

Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА
СИБИРСКОГО ОТДЕЛЕНИЯ РОССИЙСКОЙ АКАДЕМИИ НАУК
(ГПНТБ СО РАН)

УДК 001.3

Per. № НИОКТР АААА-А19-119121090008-8

Per. № ИКРБС

УТВЕРЖДАЮ

Директор ГПНТБ СО РАН

Канд. техн. наук

«27» января 2021 г.



А. Е. Гуськов

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НАУКОМЕТРИЧЕСКИЙ АНАЛИЗ ПУБЛИКАЦИОННОГО ПОТОКА
РОССИЙСКИХ ИССЛЕДОВАТЕЛЕЙ И ФАКТОРОВ ЕГО
ТРАНСФОРМАЦИИ, ИЗУЧЕНИЕ СПОСОБОВ И МЕТОДОВ
ПОВЫШЕНИЯ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ, РАЗВИТИЯ
РОССИЙСКОЙ НАУЧНОЙ ПЕРИОДИКИ
(промежуточный, за 2020 год)

Приоритетное направление IV.38. «Проблемы создания глобальных и интегрированных информационно-телекоммуникационных систем и сетей, развитие технологий и стандартов GRID».

Программа ФНИ СО РАН IV.38.1. Методы и технологии создания и интеграции гетерогенных распределенных информационно-вычислительных ресурсов для поддержки междисциплинарных научных исследований на основе сервис-ориентированной парадигмы

Руководитель НИР,
директор ГПНТБ СО РАН,
канд. техн. наук

А. Е. Гуськов

Новосибирск 2021

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель НИР,
директор, заведующий лабораторией
наукометрии, канд. техн. наук


01.02.21
подпись, дата

А. Е. Гуськов
(Введение, Раздел 3,
Заключение)

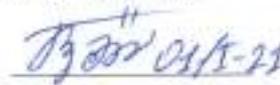
Исполнители:

Ведущий научный сотрудник, канд. техн.
наук


01.02.21
подпись, дата

Н. А. Мазов
(Раздел 6)

Старший научный сотрудник, канд. пед. наук


01/11/21
подпись, дата

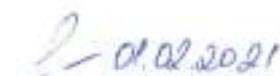
В. Н. Гуреев
(Раздел 6)

Старший научный сотрудник, канд. физ-мат.
наук


01.02.2021
подпись, дата

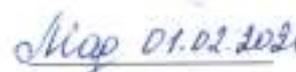
И. В. Савченко
(Разделы 1, 2, 4, 5)

Младший научный сотрудник


01.02.2021
подпись, дата

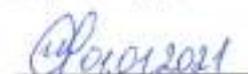
А. В. Багирова
(Раздел 2)

Младший научный сотрудник


01.02.2021
подпись, дата

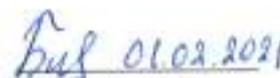
Л. Л. Садовская
(Раздел 2)

Младший научный сотрудник


01.02.2021
подпись, дата

И. В. Селиванова
(Разделы 1, 2)

Научный сотрудник, канд. техн. наук


01.02.2021
подпись, дата

П. Ю. Блинов
(Раздел 3)

Научный сотрудник


01.07.2021
подпись, дата

Д. В. Косяков
(Разделы 1, 2, 3, 4, 5)

РЕФЕРАТ

Отчет промежуточный за 2020 г. содержит 122 страниц, 49 рисунков, 14 таблиц, 71 источник, 1 приложение.

БИБЛИОМЕТРИЯ, НАУКОМЕТРИЯ, НАУЧНЫЕ ЖУРНАЛЫ, НАУЧНАЯ ПЕРИОДИКА, РЕДАКЦИОННАЯ КОЛЛЕГИЯ, ПУБЛИКАЦИОННАЯ АКТИВНОСТЬ, ТЕМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ, SCOPUS, ФРАКЦИОННЫЙ СЧЕТ.

Объект исследований: российский сектор исследований и разработок, история его развития в первых десятилетиях XXI в., его результативность и национальная инфраструктура научных коммуникаций.

Целью исследований являлись разработка методов улучшения качества наукометрических данных; структурно-динамический, секторальный и тематический анализ документопотока российских научных публикаций; выявление успешных практик развития научных журналов.

Методы исследования основаны на информетрическом подходе – количественном анализе библиографической информации с привлечением дополнительных данных об организационной структуре сектора исследований и разработок.

Важнейший результат. *Сравнение целочисленного и фракционного счета публикаций для задач оценки научной результативности.* На примере наукометрических показателей российских организаций за 2000-2018 гг. проведено аналитическое сопоставление фракционного счета, используемого в методике КБПР, с целочисленным. Показано, что со временем корреляция между фракционным и целочисленным счётом падает с 0,98 в 2002 до 0,8 в 2017. Это приводит к существенным изменениям в рейтинговании и категорировании этих организаций при использовании разных методов. Наибольшие различия фиксируются в группе ведущих университетов – участников Проекта 5-100, умеренные – среди научных организаций, а наименьшие – среди других вузов, что косвенно подтверждает влияние «наукометрического давления». Анализ методики КБПР выявил, что бенефициарами от её введения оказываются институты химического профиля и материаловедения, которые чаще публикуются в высокорейтинговых журналах и имеют небольшие коллаборации. В зону «проигрыша» попали организации, ведущие исследования в области физики высоких энергий и университеты экономического профиля.

По результатам исследования в 2020 г. было опубликовано 9 статей в рецензируемых научных журналах, в том числе 3 – в журналах, индексируемых в Web of Science и Scopus, представлено 8 докладов на международных и российских конференциях.

СОДЕРЖАНИЕ

СПИСОК ИСПОЛНИТЕЛЕЙ.....	2
РЕФЕРАТ	3
СОДЕРЖАНИЕ.....	4
СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ.....	6
ВВЕДЕНИЕ	7
1 Разработка методов улучшения качества наукометрических данных	12
1.1 Проблема публикаций без страновых аффилиаций.....	12
1.2 Тематическая классификация публикаций на основе теоретико-информационного подхода	24
1.3 Метод уточнения тематики публикаций на основе публикационной истории авторов.....	31
2 Общий анализ документопотока российских публикаций, индексируемого в Scopus	34
2.1 Полнота охвата публикационного потока в БД Scopus	35
2.2 Динамика, позиция в мире, целый и фракционный счет.....	38
2.3 Структура публикационного потока по типам документов.....	42
2.4 Тематическая структура публикационного потока.....	44
2.5 Распределение публикационного потока по квартилям научных журналов.....	47
3 Феномен «инфляции результативности» и методики фракционного счета	50
3.1 Инфляция результативности	50
3.2 Фракционный счет.....	54
3.3 Персональная результативность российских исследователей	56
3.4 Коэффициент самодостаточности, неравномерность реакции на «наукометрическое давление»	58
3.5 Комплексный балл публикационной результативности – победители и проигравшие.....	63
4 Анализ региональной и секторальной структуры документопотока	72
4.1 Региональная структура документопотока	73
4.2 Секторальная структура документопотока	76

5 Сопоставление российского публикационного потока с результатами стран – конкурентов России в мировом научном пространстве	81
5.1 Извлечение и подготовка данных, уточнение тематики и определение секторальной принадлежности	81
5.2 Структура публикационного потока по типам документов и квартилям научных журналов.....	82
5.3 Тематическая структура публикационного потока Бразилии, Франции, Индии и России	86
5.4 Секторальная структура публикационного потока Бразилии, Франции, Индии и России	88
6 Инфраструктура российской научной периодики.....	91
6.1 Анализ альтметрик как предиктора будущих цитирований научного журнала	91
6.2 Оценка публикационного вклада редколлегии в библиометрические показатели научного журнала	95
ЗАКЛЮЧЕНИЕ.....	104
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	112
ПРИЛОЖЕНИЕ. СПИСОК ПУБЛИКАЦИЙ И ДОКЛАДОВ ПО ПРОЕКТУ.....	120

СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

ББД	– Библиографическая база данных
БД	– База данных
ИНН	– Индивидуальный номер налогоплательщика
КБПР	– Комплексный балл публикационной результативности
Минобрнауки России	– Министерство науки и высшего образования России
РИНЦ	– Российский индекс научного цитирования
ФАНО России	– Федеральное агентство научных организаций России
ФЦП	– Федеральная целевая программа
АС	– Article count
A&HCI	– Arts and Humanities Citation Index
API	– Application program interface
Ar	– Article
ASJC	– All Sciences Journal Classification
CNCI	– Category Normalized Citation Impact
CP	– Conference proceedings
DCM	– Data Compression Method
ESCI	– Emerging Sources Citation Index
FC	–Fraction Count
FWCI	– Field-Weighted Citation Impact
JCR	– Journal Citation Reports
KNN	– k-nearest neighbors
LR	– Logistic regression
NB	– Naive Bayes classifier
NFC	– National Fraction Count
Re	– Review
RF	– Random forest
RSCI	– Russian Science Citation Index
SCIE	– Science Citation Index
SSCI	– Social Sciences Citation Index
SJR	– Scimago Journal Ranking
SVM	– Support vector machine
WoS	– Web of Science
WoS CC	– Web of Science Core Collection

ВВЕДЕНИЕ

В последние 10-15 лет в России выстраивается новая система государственного управления сектором исследований и разработок, основанная на активном применении наукометрических индикаторов. Требования к показателям публикационной активности присутствуют в структуре государственных заданий научных организаций и конкурсной документации научных фондов, а также являются предметом регулярных дискуссий относительно путей развития российской науки. Сейчас можно констатировать, что реформа фундаментальной науки внесла значительные изменения не только в институциональную структуру управления исследованиями, но и в систему мотивации ученых. В конечном итоге это привело к изменению объема и содержания российского «валового научного продукта».

Одной из главных целей реформирования является увеличение количества публикаций российских авторов в ведущих мировых научных журналах, индексируемых в Web of Science или Scopus. С 2013 года действительно можно наблюдать рост доли российских публикаций в мировом потоке, который за несколько лет поднялся с 1,8% до 2,3-2,6% (по разным оценкам). Вместе с тем, фиксируются многочисленные (а, иногда, и системные) негативные эффекты такой системы мотивации: увеличивается доля публикаций в хищнических журналах; расцветает рынок откровенно мусорных конференций и изданий; одной публикацией могут отчитываться много раз в разных проектах и организациях; в погоне за количеством падает качество научных статей; статус монографии (завершенный научный труд) долгое время остается неопределенным.

При этом за последние годы проведен ряд реформ, имевших целью улучшить российский научный ландшафт: запуск ФЦП "Исследования и разработки", Проект повышения конкурентоспособности ведущих российских университетов среди ведущих мировых научно-образовательных центров,

создание Российского научного фонда, поддержка "мегагрантов", реформа государственных академий наук. Однако системный анализ фактического влияния этих изменений не проводился, что не позволяет в настоящее время уверенно говорить об их успешности, а в будущем – обоснованно вводить новые научные политики с прогнозируемыми результатами.

Этот проект посвящен изучению фактического влияния перечисленных событий на индикаторы российской науки, причем речь идет не только о количественных, но и о качественных показателях. Первым объектом изучения является массив публикаций российских ученых с 2000 года в международной базе данных Scopus, который содержит основные результаты фундаментальных исследований, признанных на мировом уровне. Вторым объектом изучения является сама российская научная периодика, состояние которой также является характерным отражением положения российской фундаментальной науки.

На первом этапе выполнения проекта в 2019 году акцент был сделан на трёх направлениях:

- 1) *Разработка технологии улучшения библиографических данных публикаций БД Scopus с устранением неоднозначностей и ошибок, связанных с некорректной атрибуцией авторов и аффилиаций.* Было проанализировано качество библиографических данных на примере БД Scopus и возможности для его улучшения. Показано, что ошибки идентификации приводят к наличию дублирующих профилей у 77% организаций и 24% авторов. Выполнена оценка влияния этих ошибок на количественные показатели результативности организаций. Исследована возможность классификации научных текстов на основе метода компрессии их аннотаций. Для последующего анализа была сформирована база данных расширенных библиографических описаний публикаций российских исследователей за 2000-2018 гг. на основе БД Scopus, сформированы справочники организаций и авторов, выполнена частичная очистка данных.

2) *Разработка методов структурно-динамического анализа публикационного потока, выполнение анализа полученного массива данных с выявлением изменений в секторальном, тематическом и территориальном аспектах, обнаружение и сопоставление эволюционных и аномальных изменений в публикационном потоке с влиянием стимулирующих факторов.*

Выполнен анализ динамики позиционирования российской науки в мировом контексте, определены и описаны реперные точки изменений национальной научной политики, прослежено и проанализировано их влияние на изменение публикационной активности как российских исследователей в целом, так и в региональном и секторальном разрезе. Продемонстрировано, что опережающий рост публикационной активности обеспечен сегментом высшего образования и локализован в регионах Сибири, Урала и Поволжья. Показано, что данные государственной статистики не вполне адекватно отражают текущее состояние и историю развития кадров науки, социологические исследования также не позволяют обрисовать полную картину, в связи с этим обоснована необходимость и возможность использования в таких исследованиях библиометрических данных. Определены и классифицированы факторы, искажающих показатели национальной публикационной активности, сделаны оценки их влияния. Подробно проанализировано влияние феномена международной синхронной мобильности на показатели результативности российских организаций и российской науки в целом. Обнаружены свидетельства предположительно недобросовестных практик ряда российских организаций, финансирующих в рамках программ по привлечению ведущих зарубежных исследователей «покупку» публикаций в форме вознаграждения за указание этими исследователями российской аффилиации в качестве дополнительной в своих статьях. Выполнен анализ проведенной в 2017 г. ведомственной оценки результативности научных исследований организаций, подведомственных ФАНО России, определены недостатки методики количественной оценки, обоснована необходимость перехода на использование фракционного метода

счета публикаций, разработаны методы исследования последствий перехода с целого на фракционный счет.

3) *Анализ территориального распределения и библиометрическое исследование публикационной активности членов редколлегии научных журналов в области информационных и библиотечных наук, истории и археологии.* Разработана методика выделения ядра наиболее авторитетных журналов. Сформированы для последующего анализа базы данных членов редакционных коллегий отечественных журналов по информатике, библиотековедению и наукам о Земле. Анализ географического распределения членов редколлегии анализируемых дисциплин указывает на достаточно большое присутствие в редколлегиях зарубежных представителей, работа которых в отечественных журналах, тем не менее, зачастую носит формальный характер. Показано, что центры экспертных групп, представленных членами редколлегии анализируемых журналов, находятся преимущественно в Москве, Санкт-Петербурге и Новосибирске. Для некоторых журналов показано влияние редколлегии на рейтинговые показатели журнала, включая случаи публикации высококачественных статей, впоследствии привлекающих большое число цитирований, а также «скрытое» цитирование своего журнала в других источниках. Для таких журналов редколлегии фактически являются самой основой их существования, а также «поставщиком» основных библиометрических показателей, таких как публикуемость и цитируемость, что противоречит рекомендациям международных комитетов по публикационной этике. В то же время для ряда других изданий указанная зависимость отсутствовала.

Основываясь на этих и других полученных результатах, на второй этап проекта, выполняемый в 2020 году, были поставлены следующие задачи:

1) Структурно-динамический анализ российских публикаций в ведущих мировых изданиях за 2000-2019 гг. по данным БД Scopus и выявление качественных изменений публикационного потока, включающий:

- 1.1) Развитие методов улучшения качества наукометрических данных;
 - 1.2) Общий анализ документопотока российских публикаций
 - 1.3) Анализ феномена «инфляции результативности» и методик фракционного счета
 - 1.4) Анализ региональной и секторальной структуры документопотока
- 2) Структурно-динамический анализ публикационной активности стран-конкурентов России в мировом научном пространстве на примере Франции, Бразилии и Индии в разрезах отраслей и секторов науки за 2000-2019 гг.
- 3) Разработка системы индикаторов, характеризующих развитие ведущих российских периодических изданий, их расчет и анализ динамики изменения за период 2000-2019 гг. Выявление успешных практик развития научных журналов.

Отчет о выполнении второго этапа состоит из введения, шести разделов и заключения. В соответствии со структурой задач, решение первой задачи рассматривается в разделах 1-4, второй задачи – в разделе 5, третьей задачи – в разделе 6.

1 Разработка методов улучшения качества наукометрических данных

1.1 Проблема публикаций без страновых аффилиаций

При детальном изучении наукометрических данных нами было обнаружено, что значительное число публикаций, индексируемых в Scopus не аффилировано ни с одной страной. В неформальном рейтинге этот «нейтральный флаг» в начале 2000х занимал второе место среди всех стран, но к концу второго десятилетия сместился на седьмую позицию. Эта проблема характерна и для Web of Science – «нейтральный флаг» в этой БД занимал в 2019 г. 6-ое место. Ожидаемо, среди таких публикаций много материалов, не связанных с результатами исследований. Однако, даже при рассмотрении «научных» типов документов доля публикаций без страновой аффилиации довольно высока. Подавляющее большинство публикаций-космополитов являются результатом невнимательности авторов к аффилиационным данным, несовершенными издательскими практиками, позволяющими указывать неполную информацию об организациях или не указывать ее вовсе, а также проблемами автоматизированного сбора данных. Очевидно, что эти публикации имеют связь с конкретными организациями и странами, и их неучет существенно влияет на результаты наукометрических исследований на национальном и институциональном уровне. Масштаб проблемы постепенно уменьшается, но остается значительным. Потенциально это искажение распределено неравномерно – возможно, какие-то направления наук и/или организации «потеряли» больше чем другие. Доля научных публикаций без страновой аффилиации от общего количества публикаций в разных областях наук приведена в таблице 1.

Эта проблема уже отмечалась исследователями – первой известной нам публикацией, в которой указывается на существование проблемы, является статья Сибиллы Хинце, опубликованная в 1994 г. [1]: автор отмечает, что в среднем 17% документов из базы INSPEC за период с 1985 по 1993 годы не содержат сведения о стране, наблюдается снижение доли таких статей с 29%

в период с 1985 по 1989 год до 12% в последующие годы. Далее автор просто исключает указанные статьи из рассмотрения и при этом не анализирует погрешности, являющиеся следствием исключения данных из выборки, хотя этот шаг может привести к искажению результатов.

Таблица 1 – Доля публикаций без страновой аффилиации по областям наук в 2019 г. (типы публикаций Article, Review, Book Chapter, Conf. Review, Conf. Paper, Book)

Область наук	Всего документов	Документов без страновой аффилиации	Доля в %
Undefined	1078	368	34.1%
Arts and Humanities	127162	13016	10.2%
Veterinary	23402	1558	6.7%
Nursing	50065	2961	5.9%
Social Sciences	316281	15255	4.8%
Psychology	76126	3583	4.7%
Dentistry	16571	738	4.5%
Business, Management and Accounting	101482	3451	3.4%
Health Professions	37658	1113	3.0%
Medicine	680251	19958	2.9%
Economics, Econometrics and Finance	57883	1319	2.3%
Computer Science	496254	8559	1.7%
Engineering	718238	11793	1.6%
Earth and Planetary Sciences	164578	2598	1.6%
Energy	168809	2543	1.5%
Neuroscience	72961	977	1.3%
Decision Sciences	73080	975	1.3%
Immunology and Microbiology	80786	1066	1.3%
Pharmacology, Toxicology and Pharmaceutics	93538	1156	1.2%
Chemical Engineering	163108	1830	1.1%
Mathematics	264358	2955	1.1%
Multidisciplinary	57658	629	1.1%
Biochemistry, Genetics and Molecular Biol.	335373	3259	1.0%
Environmental Science	230947	2203	1.0%
Materials Science	367888	3489	0.9%
Chemistry	274398	2491	0.9%
Physics and Astronomy	391086	2941	0.8%
Agricultural and Biological Sciences	239419	1696	0.7%

В [2] исследуются ошибки, связанные с неполнотой данных в базах данных Scopus и Thomson Reuters, в том числе анализируется и степень отсутствия информации о стране. Оказалось, что информация о стране

отсутствовала в более чем 12 млн записей Scopus, что соответствовало 34% всех проиндексированных публикаций на 2009 год. Очевидно, что такие ошибки могут существенно влиять на результат наукометрического исследования, например, в случае расчета индекса Хирша по базе Scopus для стран в том же 2009 году [3]. Уровень пропусков информации о стране в базах WoS согласно [2] ниже, около 14%. Автор справедливо называет уровень ошибок такого рода чрезвычайно высоким, и призывает анализировать причины такого положения дел, активно обсуждать способы улучшения ситуации, и сообщать о наличии проблемы с пропуском данных при публикации разнообразных отчетов на основе данных WoS и Scopus. Несмотря на призывы к максимально широкой дискуссии, упоминания об этой проблеме достаточно скудны: согласно данным Scopus статью [2] процитировали 31 раз за прошедшие 11 лет.

В недавней работе [4] автор обнаружил, что результаты поиска в Scopus зависят от способа поиска. Автор сначала собрал библиометрические данные из Scopus (за 2012–2014 годы) для всех чилийских университетов, используя страну в качестве поискового запроса, а затем составил список учреждений, сопоставив каждому набор чисел отражающих количество документов, публикуемых за каждый год. А потом с удивлением обнаружил, что, изменив стратегию поиска на запросы по учреждению, он получил результаты отличные от первоначального. Аналогичная проверка была проведена и для WoS и привела к схожему результату. Автор связывает наличие подобных артефактов с отсутствием стандартизации в названиях чилийских университетов. Далее автор рассматривает эту проблему немного более детально, уже для организаций США. Его предположения о влиянии разницы в написании адресов и международного сотрудничества на разницу в результатах запроса не находят подтверждения.

Китайские исследователи опубликовали в 2018 году достаточно масштабное исследование [5], посвященное проблеме отсутствия информации о адресах авторов. Для анализа авторы использовали три основных индекса

цитирования WoS: SCIE, SSCI и A&HCI. Они показывают, что с 1900 по 2015 год более одной пятой публикаций, проиндексированных в WoS, полностью пропускали информацию из адресного поля. Масштабы проблемы сильно различаются в зависимости от различных факторов. Влияние некоторых факторов таких как: период времени, язык публикации и тип документов анализируется в статье. Также внимание уделено проблеме частичного отсутствия информации. Показано, что в случае SCIE и SSCI проблема с пропусками информации особенно остро стояла до начала 1970-х годов, и остается таковой для A&HCI. Авторы попытались масштабировать свой подход на БД Scopus, что привело к следующим выводам: проблема отсутствия адресной информации в Scopus настолько серьезна, что ее нельзя игнорировать, а также что существующая структура Scopus не готова к такому типу анализа в связи с отсутствием «запросов с широкими подстановочными параметрами» (broad wildcard queries).

И наконец, нельзя не упомянуть совсем свежее исследование [6], в котором предложен метод для решения проблемы с пропущенной информацией о связи автора со страной. Предлагается заполнять пропущенную информацию, используя информацию из сторонних графов знаний, в данном случае из Wikidata (<https://www.wikidata.org/>). Перефразируя авторов, метод показал хорошие результаты, но требует существенной доработки и дополнительной проверки. С нашей точки зрения существенным недостатком и лимитирующим фактором для данного подхода, является зависимость от развития стороннего сервиса, в данном случае Wikidata.

В качестве потенциально российских публикаций мы отобрали публикации без указания аффилиационных данных, написанные авторами, хотя бы раз в публикациях 2000-2019 гг. аффилированными с российскими организациями. Часть этих публикаций может не быть аффилирована с РФ по целому ряду причин, например, авторы могли эмигрировать, предыдущие аффилиации авторов с РФ могли быть ошибочными и т. д. Для уточнения вероятности аффилиации с РФ нами был построен классификатор таких

публикаций. В первую очередь для классификации потенциально российских публикаций мы использовали простые правила (паттерны), применяя которые мы могли делать выводы о связи статьи с Российской Федерацией (РФ). Для построения таких правил требуется достаточно четкое определение аффилиации публикации со страной, а также окажутся полезны знания о структуре науки в стране, взаимосвязях отдельных параметров, которые позволят в дальнейшем делать предсказания. Начнем с определения: будем считать публикацию аффилированной со страной в том случае, если исследование выполнено с участием организации из этой страны. Дальнейшие три пункта о науке в России субъективны, основаны на личном опыте авторов и не претендуют на какую-либо точность:

- «Исследование, выполненное с участием Российской организации» не эквивалентно «Исследование, выполненное с участием сотрудников Российской организации». В современном мире возможны ситуации, когда один человек работает одновременно в нескольких организациях, зарегистрированных в разных странах. Такие «космополиты» не слишком заметны на фоне общей массы российских исследователей, но их количество не настолько мало чтобы ими пренебрегать. К тому же высокая востребованность со стороны работодателей часто подкрепляется высокой научной и публикационной результативностью.

- Российские научные фонды финансируют исследования с участием российских организаций. Информация о финансовой поддержке часто отражается в поле «Acknowledgement».

- Название страны может быть указано в информации для связи с авторами, иногда без указания организации.

- Российские ученые любят публиковаться в российских журналах, а зарубежные без российских соавторов – нет.

Теперь сформулируем правила классификации в порядке значимости:

1) Если статья аффилирована с одной или несколькими организациями и известно, что, хотя бы одна из организаций российская,

можно с уверенностью говорить об аффилиации статьи с РФ. Если же известно, что все связанные организации не имеют отношения к России, следует признать статью, не связанной с РФ.

2) В случае отсутствия информации об аффилированных организациях, или о связи этих организаций с конкретной страной, можно использовать информацию о финансовой или другой поддержке, а также данные для связи с авторами. Если в публикации есть упоминание российских научных фондов, или имеется упоминание РФ в Acknowledgement, или кто-либо из авторов указал Россию в адресе для переписки, вероятно эта публикация связана с РФ.

3) В ситуации, когда первые два пункта не позволяют классифицировать публикацию будем использовать информацию об источнике (журнале). Если журнал является российским будем считать публикацию связанной с РФ. При этом российскими будем считать не те источники, которые издаются в РФ, а те в которых редакционная коллегия состоит из преимущественно российских ученых.

Предложенные правила позволяют неплохо классифицировать потенциально российские публикации, но существуют некоторые проблемы. Во-первых, остаются не классифицированными публикации из числа потенциально российских ($\approx 20\,000$ в данном случае). Во-вторых, требуется проверка качества такой классификации, поскольку предложенные правила основаны во многом на субъективных наблюдениях авторов. Особенно это касается последнего правила, связанного с источником публикации.

Для решения обозначенных выше проблем мы создали простой алгоритм (классификатор), который предсказывает, является публикация российской или нет. Идея состоит в подсчете количества аффилиаций авторов с разными странами и последующее использование этой информации для предсказания аффилиации публикации со страной. В качестве обучающей выборки мы использовали набор публикаций, размеченных Scopus как российские (1 051 тыс.), а также потенциально российские публикации,

размеченные нами согласно 1 правилу как не имеющие отношения к РФ (592 тыс.). О каждой публикации из обучающей выборки мы уверенно можем сказать, российская она или нет. Далее мы создали список всех авторов публикаций из обучающей выборки. Каждому автору сопоставим две количественные меры: количество российских публикаций, соавтором которых он является, и количество зарубежных публикаций с его участием. Просуммировав эти меры для всех авторов произвольной публикации, мы также получим два числа. Полученные числа позволяют делать вывод о принадлежности публикации к определенному классу. В самом простом случае мы их можем просто сравнивать, у какого класса мера больше, к тому классу и относим публикацию. Ясно, что лучше использовать сравнение с весом, хотя бы потому что для обучения у нас использована не сбалансированная выборка. Весовой коэффициент мы подбирали эмпирически, выполнив серию численных экспериментов. Мы варьировали весовой коэффициент и подсчитывали количество неверных предсказаний классификатора на обучающей выборке. Затем нашли минимум полученной зависимости, результат представлен на рисунке 1. Оптимальным весовым коэффициентом оказалось число 1.1. То есть получилось, что, если число связей всех авторов с российскими публикациями в 1.1 раза выше числа связей с зарубежными, мы классифицируем статью как российскую, иначе как зарубежную.

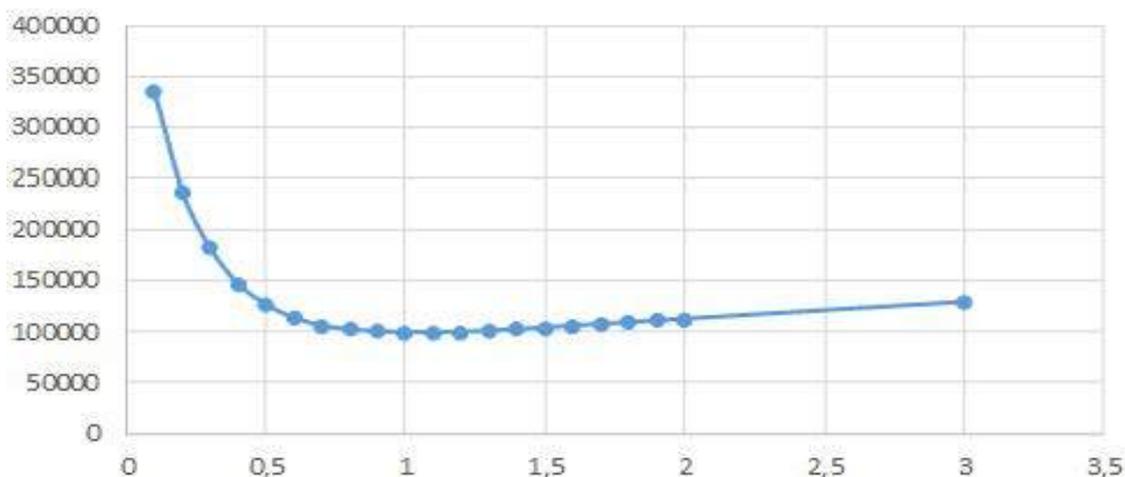


Рисунок 1 – Число неверно классифицированных документов в зависимости от множителя

Для оценки качества полученного классификатора будем использовать стандартные метрики качества, используемые в машинном обучении [7] для задач бинарной классификации. Ограничимся самыми простыми метриками: компонентами матрицы ошибок (TP, TN, FP, FN – true/false positive/negative), правильностью (A – accuracy), точностью (P – precision) и полнотой (R – recall). Для проверки мы опробовали классификатор на обучающей выборке и трех специально подготовленных наборах данных. Результаты применения классификатора представлены в таблице 2.

Таблица 2 – Результаты классификации различных наборов данных о публикациях используя простой бинарный классификатор

Набор данных	TP*	TN*	FP*	FN*	A	P	R
1	999231	545260	47339	51756	0.94	0.95	0.95
2	4295	0	0	178	0.96	1	0.96
3	1820	10707	578	112	0.95	0.76	0.94
4	9100	10707	578	560	0.95	0.94	0.94

* Позитивным классом считались российские публикации, негативным не российские.

Наборы данных: 1 - обучающая выборка, 2 - публикации из числа потенциально российских, признанные аффилированными с РФ по 1 правилу, 3 - публикации в Journal of Physics: Conference Series 2020 года, 4 - сбалансированная выборка публикаций в Journal of Physics: Conference Series 2020 года

Видно, что для всех использованных наборов данных классификатор показывает очень хорошие показатели правильности и полноты, показатель точности оказался значительно ниже для третьей выборки. Набор данных 2

был получен путем применения первого решающего правила, определяющего принадлежность статьи к стране по принятому нами определению. Набор данных 3 сильно не сбалансирован, число зарубежных публикаций более чем в 5 раз превосходит число российских. Чтобы понять связано низкое значение точности на третьей выборке с работой классификатора, или вызвано дисбалансом классов в наборе данных, выборка 3 была сбалансирована путем увеличения в 5 раз миноритарного класса используя случайный оверсэмплинг [8]. Полученный набор данных обозначен цифрой 4. Как видно выравнивание классов привело к увеличению параметра точности до значения сопоставимого с другими выборками.

Ранее мы предположили, что неучтенные публикации распределены неравномерно в пространстве признаков. Изучение степени этой неравномерности мы начнем с анализа временной зависимости. На рисунке 2 показано как менялось число публикаций со временем, и какая доля в этом количестве соответствует неучтенным публикациям. Заметно, что наиболее характерна проблема потерянных публикаций для первой пятилетки 21 века. Начиная с 2005 года, их количество, как в абсолютном, так и в процентном отношении, монотонно снижалось. Исключение составил лишь 2008 год в котором наблюдался слабый локальный максимум. Если посмотреть на общее количество публикаций РФ, то отчетливо видно, что на период с 2001 по 2005 год приходится локальный максимум публикационной активности. В 2006 происходит резкий спад числа публикаций, более чем на 7 тысяч, это соответствует примерно 15%. К показателям 2005 года РФ возвращается только в 2011-2012 годах. С 2013 по 2019 год рост числа публикаций существенно ускоряется, значение производной резко меняется. Все вышесказанное позволяет подтвердить, что масштаб проблемы постепенно уменьшается. В 2018-2019 годах, согласно нашим подсчетам, количество потерянных Россией публикаций не превосходило одного процента.

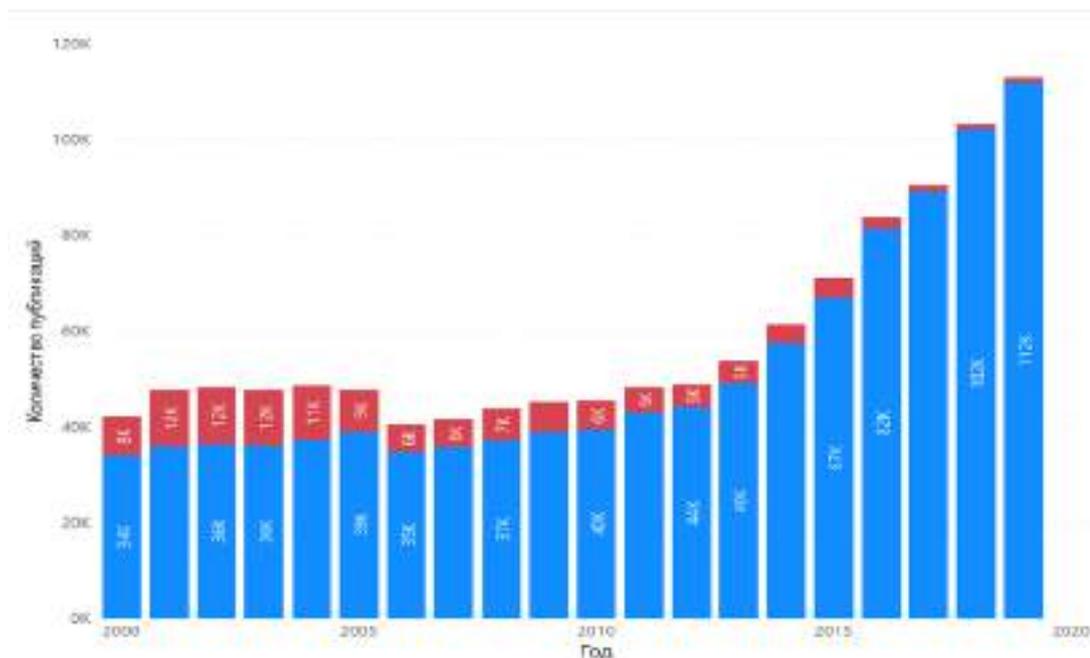


Рисунок 2 – Временная зависимость количества публикаций РФ, с разделением на учтенные (●) и неучтенные (●) в Scopus

Рассмотрим распределение неучтенных публикаций по областям знаний. Поскольку мы использовали данные Scopus воспользуемся классификацией ASJC, используемой в этой БД. Для начала поймем, что происходит на верхнем уровне, уровне отраслей знаний. На рисунке 3 представлено распределение всех публикаций РФ за рассматриваемый период (с 2000 по 2019 годы) по отраслям знаний. Из рисунка ясно, что львиная доля публикаций – 64%, относится к «Physical Sciences». Значительная доля публикаций попадает в области знаний «Life Sciences» и «Health Sciences», 12,1% и 12,5% соответственно. Самой малочисленной оказалась область «Social Sciences» – 6,3%. Оставшиеся 5% публикаций либо не отнесены ни к одной из областей, либо считаются мультидисциплинарными исследованиями. Совсем непохожую картину мы увидим, если посмотрим на распределение по отраслям знаний неучтенных публикаций (рисунок 4). Количество неучтенных публикаций в области «Health Sciences» практически совпадает с таковым для «Physical Sciences», 41,1% и 43,6% соответственно. Это говорит о существенном различии структуры распределения неучтенных публикаций в этих группах. Временная зависимость количества публикаций

из области «Physical Sciences» с разделением на учтенные и неучтенные очень похожа на общую картину (см. рисунок 2). Что неудивительно – мажоритарный класс (64% от общего количества) задает тенденции для всей выборки. В целом для физических наук характерно чуть меньшее количество неучтенных публикаций, и более ранняя тенденция к их снижению. Количество неучтенных публикаций в этой области резко, скачкообразно снизилось в период с 2005 по 2007 год, дальнейшее снижение было монотонным и повторяло общий тренд. Картина для публикаций из области «Health Sciences» существенно отличается.

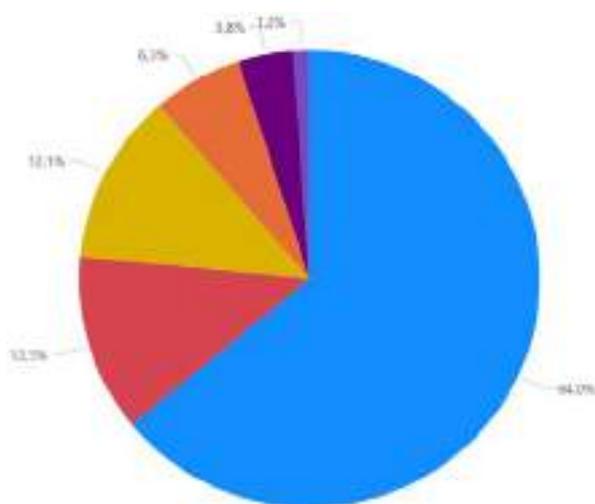


Рисунок 3 – Распределение российских публикаций за период с 2000 по 2019 год, по супергруппам классификации ASJC. ● – Physical Sciences, ● – Health Sciences, ● – Life Sciences, ● – Social Sciences, ● – Multidisciplinary, ● – отрасль знаний не указана

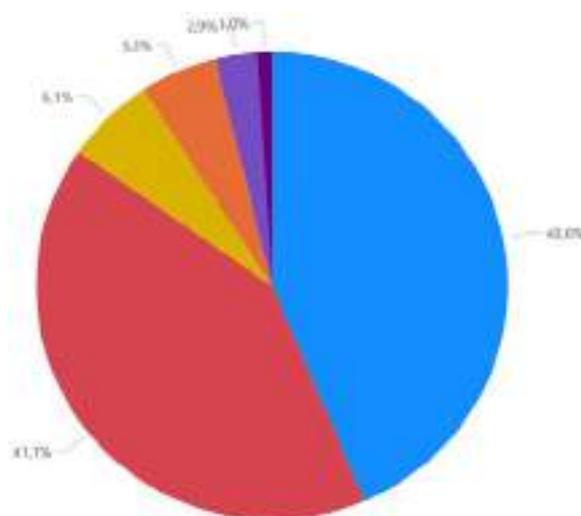


Рисунок 4 – Распределение неучтенных российских публикаций за период с 2000 по 2019 год, по супергруппам классификации ASJC. ● – Physical Sciences, ● – Health Sciences, ● – Life Sciences, ● – Social Sciences, ● – Multidisciplinary, ● – отрасль знаний не указана

Из рисунка 5 понятно, что доля неучтенных публикаций в области наук соответствующей категории «Health Sciences» существенно выше, чем в общей выборке публикаций РФ (см. рисунок 2). В период с 2006 по 2010 год количество неучтенных публикаций превосходило 50% от общего числа публикаций. Пик пришелся на 2008 год, менее 42% всех публикаций РФ в этой предметной области Scopus посчитал российскими, а 58% потеряны. Ситуация

улучшалась достаточно медленно и монотонно до 2015 года, а в период с 2015 по 2017 произошел резкий скачок в качестве учета российских публикаций, в 2018 году менее 2% публикаций потеряно, а в 2019 цифра приблизилась к 1%.

Ситуация с отраслью знаний «Life Sciences» наоборот лучше, чем в среднем по выборке: за весь рассматриваемый период число неучтенных публикаций единожды незначительно превысило 10% рубеж в 2009. В остальные годы с 2000 по 2013 число неучтенных публикаций колебалось в диапазоне 6-9%. Значительное снижение числа неучтенных публикаций происходило скачкообразно, двумя большими скачками в период с 2013 по 2017 год. В 2019 году доля неучтенных публикаций составила чуть более 0,4%.

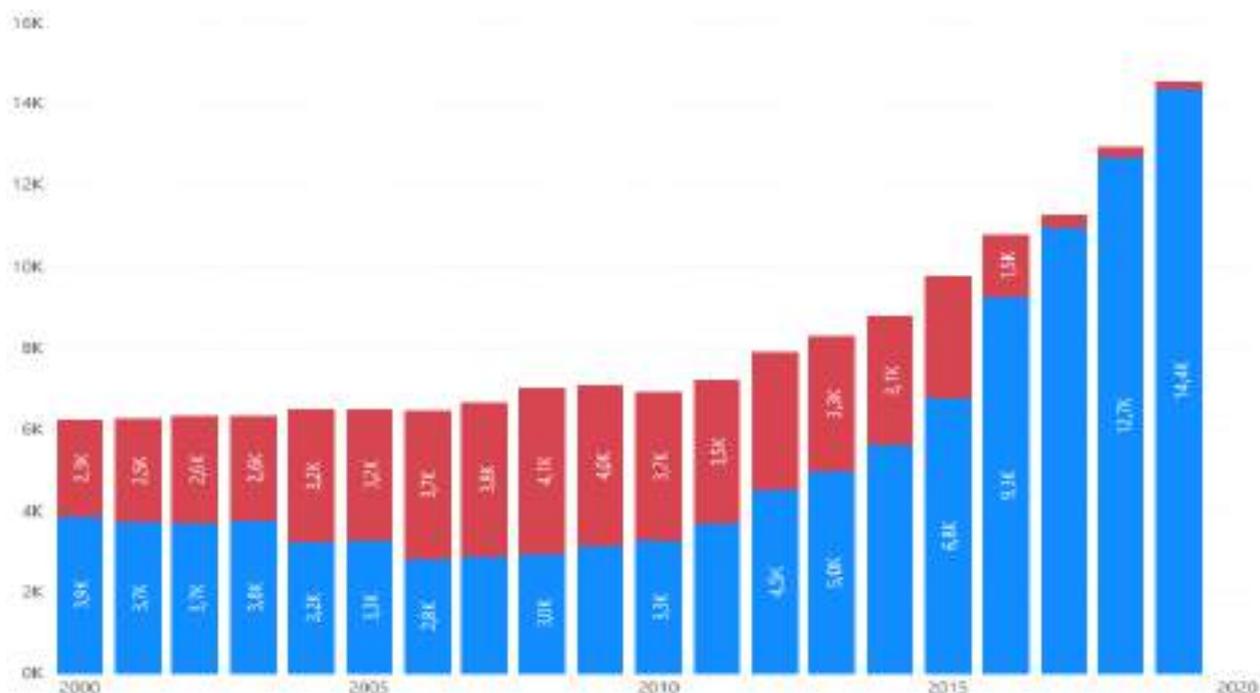


Рисунок 5 – Временная зависимость неучтенных публикаций РФ для отрасли знаний «Health Sciences». ● – учтенные в Scopus, ● – неучтенные в Scopus

Ситуация с наиболее малочисленной отраслью «Social Sciences» напоминает ситуацию «Health Sciences», но в меньших масштабах. Распределение менее стабильно, изменения не монотонны, но в целом число неучтенных публикаций превышает среднее по выборке, но раза в 1,5-2 ниже, чем для «Health Sciences». Пик неучтенных публикаций смещен в область

второй пятилетки, максимум приходится на 2008 год. Интенсивное улучшение ситуации наблюдалось с 2011 по 2016. В последние годы число неучтенных публикаций снижается медленнее, чем в других областях, возможно это связано с более стремительным ростом общего числа публикаций в этой области.

Для категории мультидисциплинарных исследований критичен период с 2000 по 2003 год, в эти годы число неучтенных публикаций из этой области составляло от 75% до 96% от общего количества. Начиная с 2004 года, ситуация кардинально улучшилось – число неучтенных публикаций ни разу не превысило 6%. И это несмотря на довольно сильные колебания общего числа публикаций в период с 2005 по 2013, например, в 2006 общее число российских публикаций из этой области снизилось более чем в 11 раз по сравнению с 2005 (с 1142 до 97). Неожиданным оказалось и поведение небольшой группы публикаций, для которых супергруппа не указана. Для этой группы число неучтенных публикаций было невелико на всем исследуемом временном отрезке, с максимумом в 7.6% от общего числа публикаций в 2000 году.

Проведенный анализ показывает, что учет неаффилированных с РФ публикации российских авторов заметно меняет характер изменения публикационного потока со временем, особенно в отдельных областях, прежде всего, в медицинских науках. Для полноценного включения этих публикаций в анализ необходимо разработать алгоритм привязки этих публикаций к аффилиациям. Мы предполагаем, что удовлетворительные результаты можно будет получить, используя аффилиационную историю авторов публикаций и дополнительные косвенные данные.

1.2 Тематическая классификация публикаций на основе теоретико-информационного подхода

Тематическая рубрикация научных публикаций повсеместно используется в разных видах наукометрического анализа, а также является основой для расчета взвешенных метрик, таких как «Field-Weighted Citation

Impact» (FWCI) в Scopus, Category Normalized Citation Impact (CNCI) в InCites и т.д. В базах данных Scopus, WoS базовая тематическая рубрикация происходит на уровне изданий, а не отдельных статей [9], т.е. всем статьям одного издания присваиваются одни и те же категории. В Scopus используется классификатор All-Science Journal Classification (ASJC). На его первом уровне расположены 4 общих научных направлений: биологические науки, физические науки, медицина, социальные и гуманитарные науки. На втором - 27 областей наук, которые, в свою очередь, разделены на 333 узких категорий. В WoS классификатор также состоит из трех уровней. На основном уровне расположены шесть областей наук, которые разделены на 39 подуровней. На третьем подуровне классификатора WoS содержится 253 категории.

Таким образом, отдельной публикации могут быть присвоены тематические рубрики, не имеющие к ней никакого отношения. Отметим, что из 24272 активных журналов, индексируемых в Scopus на июнь 2020 г. только 7825 присвоена одна тематическая рубрика, у 16447 журналов две и более рубрик, причем 83 журналам присвоена только рубрика «General». Важность присваивания публикации своей собственной, а не журнальной категории, можно также продемонстрировать и на следующем примере. Создадим запрос на отображение в БД Scopus публикаций области «Социальные науки» за период 2014-2019 гг. и отсортируем по убыванию числа цитирования. Первые две публикации этого рейтинга «Very deep convolutional networks for large-scale image recognition» и «Adam: A method for stochastic optimization» соответствуют мультидисциплинарному источнику «3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings». Названия и аннотации этих публикаций показывают, что, скорее всего, они должны принадлежать области «Компьютерные науки», но не «Социальные науки», тем самым искажая данные о самых высокоцитируемых публикациях этой области. При этом если эти две публикации имеют 6652 и 6467 цитирований соответственно, то следующая за ними публикация имеет

только «2757» цитирований. Отметим также, что из-за мультидисциплинарности источника эти публикации также встречаются в области «Компьютерные науки».

Отметим также, что некорректная классификация приводит не только к искажению результатов различных исследований, но и может заметно затруднять и ухудшать поиск нужных публикаций в ББД. Рассмотрим следующий пример. Всего по данным Scopus за 2019 год в области наук «Медицина» появилось 799381 новых статей. Однако многие из этих статей вошли в мультидисциплинарные журналы, тематика которых может быть далека от медицины. Например, к таким областям относятся «Социальные науки», «Науки об окружающей среде», «Технические науки» и др. В таблице 3 приведены потери по публикациям в каждой области наук, если из результатов поиска исключить мультидисциплинарные журналы, где одна из областей наук – «Медицина».

Необходимо отметить, что в последние годы активно развиваются методики тематической классификации на уровне отдельных статей, результаты которых появляются и в базах данных, аналитических инструментах. В 2017 году в аналитическом инструменте компании Elsevier SciVal для уточнения тематик публикаций появился новый алгоритм Topic Prominence, направленный на выявление новых актуальных тем. Публикации Scopus были сгруппированы по темам по результатам прямого анализа цитирования. Далее происходила иерархическая кластеризация публикаций. Третий этап был посвящен названию получившихся кластеров метками на основе терминов заголовков и аннотаций публикаций, находящихся в этой группе. Это решение было построено на основе работ [10,11]. Clarivate Analytics также внедряет классификационную схему на уровне отдельных публикаций – Citation Topics в аналитический продукт InCites. В отличие от Topic Prominence эта классификация иерархическая и может быть использована для анализа публикационного потока на разном уровне детальности.

Однако, эти детальные классификаторы пока не используются для классификации, поиска и расчета взвешенных индикаторов. Тематический анализ публикационного потока на основе журнальной классификации будет значительно искажен без дополнительного уточнения тематических рубрик отдельных публикаций.

Таблица 3 – Потери по областям наук без области «Медицина» в 2019 г. в БД Scopus

Область наук	Количество публикаций	Количество публикаций без области Медицина	Потери в области наук
Health Professions	41566	12645	69,60%
Nursing	58466	26332	55,00%
Immunology and Microbiology	87327	41888	52,00%
Neuroscience	80209	41774	47,90%
Pharmacology, Toxicology and Pharmaceutics	100023	67099	32,90%
Psychology	81410	57278	29,60%
Biochemistry, Genetics and Molecular Biology	358081	252392	29,50%
Dentistry	18595	13409	27,90%
Veterinary	25607	22909	10,50%
Decision Sciences	74918	69448	7,30%
Social Sciences	337054	312752	7,20%
Environmental Science	238197	221354	7,10%
Agricultural and Biological Sciences	250414	241078	3,70%
Engineering	733640	708182	3,50%
Computer Science	506379	488622	3,50%
Arts and Humanities	138350	133542	3,50%
Physics and Astronomy	399144	386543	3,20%
Chemical Engineering	167579	162823	2,80%
Chemistry	281893	275444	2,30%
Materials Science	375432	367580	2,10%
Energy	171638	168802	1,70%
Mathematics	269594	265542	1,50%
Business, Management and Accounting	106171	104543	1,50%
Economics, Econometrics and Finance	60495	59900	1,00%
Earth and Planetary Sciences	168604	167858	0,40%

Ранее участниками этого проекта был предложен метод классификации на основе сжатия данных (DCM). Его суть состоит в следующем. Пусть есть n

научных тематик: X_i , $i=1, \dots, n$; а y – это классифицируемый научный текст, научную тематику которого нужно определить зная, что он точно относится к одной из указанных тематик. Пусть φ – это «архиватор», который может быть применен для сжатия любого множества текстов. Определим функцию $C(y/x_1, \dots, x_k) = (|\varphi(x_1, \dots, x_k, y)| - |\varphi(x_1, \dots, x_k)|) / |y|$ – степень сжатия (в %) текстов y с множеством текстов (x_1, \dots, x_k) , где $|y|$ – размер текста. Таким образом, полагается, что текст y принадлежит тематике X_j , т.е. $C(y/ x_1^j, x_2^j, \dots, x_{m_j}^j) = \min_{i=1, \dots, n} C(y/ x_1^i, x_2^i, \dots, x_{m_i}^i)$.

Проведен сравнительный анализ метода классификации на основе сжатия данных с другими методами классификации текстов, а также сопоставление результатов автоматической классификации с экспертной оценкой. Перечисленные ниже методы были реализованы при помощи бесплатной библиотеки `scikit-learn` для языка программирования Python. Отметим, что в исследовании не проводился подробный анализ выбора параметров для методов из библиотеки `scikit-learn`, в том числе и по оптимальному размеру обучающей выборки. Указанные параметры были подобраны экспериментальным путем:

- Logistic regression (LR) с параметрами: `verbose=1, solver='liblinear', random_state=0, C=3, penalty='l2', max_iter=1000`;
- Naive Bayes classifier (NB) методом: `MultinomialNB()`;
- Random forest (RF) с параметрами: `n_estimators=200, max_depth=20`;
- Support vector machine (SVM) с параметрами: `C=1.0, kernel='linear', degree=8, gamma='auto'`;
- k-nearest neighbors (KNN) с параметрами: `n_neighbors=6, algorithm='brute'`.

Методы классификации были применены к полным англоязычным текстам препринтов из архива научных публикаций `arXiv.org` и аннотациям публикаций из базы данных `Scopus`. Для всех методов были использованы одни и те же обучающие выборки и тестовые файлы. Экспертная оценка

качества классификации методом DCM, была проведена на основе аннотаций англоязычных публикаций в журнале «Геология и геофизика».

На рисунке 6 приведено сравнение точности классификации различными методами в зависимости от вида источника данных. Результаты показывают, что на аннотациях и полнотекстовых документах все методы, кроме KNN, показывают точность от 84% и выше. На полнотекстовых документах наилучшую точность показал метод DCM (91%), следующий за ним – SVM (88%). На аннотациях – SVM (89%), DCM (88%) и LR (88%).

Наилучшую точность по результатам экспертной оценки показывает DCM (80%). Точность остальных методов не превышает 61%, однако стоит заметить, что экспертная оценка проводилась только для метода DCM, остальные методы являются в этом случае скорее уточнением к полученным результатам. В таблице 4 приведены основные метрики оценки качества классификации для полных текстов и аннотаций: Accuracy, Precision и Recall.

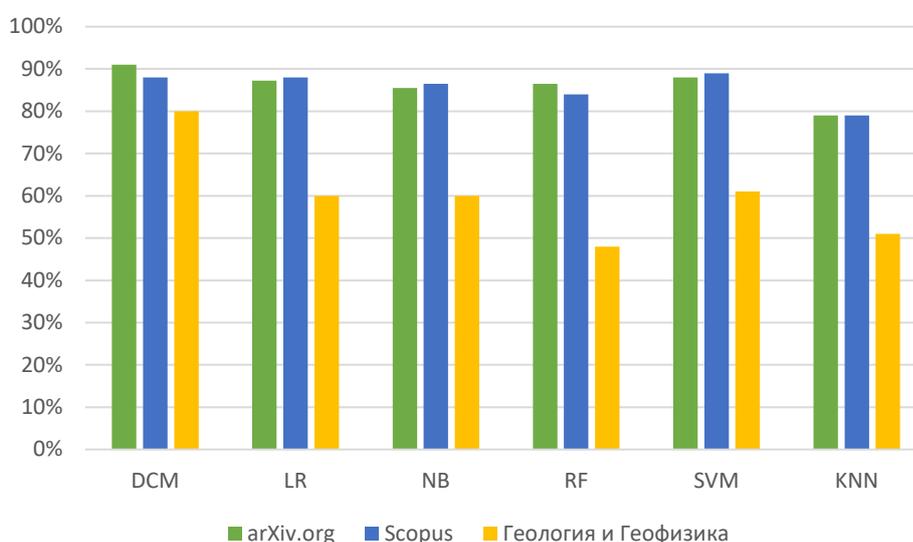


Рисунок 6 – Сравнение точности классификации различными методами в зависимости от источника данных

Таблица 4 – Сравнение основных характеристик классификации в разных методах для полных текстов и аннотаций

	Accuracy		Precision		Recall	
	Полные тексты	Аннотации	Полные тексты	Аннотации	Полные тексты	Аннотации
DCM	91%	88%	91%	87%	92%	88%
LR	87%	88%	87%	88%	88%	88%
NB	86%	87%	85%	87%	87%	87%
RF	87%	84%	86%	84%	87%	85%
SVM	88%	89%	88%	89%	89%	89%
KNN	79%	79%	79%	79%	81%	80%

Показано, что автоматическое создание ядер для классификации аннотаций публикаций, индексируемых в БД Scopus, является достаточно трудоемким процессом, а в ряде случаев невозможным вообще. Во-первых, это связано с ограничением на выгрузку данных, установленным в Scopus, и отсутствием названий категорий в Scopus Search API. Во-вторых, во многих областях наук полностью отсутствуют журналы и, соответственно, публикации, у которых указана только одна категория. Также это указывает на то, что классификация, используемая в настоящий момент в Scopus и SciVal, может быть не до конца достоверной. Например, по данным SciVal по количеству публикаций категория «Theoretical computer science» находится на втором месте среди всех публикаций области «Mathematics». Наши исследования показали, что эта категория является одной из самых малочисленных категорий как с точки зрения присутствия журналов, так и публикаций только с этой категорией. Выделения публикаций из мультидисциплинарных журналов, у которых может быть конкретно эта категория, до настоящего момента проведено не было.

Таким образом, многие исследования, основанные на использовании проклассифицированных по ASJC публикаций, могут иметь некоторые неточности. Классификация публикаций методом, основанным на сжатии данных, по всем 26 областям наук невозможна в виду их обширности, а также изначальной классификации Scopus. Часто в разных областях наук находятся

терминологически близкие категории, что затрудняет отнесение публикации к верной области. В основном в ядра попадают публикации лишь из нескольких категорий (зачастую это категория «general» или «all»), что не дает возможности сделать ядро терминологически «узким». Таким образом, проведенное исследование показывает, что применение метода классификации, основанного на сжатии данных, а также других методов, базирующихся на лексической близости, к аннотациям публикаций, индексируемых в базе данных Scopus, возможно только к ограниченному числу терминологически различающихся категорий.

1.3 Метод уточнения тематики публикаций на основе публикационной истории авторов

Рассмотренные выше методы классификации являются ресурсоемкими и в общем случае плохо подходят для уточнения классификации по тематике большого количества публикаций. В связи с этим нами был разработан алгоритм на основе анализа публикационной истории авторов. Для каждой публикации формируется вектор тематических рубрик, составленный из тематических рубрик всех публикаций каждого из авторов анализируемой публикации с расчетным количеством таких публикаций отсортированный по убыванию. Таким образом в начале вектора оказываются тематические категории, с наибольшей вероятностью соответствующие тематике анализируемой публикации. Этот метод основан на предположении, что отправка статьи в конкретный журнал эквивалентна заявке ее авторов о соответствии этой статьи тематикам этого журнала. При этом сохраняется неопределенность, связанная с мультидисциплинарными и политематическими журналами. Однако, чем больше таких заявок (длиннее история публикаций авторов), тем больше уверенности в тематической направленности анализируемой публикации.

Ограничения метода связаны с слишком короткой публикационной историей авторов (одна статья или несколько статей в одном или эквивалентных по наборам тематических рубрик журналах), а также в случае

смены тематической направленности исследований авторов. Второй случай оценить затруднительно, а вот первый легко определяется по наличию в тематическом векторе нескольких рубрик с одинаковым количеством публикаций. В проанализированном нами множестве публикаций (более 1,7 млн) при уточнении классификации по второму уровню ASJC таких нашлось чуть менее 3,7 %.

Для определения итоговой тематической рубрики анализируемой публикации могут применяться два подхода – в первом публикации назначается единственная (лидирующая) рубрика, во втором – несколько рубрик (с усечением вектора тематических рубрик по какому-либо признаку или по полному вектору) с весами. Для анализа публикационного потока российских исследователей мы применили первый подход. Полученные результаты тематической рубрикации на рисунке 7 заметно отличаются от результатов, полученных с использованием аналитических инструментов Scopus на рисунке 8. Так, значительно большей оказывается доля публикаций в областях «Физика и Астрономия», «Медицина», «Науки о Земле», заметно меняется порядок приоритетов областей исследований.

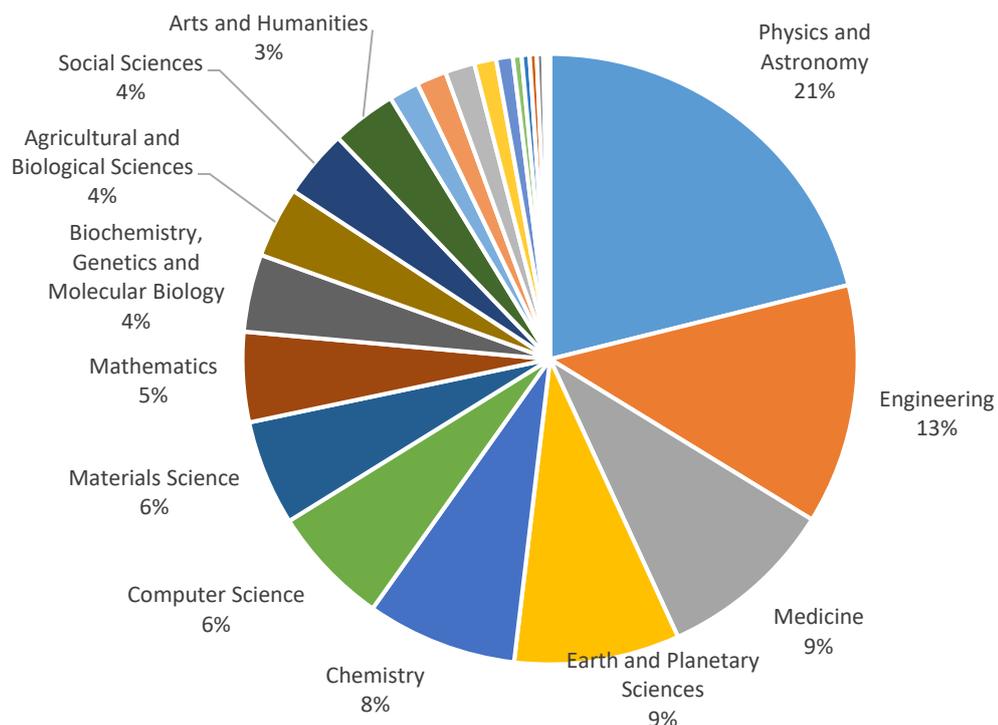


Рисунок 7 – Результаты уточненной тематической классификации российских публикаций за 2019 г.

Documents by subject area

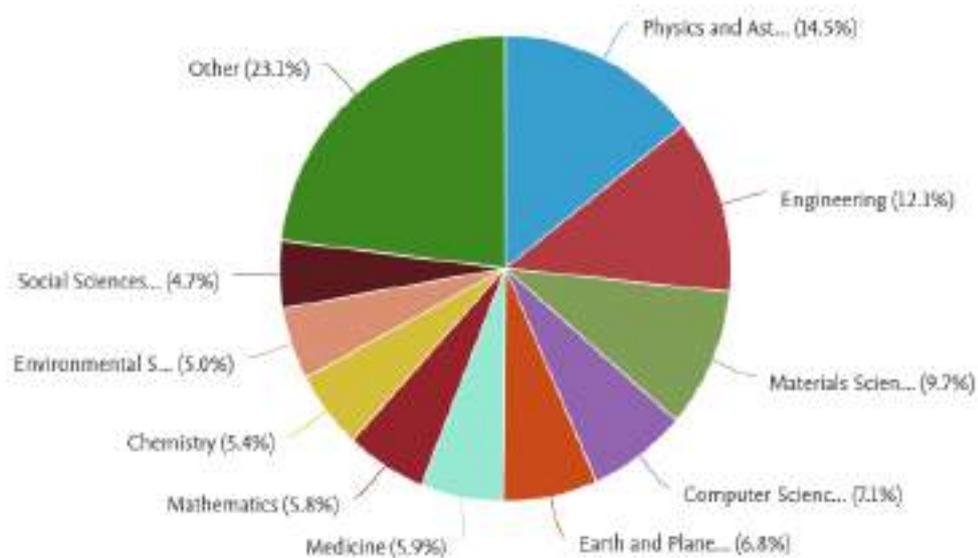


Рисунок 8 – Тематическая структура российских публикаций за 2019 г. в аналитическом интерфейсе Scopus

2 Общий анализ документопотока российских публикаций, индексируемого в Scopus

Исследование документопотока российских публикаций выполнено на основе данных о публикациях в изданиях, индексируемых в БД Scopus. Выбор этого источника обусловлен высоким качеством библиографической информации [12], а также следующими обстоятельствами:

- Доступностью программных интерфейсов для извлечения данных в рамках национальной подписки.
- Полнотой охвата, индексированием значительного количества высококачественных журналов на национальных языках, изданий в области общественных и гуманитарных наук.
- Наличием идентификаторов авторов и организаций, что делает возможным некоторые используемые виды анализа.

Анализ [13] показывает, что, несмотря на постоянные усилия команды Elsevier по повышению качества библиографической информации в Scopus, тем не менее, присутствует значительное количество ошибок идентификации авторов и их аффилиаций, вносящих искажения в результаты анализа. При обработке данных и выполнении расчетов был предпринят ряд шагов для снижения влияния этих ошибок. Так, для аффилиаций была проведена дедупликация профилей с привязкой организаций к уникальному идентификатору – российскому индивидуальному номеру налогоплательщика (ИНН).

Часть инструментов государственной научной политики ориентирована на использование БД Web of Science (WoS) в качестве основного источника данных о научных публикациях. В частности, в задачах оценки результативности как мера качества научного журнала используется квартиль журнала в рейтинге Journal Citation Reports (JCR). В связи с этим, для качественного анализа документопотока данные о квартилях журналов в JCR были привязаны к данным публикаций, полученным из БД Scopus.

Дополнительно были использованы данные рейтинга Scimago Journal Ranking (SJR), ориентированного на БД Scopus.

2.1 Полнота охвата публикационного потока в БД Scopus

Основными показателями, характеризующими публикационную результативность российской науки в Национальном проекте «Наука», ведомственной и федеральной оценке результативности и других инструментах национальной научной политики выступают количество публикаций, проиндексированных в БД WoS и/или Scopus.

Традиционно считается, что Web of Science – более престижная БД, качество журналов в ней выше, отбор строже. С этим принято связывать относительно худшие позиции России в рейтинге стран по данным этой БД. Однако по общему количеству индексируемых документов Scopus и Web of Science Core Collection (WoS CC) отличаются незначительно (рисунок 9 –), особенно после создания индекса Emerging Sources Citation Index в 2015 г. и включения его в Core Collection.

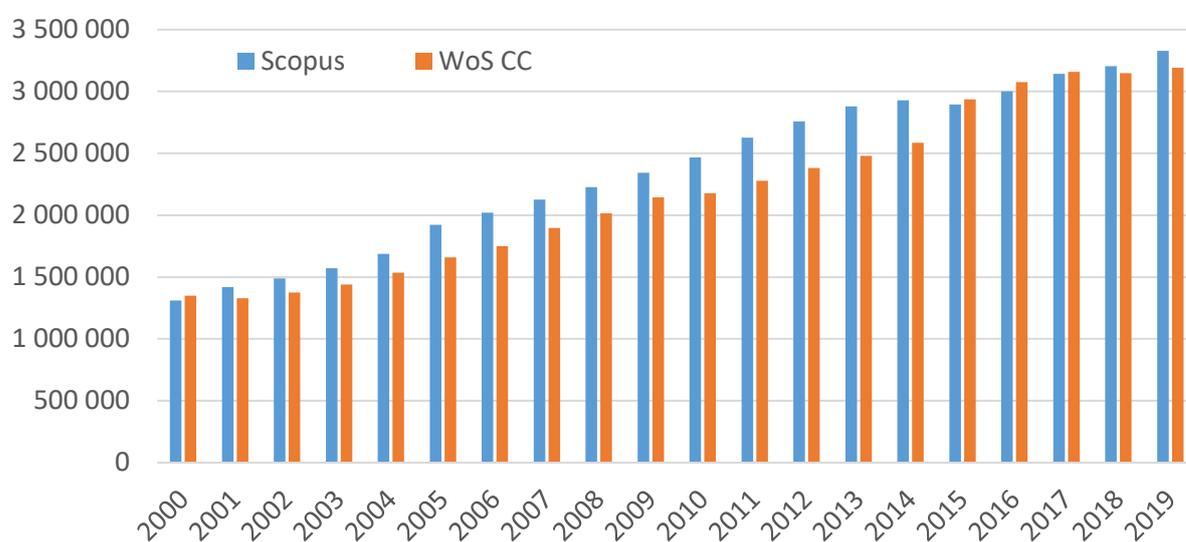


Рисунок 9 – Количество документов, проиндексированных в БД Scopus и WoS CC по годам

Нет существенных отличий и по индексации различных типов источников – доля научных журналов в Scopus постепенно росла последние

годы и составляет в 2019 г. 80%, в WoS CC доля журнальных индексов также около 80%. На труды конференций в Scopus приходится около 15%, доля соответствующих индексов в WoS CC даже чуть больше – 17%. Данные о структуре баз данных представлены в таблице 5.

В БД WoS CC в меньшей степени представлены многие развивающиеся страны. Так, публикации с аффилиацией США занимают 19,12% общего количества проиндексированных в WoS CC в 2019 г. и только 16,22% – в Scopus, Великобритании – 6,05% / 5,05%, Германии – 4,49% / 4,39%, Италии – 3,09% / 3,00%, Франции – 2,89% / 2,85%, Канады – 3,03 % / 2,74%. Если посмотреть на страны БРИКС, то у четырех стран будут обратные соотношения: КНР – 14,28% (WoS CC) / 16,34% (Scopus), Индия – 3,32% / 4,57%, Россия – 2,05% / 2,72%, ЮАР – 0,65% / 0,66%, только у Бразилии соотношение в пользу WoS – 2,09% / 2,04%. Можно предположить, что наблюдается разница в дисциплинарной структуре и отборе источников в этих БД с большей ориентированностью WoS CC на издания стран первого мира, прежде всего Великобритании и Северной Америки, которая приводит к меньшей «инклюзивности» этой базы данных.

Таблица 5 – Структура баз данных WoS CC и Scopus за последние 5 лет

БД	Индекс / тип источника	2015	2016	2017	2018	2019
WoS CC	Science Citation Index Expanded	56.9%	56.2%	55.7%	58.0%	61.7%
	Social Sciences Citation Index	8.9%	9.2%	9.2%	9.9%	11.1%
	Arts and Humanities Citation Index	3.8%	3.5%	3.4%	3.2%	3.2%
	Conference Proceedings Citation Index - Science	16.2%	16.5%	16.9%	15.6%	12.3%
	Conference Proceedings Citation Index - Social Sciences and Humanities	1.6%	1.7%	1.6%	1.1%	0.6%
	Book Citation Index - Science	1.6%	1.8%	1.9%	1.3%	0.8%
	Book Citation Index - Social Sciences and Humanities	2.6%	2.5%	2.9%	2.0%	1.2%
	Emerging Sources Citation Index	8.5%	8.5%	8.5%	8.9%	9.2%
Scopus	Journals	78.8%	78.6%	77.8%	78.4%	80.2%
	Conference Proceedings	12.5%	13.1%	13.5%	14.2%	14.7%
	Books	7.9%	7.5%	8.0%	6.8%	4.7%
	Other	0.8%	0.7%	0.6%	0.6%	0.4%

Так как покрытие WoS CC заметно отличается от покрытия Scopus, это ставит вопрос о полноте полученных из Scopus списков публикаций российских организаций. Анализ показывает, что в Scopus не индексируется почти 3,5 тысячи наименований журналов, входящих в основные индексы Web of Science. Наибольшая разница как в количестве наименований, так и в российских статьях, относится к индексу Emerging Sources Citation Index. Общая информация по изданиям и российским статьям в них, которые индексируются в Web of Science и отсутствуют в Scopus приведена в таблице 62.2.

В индексы Conference Proceedings Citation Index- Science (CPCI-S), Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH), Book Citation Index– Science (BKCI-S), Book Citation Index– Social Sciences & Humanities (BKCI-SSH) также включено более 22 тыс. публикаций с российскими аффилиациями. Не менее половины этих публикаций индексируется и в Scopus. На платформе Web of Science совместно с Научной

электронной библиотекой eLibrary.ru развернут индекс Russian Science Citation Index в который в 2019 г. входили 777 журналов (чуть более 53 тыс. публикаций). Из них 291 индексируется в Scopus, проиндексировано 12 тыс. статей. Таким образом, полный и корректный анализ публикационного потока невозможен без данных из Web of Science. Однако российские публикации из индексов WoS, не проиндексированные в Scopus, в основном относятся к RSCI и ESCI, в журналы верхних квартилей по JCR попадает очень немного отсутствующих в Scopus публикаций. В связи с этим, полученные из Scopus данные позволяют провести достаточно полный анализ публикационного потока российских исследователей и его сравнение с результатами ученых из других стран.

Таблица 6 – Количество изданий и российских статей за 2019 г. в них по основным индексам Web of Science Core Collection, не индексируемых в Scopus

Индекс WoS	квартиль по JCR2018	изданий	статей
Arts & Humanities Citation Index (A&HCI)	Q3	3	0
	Q4	4	0
	-	38	90
ESCI	-	3113	8153
Science Citation Index Expanded (SCI-EXPANDED)	Q1	37	146
	Q2	26	71
	Q3	25	151
	Q4	33	13
	-	121	135
Social Sciences Citation Index (SSCI)	Q1	12	1
	Q2	11	5
	Q3	9	1
	Q4	14	0
	-	25	7
Всего		3471	8773

2.2 Динамика, позиция в мире, целый и фракционный счет

Данные Scopus за 2019 год демонстрируют несколько замечательных событий – впервые Китайская народная республика обошла Соединенные Штаты по суммарному годовому количеству публикаций и вышла на первое место, Индия обошла Германию и выдвинулась на четвертое место, а Россия

продолжает уверенное восхождение, поднимаясь на одну позицию вот уже пятый год подряд (рисунок 10) и превысив значение в 115 тыс. публикаций. Как уже неоднократно отмечалось, быстрый рост количества российских публикаций начался в 2013 г. и был, по всей видимости, вызван прежде всего поставленной в майских указах 2012 г. целью увеличения доли России в базе данных Web of Science до 2,44% и подкрепленном целями майского указа 2018 г. Количество российских публикаций в базе данных Scopus растет более чем на 10% ежегодно (за исключением 2017 г. – 9,3%), пиковый рост был достигнут в 2016 г. – 21,4%. На девятом месте рейтинга сейчас находится Канада с 116 тыс. публикаций, на восьмом – Франция (более 120 тыс.), эти страны заметно отстают от РФ в темпах роста, что позволяет надеяться на дальнейшее «восхождение» российской науки в мировом рейтинге.

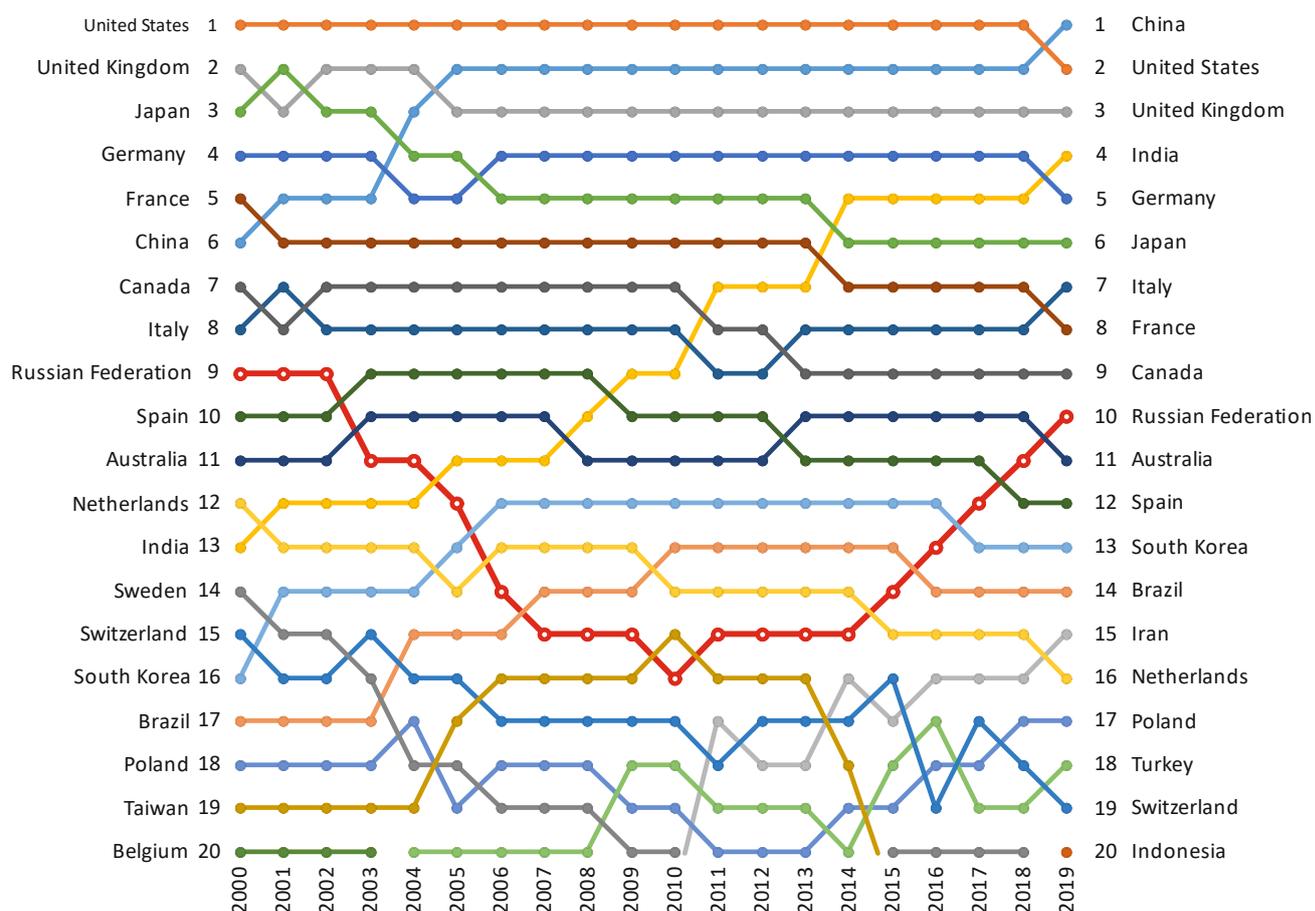


Рисунок 10 – Динамика странового рейтинга по количеству публикаций в БД Scopus

Наши результаты в БД Web of Science Core Collection (WoS CC) выглядят заметно скромней – пока что мы занимаем 14 место по результатам 2019 г. (рисунок 11). Судя по всему, в этой БД еще далеко не закончена индексация трудов многих конференций и ситуация может измениться, хоть и незначительно.

Приведенные выше страновые рейтинги рассчитаны на основании целого счета – метода, при котором каждая публикация считается как целая единица для каждой страны в списке аффилиаций авторов этой публикации. Во многом более корректным подходом является фракционный счет, учитывающий вклад каждого автора (равными долями) и, соответственно, каждой аффилиации этого автора. Подробный разбор метода и особенностей фракционного счета приведен нами в работе [14]. Доля российских публикаций в мировом потоке, рассчитанная фракционным счетом, отличается от результатов целого счета незначительно, что отражает достаточно низкий уровень международных коллабораций российских исследователей (таблица 7Таблица 7 –).

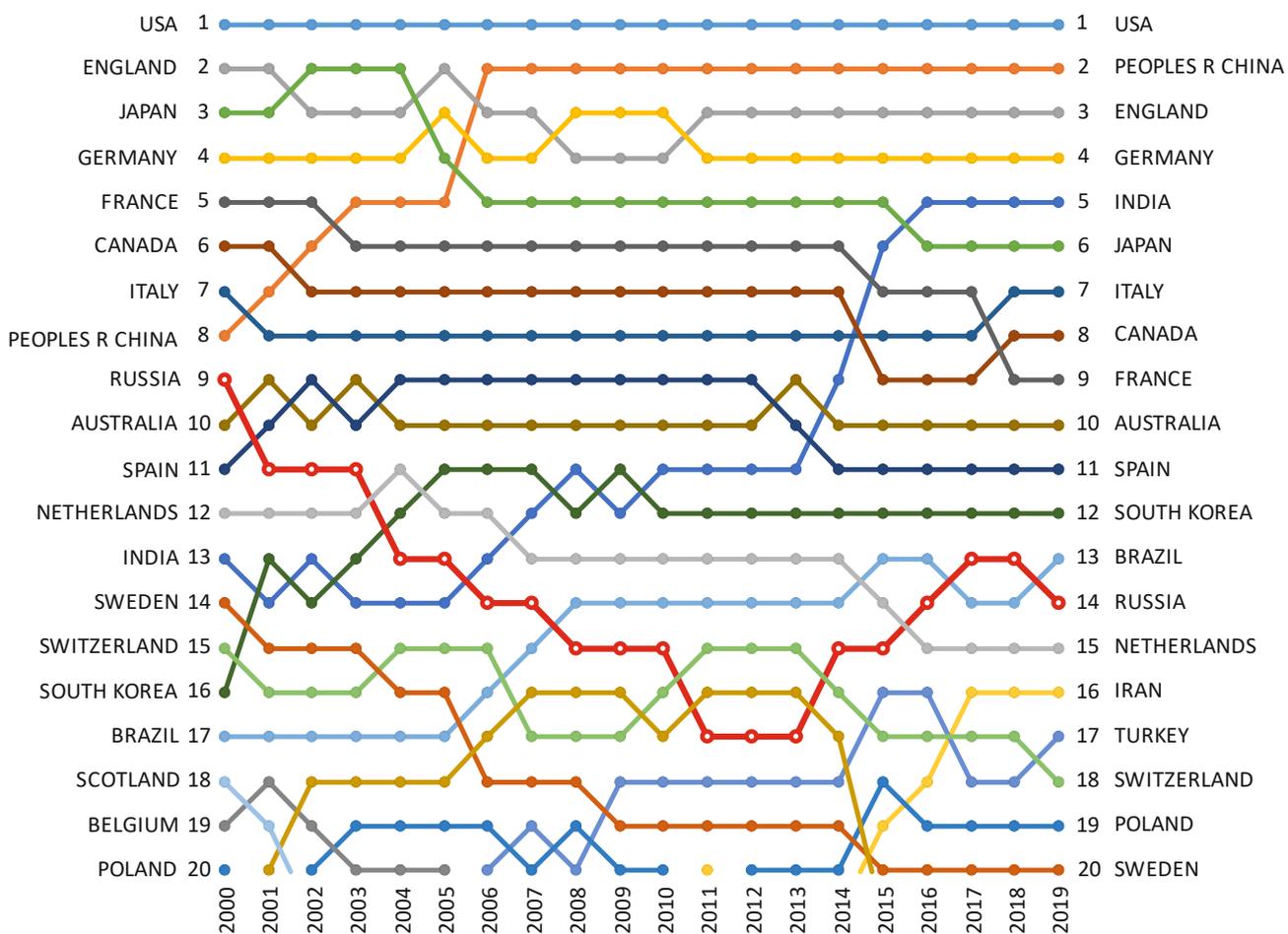


Рисунок 11 – Динамика странового рейтинга по количеству публикаций в БД Web of Science Core Collection

Таблица 7 – Динамика публикационной активности российских исследователей целым и фракционным счетом и доля РФ в мировом публикационном потоке по данным Scopus.

Год	Целый счет	Фракционный счет	Разница, в % от целого счета	Мир	Доля целым счетом	Доля фракционным счетом
2000	34 732	34 151	1.67%	1 309 338	2.65%	2.61%
2001	36 763	36 032	1.99%	1 419 920	2.59%	2.54%
2002	37 488	36 788	1.87%	1 489 605	2.52%	2.47%
2003	37 541	36 637	2.41%	1 572 288	2.39%	2.33%
2004	38 258	37 395	2.26%	1 688 093	2.27%	2.22%
2005	40 281	39 142	2.83%	1 921 823	2.10%	2.04%
2006	36 130	35 054	2.98%	2 019 999	1.79%	1.74%
2007	37 045	36 162	2.38%	2 125 176	1.74%	1.70%
2008	38 832	37 737	2.82%	2 227 805	1.74%	1.69%
2009	39 912	39 106	2.02%	2 343 401	1.70%	1.67%
2010	40 698	39 572	2.77%	2 465 978	1.65%	1.60%
2011	44 173	43 037	2.57%	2 626 431	1.68%	1.64%
2012	45 253	44 064	2.63%	2 757 363	1.64%	1.60%
2013	50 253	49 199	2.10%	2 878 690	1.75%	1.71%
2014	58 899	57 539	2.31%	2 927 603	2.01%	1.97%
2015	68 274	67 250	1.50%	2 894 618	2.36%	2.32%
2016	82 899	81 779	1.35%	2 999 639	2.76%	2.73%
2017	90 573	89 547	1.13%	3 143 159	2.88%	2.85%
2018	104 243	102 722	1.46%	3 203 977	3.25%	3.21%
2019	115 656	112 237	2.96%	3 329 707	3.47%	3.37%

2.3 Структура публикационного потока по типам документов

Основной объем российских публикаций – это статьи (Ar – article) и обзоры (Re – review). В процессе подготовки [15] нами было выявлено, что определение типа публикации в БД Scopus не всегда верное, часть исследовательских статей может быть помечена типом Review и наоборот. Кроме того, растущее многообразие видов обзоров значительно изменяет представление об этом типе документа как о вторичном, неоригинальном, так как такие виды, как аналитические и систематические обзоры часто носят характер самостоятельных исследований. В связи с этим мы рассматриваем эти типы документов вместе. Второе место занимают материалы конференций (CP – conference proceedings), опубликованные в сборниках трудов,

специализированных сериях и журналах. Динамика структуры документопотока по типам представлена в таблице 8.

Таблица 8 – Динамика структуры потока российских публикаций по типам документов

Год	Ar+Re	CP	Другое	Ar+Re	CP	Другое
2000	31757	2784	191	91.4%	8.0%	0.5%
2001	32994	3485	284	89.7%	9.5%	0.8%
2002	31873	5197	418	85.0%	13.9%	1.1%
2003	31186	5843	512	83.1%	15.6%	1.4%
2004	31060	6637	561	81.2%	17.3%	1.5%
2005	31127	8464	690	77.3%	21.0%	1.7%
2006	28068	7315	747	77.7%	20.2%	2.1%
2007	29727	6541	777	80.2%	17.7%	2.1%
2008	31836	5959	1037	82.0%	15.3%	2.7%
2009	32975	5869	1068	82.6%	14.7%	2.7%
2010	32813	6639	1246	80.6%	16.3%	3.1%
2011	35832	6826	1515	81.1%	15.5%	3.4%
2012	35325	7964	1964	78.1%	17.6%	4.3%
2013	40368	7557	2328	80.3%	15.0%	4.6%
2014	45706	11286	1907	77.6%	19.2%	3.2%
2015	53833	12354	2087	78.8%	18.1%	3.1%
2016	61409	18660	2830	74.1%	22.5%	3.4%
2017	65227	22624	2722	72.0%	25.0%	3.0%
2018	71847	29185	3211	68.9%	28.0%	3.1%
2019	77522	34872	3262	67.0%	30.2%	2.8%

Необходимо отметить, что заметная часть российского роста обеспечена материалами конференций, доля которых в общем потоке российских публикаций превысила в 2019 г. отметку в 30%. Очевидно, что столь высокая зависимость от конференций может заметно повлиять на показатели результативности в 2020 г. в связи с пандемией коронавируса нового типа, особенно если учесть, что среднемировое значение в БД Scopus заметно ниже – чуть более 16%. Высокая доля публикаций в трудах конференций не является уникальной, резкий рост этой доли был характерен для Китая до 2011 г. (почти 34%) и Индии (в 2018 г. – более 25%). Китай в последние годы пересматривает принципы стимулирования и оценки научной результативности, стремясь превратить количественный рост в качественный. Российская научная

политика последнего времени, судя по целям нацпроекта «Наука» и введению КБПР, также ориентируется на качественный рост.

2.4 Тематическая структура публикационного потока

Для уточнения и анализа тематической структуры публикационного потока мы использовали описанный в главе 1.3 метод на основе публикационной истории авторов. Уточненная классификация на рисунке 12 позволяет выполнить визуальный анализ изменения тематической структуры российской науки. Это изменение происходит в основном не из-за сжатия отдельных направлений исследований, а вследствие разных темпов роста. Обращает на себя внимание снижение доли естественных наук (Physical Sciences) в основном за счет медленного роста физики и астрономии, химии, заметный рост компьютерных наук; рост доли наук о здоровье (Health Sciences) преимущественно за счет медицины; резкий рост общественных и гуманитарных наук (Social Sciences). Можно констатировать, что, вероятно, за счет расширения индексации в Scopus российских журналов, трудов российских конференций и повышения публикационной активности российских исследователей в зарубежной печати происходит постепенное приближение структуры части российской науки, которая видима на международном уровне, к ведущим странам. Тем не менее, пока основным научным направлением остаются естественные науки (Physical Sciences), обеспечивающие почти 76% всех российских публикаций, индексируемых в Scopus (рисунок 13).

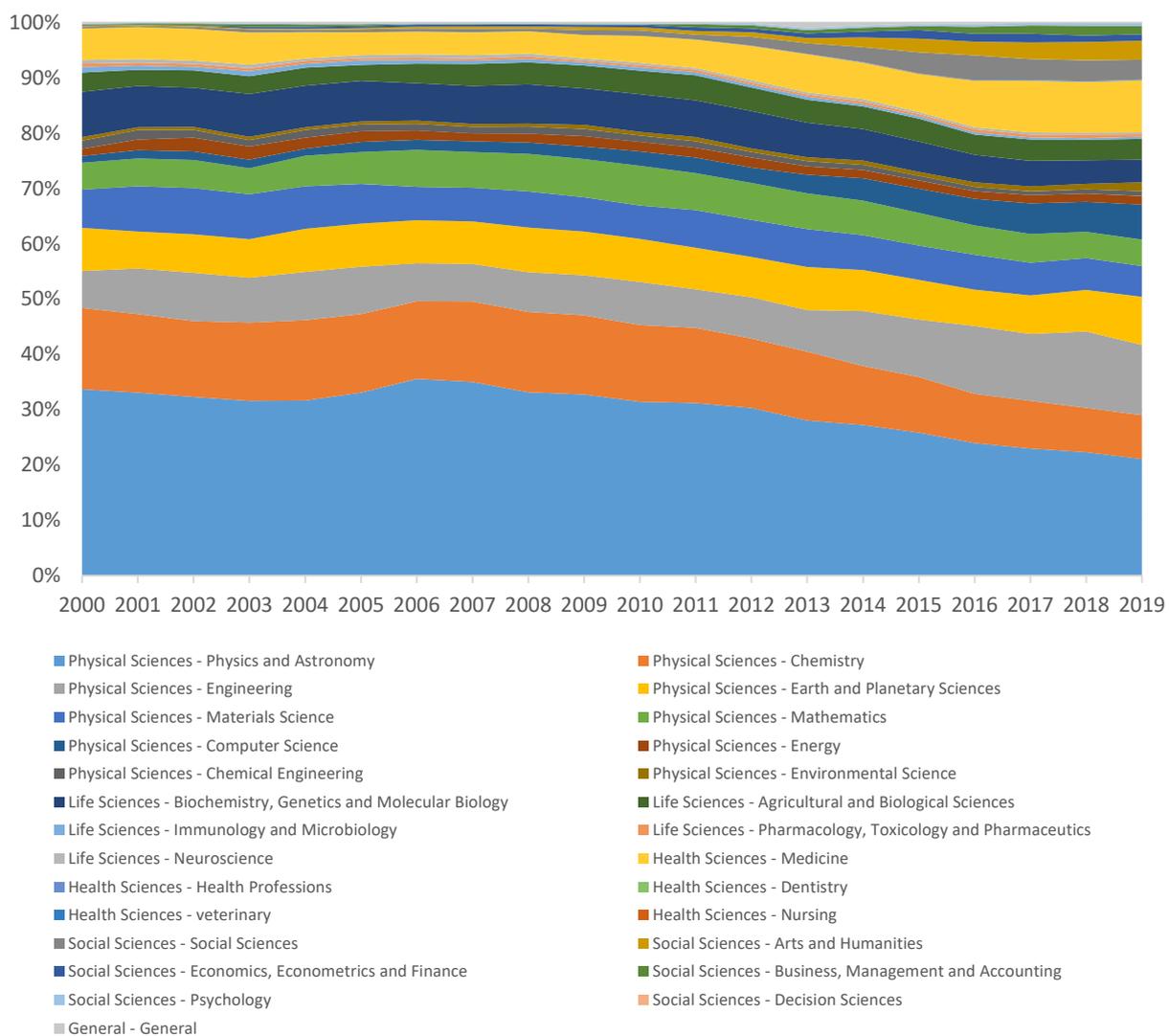


Рисунок 12 – Динамика изменения тематической структуры российской науки

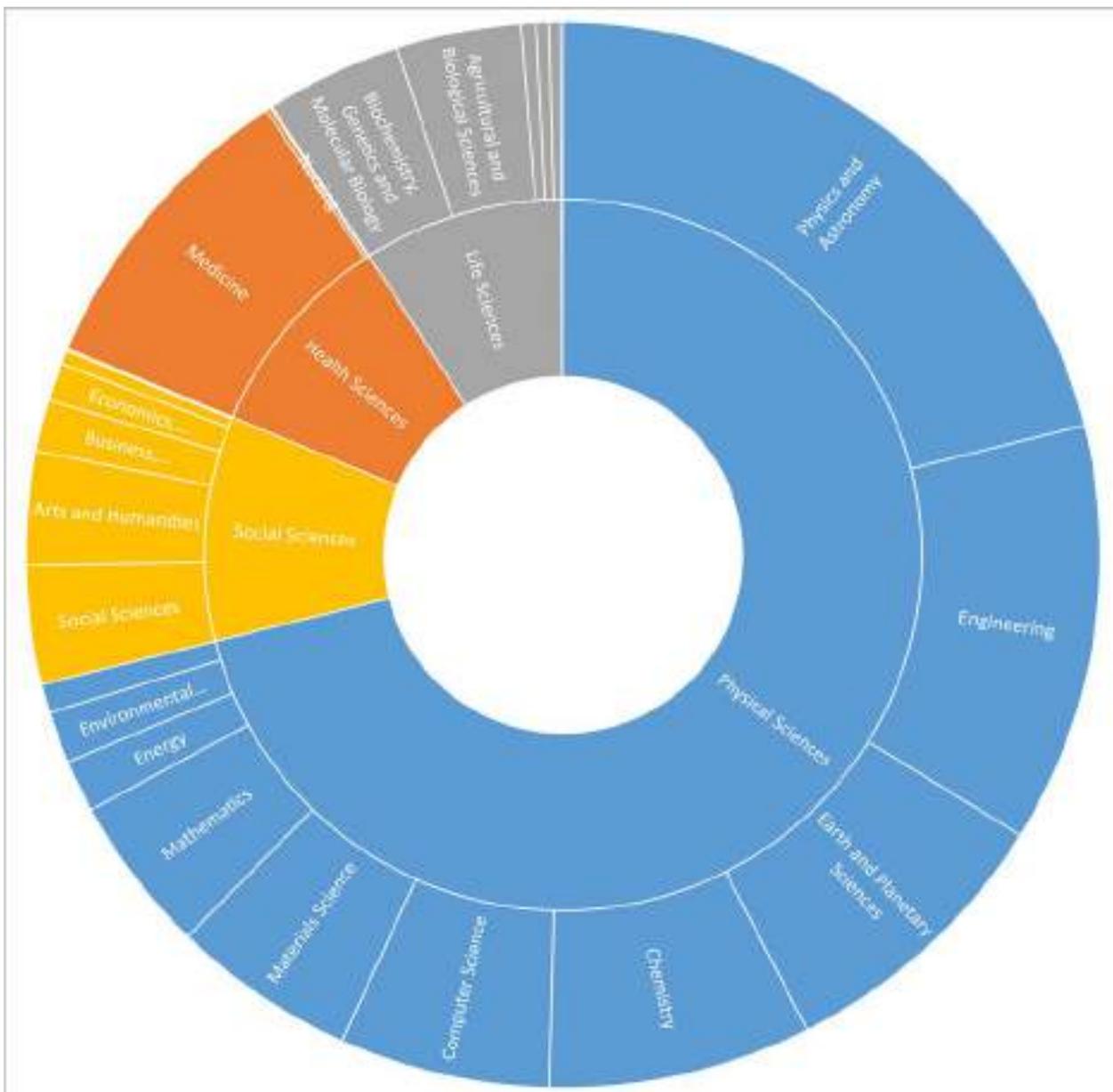


Рисунок 13 – Двухуровневая тематическая структура российского публикационного потока за 2019 г.

2.5 Распределение публикационного потока по квартилям научных журналов

Как уже было сказано выше, старт Национального проекта «Наука» и внедрение Комплексного балла публикационной результативности (КБПР) обозначили изменение ориентации национальной научной политики с стимулирования валовых количественных показателей публикационной результативности на приоритетное поощрение публикаций в высококачественных журналах, находящихся в верхних квартилях рейтинга Journal Citation Reports. В связи с выбором в качестве источника данных БД Scopus, качественный состав публикационного потока лучше характеризует рейтинг изданий Scimago Journal Ranking. Для сравнения мы приводим результаты анализа по обоим рейтингам. На рисунке 14 представлена динамика публикационного потока по рейтингам источников Scimago Journal Ranking. Активный количественный рост наблюдается во всех качественных группах, кроме журналов Q4. В процентном отношении быстрее всего растёт количество публикаций в источниках третьего квартиля (Q3) и в не входящих в рейтинг изданиях. В последние годы наблюдается ускорение роста в Q1, последовавшее после отставания в 2014-2015 гг., связанного с бурным ростом общего количества публикаций. Доля публикаций в Q1 выросла с 15,7% в 2017 г. до 18,2% в 2019 г. Этот эффект наблюдается и в группе Q2, соответственно темпы количественного роста в Q3, Q4 и в изданиях вне рейтинга постепенно падают, а их доля сокращается. Тем не менее, доля публикаций в Q1+Q2 составляет в 2019 г. всего чуть менее 36,7%, что заметно меньше среднемирового уровня. Близкая динамика наблюдается и в структуре по рейтингу Journal Citation Reports (рисунки 15 и 16), но в нем доля публикаций Q1+Q2 еще ниже (26,8% в 2019 г.), в связи с тем, что в этот рейтинг входит меньше журналов.

На рисунке 16 представлена качественная структура российского публикационного потока в 2019 г. по научным направлениям. Если судить по

доле публикаций в журналах первого квартиля, наиболее конкурентоспособными научными направлениями в России являются искусство и гуманитарные науки (29,5%), нейронауки (27,2%), биохимия, генетика и молекулярная биология (24,4%), химия (19,6%). Сравнительно низкая конкурентоспособность отличает компьютерные (2,8%), технические (3,8%) науки и энергетику (4%). В некоторой степени на это распределение влияет доля публикаций в трудах конференций, которых значительно больше в направлениях – аутсайдерах, при ограничении выборки журнальными статьями результаты в этих тематических направлениях улучшаются, но все равно остаются на сравнительно низком уровне.

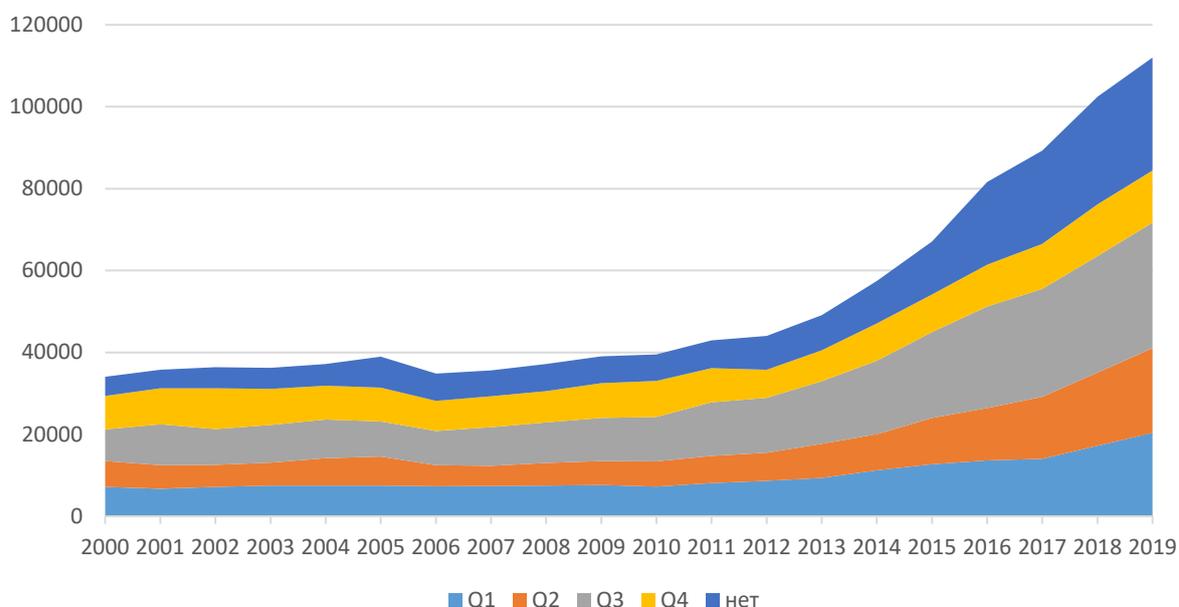


Рисунок 14 – Динамика структуры публикационного потока российских исследователей по рейтингам источников по Scimago Journal Ranking

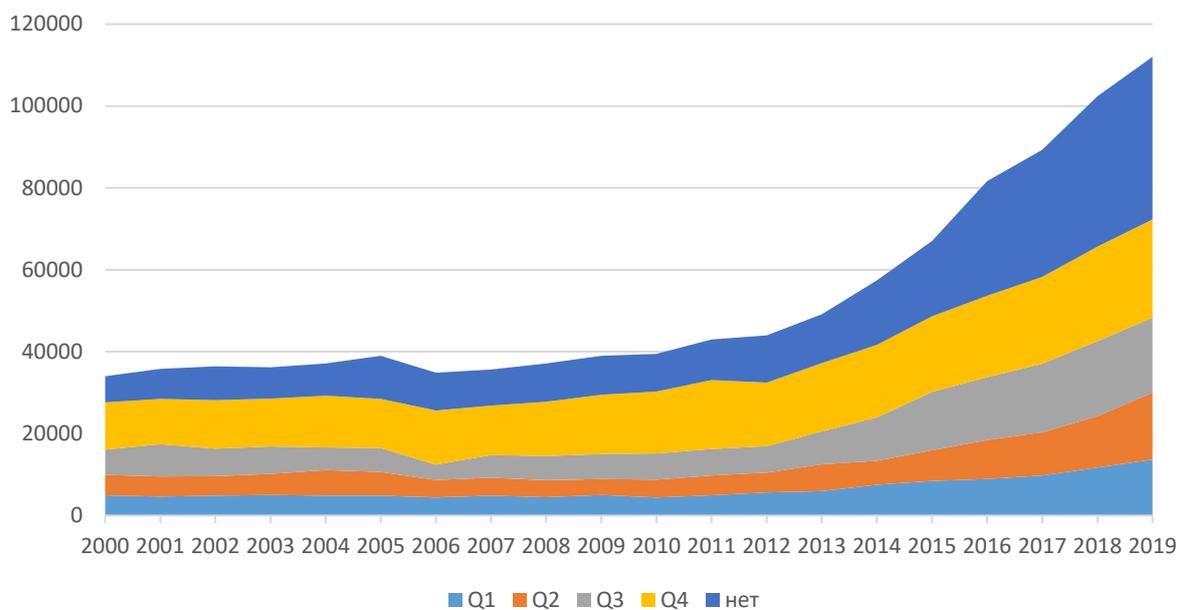


Рисунок 15 – Динамика структуры публикационного потока российских исследователей по рейтингам источников по Journal Citation Reports

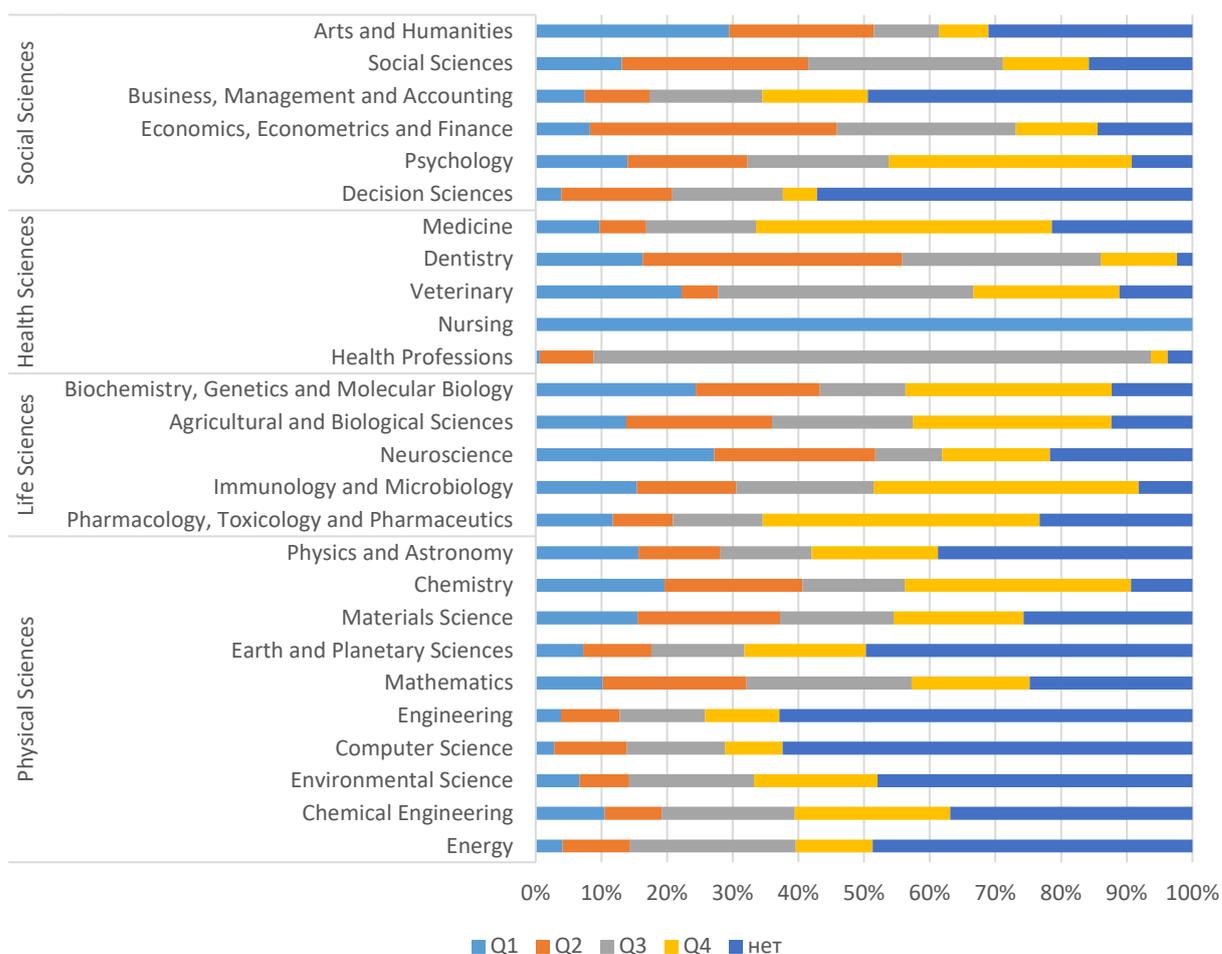


Рисунок 16 – Качественная структура российского публикационного потока в 2019г. по научным направлениям (Journal Citation Reports)

3 Феномен «инфляции результативности» и методики фракционного счета

3.1 Инфляция результативности

Самым простым индикатором результативности отдельного ученого, научной группы, организации или нации является количество публикаций в научных изданиях. Несмотря на многолетнюю критику такого подхода и его последствий [16–20], этот индикатор и его вариации, связанные с ограничениями по наличию публикаций в признанных индексах научного цитирования, типах публикаций и т.д. до сих пор является основным, а иногда и единственным во многих системах оценки результативности. На него обращают внимание при приеме на работу, продвижении, выделении финансирования, решении о закрытии научных подразделений и других важных решениях [21]. Многие исследователи отмечают, что в последние десятилетия произошли заметные изменения как в организации науки во всем мире, так и в сфере научных коммуникаций. Мы наблюдаем ежегодный рост количества научных публикаций и опережающий рост персональной и институциональной результативности связанный прежде всего с увеличением числа и масштабов научных коллабораций [22–24]. Появились супер-продуктивные авторы, публикующие по 25 и более статей в год [25–27]. Особняком стоит феномен мегаколлабораций, в рамках которых тысячи исследователей объединяются для решения большой научной задачи. Такие коллаборации могут быть очень продуктивными [28] и публиковать сотни статей в год.

Сложившаяся ситуация позволяет некоторым исследователям говорить об инфляции результативности [29–31], ведь получается, что «стоимость» одной статьи в списке публикаций ученого сейчас заметно ниже, чем 10-20 лет назад. Администраторы, управляющие наукой, видя это, ищут пути более адекватной оценки результативности. Одним из очевидных путей такого рода является использование фракционного счета, при котором каждому автору за публикацию начисляется его авторская доля, а не единица [32–34]. Этот

подход распространяется и на оценку результативности организаций. Использование фракционного счета позволяет нивелировать или сгладить часть эффектов, вызванных растущими масштабами научной коллаборации. Moed заметил, что фракционный счет отражает вклад автора или организации в исследование, в отличие от целого счета, который фиксирует участие [35]. Ряд авторов отмечает, что фракционный счет недооценивает коллаборации и предлагает модифицированные варианты, в основе которых лежит предположение о большей ценности результата, полученного в коллаборации [36]. Это связано с большей цитируемостью публикаций, написанных в коллаборациях, особенно международных [24, 37].

В России количественная оценка результативности начала применяться сначала на индивидуальном уровне с 2006 г. в связи с реализацией пилотного проекта совершенствования системы оплаты труда научных работников [38]. После масштабной реформы государственных академий наук, внедрения проектного принципа управления и изменений в подходах к администрированию государственных научных организаций были введены ключевые показатели эффективности, основанные на планируемом количестве статей в рецензируемых научных журналах, а в 2013-2016 гг. проведена комплексная оценка результативности научно-исследовательских институтов [39]. Результаты этой оценки используются в том числе и при определении объемов финансирования научных организаций. Количественный подход к оценке результативности привел к повышению интереса к наукометрическим исследованиям в России [40] и обеспокоенности научного сообщества проблемами результативности, рядовые ученые и научные администраторы стали обращать больше внимания на индексы научного цитирования, показатели журналов, метрики результативности.

Так называемые «майские» указы президента Путина, последовавшие после выборов 2013 г., закрепили ориентацию на количественные показатели результативности в национальной научной политике. Этот курс был подтвержден и усилен в «новом майском» указе после выборов 2018 г.

Научная политика России ориентирована на достижение страной пятого места по публикационной активности по данным международных индексов цитирования. Такое управляющее воздействие оказало заметное влияние на публикационную активность российских исследователей [41]. Как мы показывали в [42], в основном рост публикационной активности обеспечен увеличением количества активных авторов, преимущественно за счет преподавателей высших учебных заведений и ученых академических институтов, ранее не публиковавших результаты своих исследований в изданиях, индексируемых международными базами данных.

В то же время есть основания полагать, что некоторые организации или исследовательские группы наращивают свою публикационную активность сомнительными методами [43], часть из которых связана с коллаборациями. К таким методам стоит отнести прежде всего практику указания дополнительных аффилиаций в публикациях. Особенно это касается случая, когда организация включает статью в свой публикационный отчет потому, что несколько или даже всего один из авторов статьи указали эту организацию как дополнительную аффилиацию [44]. Зачастую это связано с относительно небольшим материальным стимулированием таких авторов со стороны этой дополнительной аффилиации и эта практика получила в литературе название «покупка академического престижа» [45].

При целом счете (когда публикация учитывается как единица в каждой из аффилированных с ней организаций) учреждения, являющиеся активными участниками множества коллабораций получают преимущество перед меньше сотрудничающими коллегами, что, потенциально, может вести к некорректной оценке их результативности. Каждая российская исследовательская организация отчитывается о количестве публикаций за год. Эта результативность может быстро расти, в значительной степени за счет публикаций во внешних коллаборациях. В условиях постоянного роста среднего количества авторов одной публикации, активного развития разного уровня коллабораций, связанного как с объективными причинами (рост

сложности и междисциплинарности научных исследований, развитие специализации), так и, вероятно, со стремлением к увеличению количественных показателей публикационной результативности этот рост только ускоряется (рисунок 17 Рисунок 17 –).

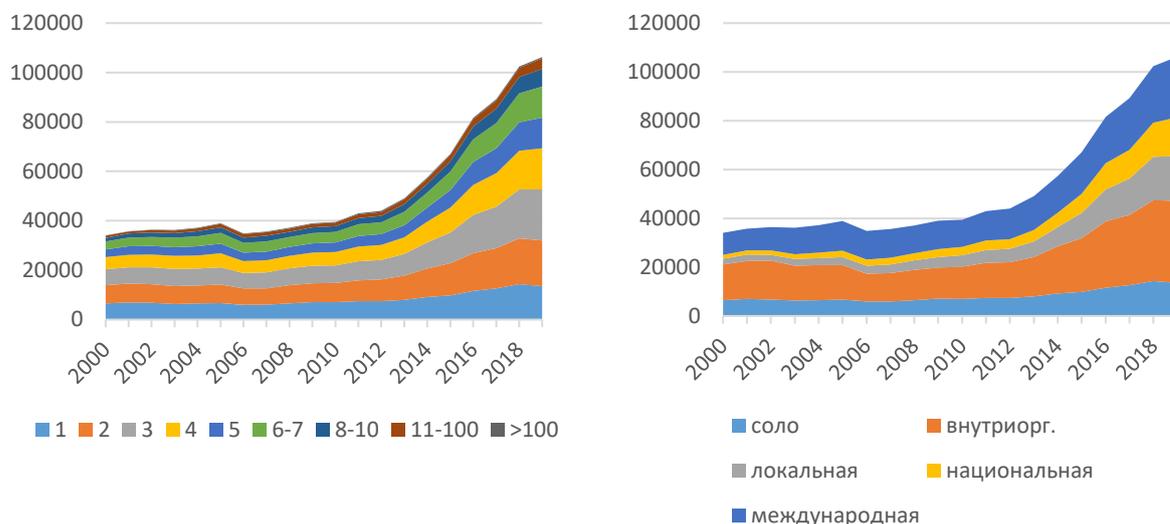


Рисунок 17 – Развитие коллабораций в РФ. а) количество публикаций по количеству авторов одной публикации, б) количество публикаций по типу коллаборации

Если не брать во внимание этот характер роста, то может показаться, что общий прогресс национальной результативности может быть получен суммированием результативности отдельных организаций. Реальные цифры, конечно, будут ниже, эту разницу можно принять за коэффициент инфляции. На рисунке 18 (а) представлена динамика количества российских публикаций, индексируемых в БД Scopus, а также «отчетного» количества публикаций – суммы количества публикаций по отдельным российским аффилиациям, а также всем (не только российским) аффилиациям авторов этих публикаций, в виде гистограмм представлен расчетный коэффициент инфляции для российских организаций. Видно, что разрыв между номинальным и отчетным количеством публикаций постоянно увеличивался весь исследуемый период времени, этот рост был особо активным в 2012-2016 гг. К концу периода коэффициент инфляции для российских организаций составил почти 70%, а

для всех аффилиаций – более 180%. На рисунке 18 (б) показана та же динамика, но только для журнальных публикаций.

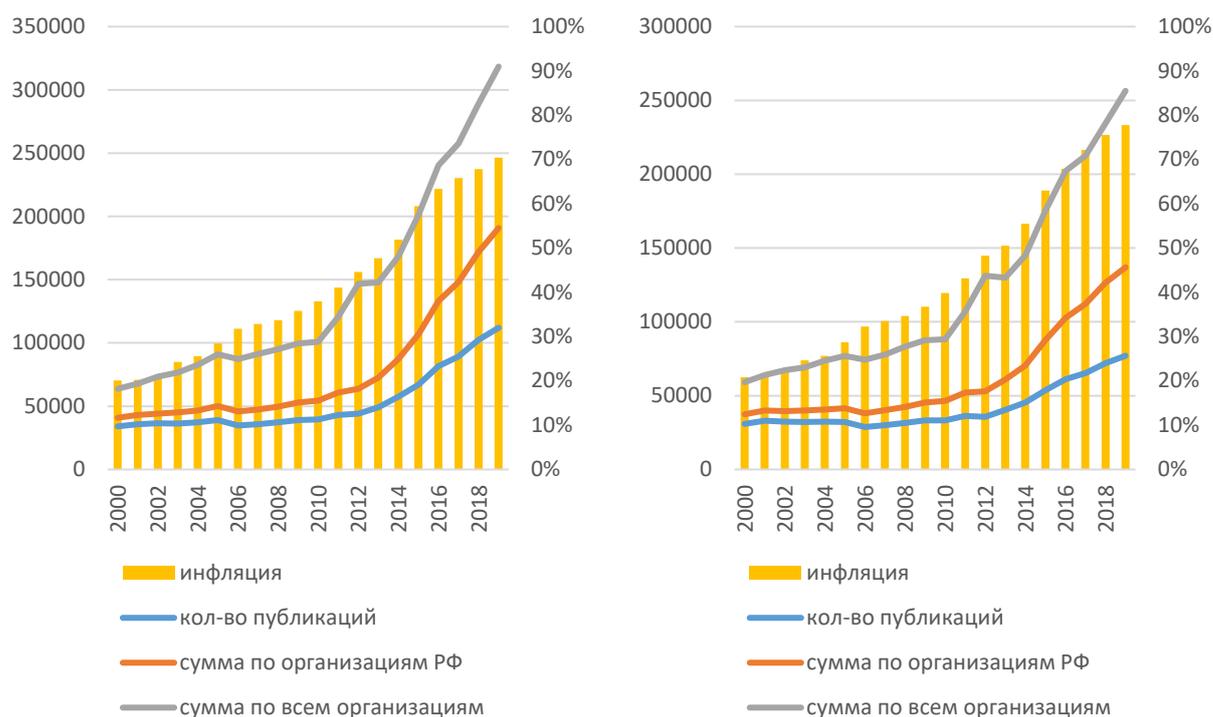


Рисунок 18 – Инфляция результативности исследовательских организаций. а) все типы документов, б) только журнальные статьи

3.2 Фракционный счет

Перечисленные выше причины (и не только они) привели к тому, что со стороны кураторов национальной научной политики начал формироваться запрос на более точные методы измерения научной результативности. Одним из таких методов является фракционный счёт публикаций.

Впервые метод фракционного счёта был предложен ещё в 1966 г. – для пропорционального учёта соавторов статей при измерении продуктивности исследователей [46]. Систематический обзор методов фракционного счёта представлен в работе [47], его преимущества показаны в [48]. В 2005 г. *Moed* впервые высказал [49] идею о том, что традиционный целочисленный счёт соавторов и организаций измеряет их участие (*participation*), а фракционный – вклад (*contribution*). Преимущества фракционного счёта перед целочисленным в типовых задачах оценки научной результативности обоснованы в работе

[50]. Вероятно, наиболее известным случаем применения фракционного счёта является Лейденский рейтинг. Его авторами было показано, что именно фракционный счёт позволяет более аккуратно сравнивать показатели университетов [51].

Российскими исследователями фракционный счёт изучался крайне мало [52], аналитическое сравнение с целочисленным счётом не проводилось. До 2020 г. методы фракционного счёта публикаций не упоминаются в официальных документах, относящихся к государственной научной политике, хотя есть свидетельства его использования Министерством науки и высшего образования РФ в 2019 г. при оценке публикационной активности подведомственных организаций.

Поясним методы расчета количества публикаций организации целым и фракционным счетом:

AC_i (*Article Count*, целочисленный счёт) – количество публикаций, у которых есть хотя бы один автор с аффилиацией организации i ;

FC_i (*Fraction Count*, фракционный счёт – сумма фракционных баллов $fs_p(i) \in [0...1]$ всех публикаций p организации i , $fs_p(i) = \frac{1}{N_p} \sum a_j \frac{Z(i, a_j)}{AF(a_j)}$, где a_j – это авторы публикации p ; $1 \leq j \leq N_p$ – количество авторов; $AF(a_j)$ – количество аффилиаций автора a_j , $Z(i, a_j)$ равно 1, если a_j имеет аффилиацию i , и 0 – если не имеет. В терминах, определённых в [50], это называется фракционным счётом уровня авторов (*author-level fractional counting*).

Примеры расчета FC приведены в таблице 9. Фракционный счет может быть также применен и к отдельному автору.

Таблица 9 – Примеры расчёта FC для одной публикации, аффилированной с российскими организациями $O1, O2$

Авторы и аффилиации публикации	AC	FC	
		$O1$	$O2$
Авт 1 ^{$O1$} , Авт 2 ^{$O1, O2, F1, F2, F3$}	1	6/10	1/10
Авт 1 ^{$O1$} , Авт 2 ^{$F1$} , Авт 3 ^{$F1, O2$}	1	1/3	1/6
Авт 1 ^{$O1, O2$} , Авт 2 ^{$O1$} , Авт 3 ^{$O1$} , Авт 4 ^{$O1$} , Авт 5 ^{$O1, F1$}	1	8/10	1/10
Авт 1 ^{$O1, O2, F1$} , Авт 2 ^{$F2$} , Авт 3 ^{$F3$} , ..., Авт 100 ^{$F100$}	1	1/300	1/300
Сумма	4	1,736	0,367

Если в фракционном счете учитывать только российские аффилиации, то можно получить результат национальным фракционным счетом (National Fraction Count – NFC). Особенность этого показателя заключается в том, что сумма NFC по организациям или авторам будет равна национальному научному продукту целым счетом. Таким образом, NFC отражает вклад организации или автора в национальный научный продукт.

3.3 Персональная результативность российских исследователей

Средняя для российских авторов персональная результативность (в количестве публикаций в год) неравномерно росла в исследуемый период с 1,93 статьи в 2000 г. до 2,32 статьи в 2019 г. (рисунок 19). Персональная результативность, рассчитанная фракционным счетом, в этом же периоде слегка упала с 0,58 до 0,56 статьи в год. При этом средний вклад в результативность России (общее количество российских статей в год деленное на общее количество активных авторов с российскими аффилиациями) падал даже заметнее (рисунок 20). Необходимо отметить, что в значительной степени рост количества российских публикаций обусловлен увеличением числа активных авторов, некоторые тенденции поэтому находятся под сильным влиянием этого фактора. Если посмотреть на публикационную активность «постоянных» авторов – ученых, публиковавших результаты своих исследований в не менее чем 5 годах в исследуемом периоде, то можно заметить еще более значительный рост персональной результативности в целом счете – с 2,5 до 3,8 публикации в год, но результативность фракционным

счетом у них тоже росла и к концу периода приблизилась к 0,9 статьи в год (рисунок 21).

Таким образом, можно сделать вывод о том, что на персональном уровне видимая инфляция результативности в большей степени связана с появлением новых авторов, часто включаемым в состав растущих авторских коллективов. Рост персональной результативности в целом счете у продуктивных авторов в основном подтверждается и расчетом по методу фракционного счета. Наши оценки показывают, что для этой категории ученых инфляция за рассматриваемый период составила около 10%.

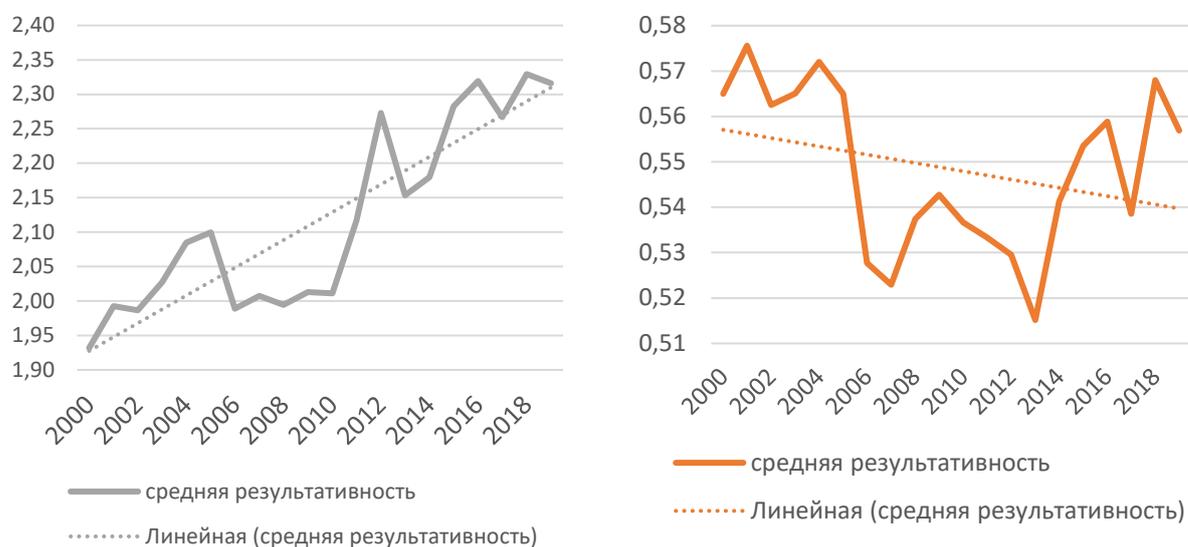


Рисунок 19 – Средняя результативность российских авторов: а) целым счетом, б) фракционным счетом



Рисунок 20 – Средний вклад исследователя в национальный результат

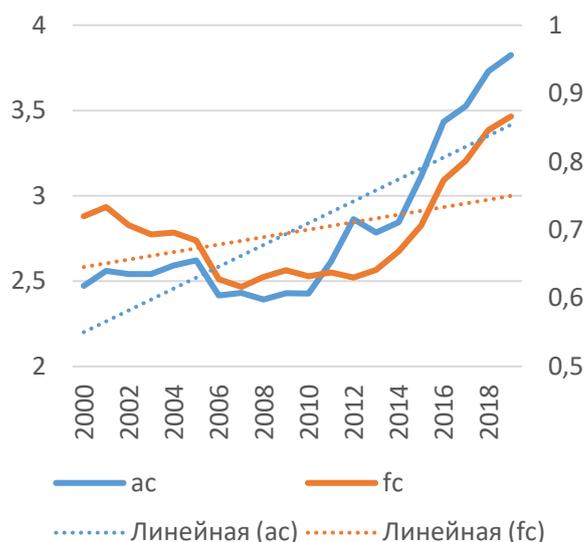


Рисунок 21 – Средняя результативность ведущих российских исследователей (активных в течение 5 и более лет). ас – целым счетом, fc – фракционным счетом

3.4 Коэффициент самодостаточности, неравномерность реакции на «наукометрическое давление»

Разницу между фракционным и целочисленным счётом можно определить с помощью соотношения FC/AC . Этот показатель можно назвать коэффициентом самодостаточности, он будет равен единице если у всех публикаций организации все авторы аффилированы только с этой организацией и уменьшаться с увеличением количества коллабораций. Этот коэффициент уменьшается на протяжении последнего времени практически для всех российских организаций (рисунок 22).

Если предположить, что усиление коллаборационной активности является одной из форм ответа на изменения в национальной научной политике, «наукометрическое давление», связанное с требованием повышения публикационной результативности, то должна быть заметна разница между разными группами организаций. Мы предположили, что наиболее подверженными наукометрическому давлению были университеты проекта 5/100 [43]. В несколько меньшей степени это касалось академических научно-

исследовательских институтов, переживших процесс ведомственной оценки результативности [53]. Наконец, наименее чувствительными к этому давлению должны оказаться ведущие вузы за исключением вузов проекта 5/100, МГУ и СПбГУ.

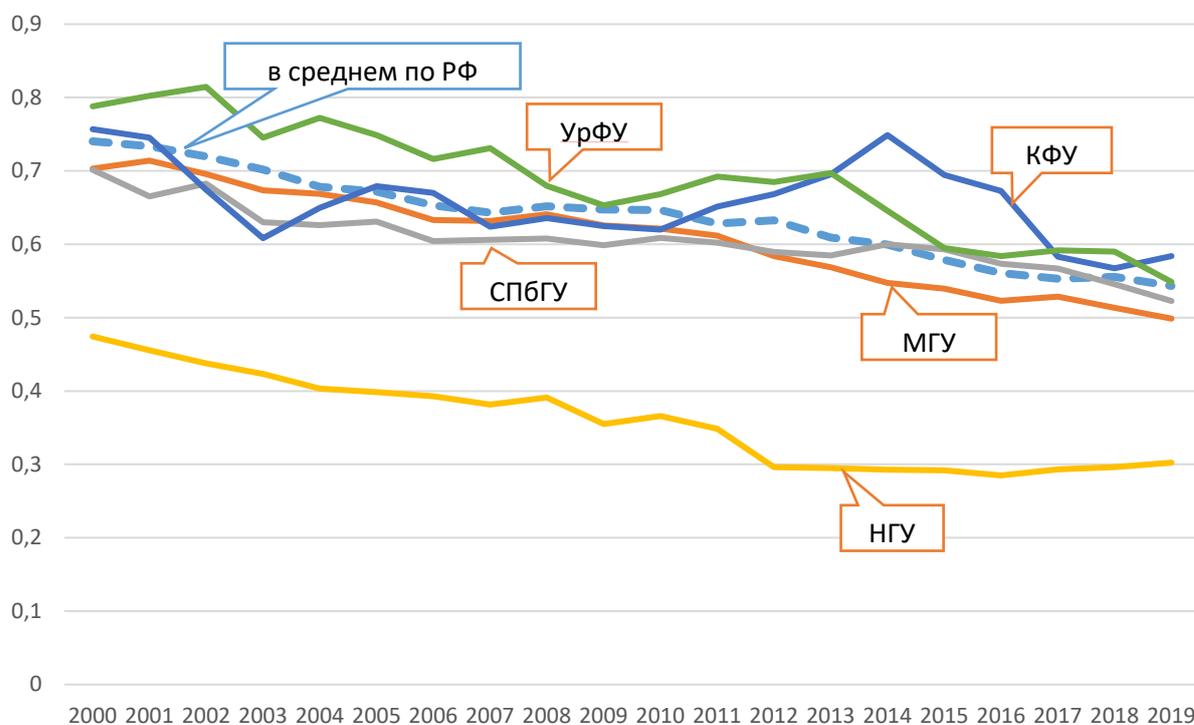


Рисунок 22 – Динамика коэффициента самодостаточности для некоторых российских организаций и среднего по РФ

Мы сформировали три выборки организаций:

1. Группа *1-LU (Leading Universities)* – 21 университет, участвующий в проекте повышения конкурентоспособности российских вузов, целью которого является вхождение пяти российских университетов в сотню ведущих мировых научно-образовательных центров (<https://www.5top100.ru/universities>). На весь проект ежегодно выделялось дополнительное финансирование в размере около 10 млрд р. (140 млн евро), которое распределялось в зависимости от достигнутых результатов. По этой причине администрация, профессорско-преподавательский состав и научные сотрудники этих университетов работали в условиях сильного

«наукометрического прессинга». В 2018 г. наибольшее число публикаций было у Новосибирского государственного университета (2 839), наименьшее – у Тюменского государственного университета (364).

2. Группа *2-RI (Research Institutes)* – 100 научных организаций под научно-методическим руководством РАН. После реформы государственных академий наук и смены учредителя у более чем 700 научных организаций (в 2013 г. им стало Федеральное агентство научных организаций) драматически изменился стиль управления научными исследованиями – он стал существенно ориентирован на количественные показатели.

Так в 2015 г. в государственное задание введён обязательный показатель – количество статей, которые должны быть опубликованы в отчётном году и проиндексированы в *WoS*, *Scopus*, а также в РИНЦ и в других российских и международных информационно-аналитических системах научного цитирования. В 2015 г. была объявлена, а в 2017–2018 гг. проведена оценка научной результативности с учётом наукометрических показателей [12].

Из 454 прошедших проверку организаций к первой категории отнесено 142, ко второй – 205, к третьей – 107. Попавшие в первую категорию могли претендовать на дополнительное финансирование, а для третьей категории разрабатывались планы ликвидации или присоединения к другим организациям. «Наукометрический прессинг» по отношению к этим организациям также присутствовал, но был менее выражен и начался на два-три года позже, чем для группы *1-LU*. В эту группу были включены 100 научных организаций, у которых в 2018 г. было наибольшее количество публикаций (от 125 до 1 493), проиндексированных в *Scopus*.

3. Группа *3-HEI (High Education Institutes)* – 100 вузов разного профиля: федеральные и региональные университеты технической, медицинской, педагогической, строительной и иной направленности, не включённые в группу *1-LU*. До 2016 г. не предъявлялось формальных требований к повышению их наукометрических показателей, хотя высокая конкуренция за дополнительное финансирование к этому располагала.

В 2016–2017 гг. на основании конкурса были определены 33 опорных вуза, относящиеся к разным регионам и получившие дополнительную финансовую поддержку. Таким образом, эти организации позже всех и в наименьшей степени столкнулись с системным «наукометрическим прессингом». В эту группу мы включили 100 вузов, у которых в 2018 г. было наибольшее число публикаций, проиндексированных в *Scopus* (min 91, max 1 406).

МГУ и СПбГУ не были включены в перечисленные выше группы в связи со своим исключительным положением, условиями финансирования и давней традицией проведения фундаментальных научных исследований.

На рисунке 23 отображена зависимость между AC и NFC в 2006 и 2018 гг. в трёх рассматриваемых группах. Проанализировав данные и за другие годы, мы отметили следующие тенденции:

- со временем коэффициент корреляции снижается, т.е. индикаторы AC и NFC всё более отличаются друг от друга. Это говорит о том, что возрастает разрыв между индикатором участия в исследовании (AC) и индикатором вклада в него (FC и NFC);

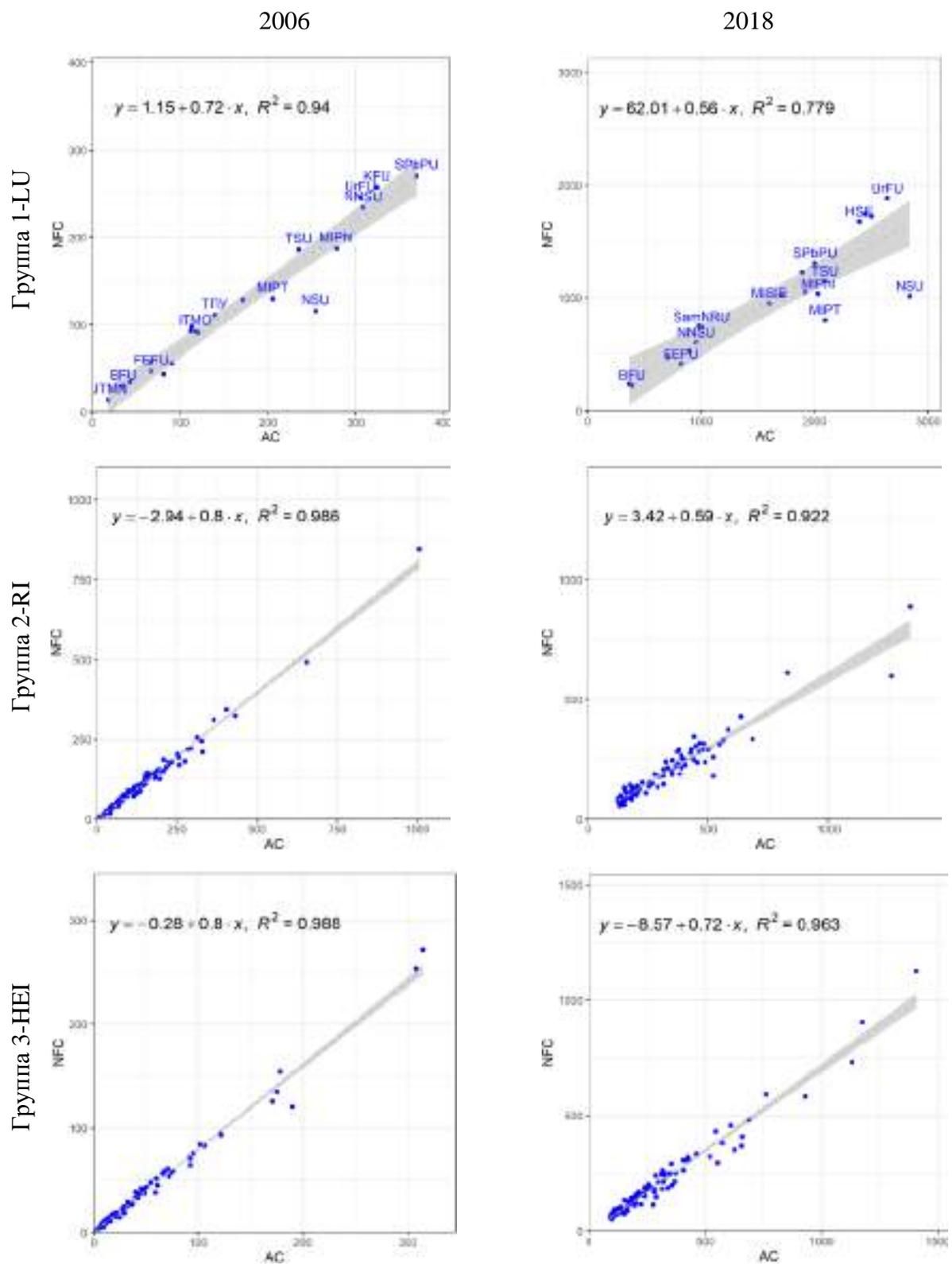


Рисунок 23 – Корреляция целочисленного (AC) и национального фракционного счета (NFC) в 2006 и 2018 гг. для трёх групп организаций

- среднее отношение NFC/AC заметно уменьшается, что говорит об увеличении среднего количества российских аффилиаций на одну статью. Это может свидетельствовать как об объективных процессах, связанных с развитием коллабораций и мультидисциплинарных исследований, так и о расширении практик недобросовестного соавторства;
- поведение коэффициента корреляции в трёх группах также различно. Наименьшие коэффициенты фиксируются в группе 1-LU. Наибольшее влияние на это оказывают НГУ и МФТИ, которые входят в крупные коллаборации, связанные с БАК и очень активно сотрудничают с другими исследовательскими институтами в области физики, химии и других естественных наук. При этом и для других университетов характерна повышенная дисперсия, что, на наш взгляд, связано с высоким «наукометрическим давлением» и различными стратегиями повышения публикационной активности, в том числе интенсивным вхождением в новые коллаборации. Наименьший разброс фиксируется в группе 3-HEI – здесь университеты менее всего участвовали в гонке за наукометрическими показателями до недавнего времени.

3.5 Комплексный балл публикационной результативности – победители и проигравшие

Рассмотренные выше обстоятельства, вероятно, оказали влияние на трансформацию системы оценки результативности Министерством науки и высшего образования, связанную с переходом на фракционный счет и учет «качества» научных журналов. Новый для министерства подход был опробован в оценке результатов 2018-2019 годов и применен для планирования результатов 2020 г. Методика получила название «Комплексный балл публикационной результативности», сокращенно КБПР.

Важной особенностью этого подхода является применение баллов в зависимости от «качества» публикации. Качество определяется квартилем журнала в Journal Citation Report компании Clarivate Analytics, индексированием в Web of Science Core Collection, Scopus, Russian Science Citation Index на платформе Web of Science. В таблице 10 указаны баллы для оценки публикаций в зависимости от индексации издания, в котором они опубликованы. После обсуждения в балльную оценку были внесены изменения, опубликованные 23 апреля 2020 г.

Таблица 10 – Баллы для оценки публикации при расчете КБПР

Издание	Баллов за публикацию (версия от 30.12.2019)	Баллов за публикацию (версия от 23.04.2020)	
		Все направления	Гуманитарные и общественные
WoS Q1	19,7	20	3
WoS Q2	7,3	10	3
WoS Q3	2,7	5	3
WoS Q4	1	2,5	3
WoS без IF	1	1	3
Scopus	1	1	3
RSCI	0,75	1	3
ВАК	0,5	0,12	1
Книги	1	1	1 за п.л.

В соответствии с этими изменениями несколько изменена шкала баллов. Для гуманитарных и общественных наук шкала баллов носит упрощенный характер: публикации, индексируемые в Scopus или WoS (включая RSCI) оцениваются в 3 балла, в журналах из списка ВАК – в 1 балл, монографии оцениваются в зависимости от объема – 1 балл за условный печатный лист.

Баллы, соответствующие «качеству» публикации по замыслу авторов методики должны частично компенсировать потери на публикациях, написанных коллективами авторов из разных организаций при фракционном счете. Особенно это касается международных коллабораций, так как по оценкам авторов методики большая их часть публикуется в журналах с высоким импакт-фактором.

Очевидно, что такая смена парадигмы должна привести к изменениям как в результатах оценки и ранжировании организаций, так и в стратегиях повышения публикационной активности организаций. Очевидным следствием может быть снижение уровня коллабораций, в тех случаях, когда они не были в достаточной мере обоснованы необходимостью привлечения дополнительных компетенций, объединения ресурсов и т.д. Особенно это касается как раз феномена дополнительных аффилиаций. Мы выделили 4 основных потенциальных последствия перехода на КБПР.

Последствие 1. Больше всего от перехода на фракционный счёт страдают участники крупных международных коллабораций: хотя именно такие исследования нередко оказываются прорывными, показатель FC будет близок к нулю. После запуска Большого адронного коллайдера в 2008 г. начал массово проявляться феномен мегаколлабораций, которые публикуют сотни статей ежегодно [54]. Этот феномен ещё слабо изучен, но можно утверждать, что мегаколлаборации также внесли свой вклад в увеличение разницы между результатами целочисленного и фракционного счёта. Организации, участвующие в коллаборациях ATLAS, CMS и др., получают очевидный бонус при подсчёте AC, проигрывают при переходе к FC. Эффект можно рассмотреть на примере Института ядерной физики СО РАН – одного из активных участников таких мегаколлабораций (рисунок 24). Заметно, что, в то время как результативность этого института целым счетом более-менее устойчиво росла, а фракционным – незначительно изменялась на гораздо более низком уровне, результат КБПР скорее имеет тенденцию к снижению. Основной причиной этого является то, что самые высокорейтинговые публикации сотрудники института готовят как раз в больших коллаборациях. Проблема «справедливости» фракционного счёта при оценке больших коллабораций уже обсуждалась в научном сообществе, были сделаны выводы о недооценке вклада отдельного автора в таких публикациях, Гуннар Сивертсен с соавторами предложили модифицированный метод счёта, более

адекватно оценивающий результативность участников больших коллабораций [39].

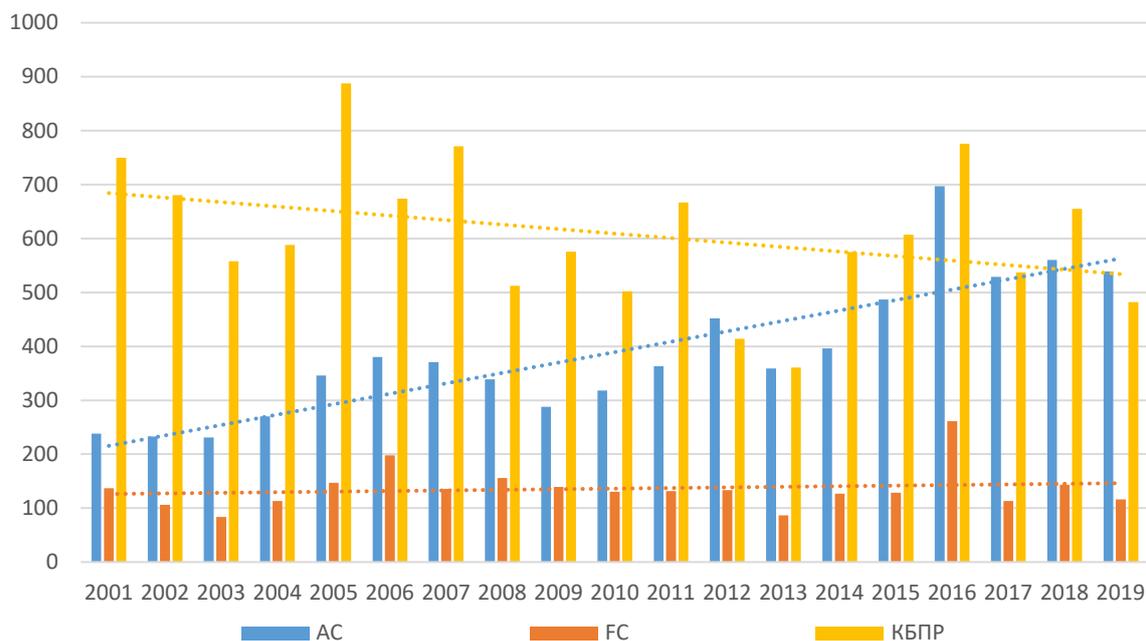


Рисунок 24 – Тренды изменения результативности ИЯФ СО РАН целым и фракционным счетом, КБПР

Более наглядно этот эффект можно рассмотреть на графике, сопоставляющем *бонус на баллах* КБПР (за счет высокого уровня журнала), который может быть выражен, как результат деления оценки КБПР на результат по фракционному счету ($KBPR / FC$) и *штраф за коллаборации* (коэффициент самодостаточности), рассчитываемый как частное FC / AC . Так как эти значения связаны, мы можем выделить зоны, в которых организации теряют при переходе на КБПР, выигрывают от этого перехода или остаются «при своих». На представлены результаты расчета для ведущих 100 организаций РФ (рисунок 25).

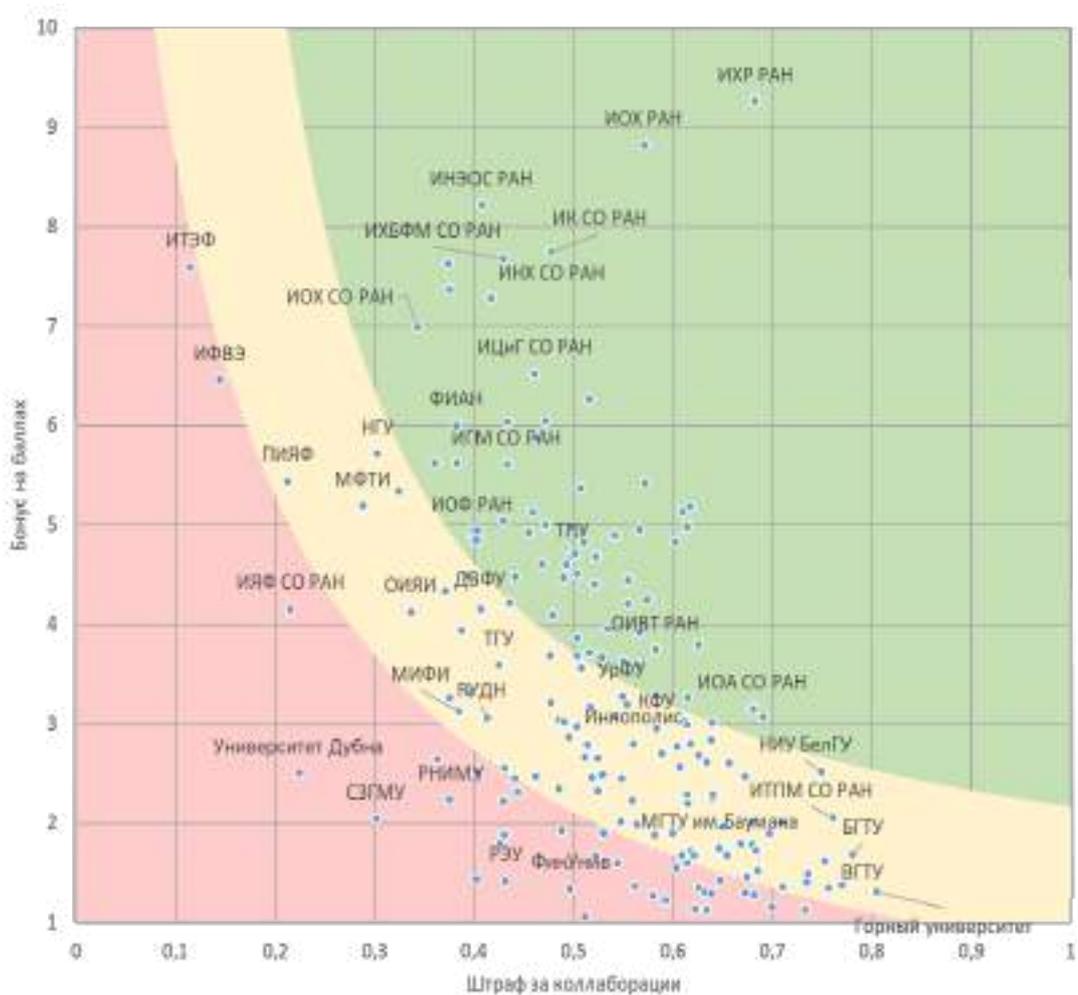


Рисунок 25 – КБПР: победители и проигравшие

Анализ этого графика позволяет сделать вывод о том, что проигрывают организации, специализирующиеся в области физики высоких энергий – активные участники мегаколлабораций. Среди выигравших много организаций, специализирующихся в области химии, материаловедения, биологии и генетики. Ранее мы показывали, что в этих областях российские ученые в значительной доле публикуются в высокорейтинговых журналах, при этом коллективы авторов публикаций преимущественно из одной или немногих организаций, в связи с чем штраф за коллаборации остается на невысоком уровне. Также заметно, что большая часть вузов попадает в нижнюю часть графика, что обусловлено тем, что авторы публикуются в основном в низкорейтинговых журналах и материалах конференций.

Последствие 2. От перехода на КБПР пострадают вузы, активно привлекающие внешних соавторов. Ряд вузов повышает публикационную активность за счет премирования внешних соавторов за «приписывание» аффилиации. Основная организация автора при таких правилах расчета КБПР может начать активно возражать против такой практики, в связи с чем эти вузы могут потерять часть публикационной активности. Для оценки масштаба этих последствий мы выполнили расчет публикационной результативности с учетом только первых аффилиаций авторов. В общем случае первая аффилиация не всегда является основным местом работы автора, но в большей части случаев это допущение справедливо, что позволяет выполнить достаточно достоверную оценку. На рисунке 26 показаны результаты расчета результативности Новосибирского государственного университета с учетом только первой аффилиации авторов публикаций и «коэффициента первенства по КБПР» ($D_{\text{КБПР}}$), рассчитанного как частное от деления полного КБПР и КБПР-1 (с учетом только первой аффилиации). Падение коэффициента самодостаточности отражает все большую зависимость НГУ от внешних соавторов.

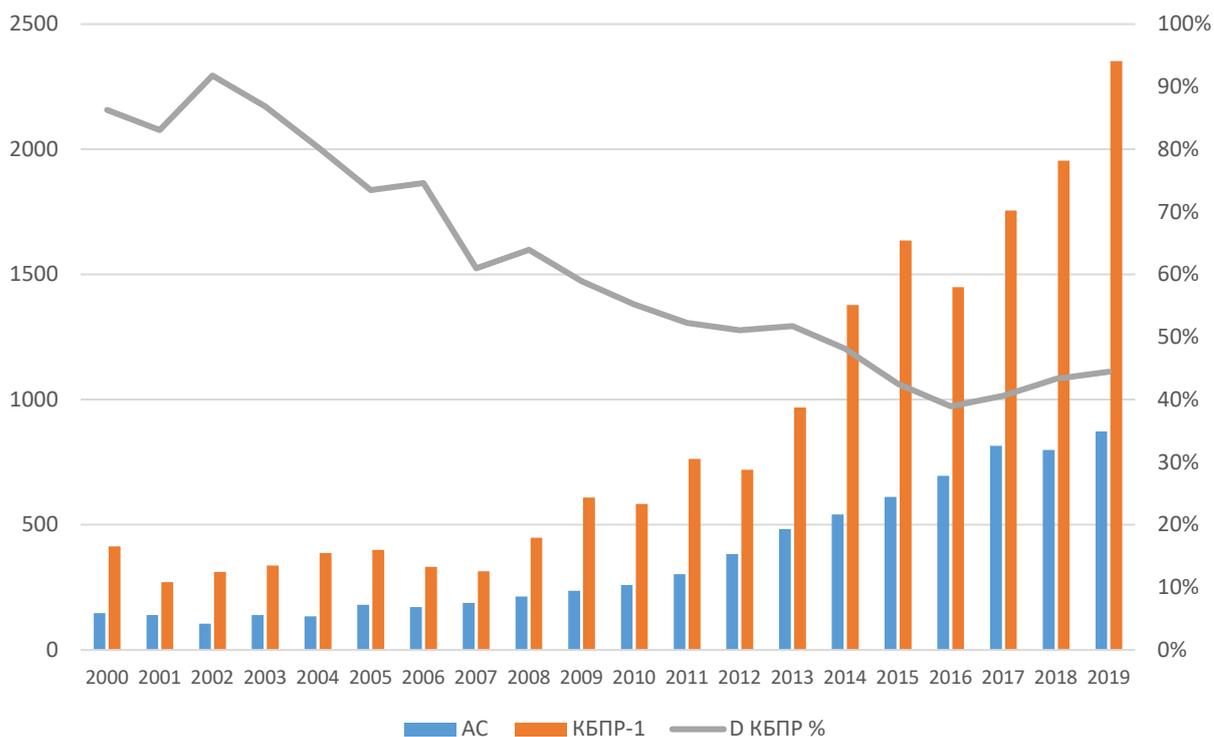


Рисунок 26 – Динамика изменения результативности НГУ целым счетом и КБПР с учетом только первой аффилиации и коэффициент «первенства» по КБПР

«Донорами» НГУ выступают научно-исследовательские институты Новосибирского городка. При расчете КБПР с учетом только первой аффилиации они выигрывают (рисунок 27).

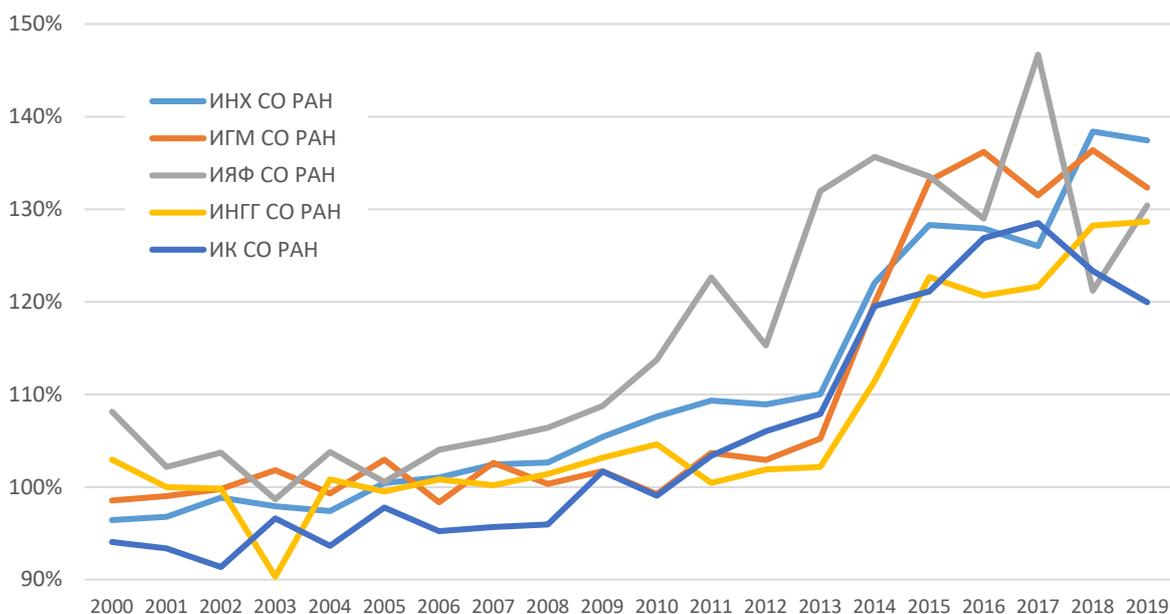


Рисунок 27 – Динамика коэффициента «первенства» по КБПР основных академических институтов – «доноров» НГУ

Высокой степенью зависимости от внешних совместителей характеризуются также МФТИ и РУДН (рисунок 28). Эти вузы могут заметно потерять в случае, если основные работодатели авторов - внешних совместителей начнут требовать от них не указывать дополнительные аффилиации. Возможным последствием может стать обострение конкуренции за ведущих исследователей между научно-исследовательскими институтами и вузами.

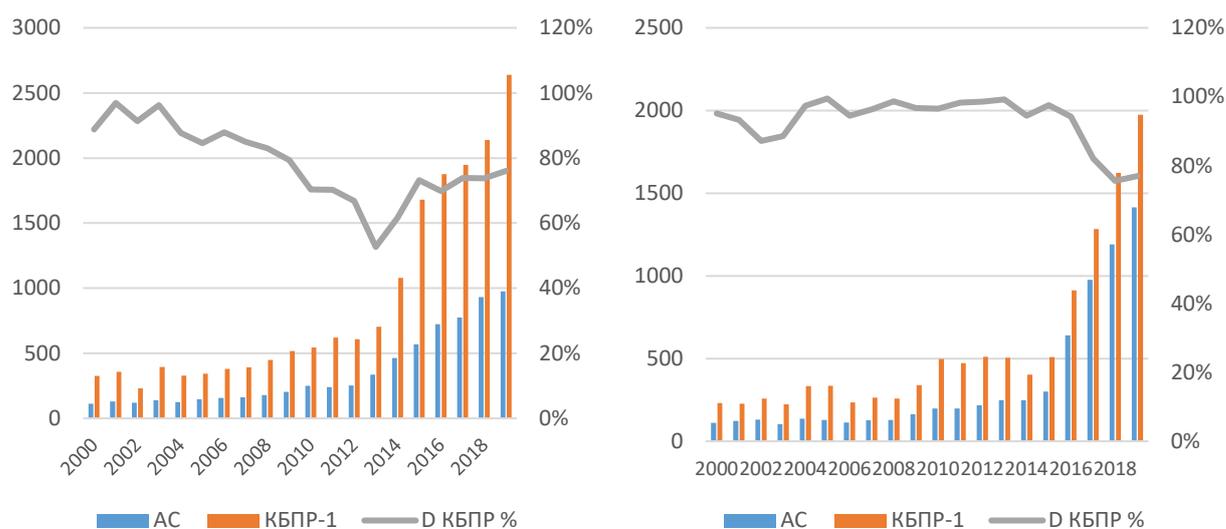


Рисунок 28 – Динамика изменения результативности МФТИ (а) и РУДН (б) целым счетом и KBPР с учетом только первой аффилиации и коэффициент «первенства» по KBPР

Последствие 3. От внедрения KBPР могут пострадать программы привлечения ведущих зарубежных ученых. В [55] мы подробно рассматривали проблему множественных аффилиаций, феномена «покупки академического престижа» и приводили примеры конкретных организаций, которые по программам привлечения ведущих зарубежных ученых сотрудничают с авторами, часто весьма продуктивными, которые указывают эту организацию в числе одной из своих аффилиаций. Эти авторы обычно не вступают в коллаборации с другими сотрудниками организации, поэтому при переходе на фракционный счет и KBPР их результативность будет невысока, что, вероятно, не будет оправдывать значительные затраты по их привлечению.

Последствие 4. От перехода на KBPР в версии от 23.04.2020 выиграют недобросовестные представители гуманитарных и общественных наук. Не

секрет, что Россия – одна из лидеров по публикациям в хищнических журналах [56]. Значительная часть этих публикаций относится к гуманитарным и общественным наукам. Некоторые вузы или отдельные подразделения этих вузов строят свои публикационные стратегии на хищнических журналах [43], причем процесс не останавливается. Так, сотрудники Института управления, экономики и финансов Казанского федерального университета активно публикуются в такого рода журналах, причем, при исключении журнала из индексации в Scopus, они переходят в другой журнал, который попадает под исключение в свою очередь (таблица 11). Крайне маловероятно, что это случайное совпадение.

Таблица 11 – Журналы с публикациями сотрудников КФУ, исключенные из индексации в Scopus

Журнал	Статей за 5 лет	Индексация
Mediterranean Journal Of Social Sciences	206	2012-2015
International Journal Of Environmental And Science Education	202	2009-2016
Social Sciences Pakistan	199	2009-2016
Journal Of Language And Literature	185	2011-2016
Man In India	128	2000-2017
Opción	123	2008-2020
Asian Social Science	120	2011-2016
Journal Of Sustainable Development	118	2014-2015

За счет плоской шкалы баллов для гуманитарных и общественных наук такие авторы и организации выигрывают у добросовестных ученых. При этом наш анализ структуры публикационного потока по квартилям журналов в разрезе тематических направлений показал, что в этом направлении наук весомая часть статей российских авторов публикуется в престижных журналах.

4 Анализ региональной и секторальной структуры документопотока

В целях регионального и секторального (вузы/научные организации/бизнес) анализа библиографические данные публикаций, полученные из БД Scopus, были дополнены: каждая российская аффилиация авторов была сопоставлена по адресу, а при его отсутствии – по данным организации с регионом (субъектом федерации), а также типом организации (сектором науки). Среди организаций выделены высшие учебные заведения, академические институты (находившиеся до реформы под руководством государственных академий наук), остальные организации отнесены к категории «прочие». Профили (идентификаторы) Scopus ведущих по публикационной активности организаций были сопоставлены с ИНН и их тип определялся по подготовленным спискам с учетом ведомственной принадлежности и организационно-правовой формы. «Длинный хвост» организационных профилей (менее 50 публикаций за 20 лет), в который попали в том числе многочисленные дублирующие профили был разобран по наименованиям.

Мы предполагаем наличие погрешностей в таком подходе, связанное с неполными или ошибочными аффилиационными данными в описаниях публикаций, размер которой еще предстоит оценить. Работа по идентификации организационных профилей и данных публикаций в Scopus продолжается как командой поддержки Elsevier, так и лабораторией наукометрии ГПНТБ СО РАН, поэтому точность такого анализа будет постепенно улучшаться. Существует также особый класс погрешностей, в основном затрагивающий региональный анализ – он связан с «зонтичными» аффилиациями (например, Российская академия наук) и территориально распределенными организационными структурами. И в том, и в другом случае автор может быть аффилирован с организацией высшего уровня. Например, автор из академического научно-исследовательского института может быть аффилирован с Российской академией наук и ошибочно отнесен к Москве;

автор из регионального филиала крупной организации – к расположению головной организации. Такие ошибки приводят к несколько завышенным показателям отдельных регионов, прежде всего, г.Москва.

Расчет структуры документопотока выполнялся полным фракционным методом с учетом всех аффилиаций всех авторов публикаций. Таким образом, полученные данные адекватно показывают публикационный вклад каждого субъекта анализа – региона, сектора и т.д. в мировую науку.

4.1 Региональная структура документопотока

Региональная структура российской науки меняется с течением времени. Это происходит, прежде всего, в связи с ускоренным развитием сектора исследований и разработок в регионах, в начале рассматриваемого периода не обладавших существенным научным потенциалом. Если в начале 2000х гг. на Москву приходилось 45,6% научного продукта международного уровня, то к концу рассматриваемого периода доля Москвы упала до 37,2%. Заметно упал также вклад Московской области - с 4,8% до 2,2%. Менее значительное падение доли пришлось и на такие развитые в научном плане регионы, как Новосибирская область (с 7,8% до 5,5%), Санкт-Петербург (с 13,2% до 12,3%), Нижегородская область (с 2,4% до 1,8%). Активно развивался научный потенциал в Республике Татарстан (рост с 2% до 3,7%), Томской области (с 2,8% до 3,7%), Ростовской области (с 1,5% до 2,2%) и многих других регионах. Значительный подъем достигнут в Тюменской, Волгоградской, Самарской, Челябинской областях, Пермском крае (рисунок 29).

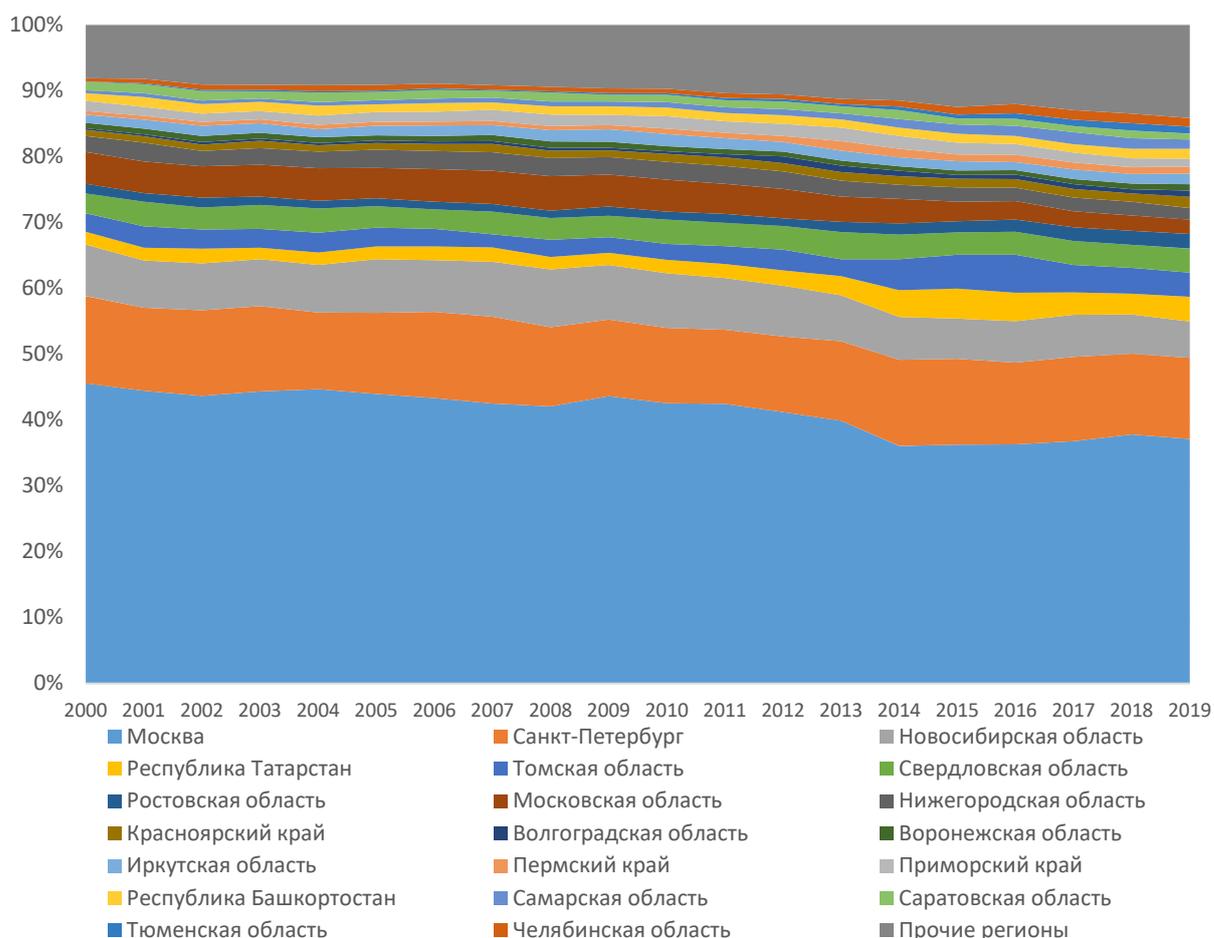


Рисунок 29 – Динамика изменения региональной структуры российской науки

Как было показано в главе 2.4, тематическая структура также претерпела заметные изменения. В основном эти изменения связаны с ускоренным развитием медицинских, общественных и гуманитарных наук. Как видно из сопоставления рисунков 30 и 31, наиболее заметным развитие медицинского направления было в Иркутской, Самарской, Свердловской области, Пермском крае, Республике Татарстан. Вклад общественных и гуманитарных наук вырос практически повсеместно из-за эффекта низкой базы.

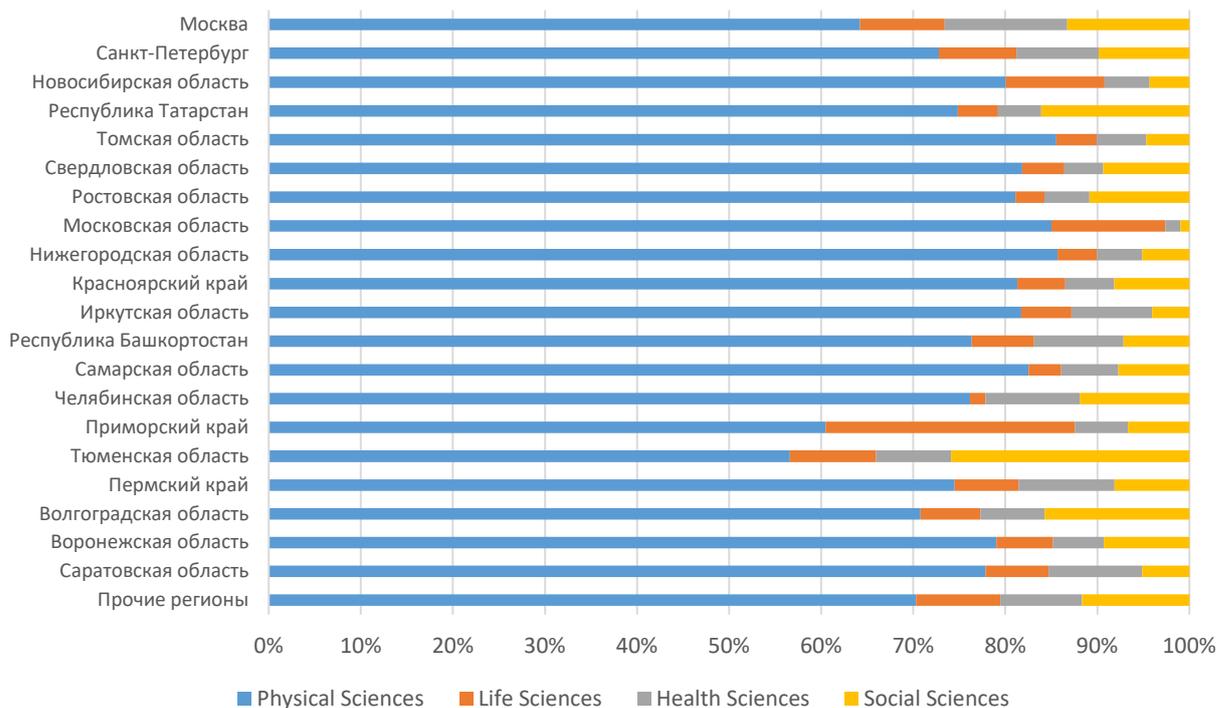


Рисунок 30 – Тематическая структура российской науки по регионам в 2019 г.

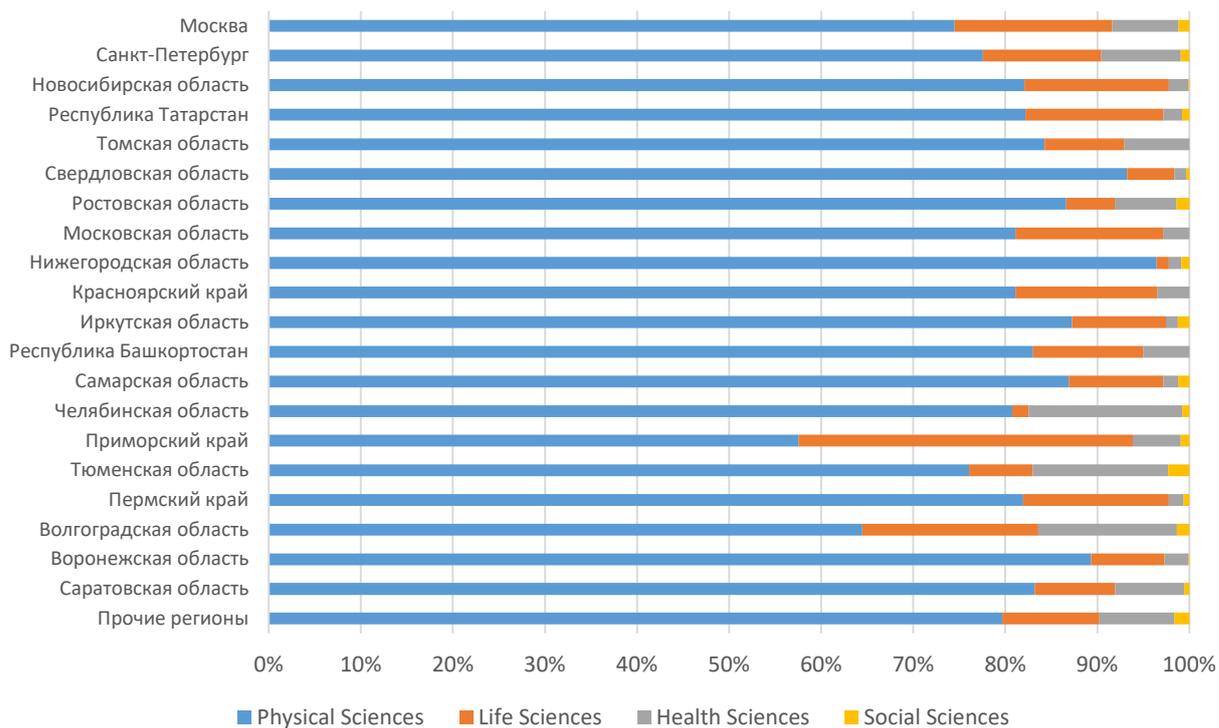


Рисунок 31 – Тематическая структура российской науки по регионам в 2000 г.

4.2 Секторальная структура документопотока

Результаты анализа динамики изменения секторальной структуры публикационного потока российских исследователей (рисунок 32) показывают заметное сжатие доли публикаций академических институтов и немного меньшее – сектора, объединяющего отраслевую, корпоративную науку, государственные научные центры разного уровня. Доля публикаций академических институтов снизилась с 39% в 2000 до 24% - в 2019. Это сжатие произошло за счет опережающего роста науки в вузах, этот сектор вырос с 29% в 2000 до 51% в 2019. Как уже было сказано выше, этот эффект является следствием государственной политики по перемещению исследовательской активности в университеты, сопровождающийся значительным количеством программ финансирования [57]. Можно также отметить, что с 2015 г. доля публикаций академических институтов начала стабилизироваться за счет активного роста публикационного потока, в том числе вследствие увеличения финансирования и усиления «наукометрического» давления последовавших за перестройкой, связанной с реформой государственных академий наук. Наиболее активный рост доли университетской науки пришелся на 2013-2016 гг. (в 2014 г. – рост на 35%, в 2016 – на 28%), природу этого роста мы рассматривали в [43]. Сектор несколько снизил динамику и в 2017-2019 гг. растет на 10-15% в год.

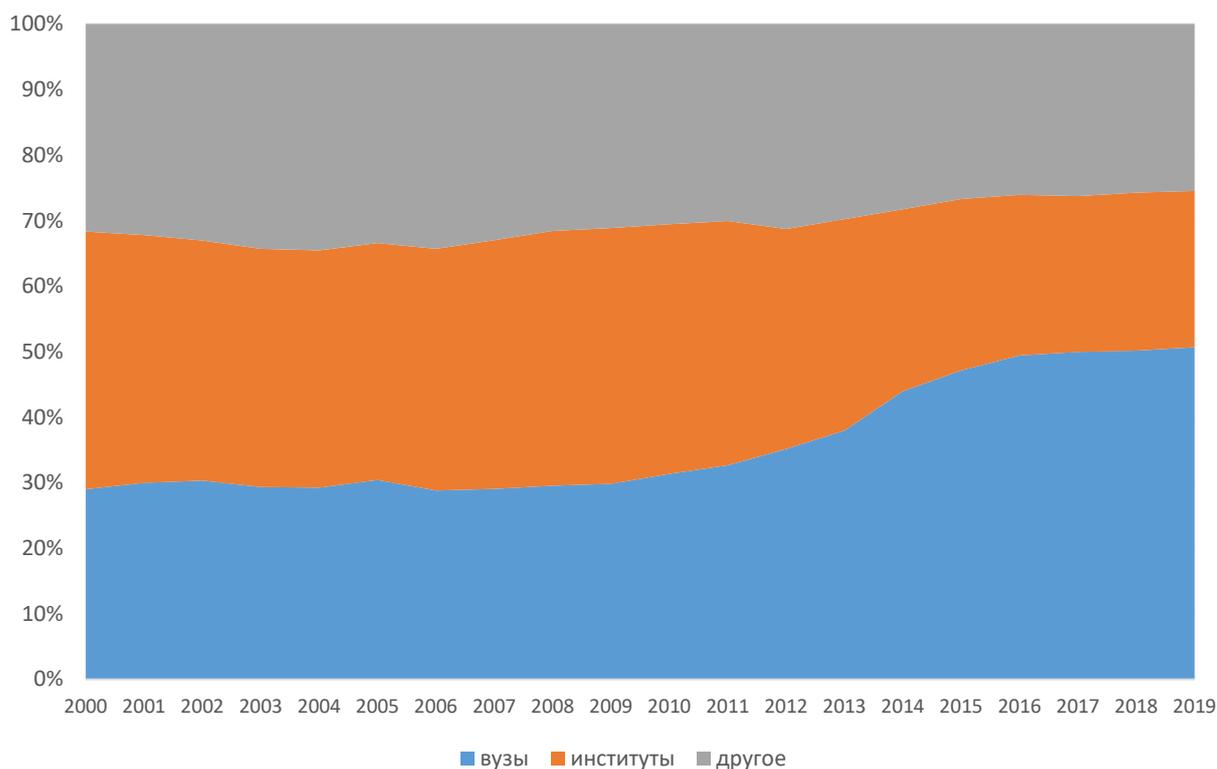


Рисунок 32 – Динамика изменения секторальной структуры российской науки

Необходимо отметить, что особые успехи вузы демонстрируют в относительно новых для российской науки областях (рисунок 33). В традиционно сильных для РФ научных направлениях по-прежнему сильны академические институты. Лидерство академической науки сохраняется в физике и астрономии (34% у институтов против 25% у вузов в 2019 г.), химии (47% / 38%), науках о Земле (61% / 14%), науках о материалах (36% / 34%), биохимии, генетики и молекулярной биологии (55% / 19%), биологии и сельском хозяйстве (55% / 24%), иммунологии и микробиологии (53% / 20%), нейронауках (57% / 15%) и некоторых других. Вузы вышли на ведущие позиции в таких областях, как инженерия / технические науки (44% у вузов против 21% у институтов в 2019 г.), математике (48% / 31%), химических технологиях (43% / 27%), энергетике (23% / 6%), компьютерных науках (39% / 25%) и других. В некоторых из этих областей вузы занимают лидирующие или ведущие позиции весь исследуемый период, но сравнительный анализ

картины за 2000 г. и 2019 г. (рисунок 34), показывает, что изменение баланса видно практически по всем направлениям исследований.

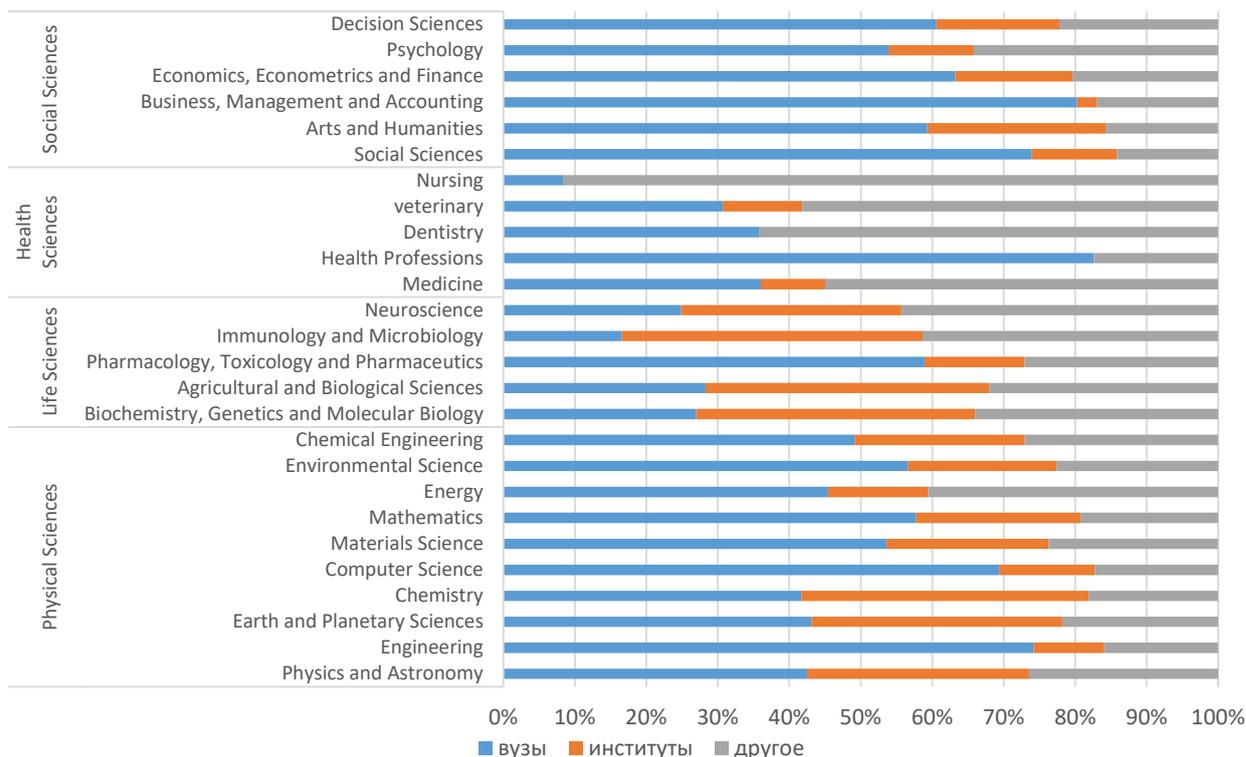


Рисунок 33 – Секторальная структура российской науки по направлениям исследований в 2019 г.

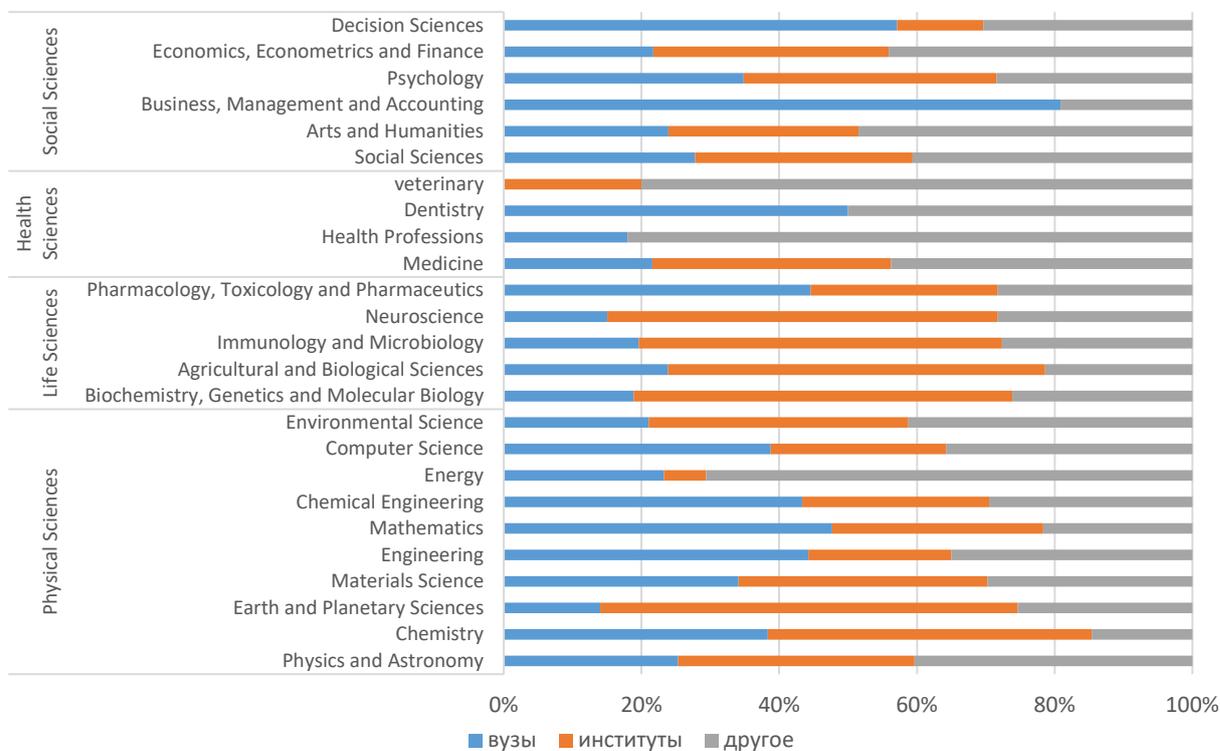


Рисунок 34 – Секторальная структура российской науки по направлениям исследований в 2000 г.

Среди ведущих в научном плане российских регионов остаются преимущественно «академические» – с преобладающим влиянием академических институтов. Такими регионами остаются Московская область с научными центрами в Черноголовке, Фрязино, Шатуре, Пушкино, Одинцово, Протвино; Новосибирская область с Новосибирским Академгородком, Приморский край с институтами ДВО РАН во Владивостоке, Иркутская область с Иркутским научным центром (рисунок 35).

Необходимо отметить, что мы рассматривали структуру публикационного потока, индексируемого в БД Scopus. При этом количество российских публикаций активно росло с 2013 г. под влиянием изменения национальной научной политики за счет различных факторов, прежде всего – увеличения количества активных исследователей, часть из которых, вероятно, начала заниматься наукой в этот период, а часть – стала публиковаться в индексируемых в этой БД изданиях. Важным фактором является и постоянное увеличение количества индексируемых в Scopus российских журналов и трудов конференций.

Основные изменения в структуре российской науки связаны с ускоренным ростом сектора университетской науки, и некоторых региональных научных и образовательных центров, что позволяет сделать вывод о решающем влиянии эффекта «низкой базы». Рост шел преимущественно там, где имелась значительная кадровая база, прежде всего в вузах, но свою роль сыграли и относительно слабые академические институты, государственные научные центры, отраслевая и корпоративная наука, заметно нарастившие публикационную активность на международном уровне. Ведущие научные организации и до этого периода активно работали и публиковались в индексируемых в Scopus изданиях, поэтому смогли нарастить результативность незначительно.

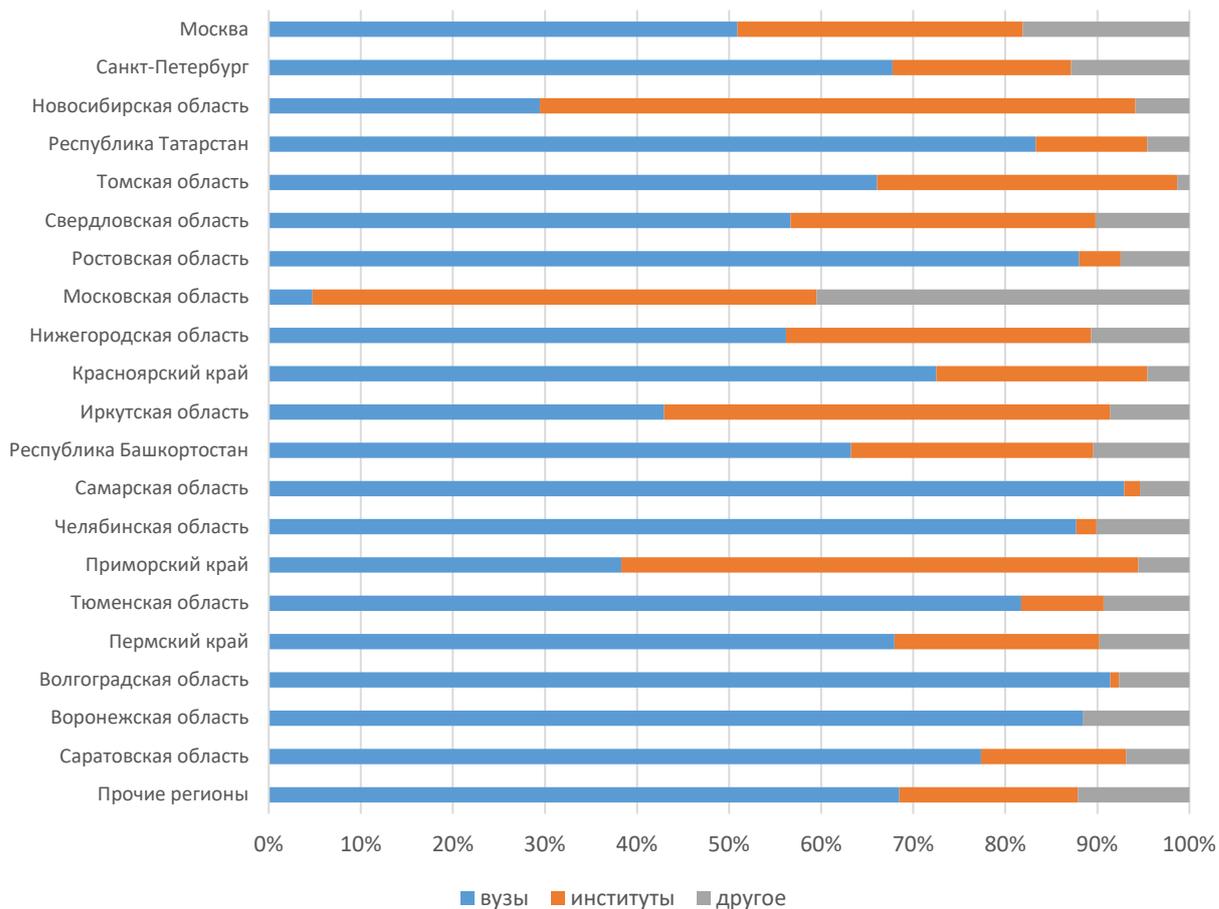


Рисунок 35 – Секторальная структура российской науки по регионам в 2019 г.

5 Сопоставление российского публикационного потока с результатами стран – конкурентов России в мировом научном пространстве

В качестве конкурентов выбраны три страны с существенно различающимися моделями организации науки:

- Франция, одна из сильных в научном плане европейских стран, занимает в 2019 г. восьмое место в рейтинге Scimago Country Rank;
- Бразилия, активно развивающаяся в научном плане страна с значительным сегментом национальной научной периодики на португальском языке, в 2019 г. занимает 14 место в рейтинге, Россия обогнала эту страну в 2016 г.;
- Индия, вышла на четвертое место в Scimago Country Rank, одна из наиболее динамично развивающихся в науке стран.

Интерес представляет сопоставление структуры научных комплексов этих стран и Российской Федерации в секторальном и тематическом разрезе.

5.1 Извлечение и подготовка данных, уточнение тематики и определение секторальной принадлежности

Анализ был выполнен на основе данных о публикациях, аффилированных с Индией, Бразилией и Францией и опубликованных в 2000-2019 годах. Тематика публикации была уточнена методом, описанным выше. Секторальная принадлежность, определялась на основе анализа наименований организаций – аффилиаций авторов с использованием справочника ключевых слов. Для выявления организаций высшего образования были использованы следующие ключевые термины: 'university', 'universidade', 'universita', 'universite', 'universitet', 'université', 'college', 'collège', 'faculdade', 'faculté', 'faculte', 'faculty', 'school', 'ecole', 'école', 'escola', 'imt'; для исследовательских организаций - 'institute', 'instituto', 'institut', 'institution', 'inst', 'laboratory', 'laboratorio', 'laboratoire', 'lab', 'laboratorio', 'laboratories', 'cea', 'recherche', 'research', 'scientifique', 'sciences', 'science', 'scientific', 'synchrotron', 'center', 'centro', 'centre', 'foundation', 'fundacao', 'fondation', 'inra', 'inria', 'insa',

'observatoire', 'observatório' и т.д.. Организации классифицировались по 6 секторам: образовательные, исследовательские, медицинские, коммерческие, государственные и остальные. Каждая организация попадает только в один сектор, в случае спорных ситуаций, когда организация может быть отнесена сразу к двум или более секторам (например, «Исследовательская больница при государственном университете») разработана система приоритетов, сектора в предыдущем предложении перечислены в порядке уменьшения приоритета. В целях сопоставления с данными о секторальной структуре российской науки были выделены образовательный и исследовательский сегмент, все остальные были объединены в «прочие». Такая классификация носит достаточно условный характер, особенно с учетом различных национальных особенностей организации сектора исследований и разработок в этих странах, но, тем не менее, позволяет получить общее представление о секторальной структуре и провести сопоставление.

5.2 Структура публикационного потока по типам документов и квартилям научных журналов

В рассматриваемый период с 2000 по 2019 год публикационный поток всех четырех стран рос. Наиболее интенсивный рост наблюдался у Индии – за 20 лет годовой публикационный поток увеличился почти в 8 раз (7,85), у Бразилии – в 6 раз, РФ в 3,4 раза и Франции в 2,1 раз. При этом в 2019 г. количество индийских публикаций (198 тыс.) более чем вдвое превысило число публикаций Бразилии (97 тыс.). Россия и Франция опубликовали в 2019 г. 120 и 129 тыс. публикаций соответственно. В 2000 г. публикационный поток Индии (25 тыс.) был выше бразильского (16 тыс.), но существенно ниже российского (35 тыс.) и французского (62 тыс.). Таким образом, в рассматриваемый период времени, Индия существенно обогнала конкурирующие с ней страны, включая Россию, по числу опубликованных работ. Россия, в свою очередь, по этому показателю сократила отставание от Франции и увеличила отрыв от Бразилии.

Как и в России, в публикационном потоке Франции, Бразилии и Индии преобладают журнальные статьи и обзоры. Доля документов Ar+Re снижается с течением времени во всех трех странах: Франции с 81,4% в 2000 г. до 74,8% в 2019, Бразилии – с 85% до 83,6%, Индии – с 89,2% до 72,9%. Минимальные значения во Франции и Бразилии приходились на 2005 г, в то время как в РФ основное снижение доли этих типов документов пришлось на последние годы (рисунок 36). Соответственно, доля публикаций в трудах конференций во Франции и Бразилии снижалась с 2005 г., в Индии – устойчиво росла до 2019 г., в РФ – росла с 2008 г., причем особо активный рост начался с 2013 г. В 2019 г. в России доля публикаций в трудах конференций – максимальная из 4 рассмотренных стран, а доля статей и обзоров – минимальная.

С точки зрения квартилей научных журналов наиболее качественная структура публикационного потока у французских исследователей, доля статей в журналах Q1 достигает 54% по Scimago Journal Ranking (рисунок 37) и 29% по Journal Citation Report (рисунок 38 Рисунок 38 –).

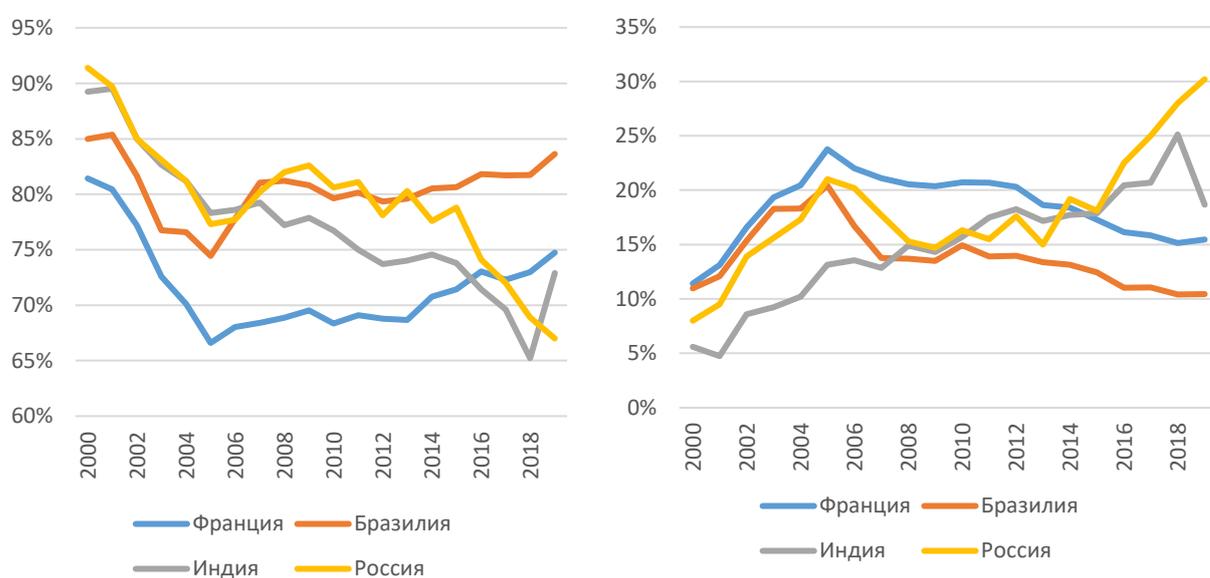


Рисунок 36 – Доля документов Ar+Re (а) и CP (б) в публикационном потоке

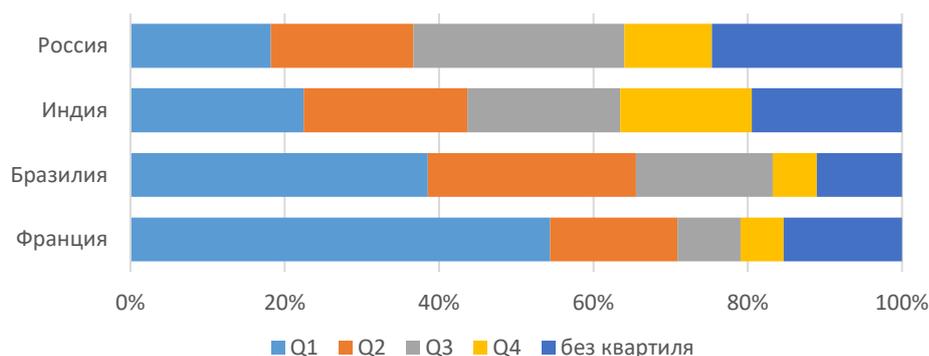


Рисунок 37 – Структура публикационного потока за 2019 г. по квантилям научных журналов по Scimago Journal Ranking

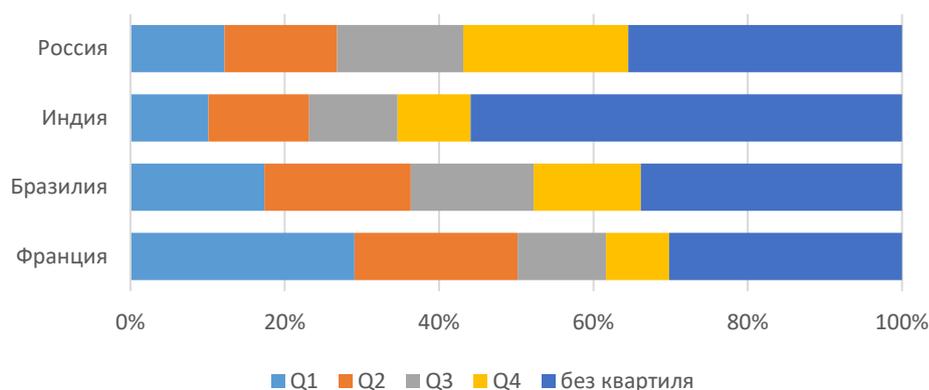


Рисунок 38 – Структура публикационного потока за 2019 г. по квантилям научных журналов по Journal Citation Report

Согласно рисункам 39 и 40, в рассмотренных странах наблюдается разная динамика изменения этой структуры. Так, доля индийских статей в Q1+Q2 падает с 2007 г. и по SJR и по JCR, но по JCR наблюдается стабилизация и даже небольшой рост в последние годы (Франция и Бразилия, наоборот, постепенно улучшают структуру публикационного потока начиная с 2006 г. Россия выглядит чуть лучше при расчетах по JCR, вероятно, в связи с исторически сложившейся преимущественной ориентацией на этот рейтинг.

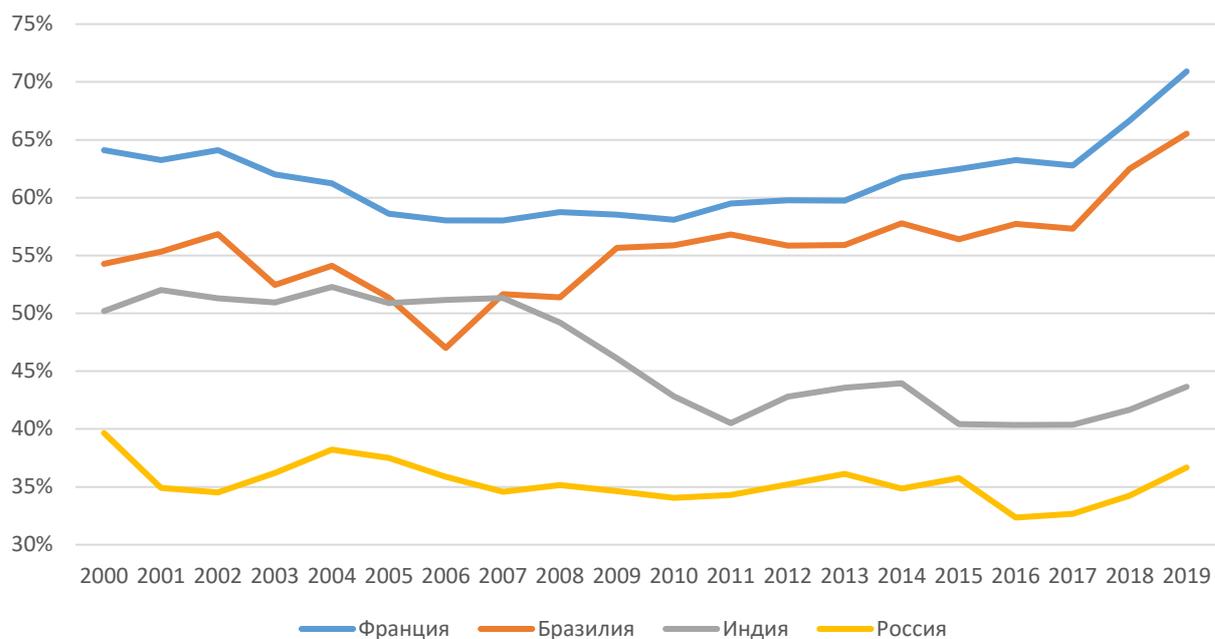


Рисунок 39 – Динамика доли публикаций в Q1+Q2 по Scimago Journal Ranking

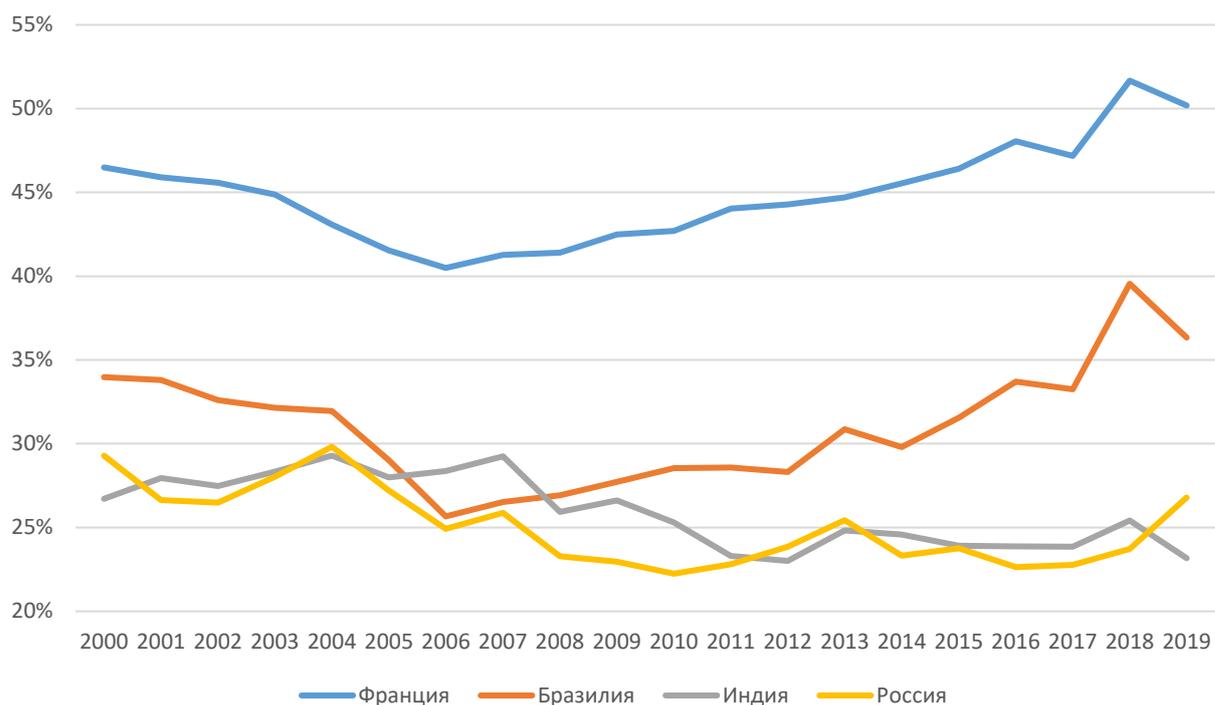


Рисунок 40 – Динамика доли публикаций в Q1+Q2 по Journal Citation Report

5.3 Тематическая структура публикационного потока Бразилии, Франции, Индии и России

Как мы отмечали выше, в российской науке доминирует направление Physical Sciences, но на протяжении рассматриваемого периода это доминирование ослабевает, прежде всего за счет ускоренного развития общественных и гуманитарных наук (рисунок 41). Можно заметить, что среди рассматриваемых стран Индия движется в обратном направлении – доля Physical Sciences в индийской науке растёт, прежде всего за счет относительно медленного развития направления Health Sciences и Life Sciences. То же явление, но в гораздо меньшей степени наблюдается и в бразильской науке. Рост доли общественных и гуманитарных наук (Social Sciences) наблюдается во всех рассматриваемых странах, но во Франции это происходит за счет уменьшения доли естественных наук.

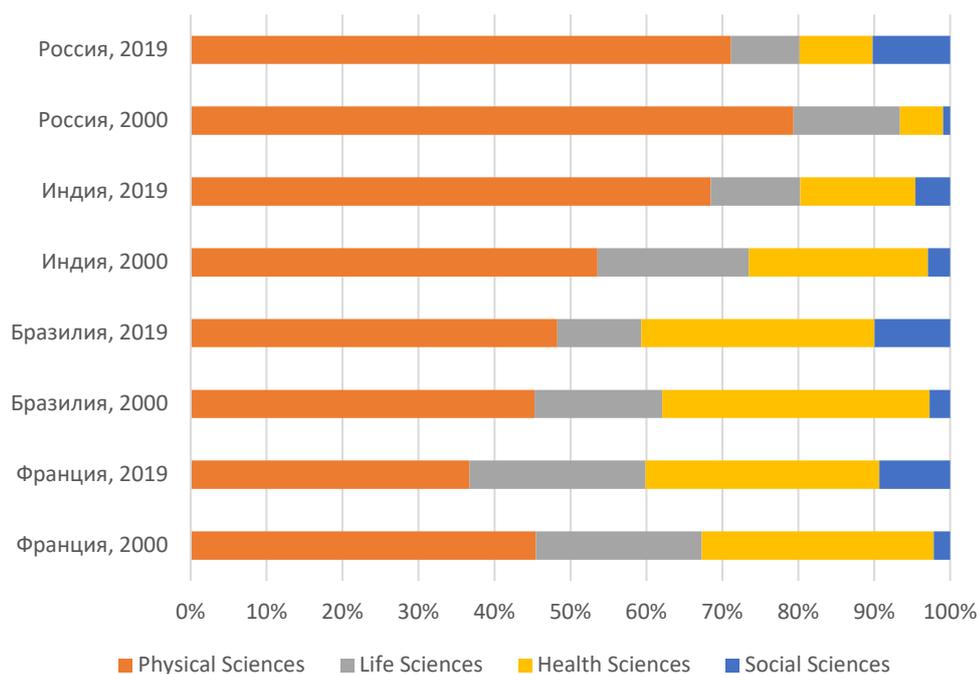


Рисунок 41 – Тематическая структура по направлениям наук в 2000 и 2019 гг.

Тематическая структура национальных наук в 2000 и 2019 гг. представлена в таблице 12. Необходимо отметить смену лидирующих областей наук в отдельных направлениях. Так в направлении естественных наук (Physical Sciences) в 2000 г. физика и астрономия были на лидирующих

позициях в Бразилии, Франции и России, в Индии эта область наук немного уступала химии. В 2019 г. она осталась лидером только в России, в трех других странах на первое место вышла область Engineering, второе место занимает Computer Science, причем в индийской науке – с заметным отрывом. В направлении наук о жизни изменения не столь драматические: лидерами являются агронауки и биология, а также биохимия, генетика и молекулярная биология, но необходимо заметить, что роль последней области падает во всех рассмотренных странах. В направлении наук о здоровье лидирующие позиции занимает медицина, но ее доля постепенно уменьшается во всех странах кроме России. Примечательные трансформации происходят в структуре направления общественных и гуманитарных наук, влияние которых в целом росло во всех странах. Если в 2000 г. везде лидировали общественные науки, то к 2019 г. в рассматриваемых странах произошли разнонаправленные движения: в Индии на первое место вышла область Business, Management and Accounting, во Франции – искусство и гуманитарные науки (Arts and Humanities), заметный рост в этой области произошел и в России. Общественные науки сохранили лидерство в Бразилии и России. Еще раз отметим, что трансформации в этой области происходили на фоне роста всего направления, поэтому они обусловлены разными темпами этого роста в разных областях.

Таким образом, на уровне направлений наук структура российской науки постепенно трансформируется в том же направлении, что и в Бразилии и Франции, к 2019 г. за счет встречного движения приблизившись к структуре индийской науки. На уровне областей наук направления трансформации тоже в целом совпадают. В связи с этим можно констатировать, что заметный потенциал развития есть в таких областях, как Computer Science, Agricultural and Biological Sciences и Medicine.

Таблица 12 – Сравнение тематической структуры по направлениям и областям наук в Бразилии, Франции, Индии и России в 2000 и 2019 гг.

Направления / области наук	Бразилия		Франция		Индия		Россия	
	2000	2019	2000	2019	2000	2019	2000	2019
Physical Sciences	45,4%	36,7%	45,3%	48,3%	53,6%	68,5%	79,4%	71,2%
Chemical Engineering	2,2%	1,5%	0,8%	0,6%	1,4%	1,0%	1,5%	0,9%
Chemistry	8,3%	5,2%	7,3%	5,1%	14,2%	6,6%	14,7%	7,9%
Computer Science	4,3%	7,4%	4,7%	9,6%	2,8%	18,7%	1,2%	6,3%
Earth and Planetary Sciences	2,5%	2,6%	3,9%	4,1%	3,4%	2,0%	7,8%	8,7%
Energy	0,7%	0,6%	0,3%	0,6%	0,6%	1,0%	1,3%	1,6%
Engineering	8,1%	7,6%	6,9%	10,3%	8,2%	22,3%	6,7%	12,7%
Environmental Science	1,1%	2,4%	1,1%	1,6%	3,1%	2,0%	0,6%	1,6%
Materials Science	3,2%	3,1%	3,8%	3,1%	6,8%	6,2%	7,0%	5,6%
Mathematics	2,4%	2,2%	4,2%	4,3%	1,9%	2,4%	4,8%	4,8%
Physics and Astronomy	12,6%	4,1%	12,2%	9,0%	11,1%	6,2%	33,7%	21,1%
Life Sciences	21,9%	23,2%	16,8%	11,1%	19,9%	11,8%	14,0%	9,0%
Agricultural and Biological Sc.	13,4%	18,2%	5,1%	4,5%	11,2%	5,1%	3,5%	3,7%
Biochemistry, Genetics and...	4,9%	2,6%	8,8%	4,9%	5,2%	3,3%	8,2%	4,1%
Immunology and Microbiology	0,8%	0,6%	0,8%	0,5%	0,6%	0,3%	1,0%	0,4%
Neuroscience	1,1%	0,6%	1,4%	0,9%	0,1%	0,1%	0,7%	0,3%
Pharmacology, Toxicology and...	1,8%	1,2%	0,7%	0,4%	2,8%	3,0%	0,6%	0,5%
Health Sciences	30,5%	30,8%	35,2%	30,7%	23,6%	15,1%	5,7%	9,6%
Dentistry	1,1%	2,6%	0,1%	0,1%	0,1%	0,5%	0,0%	0,0%
Health Professions	0,0%	0,1%	0,0%	0,1%	0,0%	0,0%	0,0%	0,1%
Medicine	26,8%	24,6%	34,6%	29,8%	20,8%	14,1%	5,7%	9,4%
Nursing	1,4%	1,7%	0,1%	0,5%	0,0%	0,0%	0,0%	0,0%
Veterinary	1,3%	1,8%	0,5%	0,3%	2,6%	0,5%	0,0%	0,0%
Social Sciences	2,2%	9,3%	2,7%	9,9%	2,9%	4,6%	0,9%	10,2%
Arts and Humanities	0,4%	1,9%	0,5%	3,1%	0,2%	0,4%	0,2%	3,4%
Business, Management and...	0,2%	1,6%	0,2%	1,5%	0,9%	2,4%	0,2%	1,6%
Decision Sciences	0,1%	0,1%	0,0%	0,1%	0,1%	0,1%	0,0%	0,1%
Economics, Econometrics and...	0,2%	0,6%	0,5%	1,6%	0,4%	0,6%	0,1%	1,2%
Psychology	0,2%	1,2%	0,4%	1,0%	0,1%	0,1%	0,2%	0,4%
Social Sciences	1,1%	3,9%	1,0%	2,7%	1,3%	1,1%	0,3%	3,6%

5.4 Секторальная структура публикационного потока Бразилии, Франции, Индии и России

В связи с особенностями методики отнесения организаций к секторам и национальной специфики организации науки наиболее достоверным является определение принадлежности к сектору высшего образования. К научно-исследовательским институтам в России отнесены только академические организации, в связи с этим доля этого сектора с учетом государственных

научных центров и отраслевых (ведомственных) институтов должна быть заметно выше. С другой стороны, исследования в области медицины в РФ ведутся в НИИ, параллельно занимающихся клинической практикой, что нехарактерно для науки в других странах.

Сопоставление структуры национальной науки в 2000 г. и 2019 г. для рассматриваемой четверки стран приведено на рисунок 42. Заметно, что к структуре российской науки как в 2000 г., так и в 2019 г. наиболее близка структура науки Франции. В российской науке к 2019 г. вузы обеспечили даже большую долю публикаций – 51%, по сравнению с 47% во Франции. Традиционно доминирует доля вузов в науке Бразилии, в 2000 г. они давали 78% национального научного продукта, а в 2019 г. их доля выросла до 83%. В целом, вклад вузов рос во всех рассмотренных странах, но в Российской Федерации темпы этого роста были наиболее впечатляющими.

В структуре французской науки заметную роль играют исследовательские клиники и корпоративная наука. В Индии и Бразилии доля корпоративной науки меньше, а медицинские исследования в большей степени проводятся в университетских клиниках, в соответствии с описанным алгоритмом определения секторальной принадлежности они относятся к сектору высшего образования. В сочетании с большей долей медицинских исследований во Франции это дает больший вес сектору «прочие».

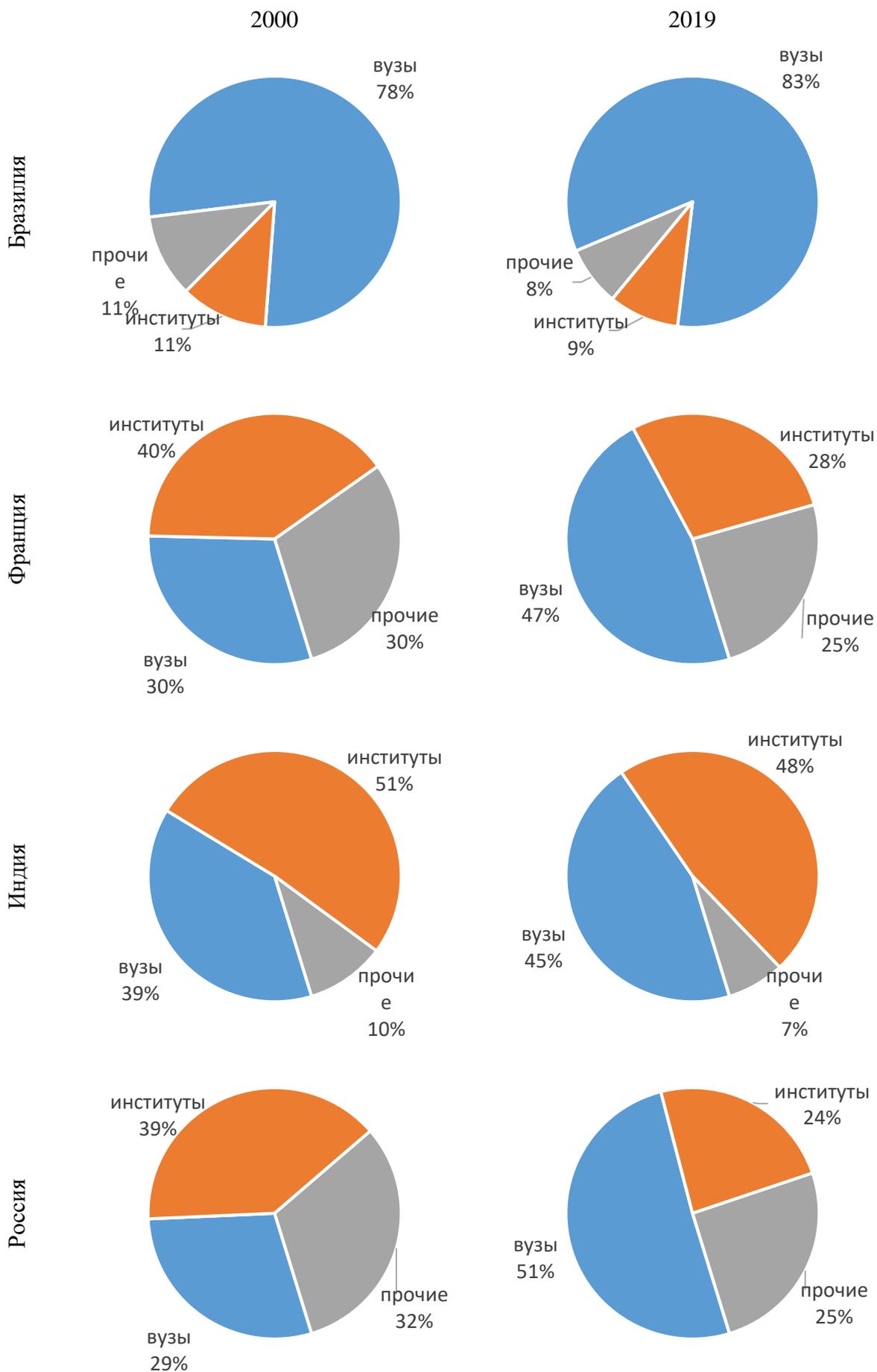


Рисунок 42 – Секторальная структура национальной науки в 2000 и 2019 гг.

6 Инфраструктура российской научной периодики

К настоящему времени разработан широкий набор библиометрических показателей для оценки качества журнала, уровня его интернационализации и влияния в научном мире. Однако в некоторых случаях необходимы более точные методы оценки периодических изданий, например, при оценке новых журналов, их отборе в международные базы данных цитирования, повышении их рейтинга, расширении их читательской аудитории или внесении корректив в редакционную политику. Для решения этих задач формируются новые подходы, включая альтметрики и различные виды анализа в составе редакционной коллегии и ее влиянии на показатели журнала. Анализ альтметрик и изучение влияния редколлегии на рейтинг журнала, сравнение библиометрических показателей редколлегии с индикаторами обычных авторов журнала, оценка исследовательской эффективности членов редколлегии являются относительно неизученными, но перспективными направлениями в более точной оценке журнального рейтинга. Работа текущего этапа по проекту заключалась, с одной стороны, в исследовании альтметрик как возможного предиктора будущих цитирований и других показателей, основанных на цитировании. С другой – был проведен анализ публикационного вклада редколлегии в рейтинговые показатели журнала.

6.1 Анализ альтметрик как предиктора будущих цитирований научного журнала

Для анализа альтметрик в приложении к оценке качества российской научной периодики была использована база данных Scopus, отражающая информацию крупнейшего агрегатора альтметрик Plum Analytics. Всего с помощью Scopus было отобрано 12 различных альтметрик Plum Analytics для 16 ведущих российских научных журналов о Земле. Эти 12 показателей распределены по пяти широким категориям в таблице 13.

Таблица 13 – Альтметрики Plum Analytics, доступные для оценки российских журналов

Использование	Сохранения	Упоминания	Социальные сети	Цитирования
Clicks	Exports-saves	Blog mentions	Shares, likes & comments	Citation indexes
Abstract views	Readers	References	+1s	
Full text views			Tweets	
Link-outs				

Выявлено, что для дальнейшего анализа подходят только две альтметрики (abstract views и full text views), так как данные по этим двум показателям были представлены в максимальной степени по всем журналам. Кроме того, эти две альтметрики позволили нам предложить собственный показатель альтметрик, характерный для российских журналов, – количество просмотренных статей в каждом анализируемом издании. Что касается остальных десяти показателей, то они доступны для того или иного журнала на непостоянной основе и не подходят для систематического анализа.

При анализе альтметрик установлено, что наиболее широко альтметриками охватываются три российских журнала по наукам о Земле, при том, что среднее число просмотров одной статьи составляет более 30: Lithology and Mineral Resources – 46,2 просмотра на статью, Petroleum Chemistry – 36,6 просмотра на статью и Soil Mechanics and Foundation Engineering – 30,2 просмотра на статью. Ввиду полноты информации для этих изданий было проведено сравнение данных на основе альтметрик с показателями на основе цитирования (рисунки 43 и 44).

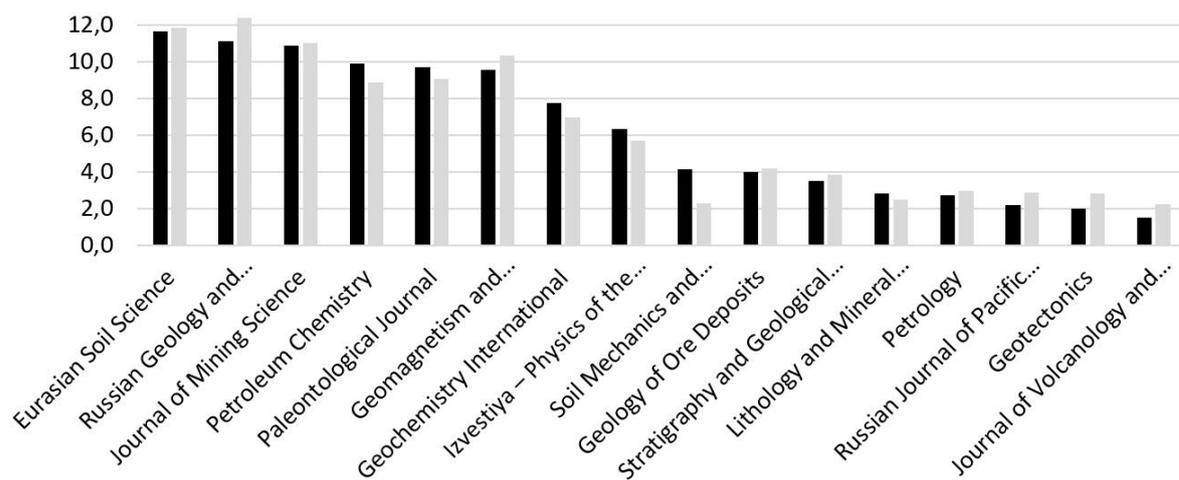


Рисунок 43 – Зависимость между нормированным числом просмотренных и процитированных статей в российских журналах по наукам о Земле по базам данных Plum Analytics / Scopus. Черные столбцы – нормированное число просмотренных статей; серые столбцы – нормированное число процитированных статей

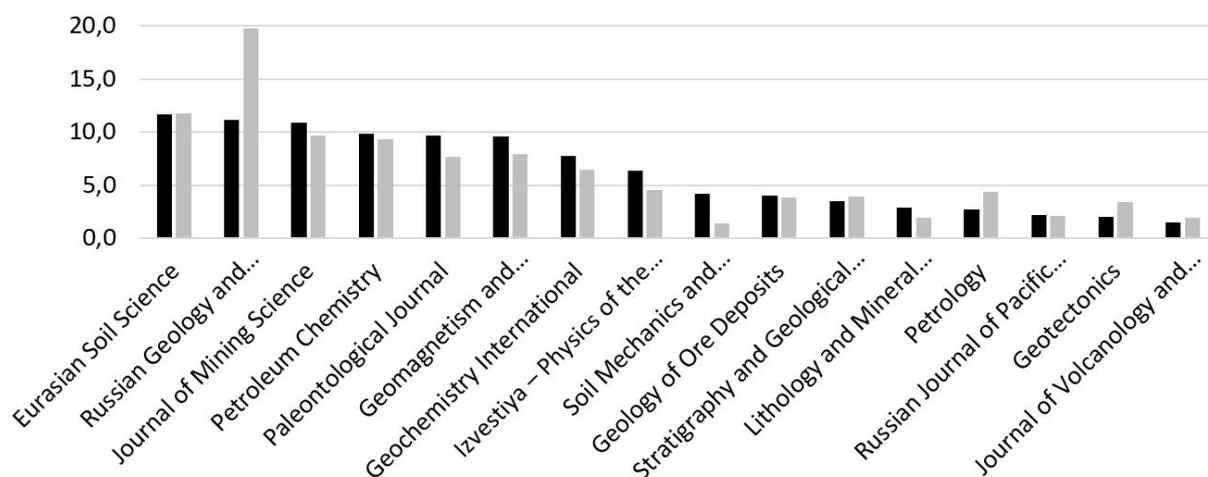


Рисунок 44 – Зависимость между нормированным числом цитирований и нормированным числом просмотренных статей по базам данных Plum Analytics / Scopus. Черные столбцы – нормированное число просмотренных статей; серые столбцы – нормированная цитируемость

Некоторые журналы показали необычную ситуацию преобладания количества цитируемых статей над просмотренными (рисунок 43). Теоретически это возможно при условии, что некоторые авторы просматривали статью, а затем цитировали ее в нескольких собственных статьях. Однако, скорее, это может быть связано с другими причинами:

а) Plum Analytics не охватывает российские источники, например, РИНЦ в то время как большинство ссылок на перечисленные журналы делаются российскими авторами, которые знакомятся с содержанием журналов из российских же баз данных;

б) ряд российских журналов охвачен альтметриками не полностью, поскольку в одних случаях среднее число просмотров одной статьи составляло 30–40, в то время как для других журналов того же качества (с сопоставимым числом цитирований) среднее число просмотров составляет лишь один или два. При этом выборочная проверка альтметрик, рассчитанных в РИНЦ для нескольких журналов, выявила у этих журналов высокое число просмотров в расчете на одну статью;

в) в то время как указатели цитирований, такие как Web of Science или Scopus, объединяют цитирование русской и переведенной версий в одном журнале, Plum Analytics, похоже, предоставляет информацию только для переводных версий и, таким образом, не учитывает альтметрики, сделанные как минимум половиной русскоязычных активных пользователей. Также при сборе альтметрик мы наблюдали отсутствие данных за некоторые периоды в отдельных журналах.

На рисунке 45 показана высокая степень зависимости между рейтингом журнала Scimago (SJR) и импакт-фактором на основе альтметрик, который мы рассчитали аналогичным образом как отношение полученных просмотров к количеству статей для трех вышеназванных журналов, в наибольшей мере охваченных альтметриками Plum Analytics.

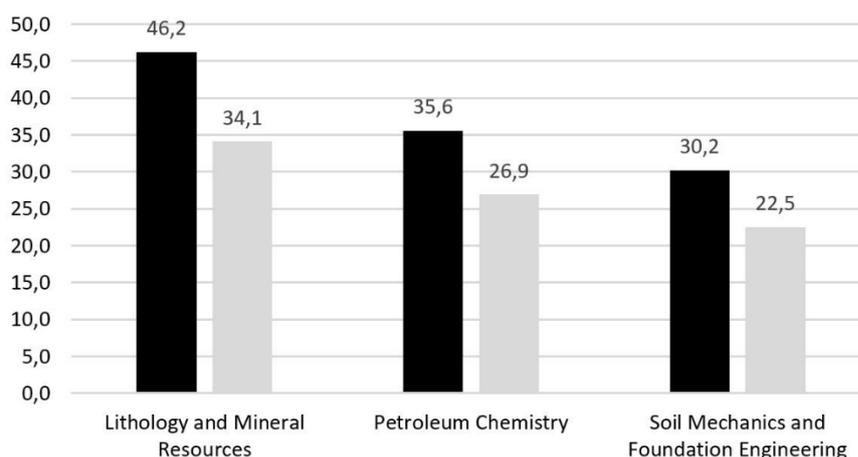


Рисунок 45 – Зависимость между «импакт-фактором», основанным на альтметриках, и рейтингом журнала Scimago. Черные столбцы – среднее число просмотров одной статьи; серые столбцы – значение SJR. Значения SJR в иллюстративных целях были умножены на 100

В случае полного охвата российских журналов системой Plum Analytics, включая полные данные о просмотрах без временных лагун и, что особенно важно, при условии включения в список источников некоторых российских баз данных, таких как РИНЦ, в будущем могут быть использованы и другие альтметрики. В целом сделан вывод, что проанализированные альтметрики уже сейчас могут быть полезны при оценке качества и значимости журнала и могут использоваться в качестве предиктора будущего числа цитируемых статей в журнале, общего количества цитирований и других традиционных показателей, основанных на цитировании, при условии, что данные по альтметрикам полные.

6.2 Оценка публикационного вклада редколлегии в библиометрические показатели научного журнала

Кроме альтметрик, на наш взгляд, недооцененным в оценке качества периодических изданий остается еще один подход на основе разностороннего анализа редакционных коллегий. На предыдущем этапе исполнения проекта было проведено исследование состава редколлегии российских журналов на примере информационно-библиотечной отрасли. В продолжение начатой работы на текущем этапе был проведен анализ оценки участия членов редколлегии в публикационном потоке отечественных журналов.

Просматриваются два направления исследований, в первом из которых анализируется корреляция авторитетности редколлегии с рейтинговыми показателями журнала (опосредованная связь), а во втором – прямое влияние редколлегии на рейтинг издания (непосредственная связь). Если первое направление достаточно широко освещено в научной литературе, то второе представляется менее изученным и включает влияние на рейтинг журнала публикаций редколлегии в собственном журнале, а также влияние цитирований редколлегией своего журнала как в нем самом (самоцитирование), так и в других источниках (скрытое самоцитирование).

Непосредственный публикационный вклад редколлегии в собственный журнал впервые был изучен в конце XX в. J. M. Campanario [58] на примере изданий общественного цикла. Было выявлено, что статьи редколлегии имеют больший объем в страницах, что предположительно связано с менее строгими для редколлегии требованиями к объему рукописей. Было показано, что импакт-фактор журнала тем выше, чем весомее доля публикаций редколлегии в журнале. Следует отметить, что причинно-следственную связь между импакт-фактором и цитируемостью статей редколлегии еще предстоит выяснить. Так, на примере журналов по проблемам коммуникации показана аналогичная положительная корреляция между цитируемостью статей редколлегии и импакт-фактором, однако тот же факт объясняется уже тем, что статьи редколлегии цитируются с большей частотой, поскольку опубликованы в журналах с более высоким импакт-фактором, тогда как в журналах с низким импакт-фактором набирают меньше цитирований [59].

В исследовании публикаций членов редколлегии в хорватских журналах проблема рассматривалась с точки зрения возможных нарушений этических норм и конфликта интересов, даны рекомендации к усилению контроля над процессами рецензирования рукописей членов редколлегий, подаваемых в собственный журнал [60]. Не было обнаружено нарушений при опубликовании результатов членами редколлегии в своих журналах в области библиотековедения и информатики: авторы связывают это с невысокой долей

таких публикаций в сравнении с ожидаемой, что обусловлено, по их мнению, стремлением членов редколлегии избежать конфликта интересов [61]. На примере международных журналов в области коммуникации показана низкая доля публикуемости редколлегии в своем журнале: в среднем 1 статья за 10 лет, не считая редакционных материалов [59].

В проведенном нами исследовании российских журналов по наукам о Земле показано, что в отдельных журналах до трети всех публикаций написаны членами редколлегии или в соавторстве с ними, а в 12 из 19 изданий собственный журнал стоит на первом месте по общему числу публикаций членов редколлегии [62]. Схожие результаты получены и в работе [63], где на выборке журналов медицинской тематики отмечалось существенное смещение публикационной активности редколлегии в сторону собственных изданий, а также в статье [64], где на примере высокорейтинговых изданий по финансам показана взаимосвязь между числом публикаций авторов в этих журналах и вероятностью их избрания в члены редколлегии. В работе [65] на выборке журналов по экономике и бухгалтерскому делу демонстрируется, что до 35 % статей членов редколлегии публикуется на страницах собственного журнала. Отдельный анализ иногда проводится на выборке исключительно зарубежных членов редколлегии. Так, на примере хорватских журналов, разделенных на две группы – входящих в Scopus и не индексируемых в нем – в работе [66] исследовался вклад зарубежных участников редколлегии в публикационную активность изданий. В одном из журналов по урологии обнаружена связь между числом публикаций редколлегии в своем журнале и ростом импакт-фактора этого издания [67].

Кроме публикаций редколлегия вносит вклад в показатели журнала и цитированиями. Как и в случае исследования вклада редколлегии в рейтинг журнала по публикациям, первыми работами по изучению вклада по цитированиям стали работы J. M. Campanario с соавторами [68,69]. Был проанализирован вклад редколлегии в значение импакт-фактора с выявлением в нем доли цитирований, исходящих от редколлегии этого же журнала.

Важным представляется обоснованное авторами разделение самоцитирований из своего журнала и «скрытых» самоцитирований, т.е. цитирований редколлегией своего журнала из других источников. Исследование таких «скрытых» цитирований представляется нам до сих пор малоизученным и заслуживающим большего внимания.

Также, как и в исследованиях по публикуемости редколлегии в собственном журнале, при рассмотрении цитирований своего журнала возникают вопросы публикационной этики и конфликта интересов. В анализе цитирований журналов с этих позиций в ряде исследований не было выявлено случаев злоупотребления редколлегией своим служебным положением [69,70]. Высокие показатели самоцитирований, исходящих от редколлегии, также, как и публикации в своем журнале, могут указывать, например, на узкую дисциплинарную направленность, и тогда они оправданы. В то же время в отдельных случаях в структуре импакт-фактора выявляется присутствие недобросовестных ссылок [71].

По-видимому, в разграничении добросовестных и недобросовестных ссылок может помочь анализ системности цитирования редколлегией своего журнала и наличие определенных моделей цитирования, при отсутствии которых даже высокие показатели и публикуемости редколлегии в своем издании, и его цитирования должны считаться правомерными. Например, нарушения можно заподозрить в случае, когда всего трое членов редакционной коллегии ответственны за треть всех ссылок, учтенных в импакт-факторе, что показано на примере *International Journal of Nonlinear Sciences and Numerical Simulation* [71].

В эмпирическом исследовании нами проанализированы 22 ведущих отечественных журнала информационно-библиотечной области, распределенные на три группы в соответствии с величиной их 2-летнего импакт-фактора по всем источникам в РИНЦ:

- импакт-фактор 0,4 – 0,799: «Библиография. Научный журнал по библиографоведению, книговедению и библиотековедению»;

«Библиосфера»; «Библиотекведение»; «Научно-техническая информация. Серия 2: Информационные процессы и системы»; «Проблемы информационной безопасности. Компьютерные системы»; «Программная инженерия»; «Системы управления и информационные технологии»;

- импакт-фактор 0,8 – 0,999: «Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии»; «Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования»; «Вестник Новосибирского государственного университета. Серия: Информационные технологии»; «Информационное общество»; «Информационные ресурсы России»; «Информационные технологии и вычислительные системы»; «Научно-техническая информация. Серия 1: Организация и методика информационной работы»; «Научные и технические библиотеки»; «Системы и средства информатики»;
- импакт-фактор выше 1: «Бизнес-информатика»; «Вычислительные технологии»; «Информатика и ее применения»; «Информатика и образование»; «Онтология проектирования»; «Прикладная информатика»

Доля редколлегии, принимающая прямое участие в публикациях и цитировании журнала. В среднем отмечено активное участие редколлегии: по всем журналам 38,5 % участников редколлегии либо публикуются в своем издании, либо цитируют его. Нам не удалось обнаружить взаимосвязи между импакт-факторами журналов и долей участников редколлегии, влияющих на библиометрические показатели издания: во всех трех группах доля активных участников была приблизительно одинаковой (рисунок 46).

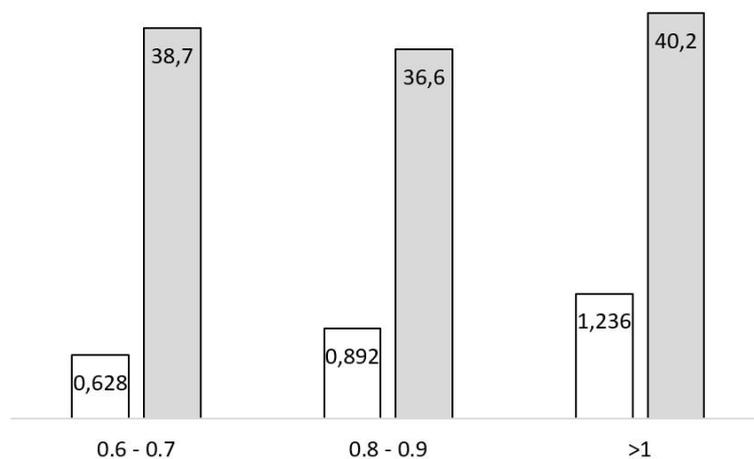


Рисунок 46 – Зависимость между импакт-фактором журнала и долей участников редколлегии, влияющих на его библиометрические показатели. Белые столбцы – усредненный импакт-фактор; серые столбцы – доля членов редколлегии (в процентах), публикующихся в журнале и (или) цитирующих его

Доля публикаций редколлегии в своем журнале и их влияние на рейтинговые показатели издания. Публикации членов редколлегии как ученых высокой квалификации преимущественно подготовлены более профессионально и цитируются более обильно, однако в отдельных случаях наблюдается обратная ситуация менее качественной подготовки рукописей. Это может быть связано с ослабленным рецензированием и использованием должностных полномочий, поскольку членам редколлегии иногда бывает значительно проще опубликовать статью в собственном журнале ввиду менее строгих критериев в сравнении с критериями, применяемыми к прочим авторам.

Мы не нашли существенной зависимости между значениями импакт-факторов и долей публикаций редколлегии в собственном журнале от общего числа публикаций в этом издании – во всех трех группах доля публикаций редколлегий в собственных журналах в среднем колебалась в пределах 16 % (рисунок 47Рисунок 47 –).

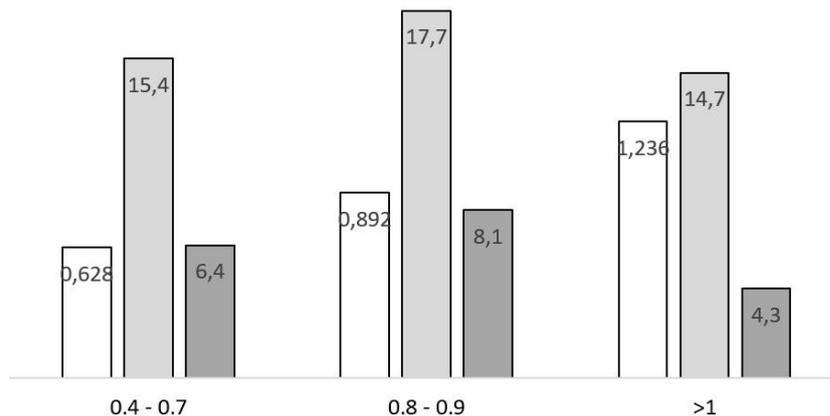


Рисунок 47 – Зависимость между импакт-фактором журнала и долей публикаций в нем редколлегии. Зависимость между импакт-фактором и долей публикаций редколлегии в соответствующем журнале от общего числа публикаций редколлегии. Белые столбцы – усредненный импакт-фактор; светло-серые столбцы – доля публикаций (в процентах) редколлегии в своем журнале; темно-серые столбцы – доля публикаций (в процентах) членов редколлегии в своем журнале от общего числа их публикаций

Доля цитирований публикаций редколлегии в общем числе цитирований журнала. Доля «скрытых» самоцитирований журнала из публикаций редколлегии в других источниках. Вклад редколлегии по цитированиям в рейтинговые показатели журнала складывается из двух компонентов: во-первых, могут цитироваться статьи, опубликованные редколлекцией в журнале; во-вторых, сама редколлегия может цитировать свой журнал в публикациях из других источников («скрытые» самоцитирования). На рисунке 48 показаны долевые значения вклада редколлегии по цитированиям, сделанным на опубликованные ими в журнале статьи; по «скрытым» самоцитированиям, а также совокупное значение по обоим показателям.

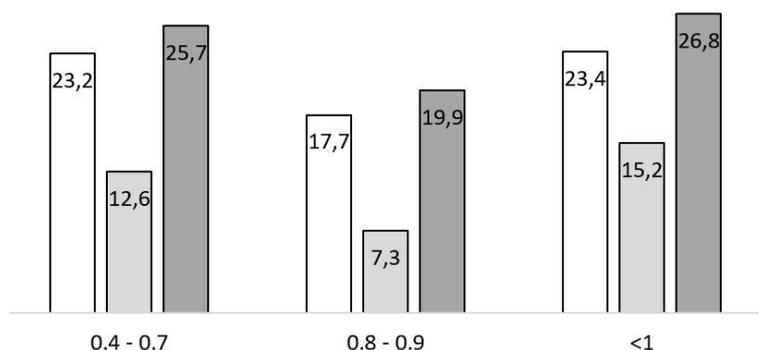


Рисунок 48 – Доля цитирований на публикации редколлегии в собственном журнале и доля цитирований, исходящих от членов редколлегии в других источниках, в общей массе цитирований журналов. Белые столбцы – доля цитирований (в процентах) статей, опубликованных редколлегией в своем журнале; светло-серые столбцы – доля (в процентах) «скрытых» самоцитирований; темно-серые столбцы – общая доля (в процентах) цитирований статей редколлегии и «скрытых» самоцитирований

По всем российским журналам информационно-библиотечной области на цитирование публикаций редколлегии приходится 21,1 % ссылок от общего числа ссылок на журнал, а на «скрытые» самоцитирования – 11,5 %. Сводное значение этих двух источников поступления ссылок на журнал составляет почти четверть – 23,9 %, что является весьма существенным вкладом редколлегии в рейтинг своего издания.

Зависимость импакт-фактора как основной журнальной метрики от публикационной активности членов его редколлегии. Поскольку основным показателем авторитетности журналов остается импакт-фактор, в завершение исследования мы проанализировали суммарное влияние редколлегии на этот показатель, с учетом проведенных выше измерений вклада редколлегии по публикациям и цитированиям. Для этого последовательно 1) из числителя импакт-фактора каждого из журналов были удалены публикации членов редколлегии; 2) соответственно из знаменателя исчезли ссылки на публикации редколлегии; 3) из знаменателя были также удалены ссылки, сделанные редколлегией на свой журнал из других источников. Таким образом, было полностью исключено прямое влияние редколлегии на публикационную активность журнала и соответствующие показатели (рисунок 49, таблица 14).

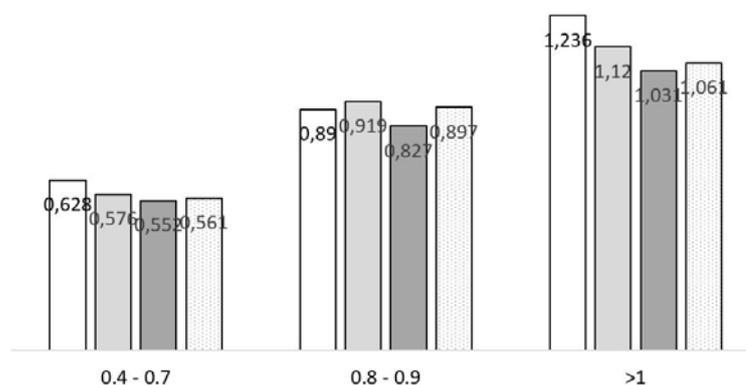


Рисунок 49 – Расчет значений импакт-факторов с учетом публикационного вклада редколлегии и без него. Белые столбцы – усредненный импакт-фактор; светло-серые столбцы – импакт-фактор без статей редколлегии; темно-серые столбцы – импакт-фактор без «скрытого» самоцитирования; заштрихованные столбцы – импакт-фактор без любого публикационного вклада редколлегии

Таблица 14 – Доля вклада редакционных коллегий по публикуемости и цитируемости в импакт-факторы журналов соответствующих групп.

Диапазон импакт-факторов по группам журналов	>1	0,8 – 0,9	0,4 – 0,7
Общая доля вклада редколлегии в ИФ журнала	14,14 %	2,82 %	10,56 %

По отдельным изданиям публикационное участие членов редакционных коллегий оказывает отрицательное влияние на показатели журналов: так, импакт-фактор журнала «Системы и средства информатики» без полного публикационного участия редколлегии был бы на 16,4 % выше текущего значения, а без публикаций редколлегии в собственном журнале – на 18,9 % выше. Напротив, в отдельных журналах наиболее и наименее авторитетных групп прослеживается существенное положительное влияние публикационной активности редколлегий на рейтинговые позиции издания. Например, общая доля вклада редколлегии в импакт-фактор журнала «Онтология проектирования» (журнал группы с высоким импакт-фактором) составляет 30,7 %, а журнала «Библиография» (журнал группы с невысоким импакт-фактором) – 34,1 %.

ЗАКЛЮЧЕНИЕ

В заключении выделим наиболее значимые научные результаты проекта в 2020 г.

Проведен анализ содержания и качества баз данных Scopus и Web of Science Core Collection, на основании которых проводятся наукометрические исследования. Показано, что по общему количеству индексируемых документов и по типам источников Scopus и Web of Science Core Collection (WoS CC) отличаются незначительно, хотя различия в составе источников остаются существенными. Выдвинута гипотеза, что в WoS CC в меньшей степени представлены развивающиеся страны и наблюдается разница в дисциплинарной структуре этих БД с большей ориентированностью WoS CC на издания стран первого мира, которая приводит к меньшей «инклюзивности» этой базы данных. Показано, что, хотя полный анализ публикационного потока невозможен без данных из WoS CC, российские публикации из этой БД, не проиндексированные в Scopus, в основном относятся к RSCI и ESCI, а в журналы верхних квартилей по JCR попадает очень немного отсутствующих в Scopus публикаций. В связи с этим, полученные из Scopus данные позволяют провести достаточно полный анализ публикационного потока российских исследователей и его сравнение с результатами ученых из других стран.

Изучена проблема нейтральных публикаций, которые не аффилированы ни с одной страной. Предложен метод страновой классификации нейтральных публикаций. С его помощью в период 2000-2019 было выявлено более 120 тысяч нейтральных публикаций, которые с большой вероятностью должны быть аффилированы с российскими авторами, причём половина из них относится к периоду 2000-2005. Наибольшая доля неучтенных публикаций в области наук относится к категории “Health Sciences”, где в 2006-2010 гг. количество неучтенных публикаций превосходило 50% от общего числа публикаций. Проведенный анализ показывает, что учет неаффилированных с

РФ публикации российских авторов в отдельных областях заметно меняет характер изменения публикационного потока со временем.

Завершено изучение метода тематической классификации публикаций на основе теоретико-информационного подхода. На примере БД Scopus выявлены ограничения метода, связанные с формированием обучающей выборки: низкая достоверность существующей классификации, отсутствие в некоторых научных дисциплинах монодисциплинарных журналов, которые бы специализировались исключительно в них, а также терминологически близкие дисциплины, отнесенные к разным областям наук. Сравнение с пятью широко применяемыми методами классификации текстовой информации показало его высокую точность на уровне 89-91%, что сопоставимо или превосходит точность других методов.

Предложен новый метод уточнения тематической классификации публикаций на основе анализа публикационной истории авторов. Основа метода – это формирование для каждой публикации вектора тематических рубрик, составленного из тематических рубрик всех публикаций каждого из авторов анализируемой публикации с расчетным количеством таких публикаций отсортированный по убыванию. Полученные результаты новой тематической рубрикации заметно отличаются от результатов, полученных с использованием аналитических инструментов Scopus: значительно большей оказывается доля публикаций в областях «Физика и Астрономия», «Медицина», «Науки о Земле»; заметно меняется порядок приоритетов областей исследований.

Проведен структурно-динамический анализ российских публикаций в ведущих мировых изданиях за 2000-2019 гг. по данным БД Scopus. На основе рейтингов источников Scimago Journal Ranking продемонстрирован активный количественный рост во всех квартильных группах, кроме журналов Q4. В процентном отношении быстрее всего растут публикации в источниках третьего квартиля (Q3) и в не входящих в рейтинг изданиях. В последние годы наблюдается ускорение роста в Q1, последовавшее после отставания в 2014-

2015 гг., связанного с бурным ростом общего количества публикаций. Тем не менее, доля публикаций в Q1+Q2 составляет в 2019г. всего чуть менее 36,7%, что значительно меньше среднемирового уровня. При этом заметная часть российского роста публикаций обеспечена материалами конференций, доля которых в общем потоке российских публикаций превысила в 2019 г. отметку в 30% при среднемировом значении 16%.

Изучен феномен инфляции научной результативности, который состоит в том, что «стоимость» одной статьи в списке публикаций ученого сейчас заметно ниже, чем 10-20 лет назад. Показано, что разрыв между номинальным и отчетным количеством публикаций постоянно увеличивался весь исследуемый период времени, этот рост был особенно активным в 2012-2016 гг. С 2000 года коэффициент инфляции для российских организаций вырос с 25% до почти 70% в 2019 г.

Одним из очевидных способов более адекватной оценки результативности является использование фракционного счета, при котором каждому автору за публикацию начисляется его авторская доля, а не единица. Показаны различия результатов наукометрического анализа: в период 2000-2019 средняя персональная результативность исследователя, рассчитанная целочисленным счетом, росла, тогда как фракционный расчет показал её падение. Показано, что наибольшие различия между целочисленным и фракционным счетом характерны для университетов из Проекта 5-100 – организаций, которые подвергались наибольшему «наукометрическому давлению», стимулирующему публикационную активность. Показано увеличение среднего количества российских аффилиаций на одну статью, что может свидетельствовать как об объективных процессах, связанных с развитием коллабораций и мультидисциплинарных исследований, так и о расширении практик недобросовестного соавторства.

Расчет доли российских публикаций в мировом потоке целочисленным и фракционным счетом показали незначительные отличия, что говорит о

достаточно низком уровне международных коллабораций российских исследователей.

Исследован Комплексный балл научной результативности (КБПР), на котором основана новая методика оценки результативности, предложенная Минобрнауки РФ. Показано, что больше всего от перехода на фракционный счёт страдают участники крупных международных коллабораций, в частности – специализирующиеся в области физики высоких энергий. В зону проигрыша попадает ряд вузов, сотрудники которых публикуются в основном в низкорейтинговых журналах и материалах конференций. Пострадают и программы привлечения ведущих зарубежных ученых, а также организации, активно привлекающие внешних совместителей. Среди выигравших много организаций, специализирующихся в области химии, материаловедения, биологии и генетики. Негативной составляющей КБПР признана плоская шкала баллов для гуманитарных и общественных наук, согласно которой недобросовестные ученые будут иметь преимущество перед авторами, публикующимися в престижных журналах.

Проанализированы изменения в региональной структуре российской науки. Если в начале 2000-ых гг. на Москву приходилось 45,6% научного продукта международного уровня, то к концу рассматриваемого периода доля Москвы упала до 37,2%. Заметно упал также вклад Московской области – с 4,8% до 2,2%. Менее значительное падение доли пришлось и на такие развитые в научном плане регионы, как Новосибирская область (с 7,8% до 5,5%), Санкт-Петербург (с 13,2% до 12,3%), Нижегородская область (с 2,4% до 1,8%). Активно развивался научный потенциал в Республике Татарстан (рост с 2% до 3,7%), Томской области (с 2,8% до 3,7%), Ростовской области (с 1,5% до 2,2%) и многих других регионах. Значительный подъем достигнут в Тюменской, Волгоградской, Самарской, Челябинской областях, Пермском крае.

Проанализирована секторальная структура публикационного потока российских исследователей и её изменения с 2000 по 2019 гг. Зафиксировано

заметное сжатие доли публикаций академических институтов (с 39% до 24%), при увеличении публикаций вузов (с 29% до 51%). Наиболее активный рост доли университетской науки пришелся на 2013-2016 гг. (в 2014 г. – рост на 35%, в 2016 – на 28%), а в последние годы динамика снизилась до 10-15% в год. Необходимо отметить, что в традиционно сильных для РФ научных направлениях по-прежнему сильны академические институты: в физике и астрономии, химии, науках о Земле, науках о материалах, биохимии, генетике и молекулярной биологии, биологии и сельском хозяйстве, иммунологии и микробиологии, нейронауках и некоторых других. Вузы вышли на ведущие позиции в таких областях, как: инженерия, технические науки, математика, химические технологии, энергетика, компьютерные науки и других.

Среди ведущих в научном плане российских регионов остаются преимущественно «академические» – с преобладающим влиянием академических институтов. Такими регионами остаются Московская область, с научными центрами в Черноголовке, Фрязино, Шатуре, Пушкино, Одинцово, Протвино; Новосибирская область с Новосибирским Академгородком, Приморский край с институтами ДВО РАН в Владивостоке, Иркутская область с Иркутским научным центром.

Подводя итоги проведенного анализа, отметим, что основные изменения в структуре российской науки связаны с ускоренным ростом сектора университетской науки, и некоторых региональных научных и образовательных центров, что позволяет сделать вывод о решающем влиянии эффекта «низкой базы». Рост шел преимущественно там, где имелась значительная кадровая база, прежде всего в вузах, но свою роль сыграли и относительно слабые академические институты, государственные научные центры, отраслевая и корпоративная наука, заметно нарастившие публикационную активность на международном уровне. Ведущие научные организации и до этого периода активно работали и публиковались в индексируемых в Scopus изданиях, поэтому смогли нарастить результативность незначительно.

Проведен структурно-динамический анализ публикационной активности стран – конкурентов России в мировом научном пространстве на примере Франции, Бразилии и Индии в разрезах отраслей и секторов науки за 2000-2019 гг. Показано, что с точки зрения секторальной структуры российская наука как в 2000 г., так и в 2019 г. наиболее близка структуре науки Франции, где на вузы приходится около половины научных публикаций, а на исследовательские организации – примерно четверть. Во всех рассмотренных странах вклад вузов систематически рос, причем в России темпы этого роста были наиболее впечатляющими (с 29% до 51%), а в Бразилии вузовские публикации доминируют сильнее всех (с 78% до 83%). В Индии и Бразилии доля корпоративной науки меньше, чем во Франции, а медицинские исследования в большей степени проводятся в университетских клиниках.

Анализ типов публикаций, аффилированных с Россией, Францией, Бразилией и Индией показал различные тренды: доля публикаций в трудах конференций во Франции и Бразилии снижалась с 2005 г., в Индии и России – наоборот росла. Наибольшая доля публикаций в журналах из Q1+Q2 принадлежит Франции. В Индии и России в 2005-2011 наблюдалось снижение этой доли до 22-24%, за которой последовала стагнация. В 2018-2019 гг. у России наметился небольшой рост доли публикаций в журналах из высоких квартилей.

Анализ публикационного потока по направлениям наук показал, что в российской науке доминирует направление Physical Sciences, но на протяжении рассматриваемого периода это доминирование ослабевает, прежде всего, за счет ускоренного развития общественных и гуманитарных наук. Индия движется в обратном направлении – доля Physical Sciences в индийской науке растет за счет относительно медленного развития направления Health Sciences и Life Sciences. То же явление, но в гораздо меньшей степени наблюдается и в бразильской науке. Рост доли общественных и гуманитарных наук (Social Sciences) наблюдается во всех рассматриваемых странах, но во Франции это происходит за счет уменьшения

доли естественных наук. Детально показаны изменения по научным дисциплинам в сравнении 2000 и 2019 гг.

На уровне направлений исследований структура российской науки постепенно трансформируется в том же направлении, что и в Бразилии и Франции, к 2019 г. за счет встречного движения приблизившись к структуре индийской науки. В связи с этим можно констатировать, что заметный потенциал развития есть в таких областях, как Computer Science, Agricultural and Biological Sciences и Medicine. В целом, наблюдается постепенное приближение структуры части российской науки, которая видима на международном уровне, к ведущим странам. Это может происходить за счет расширения индексации в Scopus российских журналов, трудов российских конференций и повышения публикационной активности российских исследователей в зарубежной печати. Тем не менее, пока основным научным направлением остаются естественные науки (Physical Sciences), обеспечивающие почти 76% всех российских публикаций, индексируемых в Scopus.

Изучение российских журналов по наукам о Земле позволило установить, что некоторые типы альтметрик, включая просмотры аннотаций и просмотры полных текстов, могут быть использованы в качестве дополнительных показателей для оценки качества журнала в случае отсутствия или недоступности традиционных библиометрических индикаторов. Кроме того, их можно применять как предиктор будущих цитирований и других показателей, основанных на цитировании.

Проведенный библиометрический анализ публикаций членов редакционных коллегий в российских журналах информационно-библиотечной области позволил количественно определить степень их влияния на рейтинговые позиции журналов, для чего были предложены дополнительные индикаторы. Показано активное участие редколлегии: по всем журналам 38,5 % участников редколлегии либо публикуются в своем издании, либо цитируют его. Публикации членов редколлегии как ученых

высокой квалификации преимущественно цитируются более обильно, однако в отдельных случаях наблюдается и обратная ситуация менее качественной подготовки рукописей. Продемонстрированы случаи, когда участие членов редакционных коллегий оказывает как положительное, так и отрицательное влияние на показатели журналов.

Полученные на текущем этапе проекта результаты могут быть полезны при пересмотре редакционной политики и формировании состава редколлегии. Например, выявленное нами отсутствие цитирований редколлегией своего журнала с большой вероятностью указывает на незнакомство участников с его контентом, что может быть исправлено путем более интенсивного привлечения редколлегии к обсуждению поступающих рукописей и замене формально привлеченных представителей более заинтересованными в работе журнала авторами. Дальнейшие библиометрические исследования в заданном направлении, с охватом более широкого спектра дисциплин могут способствовать более успешной редакционной политике отечественных журналов, направленной на повышение их рейтинговых позиций на международном издательском рынке и продвижение в международные индексы цитирований.

По результатам исследования в 2020 г. было опубликовано 9 статей в рецензируемых научных журналах, в том числе 3 – в журналах, индексируемых в Web of Science и Scopus, представлено 8 докладов на международных и российских конференциях. Подготовленная в рамках проекта диссертация Селивановой И.В. на соискание ученой степени кандидата технических наук «Методы тематической классификации научных текстов на основе теоретико-информационного подхода» принята к защите в диссертационном совете Д 219.005.02 в Федеральном государственном бюджетном образовательном учреждении высшего образования «Сибирский государственный университет телекоммуникаций и информатики» (СибГУТИ).

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Hinze S. Analysis of country specialisation in bioelectronics with special focus on German activities // *Research Evaluation*. 1994. Vol. 4, № 2. P. 107–118.
2. Jacso P. Errors of omission and their implications for computing scientometric measures in evaluating the publishing productivity and impact of countries // *Online Information Review*. 2009. Vol. 33, № 2. P. 376–385.
3. Prathap G. Visibility score for countries using SCOPUS Affiliation Identifier and the h-threshold approach // *Current Science*. 2009. Vol. 96, № 9. P. 1184–1185.
4. Krauskopf E. Call for caution in the use of bibliometric data // *Journal of the Association for Information Science and Technology*. 2017. Vol. 68, № 8. P. 2029–2032.
5. Liu W., Hu G., Tang L. Missing author address information in Web of Science - An explorative study // *Journal of Informetrics*. 2018. Vol. 12, № 3. P. 985–997.
6. Nguyen B. X., Dinneen J. D., Luczak-Roesch M. A Novel Method for Resolving and Completing Authors' Country Affiliation Data in Bibliographic Records // *Journal of Data and Information Science*. 2020. Vol. 0, № 0.
7. Fernández J. C. et al. Multi-objective evolutionary optimization using the relationship between F1 and accuracy metrics in classification tasks // *Appl Intell*. 2019. Vol. 49, № 9. P. 3447–3463.
8. He H., Garcia E. A. Learning from Imbalanced Data // *IEEE Transactions on Knowledge and Data Engineering*. 2009. Vol. 21, № 9. P. 1263–1284.
9. Wang Q., Waltman L. Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus // *Journal of Informetrics*. 2016. Vol. 10, № 2. P. 347–364.
10. Klavans R., Boyack K. W. Research portfolio analysis and topic prominence // *Journal of Informetrics*. 2017. Vol. 11, № 4. P. 1158–1174.

11. Waltman L., Eck N. J. van. A new methodology for constructing a publication-level classification system of science // *Journal of the American Society for Information Science and Technology*. 2012. Vol. 63, № 12. P. 2378–2392.
12. Baas J. et al. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies // *Quantitative Science Studies*. MIT Press, 2020. Vol. 1, № 1. P. 377–386.
13. Selivanova I. V., Kosyakov D. V., Guskov A. E. The Impact of Errors in the Scopus Database on the Research Assessment // *Scientific and Technical Information Processing*. 2019. Vol. 46, № 3. P. 204–212.
14. Гуськов А. Е., Косяков Д. В. Национальный фракционный счет и оценка научной результативности организаций // *Научные и технические библиотеки*. 2020. Vol. 1, № 9. P. 15–42.
15. Гуськов А. Е. et al. Факторы цитируемости обзоров // *Вестник Российской Академии Наук*. 2020. Vol. 90, № 12.
16. Moed H. F., Halevi G. Multidimensional assessment of scholarly research impact // *Journal of the Association for Information Science and Technology*. John Wiley and Sons Inc., 2015. Vol. 66, № 10. P. 1988–2002.
17. Abbott A. et al. Metrics: Do metrics matter? // *Nature*. 2010. Vol. 465, № 7300. P. 860–862.
18. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US states data // *PLoS ONE*. 2010. Vol. 5, № 4.
19. Halperin E. C. Publish or perish - And bankrupt the medical library while we're at it // *Academic Medicine*. Association of American Medical Colleges, 1999. Vol. 74, № 5. P. 470–472.
20. Binswanger M. How nonsense became excellence: Forcing professors to publish // *Incentives and Performance: Governance of Research Organizations*. Springer International Publishing, 2015. 19–32 p.
21. Agarwal A. et al. Bibliometrics: Tracking research impact by selecting the appropriate metrics // *Asian Journal of Andrology*. Medknow Publications, 2016. Vol. 18, № 2. P. 296–309.

22. Wuchty S., Jones B. F., Uzzi B. The increasing dominance of teams in production of knowledge // *Science*. 2007.
23. Abt H. A. The frequencies of multinational papers in various sciences // *Scientometrics*. 2007.
24. Abramo G., D'Angelo A. C., Murgia G. The relationship among research productivity, research collaboration, and their determinants // *Journal of Informetrics*. 2017.
25. Ioannidis J. P. A., Klavans R., Boyack K.W. Thousands of scientists publish a paper every five days // *Nature*. 2018. Vol. 561, № 7722. P. 167–169.
26. Larivière V., Costas R. How many is too many? On the relationship between research productivity and impact // *PLoS ONE*. Public Library of Science, 2016. Vol. 11, № 9.
27. Wager E., Singhvi S., Kleinert S. Too much of a good thing? An observational study of prolific authors // *PeerJ*. PeerJ Inc., 2015. Vol. 2015, № 8.
28. Manganote E.J.T., Schulz P.A., de Brito Cruz C.H. Effect of high energy physics large collaborations on higher education institutions citations and rankings // *Scientometrics*. Springer Netherlands, 2016. Vol. 109, № 2. P. 813–826.
29. Tarkhan-Mouravi S. Traditional indicators inflate some countries' scientific impact over 10 times // *Scientometrics*. Springer Netherlands, 2020. Vol. 123, № 1. P. 337–356.
30. Huang M. H., Lin C. S., Chen D. Z. Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact // *Journal of the American Society for Information Science and Technology*. 2011. Vol. 62, № 12. P. 2427–2436.
31. Lin C. S., Huang M. H., Chen D. Z. The influences of counting methods on university rankings based on paper count and citation count // *Journal of Informetrics*. Elsevier Ltd, 2013. Vol. 7, № 3. P. 611–621.
32. Leydesdorff L., Park H. W. Full and fractional counting in bibliometric networks // *Journal of Informetrics*. 2017. Vol. 11, № 1. P. 117–120.

33. Abramo G., D'Angelo C. A., Rosati F. Measuring institutional research productivity for the life sciences: The importance of accounting for the order of authors in the byline // *Scientometrics*. 2013. Vol. 97, № 3. P. 779–795.
34. Bornmann L. Which research institution performs better than average in a subject category or better than selected other institutions? // *Online Information Review*. 2018. Vol. 42, № 2. P. 222–237.
35. Moed H.F. Citation analysis in research evaluation // *Proceedings of ISSI 2005: 10th International Conference of the International Society for Scientometrics and Informetrics*. 2005. Vol. 2. P. 437–441.
36. Sivertsen G., Rousseau R., Zhang L. Measuring scientific contributions with modified fractional counting // *Journal of Informetrics*. Elsevier Ltd, 2019. Vol. 13, № 2. P. 679–694.
37. Abramo G., D'Angelo C.A., Solazzi M. Are researchers that collaborate more at the international level top performers? An investigation on the Italian university system // *Journal of Informetrics*. 2011.
38. Ascheulova N.A., Kolchinsky E.I. The reform of science in Russia (a historico-sociological analysis) // *Voprosy Istorii Estestvoznaniia I Tekhniki [Studies in the History of Science and Technology]*. 2010. Vol. 31, № 1. P. 95–119.
39. Kosyakov D., Guskov A. Research assessment and evaluation in Russian fundamental science // *Procedia Computer Science*. 2019. Vol. 146. P. 11–19.
40. Guskov A., Kosyakov D., Selivanova I. Scientometric research in Russia: impact of science policy changes // *Scientometrics*. 2016. Vol. 107, № 1.
41. Moed H.F., Markusova V., Akoev M. Trends in Russian research output indexed in Scopus and Web of Science // *Scientometrics*. Springer Netherlands, 2018. Vol. 116, № 2. P. 1153–1180.
42. Kosyakov D., Guskov A. Impact of national science policy on academic migration and research productivity in Russia // *Procedia Computer Science*. 2019. Vol. 146. P. 60–71.

43. Guskov A. E., Kosyakov D. V., Selivanova I. V. Boosting research productivity in top Russian universities: the circumstances of breakthrough // *Scientometrics*. Springer Netherlands, 2018. Vol. 117, № 2. P. 1053–1080.
44. Kosyakov D., Guskov A. Synchronous scientific mobility and international collaboration: Case of Russia // *17th International Conference on Scientometrics and Informetrics, ISSI 2019 - Proceedings*. International Society for Scientometrics and Informetrics, 2019. Vol. 1. P. 1319–1328.
45. Bhattacharjee Y. Saudi universities offer cash in exchange for academic prestige // *Science*. 2011. Vol. 334, № 6061. P. 1344–1345.
46. De Solla Price D. J., Beaver D. Collaboration in an invisible college // *American Psychologist*. US: American Psychological Association, 1966. Vol. 21, № 11. P. 1011–1018.
47. Egghe L., Rousseau R., Hooydonk G. V. Methods for accrediting publications to authors or countries: Consequences for evaluation studies // *Journal of the American Society for Information Science*. 2000. Vol. 51, № 2. P. 145–157.
48. Huang M.-H., Lin C.-S., Chen D.-Z. Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact // *Journal of the American Society for Information Science and Technology*. 2011. Vol. 62, № 12. P. 2427–2436.
49. Moed H. F. *Citation Analysis in Research Evaluation*. Springer Netherlands, 2005.
50. Waltman L., van Eck N.J. Field-normalized citation impact indicators and the choice of an appropriate counting method // *Journal of Informetrics*. 2015. Vol. 9, № 4. P. 872–894.
51. The Leiden ranking 2011/2012: Data collection, indicators, and interpretation - Waltman - 2012 - *Journal of the American Society for Information Science and Technology* - Wiley Online Library [Electronic resource]. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.22708> (accessed: 14.01.2021).

52. Парфенова С.Л. et al. Публикационный ландшафт российской науки: 1 (23) // Управление наукой и наукометрия. Россия, Москва: Федеральное государственное бюджетное учреждение «Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере», 2017. № 1 (23).
53. Гуськов А.Е., Косяков Д.В., Селиванова И.В. Методика оценки результативности научных организаций // Вестник Российской Академии Наук. Российская академия наук, 2018. Vol. 88, № 5.
54. Manganote E.J.T., Schulz P.A., de Brito Cruz C.H. Effect of high energy physics large collaborations on higher education institutions citations and rankings // Scientometrics. 2016. Vol. 109, № 2. P. 813–826.
55. Kosyakov D., Guskov A. Synchronous scientific mobility and international collaboration: Case of Russia // 17th International Conference on Scientometrics and Informetrics, ISSI 2019 - Proceedings. 2019. Vol. 1. P. 1319–1328.
56. Macháček V., Srholec M. Predatory publications in Scopus: Evidence on cross-country differences. 2019. Vol. 1. P. 351–362.
57. Block M., Khvatova T. University transformation: Explaining policy-making and trends in higher education in Russia // Journal of Management Development. 2017. Vol. 36, № 6. P. 761–779.
58. Campanario J.M. The competition for journal space among referees, editors, and other authors and its influence on journals' impact factors // Journal of the American Society for Information Science. 1996. Vol. 47, № 3. P. 184–192.
59. Youk S., Park H.S. Where and what do they publish? Editors' and editorial board members' affiliated institutions and the citation counts of their endogenous publications in the field of communication // Scientometrics. 2019. Vol. 120, № 3. P. 1237–1260.
60. Bošnjak L. et al. Analysis of a number and type of publications that editors publish in their own journals: case study of scholarly journals in Croatia // Scientometrics. 2011. Vol. 86, № 1. P. 227–233.

61. Walters W. H. Do Editorial Board Members in Library and Information Science Publish Disproportionately in the Journals for Which They Serve as Board Members? // *Journal of Scholarly Publishing*. University of Toronto Press, 2015. Vol. 46, № 4. P. 343–354.
62. Mazov N. A., Gureev V. N., Metelkin D. V. Bibliometric Indicators of Scientific Journals and Editorial Board Members (Based on the Example of Russian Journals on Earth Sciences) // *Sci. Tech. Inf. Proc.* 2018. Vol. 45, № 4. P. 271–281.
63. Luty J. et al. Preferential publication of editorial board members in medical specialty journals // *Journal of Medical Ethics*. Institute of Medical Ethics, 2009. Vol. 35, № 3. P. 200–202.
64. Hardin W. G. et al. Finance editorial board membership and research productivity // *Rev Quant Finan Acc.* 2008. Vol. 31, № 3. P. 225–240.
65. Lowe D. J., Van Fleet D.D. Scholarly achievement and accounting journal editorial board membership // *Journal of Accounting Education*. 2009. Vol. 27, № 4. P. 197–209.
66. Jokić M., Sirotić G. Do the International Editorial Board Members of Croatian Social Sciences and Humanities Journals Contribute to their Visibility? // *Medijska istraživanja : znanstveno-stručni časopis za novinarstvo i medije*. NAKLADA MEDIJSKA ISTRAŽIVANJA, 2015. Vol. 21, № 2. P. 5–33.
67. Mani J. et al. I Publish in I Edit? - Do Editorial Board Members of Urologic Journals Preferentially Publish Their Own Scientific Work? // *PLOS ONE*. Public Library of Science, 2013. Vol. 8, № 12. P. e83709.
68. Coslado M.A., Campanario J.M. Contribution to the impact factor of academic journals in the field of education and educational psychology of citations to articles authored by editorial board members. 2007. P. 840–841.
69. Campanario J.M., González L., Rodríguez C. Structure of the impact factor of academic journals in the field of Education and Educational Psychology: Citations from editorial board members // *Scientometrics*. 2006. Vol. 69, № 1. P. 37–56.

70. González L., Campanario J.M. Structure of the impact factor of journals included in the Social Sciences Citation Index: Citations from documents labeled “editorial material” // Journal of the American Society for Information Science and Technology. 2007. Vol. 58, № 2. P. 252–262.
71. Arnold D.N., Fowler K.K. Nefarious Numbers // Notices of the American Mathematical Society. American Mathematical Society, 2011. Vol. 58. P. 434–437.

ПРИЛОЖЕНИЕ. СПИСОК ПУБЛИКАЦИЙ И ДОКЛАДОВ ПО ПРОЕКТУ

Статьи, проиндексированные в Web of Science или Scopus

1. Mazov N. A., Gureyev V. N. Nontraditional approaches to assessing journal importance: case study of Russian journals on Earth sciences // *Serials Review*. – 2020. – V. 46. – No 1. – P. 10–20. URL: <http://dx.doi.org/10.1080/00987913.2020.1730145>.
2. Мазов Н. А., Гуреев В. Н. Публикационный вклад редколлегии в библиометрические показатели научного журнала (информационно-библиотечная область) // *Научные и технические библиотеки*. – 2020. – № 11. – С. 33–58. URL: <http://dx.doi.org/10.33186/1027-3689-2020-11-33-58>.
3. Гуськов, А. Е. Национальный фракционный счет и оценка научной результативности организаций / А. Е. Гуськов, Д. В. Косяков // *Научные и технические библиотеки*. – 2020. – № 9. – С. 15-42. DOI <https://doi.org/10.33186/1027-3689-2020-9-15-42>.

Статьи в рецензируемых журналах

1. Багирова, А. В. Потенциал междисциплинарных исследований: российские биомедицинские публикации с участием сотрудников библиотек [Текст] / А. В. Багирова, Л. Л. Садовская, П. А. Чеснялис // *Труды ГПНТБ СО РАН*. – 2020. – № 2 (6). – С. 75-82. – DOI <https://doi.org/10.20913/2618-7515-2020-2-75-82>.
2. Гуреев, В. Н. К вопросу о компетентном понимании феномена цитируемости [Текст] : рецензия на книги В. С. Лазарева / В. Н. Гуреев, Н. А. Мазов // *Управление наукой: теория и практика*. – 2020. – Т. 2, № 4. – С. 232-238. – DOI <https://doi.org/10.19181/sntp.2020.2.4.11>.
3. Гуреев В. Н., Мазов Н. А., Ильичева И. Ю. Участие редакционной коллегии в публикационном потоке научного журнала // *Труды ГПНТБ СО РАН*. –

2020. – № 4. – С. 113–119. URL: <http://dx.doi.org/10.20913/2618-7575-2020-4-113-119>.
4. Косяков, Д. В. Инфраструктура для широкомасштабного сбора вебметрических показателей [Текст] / Д. В. Косяков // Труды ГПНТБ СО РАН. – 2020. – № 1 (5). – С. 86-97. – DOI 10.20913/2618-7515-2020-1-86-97.
 5. Косяков, Д. В. Наукометрия журнала «Библиосфера»: анализ тенденций и перспектив развития [Текст] / Д. В. Косяков, И. В. Селиванова, О. Л. Лаврик // Библиосфера. – 2020. – № 3. – С. 3-13. – DOI <https://doi.org/10.20913/1815-3186-2020-3-3-13>.
 6. Селиванова, И. В. Ограничения применения метода на основе сжатия данных к классификации аннотаций публикаций, индексируемых в Scopus // Вестник НГУ. Серия Информационные технологии. – 2020. – Т. 18, № 3. – С. 57-68. – DOI 10.25205/1818-7900-2020-18-3-57-68.

Доклады на научных конференциях

1. Гуреев В. Н., Мазов Н. А., Ильичева И. Ю. Участие редакционной коллегии в публикационном потоке научного журнала // Международная научно-практическая конференция «Наука, технологии и информация в библиотеках» (LIBWAY-2020) (14–17 сентября 2020 г., Новосибирск). – Новосибирск: ГПНТБ СО РАН, 2020. – URL: <https://www.libway.ru/2020/conference/program/5f085ca2d2cf29f414fe61a1>.
2. Гуреев, В. Н. Модели авторства членов редакционных коллегий: этический аспект [Текст] / В. Н. Гуреев, Н. А. Мазов // Информационные технологии, компьютерные системы и издательская продукция для библиотек : сборник докладов двадцать четвертой международной конференции «LIBCOM-2020» (24-27 ноября 2020 г., Москва). – Москва, 2020. – С. 41-46. – DOI 10.33186/978-5-85638-231-9-2020-41-46.
3. Гуськов А. Е. Обзоры научной литературы: динамика публикаций и анализ факторов цитируемости / А. Е. Гуськов, Д. В. Косяков, А. В. Багирова. – Текст : непосредственный // Наука, технологии и информация в

- библиотеках (LIBWAY-2020) : Международная научно-практическая онлайн-конференция (Новосибирск, 14-17 сентября 2020 г.). – Новосибирск, 2020.
4. Косяков Д. В. Комплексный балл публикационной результативности и некоторые возможные последствия его использования // Специальное онлайн-мероприятие «День НЭИКОН в дистанте» 07-08 июля 2020 г.
 5. Косяков Д. В. Эволюция инфраструктуры научных коммуникаций в условиях цифровой трансформации и открытый доступ к информации // Семинар «Открытая наука России» в рамках конференции LIBWAY-2020 17 сентября 2020 г.
 6. Косяков Д. В. Метаданные научных публикаций — значимый элемент инфраструктуры открытой науки // Онлайн-семинар «Открытая наука России 2020» 22 октября 2020 г.
 7. Мазов, Н. А. Роль и возможности библиотеки в изучении приоритетных научных направлений организации // Информатизация образования и методика электронного обучения: цифровые технологии в образовании : материалы IV международной научной конференции (6-9 октября 2020 г., Красноярск). – Красноярск, 2020. – Ч. 2. – С. 336-340.
 8. Селиванова И. В. Сравнение экспертной, журнальной и автоматической классификации на примере аннотаций и полных текстов научных публикаций / И. В. Селиванова, А. Е. Гуськов, Д. В. Косяков // Наука, технологии и информация в библиотеках (LIBWAY-2020) : Международная научно-практическая онлайн-конференция (Новосибирск, 14-17 сентября 2020 г.). – Новосибирск, 2020.