

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.6-047.44:070.4

НЕГБОЕВ
Аслиддин Нигмат Угли

**МЕТОДЫ И АЛГОРИТМЫ АНАЛИЗА ДАННЫХ ПУБЛИКАЦИОННОЙ
АКТИВНОСТИ НАУЧНОГО КОЛЛЕКТИВА**

Автореферат
на соискание степени магистра
по специальности 7-06-0612-01 «Программная инженерия»

(подпись магистранта)

Научный руководитель

Шульдова Светлана Георгиевна

(фамилия, имя, отчество)

канд. тех. наук, доцент

(ученая степень, ученое звание)

(подпись научного руководителя)

Минск 2025

ВВЕДЕНИЕ

В современной научной среде публикационная активность является одним из ключевых индикаторов продуктивности и влияния, отражая объем и качество научно-исследовательской работы отдельных ученых, коллективов и целых организаций. Публикации служат не только основным механизмом распространения научного знания, но и инструментом для оценки вклада исследователей в развитие своей области знаний, что напрямую влияет на распределение финансирования, аттестацию кадров и позиции университетов в международных рейтингах.

Стремительный рост объемов научной информации и усложнение структуры научных коммуникаций делают традиционные методы оценки, основанные на простом подсчете публикаций и цитирований, недостаточными для получения глубокого и всестороннего представления о научной деятельности. Такие подходы позволяют ответить на вопрос "сколько?", но не дают ответа на вопросы "как?" и "о чем?", оставляя за рамками анализа структуру внутренних коллабораций, тематическую специализацию исследовательских групп и потенциал междисциплинарного взаимодействия.

Это противоречие между возрастающей сложностью научной деятельности и ограниченностью классических инструментов оценки определяет актуальность настоящей диссертационной работы. Возникает необходимость в разработке и применении комплексных методик анализа, способных интегрировать различные типы данных и предоставлять многоаспектную картину научного ландшафта организации.

В связи с этим, данная работа посвящена разработке и апробации комплексного подхода к анализу данных публикационной активности, основанного на синергетическом объединении методов графового моделирования, семантического анализа и наукометрической оценки. Применение данного подхода на примере Белорусского государственного университета информатики и радиоэлектроники позволяет не только оценить его эффективность, но и получить ценные практические результаты, характеризующие структуру и содержание научной деятельности ведущего технического вуза.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Степень разработанности проблемы. В наукометрии существует значительное количество работ, посвященных отдельным методам анализа публикационной активности, включая библиометрический анализ, исследование

сетей соавторства и тематическое моделирование. Однако большинство исследований фокусируются на применении одного из этих методов. Работ, посвященных разработке и апробации комплексных методик, интегрирующих структурный (графовый) и содержательный (семантический) анализ для получения целостной картины научного ландшафта крупной организации, на данный момент недостаточно, что и определяет исследовательский пробел, который призвана заполнить настоящая диссертация.

Цель и задачи исследования.

Целью диссертационного исследования является разработка и апробация комплексного подхода к анализу данных публикационной активности сотрудников научного коллектива, направленного на объективную оценку их научной продуктивности, выявление структуры научного взаимодействия и определение ключевых тематических направлений исследований.

Для достижения поставленной цели были определены следующие задачи:

- провести анализ существующих подходов, методов и инструментов для оценки публикационной активности научных коллективов;
- разработать комплексный подход к многоаспектному анализу публикационной активности, включающий этапы сбора, интеграции, предобработки данных, наукометрического анализа, графового моделирования и семантического анализа текстов;
- адаптировать и выбрать математические модели и алгоритмы для реализации ключевых компонентов предлагаемого подхода;
- выполнить программную реализацию для проведения экспериментальных исследований;
- провести апробацию разработанного подхода на реальных данных публикационной активности научного коллектива бгуир;
- проанализировать полученные результаты, оценить их достоверность и практическую значимость, сформулировать выводы и рекомендации.

Объект и предмет исследования.

Объект исследования: публикационная активность научного коллектива Белорусского государственного университета информатики и радиоэлектроники.

Предмет исследования: методы и алгоритмы анализа данных публикационной активности для оценки научной продуктивности, структуры взаимодействия и тематической направленности исследований.

Научная гипотеза. Интеграция методов наукометрического анализа, графового моделирования сетей сотрудничества и усовершенствованного

семантического анализа текстов публикаций, при условии качественной предобработки данных, позволяет получить более достоверную, многоаспектную и интерпретируемую оценку публикационной активности научного коллектива и его внутренней структуры по сравнению с применением базовых методик.

Основные положения, выносимые на защиту:

1. Комплексный подход к анализу публикационной активности, основанный на интеграции графового моделирования и семантического анализа, который позволил получить многоаспектную оценку научной деятельности по сравнению с разрозненным применением данных методов.

2. Структура научного сотрудничества в крупном техническом университете, полученная в результате применения адаптированного алгоритма выявления сообществ Лувена на большом массиве данных (~10 000 публикаций) научного коллектива БГУИР, представляет собой ядро активно взаимодействующих исследователей, которое, в свою очередь, состоит из ряда устойчивых сообществ, имеющих выраженную тематическую специализацию.

3. Результаты сопоставления структуры научных сообществ БГУИР и их тематической специализации позволили создать детальную "карту науки" университета, и являются эффективным инструментом для поддержки принятия управленческих решений в сфере научной политики университета.

Научная новизна

Разработан комплексный подход, системно интегрирующий графовый и семантический анализ для исследования публикационной активности, позволяющий одновременно выявлять структуру коллабораций и их тематическую специализацию.

Адаптирован алгоритм выявления сообществ Лувена для анализа крупной сети соавторства научного коллектива БГУИР путем эмпирического подбора параметра resolution для получения содержательно-значимых кластеров.

Впервые на большом массиве данных (~10 000 публикаций) проведено сопоставление структуры научных сообществ БГУИР и их тематической специализации, что позволило создать детальную "карту науки" университета.

Теоретическая и практическая значимость.

Теоретическая значимость работы заключается в систематизации и дополнении методологического аппарата наукометрических исследований за счет демонстрации преимуществ интеграционного подхода.

Практическая значимость исследования состоит в том, что его результаты и сама методика могут быть использованы научными и административными подразделениями БГУИР и других университетов для:

- Стратегического планирования научной деятельности и поддержки приоритетных направлений;
- Стимулирования междисциплинарных исследований;
- Мониторинга и оценки эффективности научных коллективов.

Личный вклад соискателя

Все исследования, результаты которых изложены в диссертации, получены лично соискателем в процессе научных исследований и экспериментов. Вклад научного руководителя заключается в формулировке целей и задач исследования.

Апробация результатов работы

Результаты исследований, вошедшие в диссертацию, были опубликованы в электронном сборнике материалов 61-й научной конференции аспирантов, магистрантов и студентов БГУИР (г. Минск, Беларусь, 2025 год), в научном журнале “Студенческий вестник” (Интернаука 2025. № 17(350) часть 8).

Структура и объем работы

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников и приложений. Общий объем работы составляет 76 страниц. Работа содержит 3 таблиц, 2 рисунков. Список литературы включает 36 наименований, список собственных публикаций соискателя из 3 наименований, 3 приложения.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы исследования, связанная с необходимостью перехода от традиционных наукометрических показателей к комплексным методам анализа публикационной активности. Рассмотрены проблемы, возникающие при использовании классических метрик, и сформулирована потребность в разработке интеграционного подхода, способного дать многоаспектную оценку научной деятельности современного университета.

В **общей характеристике работы** определена связь диссертации с актуальными задачами в области анализа научных данных, сформулированы цель и задачи исследования, объект и предмет, научная гипотеза и положения, выносимые на защиту. Раскрыта научная новизна и практическая значимость полученных результатов, а также приведены сведения об их апробации.

В первой главе проведен системный анализ теоретических и методологических основ исследования публикационной активности. Рассмотрена эволюция наукометрических подходов, выполнен обзор и сравнительная характеристика основных источников библиографических данных (Scopus, Web of Science, РИНЦ, lens.org), выявлены их преимущества и ограничения. Детально описан математико-статистический аппарат, релевантный для решения поставленных задач, включая основы теории графов для моделирования сетей соавторства, алгоритмы выявления сообществ (в частности, метод Лувена) и методы семантического анализа текстов (TF-IDF, NMF). Также проведен обзор и обоснование выбора программных средств для реализации исследования.

Во второй главе описана разработка и апробация предложенного комплексного подхода. Представлен четырехэтапный конвейер анализа: сбор и интеграция данных, их предобработка, многоаспектный анализ и синтез результатов. Детально изложены методы реализации каждого этапа, включая нормализацию имен авторов, лингвистическую обработку текстов, построение графа соавторства и применение алгоритмов кластеризации и тематического моделирования. На основе данных о публикационной активности БГУИР (~10 000 публикаций) было выявлено ядро научного взаимодействия из 2716 авторов, в котором идентифицировано 42 устойчивых научных сообщества. Ключевым результатом главы является визуализация структуры топ-10 крупнейших научных сообществ (Рисунок 1) и их содержательная характеристика на основе семантического анализа, представленная в виде итоговой таблицы.

Структура топ-10 научных сообществ

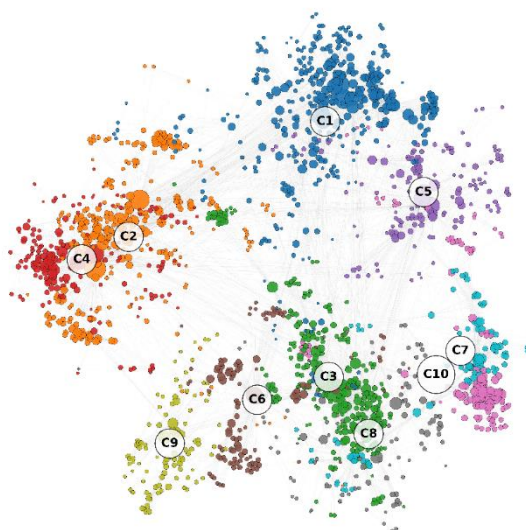


Рисунок 1 — Структура топ-10 научных сообществ БГУИР

В третьей главе раскрыта научная новизна и практическая значимость полученных результатов. Продемонстрирован синергетический эффект от

интеграции графового и семантического анализа, который позволяет получить целостную "карту науки" организации. На основе анализа этой карты сформулирован ряд конкретных практических рекомендаций для управления научной деятельностью в университете, касающихся стратегического планирования, поддержки приоритетных направлений, стимулирования междисциплинарных исследований и развития кадрового потенциала. Также в главе рассмотрены объективные ограничения проведенного исследования и намечены перспективные направления для будущей работы.

ЗАКЛЮЧЕНИЕ

В рамках данной диссертационной работы было проведено исследование, посвященное актуальной научной задаче — разработке и апробации методики анализа данных публикационной активности научного коллектива. Цель работы состояла в создании комплексного подхода, позволяющего провести многоаспектную оценку научной деятельности, дополняющую традиционные библиометрические показатели.

Для достижения поставленной цели были решены следующие задачи:

Проведен анализ теоретических и методологических основ исследования публикационной активности, который показал, что применение отдельных методов не позволяет сформировать целостное представление о структуре и содержании научной деятельности, что обосновало необходимость их интеграции.

Разработан и описан комплексный подход, основанный на синергетическом эффекте от объединения трех методов: графового моделирования для анализа сетей сотрудничества, семантического анализа для выявления тематической специализации и наукометрической оценки для измерения продуктивности.

Осуществлена практическая реализация подхода с использованием языка программирования Python и современных библиотек для анализа данных. Был реализован конвейер обработки данных: от сбора и предобработки массива публикаций (~10 000 записей) до применения алгоритмов кластеризации (алгоритм Лувена) и тематического моделирования (TF-IDF и NMF).

Выполнена апробация разработанного подхода на данных Белорусского государственного университета информатики и радиоэлектроники, в результате которой была выявлена структура научного взаимодействия коллектива (2716 авторов в ядре сети), идентифицировано 42 научных сообщества и определена их тематическая специализация.

Определена практическая значимость полученных результатов, на основе которых сформулированы рекомендации для стратегического управления научной деятельностью в университете.

Основным итогом работы является подтверждение выдвинутой научной гипотезы о том, что интеграция графового, семантического и наукометрического анализа позволяет получить более полную и многоаспектную картину публикационной активности научного коллектива, чем использование этих методов по отдельности. Разработанный подход позволил не просто измерить количественные показатели, а создать структурированное представление научного ландшафта организации, где взаимосвязаны структура научных связей и содержательная сторона исследований.

Научная новизна исследования определяется разработкой и апробацией комплексной методики, системно объединяющей современные методы интеллектуального анализа данных для решения наукометрических задач. Практическая значимость работы состоит в том, что ее результаты и методика могут быть применены научными и административными подразделениями университетов в качестве инструмента для мониторинга и поддержки принятия решений в сфере научной политики.

Таким образом, цель диссертационного исследования достигнута, а полученные результаты обладают теоретической и практической ценностью, открывая перспективы для дальнейших исследований в области интеллектуального анализа научных данных.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1-А] Негбоев А.Н. Семантический анализ и цитируемость публикаций научного коллектива // Студенческий вестник: электрон. научн. журн. 2025. № 17(350). URL: <https://studvestnik.ru/journal/stud/herald/350>

[2-А] Негбоев А.Н. Анализ соавторства и международного сотрудничества в публикационной активности научного коллектива // Студенческий вестник: электрон. научн. журн. 2025. № 17(350). URL: <https://studvestnik.ru/journal/stud/herald/350>

[3-А] Негбоев А.Н. Анализ публикационной активности научных коллективов// Компьютерные системы и сети : сборник материалов 61-й научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 22–26 апреля 2025 г. / Белорусский государственный университет информатики и радиоэлектроники. / Минск, 2025. – С. 144–146