

УДК 001.89+004.912

НАУЧНЫЕ ПУБЛИКАЦИИ И ЭМБЕДДИНГ-ПРОСТРАНСТВО ЗНАНИЙ

А. Х. Мариносян¹ [0000-0003-0577-2360], С. Г. Григорьев² [0000-0002-0034-9224]

^{1, 2}Московский городской педагогический университет, г. Москва, Россия

¹a.marinosyan@yandex.ru, ²grigorsg@yandex.ru

Аннотация

Рассмотрены актуальные проблемы наукометрии, возникающие на фоне роста публикационной активности и широкого внедрения технологий генеративного искусственного интеллекта. Проанализирован существующий наукометрический инструментарий анализа научной деятельности, подразделяемый на количественные метрики и методы картографирования науки (анализ сетей цитирований, академическая генеалогия, семантический анализ и др.). Сделана попытка преодоления ограничений традиционного цитатного анализа, таких как «семантическая слепота» и уязвимость к манипуляциям. В качестве возможного решения предложена концептуальная модель, в которой единицей анализа выступает не публикация в целом, а отдельное «ключевое утверждение». Такой подход предполагает фиксацию не только содержания тезиса, но и его типа, области релевантности и характера связей с другими утверждениями (подтверждение, опровержение, уточнение, обобщение и т. д.). В контексте данного подхода предложены принципы расчета модифицированных наукометрических метрик.

Представлены результаты апробации предложенной модели на массиве из 728 статей журнала «Информатика и образование» (2016–2025 гг.). С использованием больших языковых моделей проведен анализ, результаты которого показали, что ретроспективное извлечение утверждений сталкивается с трудностями, связанными со сложившейся культурой научной коммуникации; поэтому отмечено преимущество самостоятельного формулирования ключевых утверждений (как особого типа метаданных) непосредственно авторами публикаций. Намечены возможные пути развития концепции «эмбеддинг-пространства знаний», которая в перспективе могла бы дополнить существующие подходы анализа динамики развития научных идей и теорий.

Ключевые слова: наукометрия, библиометрия, академическая генеалогия, цитатный анализ, семантический анализ, большие языковые модели, карта науки, индекс Хирша, нанопубликации.

ПОСТАНОВКА ПРОБЛЕМЫ

Современная наука переживает период роста публикационной активности, сопровождающийся качественными трансформациями в самом процессе производства научного знания. Распространение генеративного искусственного интеллекта создает дополнительное давление на традиционные механизмы научной коммуникации: большие языковые модели способны генерировать связанные научные тексты, автоматизировать поиск литературы, резюмировать источники и т. д. Это приводит к увеличению количества публикаций, но далеко не всегда к улучшению их качества. В этих условиях растет потребность в развитии наукометрических методов, способных обеспечить ориентацию в растущем массиве научных публикаций и дать сравнительно объективную оценку их вклада в развитие научного знания.

Следует учитывать, что задача наукометрических методов может быть двойной: с одной стороны, количественная оценка самих работ с целью ранжирования связанных с ними авторов, журналов и научных организаций; с другой – анализ связей между работами с целью выявления характера и динамики развития научного знания, создания «карты науки». К методам первого направления можно отнести следующие метрики.

1. *Классические цитатные метрики.* К ним относятся простой подсчет цитирований (total citation count); индекс Хирша (*h*-index), определяющий наибольшее значение *h*, при котором исследователь имеет *h* публикаций с не менее чем *h* цитированиями каждая [1]; индекс *g* (*g*-index), учитывающий распределение цитирований более чувствительно [2]; *i10*-index, подсчитывающий публикации с не менее чем 10 цитированиями [3].

2. *Нормализованные цитатные метрики.* Field-Weighted Citation Impact (FWCI) нормализует цитируемость по дисциплине, возрасту публикации и типу документа, позволяя сравнивать влияние работ из разных областей [4]. Relative Citation Ratio (RCR) сравнивает цитируемость публикации с медианной цитируемостью в соответствующих области и временном периоде [5].

3. *Альтернативные метрики (альтметрики)*. Это набор показателей, включающий упоминания на различных платформах, скачивания, сохранения в библиографических менеджерах. Например, Altmetric Attention Score агрегирует упоминания публикации в социальных сетях, новостных изданиях, политических документах и блогах, предоставляя показатель общественного внимания [6, 7].

4. *Журнальные метрики*. Импакт-фактор журнала (Journal Impact Factor, JIF) вычисляется как отношение цитирований статей за определенный год к числу статей, опубликованных журналом в предыдущие два года или в предыдущие пять лет (в зависимости от типа импакт-фактора) [8]. Кроме того, используются CiteScore [9], SJR (SCImago Journal Rank) [10], SNIP (Source Normalized Impact per Paper) [11] и Eigenfactor Score [12] для оценки престижа и влияния журналов.

5. *Семантический анализ цитирований*. Новейшие подходы используют обработку естественного языка для классификации намерения цитирования (supporting, contrasting, mentioning), что позволяет различать позитивные, негативные и нейтральные ссылки [13].

6. *Метрики воспроизводимости* оценивают надежность результатов через успешность репликаций [14], доступность данных и кода, стабильность размеров эффектов при повторных исследованиях [15].

7. *Фреймворки качества доказательств*, такие как GRADE [16], оценивают достоверность выводов на основе риска систематических ошибок, согласованности результатов, прямоты доказательств, точности оценок и риска публикационной предвзятости (publication bias).

Ко второму направлению, связанному с картографированием научного знания, можно отнести следующие методы.

1. *Анализ цитатных сетей*. Метод совместного цитирования (co-citation analysis) объединяет публикации, часто цитируемые вместе, в тематические кластеры [17]. Библиографическое сопряжение (bibliographic coupling) группирует работы по общим ссылкам в списках литературы. Эти методы позволяют визуализировать интеллектуальные связи и выявлять направления исследований.

2. *Контент-анализ и текстовое моделирование.* Анализ совместной встречаемости ключевых слов (co-word analysis) и терминов выявляет концептуальную структуру областей знания. Вероятностное тематическое моделирование (например, Latent Dirichlet Allocation) позволяет идентифицировать скрытые темы в больших корпусах научных текстов [18].

3. *Академическая генеалогия.* Данный подход картографирует науку через отношения «научный руководитель – диссертант», отслеживая передачу знания и развитие научных школ. Анализ генеалогических деревьев (например, через базу Mathematics Genealogy Project или Academic Family Tree) открывает возможность выявить наиболее плодотворные линии преемственности, оценить долгосрочное влияние отдельных ученых через их академических «потомков» и проследить географические и институциональные взаимосвязи идей [19, 20]. Современные исследования развивают этот подход и переходят от анализа формальных связей к выявлению содержательной преемственности – «интеллектуальной генеалогии». В частности, разрабатываются методы построения эмбединг-пространств на основе иерархических классификаторов, что позволяет векторизовать тематические профили диссертаций. Такой подход дает возможность метрически оценивать семантическую близость идей и отслеживать конвергенцию или дивергенцию научных направлений на уровне содержания исследований [21–23].

4. *Анализ коллаборационных сетей.* Подход основан на изучении паттернов соавторства и институциональных аффилиаций. Посредством построения и анализа сетей соавторства [24, 25] становится возможным визуализировать социальную структуру научного сообщества, идентифицировать ключевых исследователей («хабы» коллабораций), анализировать международные научные связи и динамику формирования исследовательских групп. Данный подход раскрывает социальные механизмы производства знания и позволяет оценить степень междисциплинарного взаимодействия.

5. *Методы выявления темпоральной динамики.* Эти методы направлены на изучение науки в динамике. Они включают идентификацию новых тем (emerging topics), обнаружение значимых работ с отложенным признанием [26], анализ жизненного цикла научных парадигм.

6. *Семантические подходы на основе NLP и машинного обучения.* Используются векторные представления текстов (word embeddings, sentence transformers) и графов знаний (knowledge graph embeddings) для построения семантических карт научного знания, где близость в пространстве отражает концептуальное родство. Построение таких эмбеддинг-пространств может рассматриваться как фундамент для перехода от картографирования текстов к картографированию смысловых единиц знания.

Как можно заметить, анализ цитатных связей занимает особое положение, поскольку применяется по обоим направлениям. С одной стороны, цитирование служит основой для вычисления практически всех основных количественных показателей влияния. С другой стороны, сети цитирования образуют структуры, анализ которых позволяет выявлять тематические кластеры, отслеживать эволюцию исследовательских направлений, идентифицировать междисциплинарные взаимодействия и обнаруживать возникающие исследовательские области.

Вместе с тем цитатный анализ, являясь центральным инструментом наукометрии, имеет принципиальные ограничения. Эти ограничения Д. В. Косяковым классифицируются следующим образом [27].

1. *Проблема семантики цитирования.* Традиционный цитатный анализ трактует все ссылки как эквивалентные индикаторы интеллектуального влияния, игнорируя различия в их функциях. Автор может цитировать работу, признавая фундаментальную теорию, использовать описанный метод или набор данных, полемизировать с конкурирующей интерпретацией, критиковать методологические недостатки, делать ритуальную ссылку на классика или добавлять компромиссную ссылку по настоянию рецензентов. Однако в наукометрических расчетах все цитаты равны.

2. *Проблема инфляции цитирований.* На протяжении последних десятилетий средний размер пристатейных списков литературы систематически растет. Это происходит параллельно с общим ростом объема публикуемой литературы, создавая эффект девальвирования единичного цитирования. Более поздние работы имеют преимущество в накоплении цитирований не обязательно в силу большей научной значимости, а потому, что научная система в целом про-

изводит больше цитат. Кроме того, среднее количество цитат сильно варьируется от области исследования, типа научной работы, языка публикации и пр.

3. *Проблема манипуляций цитированием.* По закону Гудхарта, когда измеримая величина становится целевым показателем, она перестает быть надежным индикатором [28, с. 308]. Цитируемость превратилась в критерий оценки академической карьеры, что породило многочисленные формы нарушений: «цитатные картели» (неформальные соглашения между группами авторов или журналов о взаимном цитировании), принуждение к цитированию со стороны редакторов и рецензентов, коммерческие сервисы по покупке цитирований и т. д.

4. *Проблема «фабрик статей» (paper mills).* Коммерческая работа по подготовке многочисленных публикаций не вносит вклад в развитие знания, она искажает цитатную структуру, на основе которой делаются выводы об авторитетности работ и значимости исследовательских направлений.

Поэтому предлагаемые наукометрические методы должны оцениваться, в частности, с позиции того, насколько они решают обозначенные проблемы или могут быть защищены от них. Кроме того, необходимо понимать конечную цель – идеальный инструментарий, к наличию которого следует стремиться, даже если возможность его реализации на данном этапе находится под вопросом. Такой конечной целью имеет смысл обозначить задачу построения эмбединг-пространства знаний – формализованной репрезентации научного знания, обладающей математическими свойствами. Под эмбединг-пространством понимается многомерная структура, в которой элементы знания представлены векторами или другими математическими объектами, положение и свойства которых отражают их содержание, а топологические характеристики самого пространства (плотность, связность, кривизна и т. д.) несут информацию о структуре предметной области. К такому пространству могли бы быть применимы математические методы для понимания динамики развития знания. В более сильной версии это пространство могло бы поддерживать операции определения истинности и полезности знания. Это означало бы переход от социологии – науки, измеряющей коммуникативные паттерны, к собственно «эпистемометрии» – количественному анализу самого знания. Но это общая

задача, в направлении решения которой, как можно предположить, будет развиваться наукометрия. В настоящей статье рассматривается метод анализа связей между публикациями, основанный на обозначенном понимании проблемы.

МЕТОДЫ

Рассмотренные выше проблемы цитатного анализа указывают на концептуальную трудность: традиционный цитатный анализ оперирует единицами научной деятельности (публикациями), а не единицами научного знания как такового. Публикация может содержать множество утверждений различной природы – от методологических описаний до ключевых научных выводов, от воспроизведения общеизвестных фактов до новаторских гипотез. Цитирование публикации целиком не позволяет различить, какие именно содержащиеся в ней утверждения признаются, используются, критикуются или игнорируются последующими исследователями.

Логичным решением является обращение к утверждению как базовой единице наукометрического анализа. Под утверждением понимается минимальная семантическая единица научного содержания, которая может быть оценена на истинность, полезность или релевантность независимо от того текстового контекста, в котором она впервые появилась.

Идея смещения фокуса с публикации на утверждение не нова. Систематическую попытку ее реализации предприняла Concept Web Alliance, предложив модель нанопубликаций (*nanopublications*) [29]. Нанопубликация определяется как «ключевое научное утверждение с сопутствующим контекстом», достаточным для его валидации и интерпретации. Архитектура модели нанопубликаций [29] включает следующие компоненты:

- *Concept* (концепт) – наименьшая неделимая и однозначная единица мысли, имеющая уникальный идентификатор.
- *Triple* (триплет) – кортеж (упорядоченное множество) из трех концептов в форме «субъект – предикат – объект»;
- *Statement* (утверждение) – триплет, которому присвоен уникальный идентификатор и который может быть процитирован;

- *Annotation* (аннотация) – единичный элемент метаданных, формирующий контекст утверждения (включая его происхождение, статус, оценку достоверности, совершенные над ним действия и т. д.);
- *nanopublication* (нанопубликация) – набор аннотаций, относящихся к одному утверждению.

Аннотация в формальном плане представляет собой триплет («субъект – предикат – объект»), в котором субъектом выступает само базовое утверждение. Ввиду жесткой структуры триплета каждая аннотация фиксирует один тип отношений. Поэтому для полноценного описания контекста требуется сформировать множество таких аннотаций. Чтобы это множество аннотаций получило статус нанопубликации, оно должно содержать набор метаданных, который признан минимально допустимым сообществом ученых (например, одни сообщества ученых могут требовать наличия аннотации с заключением рецензента, а другие – нет). Модель также предполагает, что одно и то же утверждение может встречаться в множестве источников. Все нанопубликации, относящиеся к идентичному утверждению, образуют S-Evidence – совокупность свидетельств об этом утверждении. Техническая реализация модели основана на технологиях Semantic Web. Утверждения представляются как Named Graphs в RDF – каждое утверждение и каждый набор аннотаций получают URI и могут быть процитированы, агрегированы и обработаны вычислительными агентами.

Несмотря на концептуальную привлекательность, модель [29] имеет ряд особенностей, которые представляются неоптимальными. *Во-первых*, обязательное представление всех утверждений в форме триплетов («субъект – предикат – объект») оказывается избыточным ограничением. Например, математическое утверждение «для любого простого числа $p > 2$ существует такое простое q , что $p < q < 2p$ » не имеет смысла искусственно редуцировать к триплетной форме. Аналогично, утверждения социальных и гуманитарных наук часто имеют сложную модальную структуру, не вписывающуюся в схему «субъект – предикат – объект».

Во-вторых, модель [29] изначально ориентирована на нанопубликации как отдельный формат публикации, параллельный традиционным статьям. Целесообразна модель, которая может работать с утверждениями как компо-

нентами любых типов научных текстов. Стоит отметить, что идея семантического обособления ключевых утверждений уже нашла воплощение в диссертационных исследованиях, для которых существует обязательное требование формулирования «положений, выносимых на защиту».

В-третьих, типология метаданных в модели [29] сфокусирована на происхождении и атрибуции, но недостаточно развита в отношении эпистемических и логических связей между утверждениями. Система не предлагает стандартизованного способа выразить, что одно утверждение опровергает другое, является его частным случаем, предусловием или альтернативной интерпретацией. Между тем именно такие связи критичны для построения наукометрии, учитывающей многообразие семантических отношений.

С учетом указанных ограничений предлагается ввести понятие «ключевое утверждение» (key statements, KS). Каждое ключевое утверждение, ассоциированное с научной публикацией, включает следующие компоненты:

- *собственно утверждение (assertion)*. Это текстовая формулировка научного содержания на естественном языке, которая должна быть семантически самодостаточной – понятной без обращения к полному тексту статьи. Для каждого утверждения должен быть определен уникальный идентификатор (например, DOI), позволяющий ссылаться на него из других работ;
- *область релевантности*. Этот компонент определяет, для каких научных дисциплин или поддисциплин данное утверждение представляет интерес и в развитие каких областей знания оно вносит вклад. Для кодирования могут использоваться существующие научные классификаторы: классификатор научных специальностей ВАК, УДК и т. д.;
- *тип/класс утверждения*. Различные типы научных утверждений имеют разные критерии валидации, что требует различных алгоритмов работы с ними. Например, можно выделить следующие типы: *математическое утверждение* (валидируется через формальное доказательство); *естественнонаучный закон* (валидируется через согласованность с эмпирическими данными и предсказательную силу); *эмпирический факт/наблюдение* (валидируется через воспроизводимость измерений); *методологическое утверждение* (валидируется через эффективность применения в последующих исследовани-

ях); *интерпретативное или концептуальное утверждение* (характерно для социально-гуманитарных наук; валидируется через аргументированность, согласованность с источниками и признание экспертным сообществом). Типология может быть расширена или детализирована для конкретных дисциплин;

- *значимость*. Опциональный компонент, позволяющий авторам проранжировать сформулированные ими утверждения, если публикация содержит несколько ключевых утверждений различной важности;

- *авторство*. По умолчанию авторами утверждения считаются все авторы публикации, в которой оно впервые сформулировано. Однако в случае коллабораций с распределенным вкладом может быть целесообразно явное указание авторов конкретного утверждения;

- *механизм верификации/фальсификации*. Опциональный компонент. Если для ключевого утверждения фальсифицируемость неочевидна изначально, то желательно, чтобы она была пояснена авторами утверждения.

Ключевые утверждения образуют два типа связей: с публикациями и между собой. *Связи утверждений с публикациями* могут быть следующих типов:

- *авторство утверждения*. Публикация является источником, в котором впервые появилось ключевое утверждение. Это первичная связь, фиксирующая происхождение утверждения;

- *оценка утверждения*. Публикация содержит оценку ранее сформулированного ключевого утверждения другой работы. Оценка может быть представлена в двух формах. *Дискретная оценка*: +1 – публикация подтверждает, валидирует, использует утверждение как доказанный факт; 0 – нейтральное упоминание (утверждение цитируется для контекста, обзора, без оценки истинности); –1 – публикация критикует, оспаривает или опровергает утверждение. *Континуальная оценка* позволяет выразить частичное согласие (0, +1) или частичное несогласие (–1, 0);

- *функциональная связь (опциональная)*. Уточняет, каким образом публикация использует утверждение: утверждение применяется как методологический инструмент; утверждение является источником данных; утверждение само выступает предметом анализа; утверждение используется как концепту-

альная основа; предлагается альтернативная интерпретация.

Связи между утверждениями могут быть классифицированы следующим образом:

- *обобщение/генерализация*. Утверждение KS_2 является обобщением KS_1 , если оно распространяет выводы KS_1 на более широкий класс объектов или условий;
- *уточнение/спецификация*. Обратное отношение: KS_2 является частным случаем более общего KS_1 , добавляя дополнительные ограничения или детализируя условия применимости;
- *семантическая идентичность*. Два утверждения имеют различные формулировки, но выражают одно и то же содержание. Это критично для избегания дублирования при подсчете метрик;
- *предусловие*. Истинность KS_2 логически или эмпирически зависит от истинности KS_1 ;
- *опровержение*. Истинность KS_2 несовместима с истинностью KS_1 .

В отличие от связей между публикациями и ключевыми утверждениями, которые может заявить автор статьи, связи между утверждениями требуют согласования. Минимальным требованием является согласие авторов обоих связываемых утверждений. Идеальным вариантом была бы валидация связи сообществом через определенную процедуру (например, схожую с рецензированием), хотя конкретные алгоритмы такой валидации выходят за рамки настоящей работы.

Помимо атрибутов, устанавливаемых автором, каждое утверждение должно иметь динамически обновляемый статус, определяемый научным сообществом на основе накопленных оценок: «*доказано/подтверждено*» (для математических и естественно-научных утверждений с проверенным доказательством или независимо воспроизводимыми результатами); «*гипотеза*» (утверждение имеет теоретическое обоснование, но недостаточно эмпирических подтверждений); «*дискуссионное*» (актуальные утверждения, которые не приняты, но и не опровергнуты сообществом); «*опровергнуто*» (существует доказательство ложности, и дальнейшее его исследование не нужно); «*устаревшее*» (утверждение было валидным в определенном контексте, но более не

используется в силу появления более точных или общих формулировок); «некорректное» (отвергнутое научным сообществом в силу нарушения правил, установленных для ключевых утверждений; например, по причине того, что данное утверждение по смыслу полностью дублировало предыдущее).

Предлагаемая нами модель допускает два способа/режима реализации.

- *Самостоятельное выделение авторами.* При подготовке рукописи авторы самостоятельно выделяют ключевые утверждения своей работы (аналогично тому, как сейчас составляется список ключевых слов), формулируют их в структурированном формате, указывают типы и области релевантности. Одновременно авторы составляют список ключевых утверждений из предыдущих работ, которые они оценивают (положительно, отрицательно или нейтрально), и опционально указывают функциональные связи.

- *Ретроспективное извлечение.* Для массива уже существующих публикаций ключевые утверждения извлекаются апостериорно. Это может выполняться: вручную экспертами-предметниками в процессе составления систематических обзоров или реализации соответствующих проектов; полуавтоматически – эксперты валидируют и редактируют результаты работы систем извлечения на основе больших языковых моделей; автоматически с использованием языковых моделей. Ретроспективное извлечение менее точно, но оно позволяет охватить исторический массив знания и создать критическую массу данных для запуска системы.

Логичным продолжением обсуждаемой модели является агрегация согласованных друг с другом утверждений в концепции. Концепция представляет собой связный граф утверждений, объединенных отношениями предусловия, уточнения и логической зависимости. Это важно для предотвращения манипуляций: без механизма объединения существует риск создания множества близких по смыслу ключевых утверждений для искусственного накручивания числа оценок. Конкретные алгоритмы организации концепций требуют отдельного исследования.

В случае принятия ключевых утверждений за единицу анализа требуется модификация метрик, основанных на подсчете цитирований. Предметом подсчета становится не «статья процитировала статью», а «публикация оценила

утверждение». Логика предлагаемых изменений может быть продемонстрирована на примере индекса Хирша. Классический индекс Хирша определяет, что исследователь имеет индекс h , если у него есть h публикаций, каждая из которых процитирована не менее h раз [1]. В логике ключевых утверждений этот принцип сохраняется, но вместо простого подсчета цитирований используется чистый баланс оценок – разность между поддерживающими и критикующими оценками. Формально h_{KS} можно задать выражением:

$$h_{KS} = \max \{ h : |\{ KS_i \in \mathcal{V} : B_i \geq h \}| \geq h \},$$

где $B_i = \sum_j w_{ij}$ – чистый баланс оценок утверждения KS_i ($w_{ij} \in [-1, +1]$ – оценка утверждения публикацией j), а \mathcal{V} – множество валидных утверждений, т. е. утверждений, не имеющих статусов «опровергнуто», «устаревшее» или «некорректное».

По аналогичной логике – замены числа цитирований на чистый баланс оценок с фильтрацией невалидных утверждений – возможна модификация и других: цитатных метрик (g -index [2], $h(2)$ -index [30], $i10$ -index [3] и т. д.).

Система ключевых утверждений открывает возможности и для новых метрик: *средняя оценка утверждений* (т. е. среднеарифметическое значение всех оценок ключевого утверждения, полученных в публикациях), *дисперсия оценок* и т. д.

В то же время чистый баланс оценок отражает консенсус научного сообщества, но консенсус не тождественен истине. Более того, различные типы научных утверждений имеют принципиально разные критерии валидности: математическую теорему доказывают дедуктивно, эмпирический факт проверяют воспроизведением измерений, а концептуальное утверждение гуманитарных наук оценивают через согласованность интерпретаций и аргументированность. Поэтому необходима разработка формализованных алгоритмов верификации и фальсификации, специфичных для каждого типа утверждений. Важно, чтобы результат формальной верификации имел приоритет перед балансом оценок сообщества.

Внимания заслуживает проблема защиты предлагаемой системы от манипуляций, когда создается сеть малозначимых публикаций, авторы которых взаимно положительно оценивают ключевые утверждения друг друга для ис-

кусственного повышения метрик. Для противодействия подобным стратегиям может оказаться целесообразным применить адаптированный алгоритм PageRank, изначально разработанный для ранжирования веб-страниц по их авторитетности [31]. Таким образом, оценка утверждения публикацией (w_{ij}) может учитываться с весовым коэффициентом, соответствующим среднему h_{KS} авторов оценивающей публикации. Это означает, что положительная оценка публикации авторов с высоким индексом h_{KS} будет вносить больший вклад в метрику утверждения.

Помимо анализа статуса самих утверждений данная модель позволяет оценивать качество экспертной работы исследователей, выставляющих эти оценки. Логика такого подхода может быть, к примеру, основана на сопоставлении индивидуальной позиции автора с динамикой изменения научного консенсуса. Идея заключается в поощрении исследователей, чьи оценки, изначально расходясь с мнением большинства, с течением времени оказываются более близкими к устоявшейся научной картине мира. Формально предиктивный показатель P_k ученого k , оценившего множество утверждений, можно записать через суммирование разностей расстояний:

$$P_k = \sum_i (|w_{ki} - C_i(t_0)| - |w_{ki} - C_i(t_{curr})|),$$

где w_{ki} – оценка, выставленная автором k ключевому утверждению i (KS_i); $C_i(t_0)$ – консенсусная оценка научного сообщества утверждению i в момент выставления оценки ученым; $C_i(t_{curr})$ – текущая консенсусная оценка утверждения i . Суммирование производится по всем утверждениям i , которые были оценены данным автором. Положительное значение P_k указывает, что преимущественно имело место то, что позиция оценившего утверждения ученого, даже если она была непопулярной или отличной от мейнстримной в момент оценивания, со временем стала ближе к общепринятой («опережение консенсуса»). Отрицательное значение свидетельствует о том, что преимущественно оценки ученого к настоящему времени все больше разошлись с обобщенными позициями научного сообщества. Однако следует подчеркнуть, что данный подход и его возможные модификации (например, использование взвешенных по авторитетности оценок или нормализация значений) не могут быть абсолютизирована-

ны. Исследователи в ряде случаев – и это оправдано логикой научного поиска – имеют собственную позицию и не стремятся к совпадению с мнением большинства, даже в долгосрочной перспективе. Поэтому эпистемологическая ценность такой метрики будет зависеть от предметной области: вероятно, она будет выше в эмпирических науках, таких как доказательная медицина, где важна конвергенция результатов к объективному факту, и существенно ниже в гуманитарных дисциплинах, особенно в философии, где оригинальность авторской интерпретации и разнообразие подходов составляют сущность исследовательской деятельности.

Логично также предположить, что в модели ключевых утверждений рецензенты журналов, принимающие статьи с эксплицитно выделенными ключевыми утверждениями, могли бы не только рецензировать текст в целом, но и верифицировать соответствие каждого утверждения принятым алгоритмам их использования, формулирования и оформления.

АПРОБАЦИЯ

Как было отмечено выше, предлагаемая модель ключевых утверждений допускает два способа формирования: самостоятельное выделение авторами при подготовке рукописи и ретроспективное извлечение из уже опубликованных работ. Для первичной апробации и оценки модели доступен только способ ретроспективного анализа, позволяющий работать с существующим массивом научных текстов без изменения публикационной практики.

В качестве эмпирической базы был выбран архив журнала «Информатика и образование», включающий номера с № 1/2016 по № 4/2025, охватывающий 728 статей. Выбор данного издания обусловлен несколькими факторами: во-первых, журнал представляет междисциплинарную область на стыке информатики и педагогики, что позволяет оценить применимость модели к утверждениям различной эпистемологической природы; во-вторых, относительно компактный формат статей облегчает извлечение ключевых утверждений; в-третьих, наличие полнотекстового электронного архива делает возможной автоматизированную обработку.

Процедура анализа включала три последовательных этапа.

Этап 1. Автоматизированное извлечение ключевых утверждений. С использованием большой языковой модели Gemini 2.5 Flash для каждой из 728 статей был составлен список ключевых утверждений [32]. Модель получала инструкцию идентифицировать основные научные выводы, методологические утверждения и эмпирические обобщения, содержащиеся в тексте, и формулировать их как самостоятельные, семантически завершенные высказывания. На этом этапе контроль качества не осуществлялся – задачей было максимально полное покрытие потенциально значимых утверждений.

Этап 2. Идентификация семантических связей. Полученный массив утверждений (порядка нескольких тысяч) был обработан различными языковыми моделями с большим контекстным окном для выявления имеющих общий предмет исследования утверждений, описывающих сходные феномены, применяющих аналогичные методы или относящихся к одним научным вопросам. Модель сформировала кластеры потенциально связанных утверждений, которые затем проходили экспертную валидацию. Эксперты (авторы исследования) отбирали те кластеры, где связь между утверждениями действительно существенна и возможна содержательная оценка согласованности или противоречия позиций разных авторов. Затем формулировалось обобщенное утверждение, объединяющее близкие по смыслу ключевые утверждения отдельных статей.

Этап 3. Экспертная оценка связей между ключевыми утверждениями. Для отобранных кластеров связанных утверждений эксперт возвращался к полным текстам соответствующих статей и анализировал характер отношений между утверждениями: подтверждает ли статья обобщенное ключевое утверждение (валидация), опровергает их (критика), упоминает нейтрально. В качестве нейтральной оценки рассматривались и упоминание без оценки истинности, и смешанное отношение (согласие с одними аспектами и несогласие с другими).

В результате анализа была составлена сводная таблица, включающая 20 обобщенных ключевых утверждений (табл. 1).

Табл. 1. Результаты анализа оценок обобщенных ключевых утверждений в статьях журнала «Информатика и образование» (2016–2025 гг.).

№	Обобщенное утверждение	Кол-во статей	Поддерживается	Нейтрально/Отсутствует однозначная поддержка или критика	Опровергается/Критикуется
1	Дистанционное обучение повышает качество и результаты образования	17	9	5	3
2	Электронные учебники эффективнее традиционных	14	10	2	2
3	Искусственный интеллект и нейросети повышают эффективность работы педагога	13	9	3	1
4	Интеграция математики и информатики (компьютерное моделирование) улучшает знания по обоим предметам	20	20	0	0
5	Геймификация учебного процесса повышает мотивацию и вовлеченность	12	8	3	1
6	Python – лучший выбор для начала обучения программированию и ИИ	11	9	0	2
7	Цифровые технологии – ключевой фактор инклюзии (для ОВЗ)	11	10	1	0
8	Интернет и гаджеты оказывают негативное влияние на когнитивные способности детей	6	6	0	0
9	Автоматизированное тестирование повышает объективность оценки	16	13	2	1
10	Проектный метод является ключевым для формирования проф. компетенций	17	16	1	0
11	Большие данные (Big Data) позволяют точно прогнозировать успеваемость	8	7	1	0
12	Виртуальная (VR) и дополненная (AR) реальность повышают качество обучения	10	9	0	1
13	Облачные технологии – оптимальное решение для инфраструктуры образования	13	13	0	0

14	Развитие цифровых компетенций педагога отстает от требований среды	13	13	0	0
15	Образовательная робототехника эффективно развивает инженерное мышление	12	11	1	0
16	ЕГЭ объективно оценивает качество обучения информатики	4	1	1	2
17	Социальные сети могут эффективно использоваться в учебном процессе	7	4	1	2
18	Необходима фундаментализация ИТ-образования (в противовес прикладному)	8	8	0	0
19	Информационная безопасность должна быть обязательной частью программы	10	10	0	0
20	Визуализация (интеллект-карты, инфографика) критически важна для усвоения	13	12	1	0

Таким образом, обобщенные утверждения поддерживаются в 84.3 %, рассматриваются нейтрально – 9.3 %, критикуются – в 6.4 % от общего числа статей выборки. Полученные данные позволяют сделать несколько выводов относительно структуры научной коммуникации в исследуемой области и ограничений ретроспективного подхода к выделению ключевых утверждений.

1. Анализ показал, что авторы, публикуя исследования, сознательно стремятся к тому, чтобы их работа заняла определенную нишу в исследовательском ландшафте, избегая формулировок, которые можно было бы напрямую противопоставить утверждениям предшественников. Даже когда содержательно работы касаются одного феномена и могли бы давать конкурирующие интерпретации, авторы используют различные термины, подчеркивают специфичность своего контекста или перспективы, затрудняющие формальное сравнение утверждений. Это явление можно интерпретировать двояко. С одной стороны, оно отражает объективную специфику педагогических феноменов, где результаты действительно сильно зависят от контекста. С другой стороны, это может быть стратегией минимизации риска критики и максимизации вероятности публикации: формулируя утверждение как «уникальное» для своих условий, автор снижает вероятность прямого сопоставления с противоречащи-

ми данными и облегчает рецензентам принятие работы как «нового вклада».

2. В целом авторы статей не стремятся открыто критиковать работы друг друга. Хотя выборка из одного журнала не может быть признана репрезентативной для всей науки, полученные данные показывают, что случаи явной критики конкретных работ крайне редки по сравнению с общим числом случаев рассмотрения связанных утверждений. Преобладающим паттерном является либо подтверждающее цитирование (использование предыдущих выводов как обоснования собственного подхода), либо нейтральное упоминание.

3. Попытка ретроспективно извлечь из педагогических статей ключевые утверждения и установить между ними оценочные связи столкнулась с указанными выше трудностями, которые обусловлены, вероятно, не столько с недостатками алгоритмов ретроспективного анализа, сколько с культурой научных публикаций (по крайней мере в определенных областях научного знания). Поэтому можно предположить, что для реализации предлагаемой модели существенно эффективнее было бы, чтобы выделение ключевых утверждений выполнялось самими авторами.

ЗАКЛЮЧЕНИЕ

Настоящая работа исходит из предположения, что для преодоления ограничений наукометрии, основанной на цитатном анализе (семантическая слепота, инфляция цитирований, уязвимость к манипуляциям и т. д.), целесообразно проанализировать возможность перехода к иной единице наукометрического анализа (от публикации в целом к ключевому утверждению как минимальной эпистемической единице). В статье предложена модель, модифицирующая концепцию нанопубликаций [29], адаптируя ее для широкого спектра научных дисциплин через отказ от обязательной триплетной структуры, введение развернутой типологии утверждений и связей между ними, а также через разработку системы метрик, учитывающих полярность оценок и приоритет формальных процедур верификации.

Апробация модели на архиве журнала «Информатика и образование» (728 статей, 2016–2025 гг.) выявила существенное ограничение ретроспективного подхода: сложившаяся культура научной коммуникации в исследованной области характеризуется уходом от прямой конфронтации с тезисами оппонен-

тов и низкой частотой явной критики. Это не техническая проблема извлечения утверждений (комбинация больших языковых моделей и экспертной валидации показала себя работоспособной), а следствие того, что существующие наукометрические системы не поощряют борьбу за установление истинности утверждений через систематическую проверку и критику. В результате научная литература растет преимущественно экстенсивно (введение новых утверждений через расширение вариативности, применение к другому контексту, детализацию существующих положений), а не интенсивно (через проверку и опровержение существующих тезисов).

Для полноценной реализации обсуждаемой модели (т. е. построения полноценного эмбединг-пространства знаний) необходимо ее развитие по нескольким направлениям.

1. Целесообразна разработка формализованной иерархической классификации научного знания, интегрированной с технологиями векторных представлений. Требуется структурированное пространство, где можно количественно измерять различные типы близости (утверждения могут быть близки, например, по предмету исследования, методологии, результатам), по уровню абстракции, сфере охвата (одно является обобщением другого).

2. По мере распространения агентных архитектур искусственного интеллекта – систем, где большие языковые модели интегрированы с инструментальными фреймворками для взаимодействия с формальными средами (AlphaFold, LEAN и т. д.), эмбединг-пространство знаний может рассматриваться, с одной стороны, как алгоритм, направляющий и организующий работу мультиагентных нейро-символических ИИ-систем, с другой – как множество оценок, на основе которых можно осуществлять обучение с подкреплением (т. е. оценка сгенерированного объекта осуществляется по количественным показателям, связанным с положением этого объекта в эмбединг-пространстве).

3. Оценка значимости научного утверждения не может быть полностью внутренней для науки – она должна учитывать связь знания с социальными, экономическими, технологическими и иными запросами общества. Научное утверждение может быть внутренне корректным и эпистемически валидным, но иметь различную ценность в зависимости от того, какие социальные запро-

сы, экономические приоритеты, культурные и этические установки оно затрагивает. Следовательно, помимо эмбединг-пространства знаний необходимо строить формализованные представления других сфер (например, пространства политических и экономических целей, образовательных стандартов, технологических дорожных карт) и задавать между ними отображения. Формализация алгоритма такого отображения позволит задавать в пространстве знаний векторные поля, отражающие ценности различных сфер.

4. Перспективной, но наиболее сложной задачей является разработка такого формализма структуры эмбединг-пространства знаний, который обладал бы внутренним механизмом проверки когерентности. Речь идет о способности автоматически выявлять, например, рассогласования между утверждениями (логические противоречия, скрытые конфликтующие предпосылки), зоны недостаточной плотности (пробелы в исследованиях, недостаток эмпирического обоснования) или, наоборот, дублирование усилий. В идеале эмбединг-пространство знаний должно быть не только отображением того, что уже сделано, но и инструментом генерирования новых знаний.

СПИСОК ЛИТЕРАТУРЫ

1. *Hirsch J.E.* An Index to Quantify an Individual's Scientific Research Output // Proceedings of the National Academy of Sciences of the United States of America. 2005. Vol. 102. No. 46. P. 16569–16572.
<https://doi.org/10.1073/pnas.0507655102>
2. *Egghe L.* Theory and Practice of the g-index // Scientometrics. 2006. Vol. 69. No. 1. P. 131–152. <https://doi.org/10.1007/s11192-006-0144-7>
3. *Connor J.* Google Scholar Citations Open To All // Google Scholar Blog. 2011. November 11. URL: <https://scholar.googleblog.com/2011/11/google-scholar-citations-open-to-all.html> (дата обращения: 12.01.2026)
4. *Colledge L.* Snowball Metrics Recipe Book. Edition 2. 2014. URL: <https://arma.ac.uk/wp-content/uploads/2021/08/Snowball-Metrics-Recipe-Book-edition-2.pdf> (дата обращения: 12.01.2026)
5. *Hutchins B.I., Yuan X., Anderson J.M., Santangelo G.M.* Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level // PLOS Biology. 2016. Vol. 14. No. 9. e1002541.

<https://doi.org/10.1371/journal.pbio.1002541>

6. *Priem J., Taraborelli D., Groth P., Neylon C.* Altmetrics: A Manifesto. 2011. URL: <https://digitalcommons.unl.edu/scholcom/185/> (дата обращения: 12.01.2026)

7. *García-Villar C.* A Critical Review on Altmetrics: Can We Measure the Social Impact of Research? // *Insights into Imaging*. 2021. Vol. 12. No. 1. Article 92. <https://doi.org/10.1186/s13244-021-01033-2>

8. *Bornmann L., Marx W., Gasparyan A.Y., Kitas G.D.* Diversity, Value and Limitations of the Journal Impact Factor and Alternative Metrics // *Rheumatology International*. 2012. Vol. 32. No. 7. P. 1861–1867. <https://doi.org/10.1007/s00296-011-2276-1>

9. *Teixeira da Silva J.A.* CiteScore: Advances, evolution, applications, and limitations // *Publishing Research Quarterly*. 2020. Vol. 36. No. 3. P. 459–468. <https://doi.org/10.1007/s12109-020-09736-y>

10. *González-Pereira B., Guerrero-Bote V.P., Moya-Anegón F.* A new approach to the metric of journals' scientific prestige: The SJR indicator // *Journal of Informetrics*. 2010. Vol. 4. No. 3. P. 379–391. <https://doi.org/10.1016/j.joi.2010.03.002>

11. *Moed H.F.* Measuring contextual citation impact of scientific journals // *Journal of Informetrics*. 2010. Vol. 4. No. 3. P. 265–277. <https://doi.org/10.1016/j.joi.2010.01.002>

12. *Bergstrom C.* Eigenfactor: Measuring the value and prestige of scholarly journals // *College & Research Libraries News*. 2007. Vol. 68. No. 5 P. 314–316. <https://doi.org/10.5860/crln.68.5.7804>

13. *scite.ai.* Bringing Smart Citations to Rankings. June 16, 2025. URL: <https://scite.ai/blog/smart-citations-rankings> (дата обращения: 12.01.2026)

14. *Open Science Collaboration.* Estimating the Reproducibility of Psychological Science // *Science*. 2015. Vol. 349. No. 6251. Article aac4716. <https://doi.org/10.1126/science.aac4716>

15. *Goodman S.N., Fanelli D., Ioannidis J.P.A.* What Does Research Reproducibility Mean? // *Science Translational Medicine*. 2016. Vol. 8. No. 341. Article 341ps12. <https://doi.org/10.1126/scitranslmed.aaf5027>

16. *Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schünemann H.J.* GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations // *British Medical Journal*. 2008. Vol. 336. No. 7650. P. 924–926. <https://doi.org/10.1136/bmj.39489.470347.AD>

17. *Small H.* Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents // *Journal of the American Society for Information Science*. 1973. Vol. 24. No. 4. P. 265–269. <https://doi.org/10.1002/asi.4630240406>

18. *Leydesdorff L., Nerges A.* Co-word maps and topic modeling: A comparison using small and medium-sized corpora ($N < 1,000$) // *Journal of the Association for Information Science and Technology*. 2017. Vol. 68. No. 4. P. 1024–1035. <https://doi.org/10.1002/asi.23740>

19. *Jackson A.* A Labor of Love: The Mathematics Genealogy Project // *Notices of the American Mathematical Society*. 2007. Vol. 54. No. 8. P. 1002–1003.

20. *Rossi L., Freire I.L., Mena-Chalco J.P.* Genealogical Index: A Metric to Analyze Advisor-Advisee Relationships // *Journal of Informetrics*. 2017. Vol. 11. No. 2. P. 564–582. <https://doi.org/10.1016/j.joi.2017.04.001>

21. *Лернер И.М., Мариносян А.Х., Григорьев С.Г., Юсупов А.Р., Аникьева М.А., Гарифуллина Г.А.* Подход к формированию интеллектуальной академической генеалогии с использованием больших языковых моделей // *Электромагнитные волны и электронные системы*. 2024. Т. 29. № 4. С. 108–120. EDN: BSNYZW. <https://doi.org/10.18127/j5604128-202404-09>

22. *Григорьев С.Г., Лернер И.М., Мариносян А.Х., Арутюнова Н.К., Григорьева М.А.* К вопросу подбора учебно-методической информации для реализации адаптивной электронной образовательной среды: алгоритм априорной классификации авторов // *Информатика и образование*. 2025. Т. 40. № 2. С. 66–78. EDN: YDTFTG. <https://doi.org/10.32517/0234-0453-2025-40-2-66-78>

23. *Мариносян А.Х., Григорьев С.Г., Лернер И.М., Аникьева М.А.* Автоматизированное сравнение научных исследований на базе академической генеалогии // *Информатика и образование*. 2025. Т. 40. № 6. С. 16–27. <https://doi.org/10.32517/0234-0453-2025-40-6-16-27>

24. *Newman M.E.J.* Coauthorship Networks and Patterns of Scientific Col-

laboration // Proceedings of the National Academy of Sciences. 2004. Vol. 101. Suppl. 1. P. 5200–5205. <https://doi.org/10.1073/pnas.0307545100>

25. Hou H., Kretschmer H., Liu Z. The Structure of Scientific Collaboration Networks in *Scientometrics* // *Scientometrics*. 2008. Vol. 75. No. 2. P. 189–202. <https://doi.org/10.1007/s11192-007-1771-3>

26. van Raan A.F.J. Sleeping Beauties in Science // *Scientometrics*. 2004. Vol. 59. No. 3. P. 467–472. <https://doi.org/10.1023/B:SCIE.0000018543.82441.f1>

27. Косяков Д.В. Можно ли нарисовать карту знания? История, подходы и революция ИИ в научной картографии // Компания «Пульс Науки». 2025. URL: https://vkvideo.ru/video-224951981_456239150 (дата обращения 12.01.2026)

28. Strathern M. ‘Improving ratings’: Audit in the British University system. *European Review*. 1997. Vol. 5. No. 3. P. 305–321. [https://doi.org/10.1002/\(SICI\)1234-981X\(199707\)5:3<305::AID-EURO184>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4)

29. Groth P., Gibson A., Velterop J. The Anatomy of a Nanopublication // *Information Services & Use*. 2010. Vol. 30. No. 1–2. P. 51–56. <https://doi.org/10.3233/ISU-2010-0613>

30. Kosmulski M. A New Hirsch-Type Index Saves Time and Works Equally Well as the Original h-index // *ISSI Newsletter*. 2006. Vol. 2. No. 3. P. 4–6.

31. Page L., Brin S., Motwani R., Winograd T. The PageRank citation ranking: Bringing order to the web // Stanford InfoLab. Technical Report. 1999. URL: <http://ilpubs.stanford.edu:8090/422/> (дата обращения 12.01.2026).

32. Gemini Team, Google. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities // arXiv. URL: <https://arxiv.org/abs/2507.06261>

SCIENTIFIC PUBLICATIONS AND THE EMBEDDING SPACE OF KNOWLEDGE

A. Kh. Marinosyan¹ [0000-0003-0577-2360], S. G. Grigoriev² [0000-0002-0034-9224]

^{1, 2}*Moscow City University, Moscow, Russia*

¹a.marinosyan@yandex.ru, ²grigorsg@yandex.ru

Abstract

The article examines current challenges in scientometrics arising from the surge in publication activity and the widespread adoption of generative artificial intelligence. The existing scientometric toolkit for analyzing research activity is reviewed, categorized into quantitative metrics and science mapping methods (citation network analysis, academic genealogy, semantic analysis, etc.). An attempt is made to overcome the limitations of traditional citation analysis, such as “semantic blindness” and vulnerability to manipulation. As a potential solution, a conceptual model is proposed where the unit of analysis shifts from the publication as a whole to an individual “key statement”. This approach involves recording not only the statement’s content but also its type, area of relevance, and its logical relationship with other claims (confirmation, refutation, clarification, generalization, etc.). Within this framework, principles for calculating modified scientometric metrics are introduced.

The proposed model was tested on a corpus of 728 articles from the Russian journal *Informatics and Education* (2016–2025). An analysis conducted using large language models revealed that retrospective extraction of statements faces significant hurdles due to established cultures of scientific communication. Consequently, the study highlights the advantages of having authors formulate key statements themselves as a distinct type of metadata. In conclusion, the paper outlines development paths for the concept of an “embedding space of knowledge,” which could eventually complement existing approaches to analyzing the evolution of scientific ideas and theories.

Keywords: scientometrics, bibliometrics, academic genealogy, citation analysis, semantic analysis, large language models, science mapping, h-index, nanopublications.

REFERENCES

1. *Hirsch J.E.* An Index to Quantify an Individual's Scientific Research Output // Proceedings of the National Academy of Sciences of the United States of America. 2005. Vol. 102. No. 46. P. 16569–16572.
<https://doi.org/10.1073/pnas.0507655102>
2. *Egghe L.* Theory and Practice of the g-index // Scientometrics. 2006. Vol. 69. No. 1. P. 131–152. <https://doi.org/10.1007/s11192-006-0144-7>
3. *Connor J.* Google Scholar Citations Open To All // Google Scholar Blog. 2011. November 11.
URL: <https://scholar.googleblog.com/2011/11/google-scholar-citations-open-to-all.html> (date accessed: 12.01.2026)
4. *Colledge L.* Snowball Metrics Recipe Book. Edition 2. 2014. URL: <https://arma.ac.uk/wp-content/uploads/2021/08/Snowball-Metrics-Recipe-Book-edition-2.pdf> (date accessed: 12.01.2026)
5. *Hutchins B.I., Yuan X., Anderson J.M., Santangelo G.M.* Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level // PLOS Biology. 2016. Vol. 14. No. 9. e1002541.
<https://doi.org/10.1371/journal.pbio.1002541>
6. *Priem J., Taraborelli D., Groth P., Neylon C.* Altmetrics: A Manifesto. 2011. URL: <https://digitalcommons.unl.edu/scholcom/185/> (date accessed: 12.01.2026)
7. *García-Villar C.* A Critical Review on Altmetrics: Can We Measure the Social Impact of Research? // Insights into Imaging. 2021. Vol. 12. No. 1. Article 92. <https://doi.org/10.1186/s13244-021-01033-2>
8. *Bornmann L., Marx W., Gasparyan A.Y., Kitas G.D.* Diversity, Value and Limitations of the Journal Impact Factor and Alternative Metrics // Rheumatology International. 2012. Vol. 32. No. 7. P. 1861–1867.
<https://doi.org/10.1007/s00296-011-2276-1>
9. *Teixeira da Silva J.A.* CiteScore: Advances, evolution, applications, and

limitations // Publishing Research Quarterly. 2020. Vol. 36. No. 3. P. 459–468.

<https://doi.org/10.1007/s12109-020-09736-y>

10. *González-Pereira B., Guerrero-Bote V.P., Moya-Anegón F.* A new approach to the metric of journals' scientific prestige: The SJR indicator // *Journal of Informetrics*. 2010. Vol. 4. No. 3. P. 379–391.

<https://doi.org/10.1016/j.joi.2010.03.002>

11. *Moed H.F.* Measuring contextual citation impact of scientific journals // *Journal of Informetrics*. 2010. Vol. 4. No. 3. P. 265–277.

<https://doi.org/10.1016/j.joi.2010.01.002>

12. *Bergstrom C.* Eigenfactor: Measuring the value and prestige of scholarly journals // *College & Research Libraries News*. 2007. Vol. 68. No. 5 P. 314–316.

<https://doi.org/10.5860/crln.68.5.7804>

13. *scite.ai.* Bringing Smart Citations to Rankings. June 16, 2025.

URL: <https://scite.ai/blog/smart-citations-rankings> (date accessed: 12.01.2026)

14. *Open Science Collaboration.* Estimating the Reproducibility of Psychological Science // *Science*. 2015. Vol. 349. No. 6251. Article aac4716.

<https://doi.org/10.1126/science.aac4716>

15. *Goodman S.N., Fanelli D., Ioannidis J.P.A.* What Does Research Reproducibility Mean? // *Science Translational Medicine*. 2016. Vol. 8. No. 341. Article 341ps12. <https://doi.org/10.1126/scitranslmed.aaf5027>

16. *Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schünemann H.J.* GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations // *British Medical Journal*. 2008. Vol. 336. No. 7650. P. 924–926. <https://doi.org/10.1136/bmj.39489.470347.AD>

17. *Small H.* Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents // *Journal of the American Society for Information Science*. 1973. Vol. 24. No. 4. P. 265–269.

<https://doi.org/10.1002/asi.4630240406>

18. *Leydesdorff L., Nerghees A.* Co-word maps and topic modeling: A comparison using small and medium-sized corpora ($N < 1,000$) // *Journal of the Association for Information Science and Technology*. 2017. Vol. 68. No. 4. P. 1024–1035.

<https://doi.org/10.1002/asi.23740>

19. *Jackson A.* A Labor of Love: The Mathematics Genealogy Project // Notices of the American Mathematical Society. 2007. Vol. 54. No. 8. P. 1002–1003.

20. *Rossi L., Freire I.L., Mena-Chalco J.P.* Genealogical Index: A Metric to Analyze Advisor-Advisee Relationships // Journal of Informetrics. 2017. Vol. 11. No. 2. P. 564–582. <https://doi.org/10.1016/j.joi.2017.04.001>

21. *Lerner I.M., Marinosyan A.Kh., Grigoriev S.G., Yusupov A.R., Anikieva M.A., Garifullina G.A.* An approach to the formation of intellectual academic genealogy using large language models // Journal Electromagnetic Waves and Electronic Systems. 2024. Vol. 29. No. 4. P. 108–120. EDN: BSNYZW. <https://doi.org/10.18127/j5604128-202404-09> (In Russ.)

22. *Grigoriev S.G., Lerner I.M., Marinosyan A.Kh., Arutyunova N.K., Grigorieva M.A.* On the issue of educational and methodological information selection for implementing an adaptive learning management system: Algorithm of a priori authors classification // Informatics and Education / Informatika i obrazovanie. 2025. Vol. 40. No. 2. P. 66–78. EDN: YDTFTG. <https://doi.org/10.32517/0234-0453-2025-40-2-66-78> (In Russ.)

23. *Marinosyan A.Kh., Grigoriev S.G., Lerner I.M., Anikieva M.A.* Automated comparison of scientific research based on academic genealogy // Informatics and Education / Informatika i obrazovanie. 2025. Vol. 40. No. 6. P. 16–27. <https://doi.org/10.32517/0234-0453-2025-40-6-16-27> (In Russ.)

24. *Newman M.E.J.* Coauthorship Networks and Patterns of Scientific Collaboration // Proceedings of the National Academy of Sciences. 2004. Vol. 101. Suppl 1. P. 5200–5205. <https://doi.org/10.1073/pnas.0307545100>

25. *Hou H., Kretschmer H., Liu Z.* The Structure of Scientific Collaboration Networks in Scientometrics // Scientometrics. 2008. Vol. 75. No. 2. P. 189–202. <https://doi.org/10.1007/s11192-007-1771-3>

26. *van Raan A.F.J.* Sleeping Beauties in Science // Scientometrics. 2004. Vol. 59. No. 3. P. 467–472. <https://doi.org/10.1023/B:SCIE.0000018543.82441.f1>

27. *Kosyakov D.V.* Can a knowledge map be drawn? History, approaches, and the AI revolution in scientific cartography // “Pulse of Science” Company. 2025. URL: https://vkvideo.ru/video-224951981_456239150 (date accessed: 12.01.2026) (In Russ.)

28. *Strathern M.* 'Improving ratings': Audit in the British University system. *European Review*. 1997. Vol. 5. No. 3. P. 305–321.
[https://doi.org/10.1002/\(SICI\)1234-981X\(199707\)5:3<305::AID-EURO184>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4)
29. *Groth P., Gibson A., Velterop J.* The Anatomy of a Nanopublication // *Information Services & Use*. 2010. Vol. 30. No. 1–2. P. 51–56.
<https://doi.org/10.3233/ISU-2010-0613>
30. *Kosmulski M.* A New Hirsch-Type Index Saves Time and Works Equally Well as the Original h-index // *ISSI Newsletter*. 2006. Vol. 2. No. 3. P. 4–6.
31. *Page L., Brin S., Motwani R., Winograd T.* The PageRank citation ranking: Bringing order to the web // *Stanford InfoLab. Technical Report*. 1999.
URL: <http://ilpubs.stanford.edu:8090/422/> (date accessed: 12.01.2026).
32. *Gemini Team, Google.* Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities // *arXiv*. URL: <https://arxiv.org/abs/2507.06261>

СВЕДЕНИЯ ОБ АВТОРАХ



МАРИНОСЯН Андреас Хачатурович – аспирант Института цифрового образования Московского городского педагогического университета. Область научных интересов: наукометрия, обработка естественных языков, архитектуры языковых моделей.

Andreas Khachaturovich MARINOSYAN – PhD student at the Institute of Digital Education, Moscow City University. Research interests: scientometrics, natural language processing, language model architectures.

email: a.marinosyan@yandex.ru

ORCID: 0000-0003-0577-2360



ГРИГОРЬЕВ Сергей Георгиевич – профессор Института цифрового образования Московского городского педагогического университета, чл.-корр. РАО, профессор, д. т. н. Область научных интересов: информатизация образования, методология научных исследований, математические методы в научных исследованиях, языки и методы программирования.

Sergey Georgievich GRIGORIEV – professor at the Institute of Digital Education, Moscow City University, Corresponding Member of the RAE, Professor, Doctor of Technical Sciences. Research interests: informatization of education, methodology of scientific research, mathematical methods in scientific research, programming languages and methods.

email: grigorsg@yandex.ru

ORCID: 0000-0002-0034-9224

Материал поступил в редакцию 29 января 2026 года