science of science. *Science*. 2018; 359(6379): eaao0185. DOI: 10.1126/science.aao0185.
- [11] **Gates AJ, Barabási AL.** Reproducible science of science at scale: pySciSci. *Quantitative Science Studies*. 2023; 4(3): 700-710. DOI: 10.1162/qss\_a\_00260.
- [12] **Nalimov VV, Mulchenko ZM.** Scientometrics. Study of the development of science as an information process [In Russian]. Moscow: Nauka, 1969. 192 p.
- [13] **Granovsky YV.** Is It Possible to Measure Science? V.V. Nalimov's Research in Scientometrics. *Scientometrics*. 2001; 52: 127–150. DOI: 10.1023/A:1017991017982.
- [14] **Barabási AL, Albert R.** Emergence of scaling in random networks. *Science*. 1999; 286(5439): 509-512. DOI: 10.1126/science.286.5439.509.
- [15] **Watts DJ, Strogatz SH.** Collective dynamics of 'small-world' networks. *Nature*. 1998; 393(6684): 440-442. DOI: 10.1038/30918.
- [16] **Girvan M, Newman MEJ.** Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002; 99(12): 7821-7826. DOI: 10.1073/pnas.122653799.
- [17] **Molontay R, Nagy M.** Twenty years of network science: A bibliographic and co-authorship network analysis //Big data and social media analytics: trending applications. Cham: Springer International Publishing, 2021. P.1-24. DOI: 10.1007/978-3-030-67044-3\_1.
- [18] **Costa LDF, Rodrigues FA, Travieso G, Villas Boas PR.** Characterization of complex networks: A survey of measurements. *Advances in physics*. 2007; 56(1): 167-242. DOI: 10.1080/00018730601170527.
- [19] **Saxena A, Iyengar S.** Centrality measures in complex networks: A survey //arXiv preprint arXiv:2011.07190. 2020. DOI: 10.48550/arXiv.2011.07190.
- [20] **Dias A, Rutes S, Lima L, Campra E, Silva M, Bragança de Sousa M, Porto G.** Network centrality analysis in management and accounting sciences. *RAUSP Management Journal*. 2020; 55: 207-226. DOI: 10.1108/RAUSP-02-2019-0021.
- [21] **D'Ippoliti C.** "Many-Citedness": Citations Measure More Than Just Scientific Quality. *Journal of Economic Surveys*. 2021; 35(5): 1271-1301. DOI: 10.1111/joes.12416.
- [22] **Lancichinetti A, Fortunato S.** Consensus clustering in complex networks. *Scientific reports*. 2012; 2(1): 336. DOI: 10.1038/srep00336.
- [23] **Ozcan S, Boye D, Arsenyan J, Trott P.** A scientometric exploration of crowdsourcing: Research clusters and applications. *IEEE Transactions on Engineering Management*. 2020; 69(6): 3023-3037. DOI: 10.1109/TEM.2020.3027973.
- [24] **Lund B, Ma J.** A review of cluster analysis techniques and their uses in library and information science research: *k*-means and *k*-medoids clustering. *Performance Measurement and Metrics*. 2021; 22(3): 161-173. DOI: 10.1108/PMM-05-2021-0026.
- [25] **Madani F.** 'Technology Mining' bibliometrics analysis: applying network analysis and cluster analysis. *Scientometrics*. 2015; 105(1): 323-335. DOI: 10.1007/s11192-015-1685-4.
- [26] **McLaren CD, Bruner MW.** Citation network analysis. *International Review of Sport and Exercise Psychology*. 2022; 15(1): 179–198. DOI: 10.1080/1750984X.2021.1989705.
- [27] **Yang S, Wang F.** Visualizing information science: Author direct citation analysis in China and around the world. *Journal of Informetrics*. 2015; 9(1): 208-225. DOI: 10.1016/j.joi.2015.01.001.
- [28] **Busygina TV, Yuklyavskaya AV.** A Scientometric Analysis of the Literature on the Topic "Academic Social Networks". *Bibliosphere*. 2022; 3: 101–122. DOI: 10.20913/1815-3186-2022-3-101-122.
- [29] **Barabási AL. et al.** Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*. 2002; 311(3-4): 590-614. DOI: 10.1016/S0378-4371(02)00736-7.
- [30] **Van Noorden R.** Online collaboration: Scientists and the social network. *Nature news*. 2014; 512(7513): 126-129. DOI: 10.1038/512126a.
- [31] **Valizadeh-Haghi S, Shahbodaghi A, Nasibi-Sis H.** ResearchGate social network: Opportunities and challenges. *Journal of Medical Library and Information Science*. 2021; 2: 1-6. DOI: 10.22037/jmlis.v2i.32545.
- [32] **Ebrahimzadeh S. et al.** Triggers and strategies related to the collaborative information-seeking behaviour of researchers in ResearchGate. *Online Information Review*. 2020; 44(5): 1077-1096. DOI: 10.1108/OIR-12-2019-0380.
- [33] **Yan W. et al.** How does scholarly use of academic social networking sites differ by academic discipline? A case study using ResearchGate. *Information Processing & Management*. 2021; 58(1): 102430. DOI: 10.1016/j.ipm.2020.102430.
- [34] **Janavi E, Nadi-Ravandi S, Batooli Z.** Impact of ResearchGate on Increasing Citations and Usage Counts of Hot Papers in Clinical Medicine Indexed in Web of Science. *Webology*. 2020; 17(1): 130-139. DOI: 10.14704/WEB/V17I1/a212.

- [35] **Dikshit A, Pradhan B, Santosh M.** Artificial neural networks in drought prediction in the 21st century – A scientometric analysis. *Applied Soft Computing*. 2022; 114: 108080. DOI: 10.1016/j.asoc.2021.108080.
- [36] **Nemati-Anaraki L, Razmgar M, Moradzadeh M.** Scientific impact of Iran University of Medical Sciences researchers in ResearchGate, Google Scholar, and Scopus: an altmetrics study. *Medical Journal of the Islamic Republic of Iran*. 2020; 34: 142. DOI: 10.47176/mjiri.34.142.
- [37] **Boudry C, Durand-Barthez M.** Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia. edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. *Plos one*. 2020; 15(9): e0238583. DOI: 10.1371/journal.pone.0238583.
- [38] **Sánchez-Jiménez R, et al.** Analysis of the distribution of authorship by gender in scientific output: A global perspective. *Journal of Informetrics*. 2024; 18(3): 101556. DOI: 10.1016/j.joi.2024.101556.
- [39] **Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR.** Bibliometrics: Global gender disparities in science. *Nature*. 2013; 504(7479): 211-213. DOI: 10.1038/504211a.
- [40] **Bornmann L, Mutz R, Daniel HD.** Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and technology*. 2008; 59(5): 830-837. DOI: 10.1002/asi.20806
- [41] **Bornmann L, Daniel HD.** What do citation counts measure? A review of studies on citing behavior. *Journal of documentation*. 2008; 64(1): 45-80. DOI: 10.1108/00220410810844150
- [42] **Costas R, Zahedi Z, Wouters P.** Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*. 2015; 66(10): 2003-2019. DOI: 10.1002/asi.23309.
- [43] **Jalali ZS, Introne J, Soundarajan S.** Social stratification in networks: insights from co-authorship networks. *Journal of the Royal Society Interface*. 2023; 20(198): 20220555. DOI: 10.1098/rsif.2022.0555.
- [44] **Xue H.** Analysis of Effects on Scientific Impact Indicators Based on Coevolution of Coauthorship and Citation Networks. *Information*. 2024; 15(10): 597. DOI: 10.3390/info15100597.
- [45] **Rivera RG, et al.** Using scientometrics to mapping Latin American research networks in emerging fields: the field networking index. *Scientometrics*. 2024; 129(4): 2309-2335. DOI: 10.1007/s11192-024-04970-z.
- [46] **Waltman L.** A review of the literature on citation impact indicators. *Journal of informetrics*. 2016; 10(2): 365-391. DOI: 10.1016/j.joi.2016.02.007.
- [47] **Mingers J, Leydesdorff L.** A review of theory and practice in scientometrics. *European journal of operational research*. 2015; 246(1): 1-19. DOI: 10.1016/j.ejor.2015.04.002.
- [48] **Guskov AE, Shrayberg YaL.** Challenges to develop scientometric studies [In Russian]. *Scientific and Technical Libraries*. 2023;(2):37-58. DOI: 10.33186/1027-3689-2023-2-37-58.
- [49] **Guskov AE, Selivanova IV, Kosyakov DV.** Methodology for assessing the performance of scientific organizations [In Russian]. *Bulletin of the Russian Academy of Sciences*. 2018; 88(5): 430-443. DOI: 10.7868/S0869587318050092.
- [50] **Guskov AE.** Conceptual model of the system of scientometric monitoring of the effectiveness of scientific activity [In Russian]. *Scientific and technical information. Series 2: Information processes and systems*. 2022; 12: 14-22. DOI: 10.36535/0548-0027-2022-12-4.
- [51] **Bobrov LK.** Knowledge management reflected in the Russian Science Citation Index [In Russian]. *Computational Technologies*. 2021; 4(26): 53-72. DOI: 10.25743/ICT.2021.26.4.006.
- [52] **Zolotarev DV.** Using the results of fractional counting of scientific articles in making management decisions [In Russian]. *Science. Innovations. Education*. 2016; 11(1): 105-114.
- [53] **Loiko VI, Lutsenko AI, Orlov EV.** Modern approaches in scientometrics: monograph [In Russian]. Krasnodar: KubSAU, 2017. 532 p.
- [54] **Marshakova-Shaikovich IV.** Scientific cooperation between Russia and EU countries: a bibliometric analysis [In Russian]. *Bulletin of the Russian Academy of Sciences*. 2010; 80(2): 124-130.
- [55] **Mokhnacheva YuV, Tsvetkova VA.** Russia in the global array of scientific publications [In Russian]. *Bulletin of the Russian Academy of Sciences*. 2019; 89(8): 820-830. DOI: 10.31857/S0869-5873898820-830.
- [56] **Gokhberg L, Sagieva G.** Russian Science: Bibliometric Indicators [In Russian]. *Foresight and STI Governance*. 2007; 1(1): 44-53. DOI: 10.17323/1995-459X.2007.1.44.53.
- [57] **Pechnikov AA.** Math-Net.Ru as a mirror of the academic ranking of RSCI journals [In Russian]. *Information Society*. 2024; 1: 116-125.
- [58] **Pechnikov AA.** The study of thematic communities within the Russian mathematicians co-authorship graphs [In Russian]. *Scientific and Technical Libraries*. 2025; (1): 33-55. DOI: 10.33186/1027-3689-2025-1-33-55.
- [59] **Pislyakov VV.** Self-citation and its impact on research evaluation: Literature review. Part II. *Scientific and Technical Libraries* [In Russian]. 2022;(3):85-104. DOI: 10.33186/1027-3689-2022-3-85-104.
- [60] **Maltseva DV, Vaschenko VA, Kapustina LV.** Methodology for processing Russian-language bibliographic data for building collaboration networks (based on the eLibrary database) [In Russian]. *Sociology*: 4M. 2022; 54-55: 45-78. DOI: 10.19181/4m.2022.31.1-2.2.
- [61] **Novikov DA, Gubko MV.** Scientometrics and Expertise in Science Management: Preface [In Russian]. *Management of Large Systems. Special Issue 44. Scientometrics and Expertise in Science Management*, 2013. P.8-13.
- [62] **Fronzetti Colladon A, D'Angelo CA, Gloor PA.** Predicting the future success of scientific publications through social network and semantic analysis. *Scientometrics*. 2020; 124(1): 357-377. DOI: 10.1007/s11192-020-03479-5.
- [63] **Mejia C, et al.** Exploring topics in bibliometric research through citation networks and semantic analysis. *Frontiers in Research Metrics and Analytics*. 2021; 6: 742311. DOI: 10.3389/frma.2021.742311.

- [64] **Kunnath SN, et al.** A meta-analysis of semantic classification of citations. *Quantitative science studies*. 2021; 2(4): 1170-1215. DOI: 10.1162/qss\_a\_00159.
- [65] **Nair SR, et al.** Clustering of research documents-a survey on semantic analysis and keyword extraction. *2021 6th International Conference for Convergence in Technology* (I2CT). IEEE, 2021. P.1-6. DOI: 10.1109/I2CT51068.2021.9418197.
- [66] **Osipov GS, Smirnov IV.** Semantic analysis of scientific texts and their large arrays [In Russian]. *Highly available systems*. 2016; 12(1): 41-44.
- [67] **Yenikolopov SN, Kuznetsova YuM, Osipov GS, Smirnov IV, Chudova NV.** The method of relational-situational text analysis in psychological research [In Russian]. *Psychology. Journal of the Higher School of Economics*. 2021; 18(4): 748-769. DOI: 10.17323/1813-8918-2021-4-748-769.
- [68] **Sochenkov I, Zubarev D, Tikhomirov I, Smirnov I, Shelmanov A, Suvorov R, & Osipov G.** Exactus Like: Plagiarism Detection in Scientific Texts. In: Ferro, N., et al. *Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science*. Springer, Cham. Vol. 9626. P. 837-840 DOI: 10.1007/978-3-319-30671-1\_76.
- [69] **Osipov G, Smirnov I, Tikhomirov I, Sochenkov I, Shelmanov A.** Exactus expert—search and analytical engine for research and development support. *Novel Applications of Intelligent Systems*. Cham: Springer International Publishing, 2016. P.269-285. DOI: 10.1007/978-3-319-14194-7\_14.
- [70] **Efimenko IV, Khoroshevsky VF, Noyons ECM.** Anticipating Future Pathways of Science, Technologies, and Innovations:(Map of Science)<sup>2</sup> Approach. *Anticipating Future Innovation Pathways Through Large Data Analysis*. Cham: Springer International Publishing, 2016. P.71-96. DOI: 10.1007/978-3-319-39056-7\_5.
- [71] **Khoroshevsky VF, Efimenko IV.** Semantic technology of mapping semantic technologies (scientometric analysis of OSTIS conferences) [In Russian]. Minsk: BSUIR. 2015. P.43-56.
- [72] **Gerasimenko N, Vatolin A, Ianina A, et al.** SciRus: Simple and Powerful Multilingual Encoder for Scientific Texts [In Russian]. *Dokl. Math.* 110 (Suppl 1), P.S193–S202. (2024). DOI: 10.1134/S1064562424602178
- [73] **Vatolin A, Gerasimenko N, Ianina A, et al.** RuSciBench: An Open Benchmark for Evaluating Semantic Vector Representations of Scientific texts in Russian and English [In Russian]. *Dokl. Math.* 110 (Suppl 1), P.S251–S260 (2024). DOI: 10.1134/S1064562424602191.
- [74] **Chekhovich YV, Khazov AV.** Analysis of duplicated publications in Russian journals. *Journal of informetrics*. 2022; 16(1): 101246. DOI: 10.1016/j.joi.2021.101246.
- [75] **Gritsay GM, et al.** Artificially generated text fragments search in academic documents. *Doklady Mathematics*. Moscow: Pleiades Publishing, 2023. Vol. 108. DOI: 10.1134/S1064562423701211.
- [76] **Meuschke N, et al.** A benchmark of pdf information extraction tools using a multi-task and multi-domain evaluation framework for academic documents. *International Conference on Information*. Cham: Springer Nature Switzerland, 2023. P.383-405. DOI: 10.1007/978-3-031-28032-0\_31.
- [77] **Maddi A, Maisonneuve M, Boukacem-Zeghamoui C.** Geographical and disciplinary coverage of open access journals: Open-Alex, Scopus, and WoS. *PLoS One*. 2025; 20(4). DOI: 10.1371/journal.pone.0320347.
- [78] **Culbert JH, et al.** Reference coverage analysis of OpenAlex compared to Web of Science and Scopus //Scientometrics. – 2025. – Vol. 130. – №. 4. – P. 2475-2492. DOI: 10.1007/s11192-025-05293-3
- [79] **Subramanian S, et al.** S2and: A benchmark and evaluation system for author name disambiguation. *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 2021. P.170-179. DOI: 10.1109/JCDL52503.2021.00029.
- [80] **Boccaletti S, et al.** The structure and dynamics of multilayer networks. *Physics reports*. 2014; 544(1): 1-122. DOI: 10.1016/j.physrep.2014.07.001.
- [81] **Di Pierro D, Ferilli S, Redavid D.** LPG-based knowledge graphs: A survey, a proposal and current trends. *Information*. 2023; 14(3). DOI: 10.3390/info14030154.
- [82] **Rao SX, Egger PH, Zhang C.** Hierarchical classification of research fields in the" web of science" using deep learning //arXiv preprint arXiv:2302.00390. – 2023. DOI: 10.48550/arXiv.2302.00390.
- [83] **Gubanov DA, Novikov DA.** Analysis of the terminological structure of control theory [In Russian]. *UBS*. 2024; 110: 181-210. DOI: 10.25728/ubs.2024.110.7.
- [84] **Blei DM, Ng AY, Jordan MI.** Latent Dirichlet allocation. *Journal of machine Learning research*. 2003; 3: 993-1022.
- [85] **Blei DM.** Probabilistic Topic Models. *Communications of the ACM*. 2012. Vol. 55(4). P. 77-84. DOI: 10.1145/2133806.2133826.
- [86] **Yau C. K, et al.** Clustering scientific documents with topic modeling. *Scientometrics*. 2014; 100(3): 767-786. DOI: 10.1007/s11192-014-1321-8.
- [87] **Blei DM, Lafferty JD.** Dynamic Topic Models. *Proceedings of the 23<sup>rd</sup> International Conference on Machine Learning*. 2006. P.113-120. DOI: 10.1145/1143844.1143859.
- [88] **Beltagy I, Lo K, Cohan A.** SciBERT: A pretrained language model for scientific text //arXiv preprint arXiv:1903.10676. – 2019. DOI: 10.48550/arXiv.1903.10676.
- [89] **Grootendorst M.** BERTopic: Neural topic modeling with a class-based TF-IDF procedure //arXiv preprint arXiv:2203.05794. – 2022. DOI: 10.48550/arXiv.2203.05794.
- [90] **Ebadi A, Auger A, Gauthier Y.** WISDOM: An AI-powered framework for emerging research detection using weak signal analysis and advanced topic modeling //arXiv preprint arXiv:2409.15340. 2024. DOI: 10.48550/arXiv.2409.15340.
- [91] **Rezapour R, et al.** Two-Stage Graph-Augmented Summarization of Scientific Documents. *Proceedings of the 1st Workshop on NLP for Science (NLP4Science)*. 2024. P.36-46. DOI: 10.18653/v1/2024.nlp4science-1.5.
- [92] **Kilicoglu H, et al.** Confirm or refute? A comparative study on citation sentiment classification in clinical research publications. *Journal of biomedical informatics*. 2019; 91: 103123. DOI: 10.1016/j.jbi.2019.103123.

- [93] **Qi R. et al.** Multi-task learning model for citation intent classification in scientific publications. *Scientometrics*. 2023; 128(12): 6335-6355. DOI: 10.1007/s11192-023-04858-4.
- [94] **Arnaout H. et al.** In-depth Research Impact Summarization through Fine-Grained Temporal Citation Analysis //arXiv preprint arXiv:2505.14838. 2025. DOI: 10.48550/arXiv.2505.14838.
- [95] **Li X. et al.** Evaluation of unsupervised static topic models' emergence detection ability. *PeerJ Computer Science*. 2025; 11. DOI: 10.7717/peerj-cs.2875.
- [96] **Emsley R.** ChatGPT: these are not hallucinations—they're fabrications and falsifications. *Schizophrenia*. 2023; 9(1). DOI: 10.1038/s41537-023-00379-4.
- [97] **Vivek N. et al.** Social network analysis as a new tool to measure academic impact of physicians. *Laryngoscope Investigative Otolaryngology*. 2025; 10(1). DOI: 10.1002/lio2.70060.
- [98] **Fortunato S.** Community Detection in Graphs. *Physics Reports*. 2010; 486(3-5): 75-174. DOI: 10.1016/j.physrep.2009.11.002.
- [99] **Lü L., Zhou T.** Link Prediction in Complex Networks: A Survey. *Physica A*. 2011; 390(6): 1150-1170. DOI: 10.1016/j.physa.2010.11.027.
- [100] **Holme P., Saramäki J.** Temporal Networks. *Physics Reports*. 2012; 519: 97-125. DOI: 10.1016/j.physrep.2012.03.001.
- [101] **Chebotarev PYu, Gubanov DA.** How to Choose the Most Appropriate Centrality Measure? A Decision-Tree Approach. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. Piscataway, N.J., United States: IEEE, 2024. DOI: 10.1109/TSMC.2024.3510633.
- [102] **Freeman LC.** Centrality in Social Networks: Conceptual Clarification. *Social Networks*. 1978; 1: 215-239. DOI: 10.1016/0378-8733(78)90021-7.
- [103] **Chkhartishvili AG, Gubanov DA.** Influence Levels of Users and Meta-Users of a Social Network. *Automation and Remote Control*. 2018; 79(3): 545-553. DOI: 10.1134/S0005117918030128.
- [104] **Reia SM, Silva FN, de Arruda HF.** Science of Science: a complex network perspective. *Frontiers in Research Metrics and Analytics*. 2025; 10. DOI: 10.3389/frma.2025.1595966.
- [105] **Miyauchi A, Kawase Y.** Z-score-based modularity for community detection in networks. *PloS one*. 2016; 11(1). DOI: 10.1371/journal.pone.0147805.
- [106] **Traag VA, Waltman L, van Eck NJ.** From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*. 2019; 9: 5233. DOI: 10.1038/s41598-019-41695-z.
- [107] **Gubanov DA, Melnicuk VS.** Constructing Scientific Publication Profiles Based on Texts and Coauthorship Connections (A Case Study in Control Theory and Its Applications) [In Russian]. *Control Sciences*, 2025; 1: 39-44.

## About the authors

**Dmitry Alekseevich Gubanov**, (b. 1984) graduated from Chuvash State University in 2006, Dr. Sci. (2021). Leading Researcher at the V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences. He is a co-author of more than 140 publications in the field of information influence and control in social networks, and the analysis of complex networks. ORCID: 0000-0002-0099-3386; Author ID (RSCI): 554667; Researcher ID (WoS): N-6957-2017; Author ID (Scopus): 51261227400. *gubanov@ipu.ru*.

**Liudmila Yurjevna Zhilyakova**, (b. 1970) graduated from Rostov State University in 1992, Dr. Phys.-Math. Sci. (2013). Leading Researcher at the V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences. She is the author and a co-author of more than 150 publications in the field of dynamic network models, and analysis of interactions of heterogeneous agents. ORCID: 0000-0002-8987-3777; Author ID (RSCI): 175735; Researcher ID (WoS): Q-6865-2016; Author ID (Scopus): 6508231167. *zhilyakova@ipu.ru*.

**Oleg Petrovich Kuznetsov** (b.1936) graduated from Lomonosov Moscow State University, Faculty of Philosophy (1958), Faculty of Mechanics and Mathematics (1966). He is a Doctor of Technical Sciences, a professor, and a Chief Researcher at the Institute of Control Sciences of the Russian Academy of Sciences. He is the author of more than 170 publications, including 4 monographs. Author ID (RSCI): 24; Author ID (Scopus): 36966033000; Researcher ID (WoS) J-5275-2018; ORCID 0000-0002-5061-3855. *olpkuz@yandex.ru*.

**Alexander Gedevanovich Chkhartishvili** (b. 1970) graduated from Lomonosov Moscow State University in 1991, Dr. Phys.-Math. Sci. (2006). Chief Researcher at the V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences. He is a co-author of more than 150 publications in the field of theory of control in organizations, modeling of awareness and reflection, modeling of information control and confrontation, game theory, and social network analysis. ORCID: 0000-0002-2970-1244; Author ID (RSCI): 12680; ResearcherID (WoS): M-7933-2013; Author ID (Scopus): 6506434827. *sandro\_ch@mail.ru*.

Received September 17, 2025. Revised October 16, 2025. Accepted October 20, 2025.