

Дискурс российской социологической науки 2020-2025 гг. на примере интеллектуального анализа научных текстов

Discourse of Russian sociology in 2020-2025: a case study of scientific text mining

Медведев М.М.

Студент 2 курса магистратуры, Факультет истории, социологии и международных отношений, Социология, ФГБОУ ВО «Кубанский государственный университет», г. Краснодар
e-mail: maxim.med2002@gmail.com

Medvedev M.M.

2nd year Master's Degree Student, Faculty of History, Sociology and International Relations, Sociology, Kuban State University, Krasnodar
e-mail: maxim.med2002@gmail.com

Бойченко С.А.

Студент 3 курса бакалавриата, направление «Социология», Факультет истории, социологии и международных отношений, ФГБОУ ВО «Кубанский государственный университет», г. Краснодар

Boychenko S.A.

3rd year Bachelor's Degree Student, majoring in Sociology, Faculty of History, Sociology, and International Relations, Kuban State University, Krasnodar

Аннотация

В статье представлены результаты комплексного дискурс-анализа корпуса из 595 научных публикаций по социологии за 2020–2025 гг., отобранных в электронной библиотеке CyberLeninka. Цель исследования – диагностика текущего состояния, структурных особенностей и динамики тематических трендов в российской социологической науке с применением количественных методов компьютерной лингвистики. Методология включает автоматический сбор данных, лингвистическую предобработку (лемматизация UDPipe), тематическое моделирование методом латентного размещения Дирихле (LDA) и анализ семантических связей на основе векторных представлений GloVe с последующей визуализацией семантических сетей. Результаты выявили устойчивое ядро дисциплинарного дискурса, сформированное теоретико-методологическими, историческими и образовательными понятиями. Динамический анализ показал значительный рост релевантности тем, связанных с управлением, цифровизацией, данными и технологиями, начиная с 2023 г. Ключевым выводом является обнаружение выраженного семантического разрыва между этим традиционным гуманитарным ядром и формирующимся кластером «Цифровизация и ИИ». Это свидетельствует о слабой концептуальной интеграции цифровых методов и проблематики в основной дискурс российской социологии, которая сохраняет преимущественно теоретико-рефлексивный характер. Исследование вносит вклад в наукометрию и социологию знания, предлагая инструментарий для объективного картирования и рефлексии траекторий развития социальных наук в условиях цифровой трансформации.

Ключевые слова: интеллектуальный анализ текста, социология науки, научные тексты, цифровизация, тематическое моделирование.

Abstract

The article presents the results of a comprehensive discourse analysis of a corpus of 595 scientific publications in sociology from 2020–2025, selected from the CyberLeninka online library. The aim of the study is to diagnose the current state, structural features, and dynamics of thematic trends in Russian sociological science using quantitative methods of computational linguistics. The methodology includes automatic data collection, linguistic preprocessing (lemmatization using UDPipe), topic modeling via Latent Dirichlet Allocation (LDA), and analysis of semantic relationships based on GloVe word embeddings with subsequent visualization of semantic networks. The results revealed a stable core of disciplinary discourse formed by theoretical-methodological, historical, and educational concepts. Dynamic analysis showed a significant increase in the relevance of topics related to management, digitalization, data, and technologies starting from 2023. The key finding is the identification of a pronounced semantic gap between this traditional humanities core and the emerging «Digitalization and AI» cluster. This indicates a weak conceptual integration of digital methods and topics into the mainstream discourse of Russian sociology, which retains a predominantly theoretical-reflective character. The study contributes to scientometrics and sociology of knowledge by offering a toolkit for objective mapping and reflection on the development trajectories of social sciences in the context of digital transformation.

Keywords: text mining, sociology of science, scientific texts, digitalization, topic modeling.

В условиях цифровой революции и экспоненциального роста объемов научной информации, навигация в современном знании без анализа глубинных дискурсивных структур становится подобна движению без карты. В то же время, усиливающаяся политизация и идеологизация науки, особенно в социальных дисциплинах, делают критически важным научную и дискурсивную рефлексию для демистификации механизмов производства истины и защиты эпистемологической автономии научного поля.

Современное научное знание существует не в вакууме абстрактных идей, но формируется, легитимируется и транслируется через сложную систему языковых практик, текстов и коммуникативных событий – т.е. через научный дискурс. Изучение этого дискурса с позиций дискурсивного анализа, укорененного в работах М. Фуко, Т. А. ван Дейка и Т. ван Левена, представляется эпистемологически необходимым инструментом для рефлексии и прогнозирования траекторий развития самой науки [2]. Если традиционная социология знания фокусировалась на институциональных и социальных детерминантах, то дискурсивный подход позволяет сделать видимыми сами механизмы производства истины, проследить, как через язык конструируются объекты исследования, формируются дисциплинарные границы и властные отношения внутри академического поля.

Таким образом, дискурсивный анализ научных текстов служит двум взаимосвязанным целям. Во-первых, это цель рефлексии: он дает науке зеркало, в котором она может увидеть свои собственные языковые и концептуальные контуры, осознать свою историческую и социальную обусловленность, подвергнуть критике собственные естественные предпосылки. Во-вторых, это цель прогноза: выявляя нарождающиеся тренды, маргинальные, но набирающие силу концепты и области семантической плотности, анализ дискурса позволяет строить обоснованные предположения о будущих направлениях развития дисциплины. В эпоху цифровизации и взрывного роста объемов научной информации такой подход становится ключевым для навигации в сложном ландшафте знания и для осмысленного управления его эволюцией. Данное исследование, фокусирующееся на дискурсе российской социологии, является попыткой применить этот мощный аналитический аппарат для диагностики текущего состояния и потенциальных векторов развития одной из центральных социальных наук.

Нами при помощи написанного скрипта-парсера на языке Python с использованием библиотеки Selenium был собран корпус научных статей гуманитарной тематики. Выборка составила 595 публикаций по ключевому слову «социолог(ия)», размещенных на электронной библиотеке CyberLeninka с 2020 по 2025 г.: 100 статей за 2025 г.; 100 статей за 2024 г.;

100 статей за 2023 г.; 100 статей за 2022 г.; 100 статей за 2021 г.; 95 статей за 2020 г. Работа следует стандартному пайплайну анализа текстовых данных: предобработка текста, лингвистическая аннотация (включая токенизацию и лемматизацию), фильтрация и нормализация, построение модели и визуализация результатов. Используется язык R с рядом специализированных пакетов (`udpipe`, `tidytext`, `topicmodels`, `text2vec` и др.).

Первым этапом является лингвистическая разметка текстов с помощью `UDPipe` – инструмента для автоматического морфологического анализа и синтаксического разбора, обеспечивающего «точную лемматизацию и определение частеречной принадлежности для более чем 50 языков, включая русский, без использования внешних ресурсов» [7, с. 4291]. В коде это реализовано одноименной функцией, которая аннотирует текст, выделяя леммы (словарные формы) и части речи (UPOS). Далее проводится фильтрация по частям речи: оставляются только существительные (NOUN), прилагательные (ADJ) и имена собственные (PROPN), что является стандартной практикой для выделения смысловых единиц в тематическом моделировании.

Важным этапом является формирование списка стоп-слов. Помимо базового списка, включающего общие слова («социология», «исследование», «человек» и др.), применяется расширенный метод на основе TF-IDF (Term Frequency – Inverse Document Frequency). Этот статистический показатель позволяет оценить важность термина в контексте документа и всего корпуса [3, с. 72]. Формально TF-IDF вычисляется как произведение частоты термина в документе (TF) и обратной частоты документов, содержащих этот термин. В коде идентифицируются термины с низким средним TF-IDF, но высокой частотой встречаемости по всему корпусу, что указывает на их слабую описательную способность. Эти термины добавляются в список стоп-слов, что позволяет очистить данные от общеупотребительных, но неинформативных для тематического анализа слов. Это важный шаг, так как качество тематических моделей критически зависит от тщательной предобработки текстов, включая удаление слов, не несущих смысловой нагрузки [5].

Для тематического моделирования использован метод латентного размещения Дирихле (LDA). LDA – это генеративная вероятностная модель, которая представляет документы как смесь скрытых (латентных) тем, каждая из которых характеризуется распределением вероятностей над словами [4]. Алгоритм исходит из предположения, что каждый документ корпуса порождается следующим процессом: сначала выбирается распределение тем для документа (из распределения Дирихле), затем для каждого слова в документе выбирается конкретная тема (в соответствии с этим распределением), и внутри этой темы выбирается слово (из мультиномиального распределения, специфичного для темы) [1]. Обучение модели заключается в поиске таких распределений слов по темам (β) и тем по документам (γ), которые максимизируют правдоподобие наблюдения всего корпуса. В коде модель обучается функцией `LDA()` из пакета `topicmodels` на документа-терминной матрице (DTM), построенной из лемм. Выбор количества тем ($K=12$) был предварительно обоснован с помощью функции `searchK()` из пакета `stm`, которая оценивает качество модели при разном числе тем по нескольким метрикам (семантическая связность, перплексия на отложенной выборке и др.). Оптимальное K выбирается на основе компромисса между этими метриками.

После построения LDA-модели проводится анализ семантических связей между ключевыми терминами с помощью алгоритма `GloVe` (Global Vectors for Word Representation) [6]. В отличие от LDA, которая моделирует темы на уровне документов, `GloVe` строит распределённые векторные представления слов (эмбеддинги) на основе статистики их совместной встречаемости в скользящем окне контекста. Основная идея `GloVe` заключается в том, что отношения между векторами слов должны отражать их семантические и синтаксические отношения. Алгоритм минимизирует функцию потерь, которая зависит от отношения вероятностей совместной встречаемости слов. Полученные векторы затем используются для расчёта косинусной близости между терминами, что позволяет построить семантическую сеть (граф), где вершины – это термины, а рёбра – связи между ними, вес которых отражает семантическую близость. Визуализация этой сети с помощью `ggraph`

позволяет выявить кластеры терминов, соответствующие смысловым областям, и проверить, насколько они согласуются с темами, выделенными LDA.

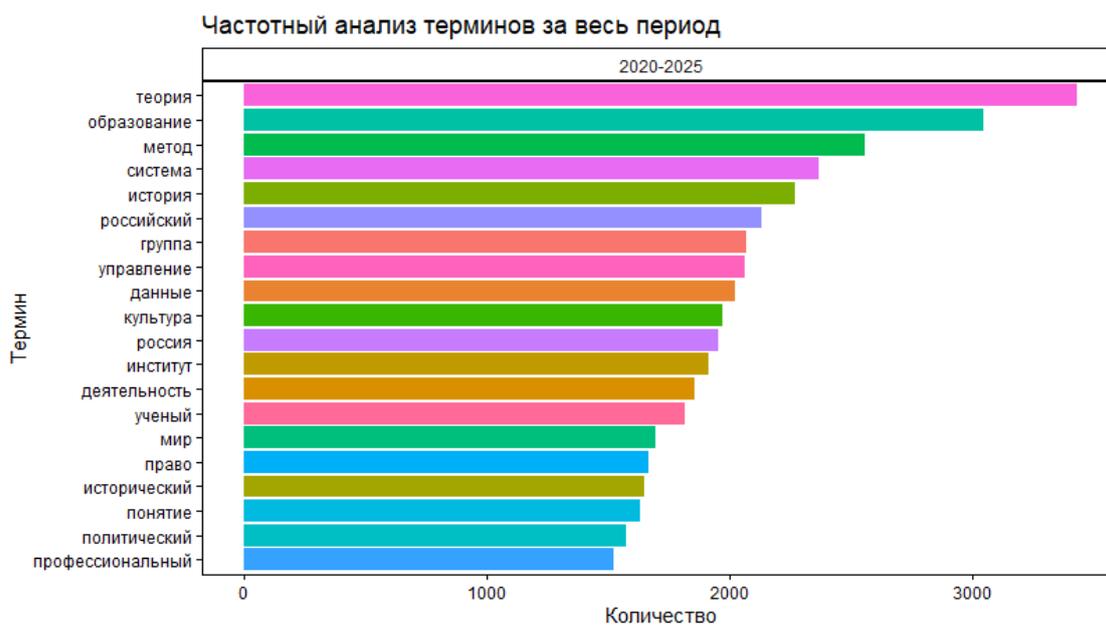


Рис. 1. Частотный анализ терминов за весь период

Перейдем к описанию результатов исследования. На рис. 1 представлена горизонтальная столбчатая диаграмма, отображающая абсолютную частоту встречаемости наиболее употребительных терминов (лемм) во всём корпусе статей за период с 2020 по 2025 г. Метод является базовым для количественного контент-анализа и служит для выявления ядра терминологического аппарата исследуемой области. После предобработки и фильтрации стоп-слов для каждого термина подсчитывается общее количество его вхождений (токенов) во всех документах. На графике отображены 20 наиболее частотных лемм, не входящих в стоп-лист, в порядке убывания частоты. Шкала по оси абсцисс показывает абсолютное количество употреблений в корпусе из 595 статей. Лидером по частоте является термин «теория», что коррелирует с гуманитарным характером выборки, где теоретический аппарат является объектом постоянной рефлексии. За ним с заметным отрывом следуют «образование», «метод», «система», «история», «российский», «группа», «управление», «данные», «культура». Наличие в топ-20 терминов «россия» и «российский» указывает на сильную национально-ориентированную специфику публикаций в выбранной базе. Присутствие терминов «данные» и «управление» отражает влияние междисциплинарных трендов, связанных с цифровизацией и менеджериализацией социального знания. Важно отметить, что частотный анализ, хотя и выявляет ядро лексикона, не учитывает контекст и совместную встречаемость терминов, что является задачей последующих методов.

Полученная частотная картина является лексическим срезом актуальной социологической проблематики в русскоязычном академическом поле первой половины 2020-х годов. Доминирование «теории» и «метода» подтверждает классическую дихотомию теоретической и методологической рефлексии в социологии как науке. Высокая частотность «образования» и «управления» сигнализирует о продолжающейся институционализации социологии как прикладной дисциплины, вовлеченной в решение задач образовательной политики и организационного менеджмента. Термин «данные» маркирует продолжающуюся «квантитативную революцию» и рост популярности data science. Наличие «истории» и «культуры» подчеркивает устойчивую связь социологии с исторической наукой и культурологией в рамках постнеклассической исследовательской парадигмы. Таким образом, частотный анализ выявляет гибридный характер дисциплины, балансирующей между

фундаментальной теоретической работой, историческим анализом и ответом на запросы со стороны смежных социально-управленческих практик.

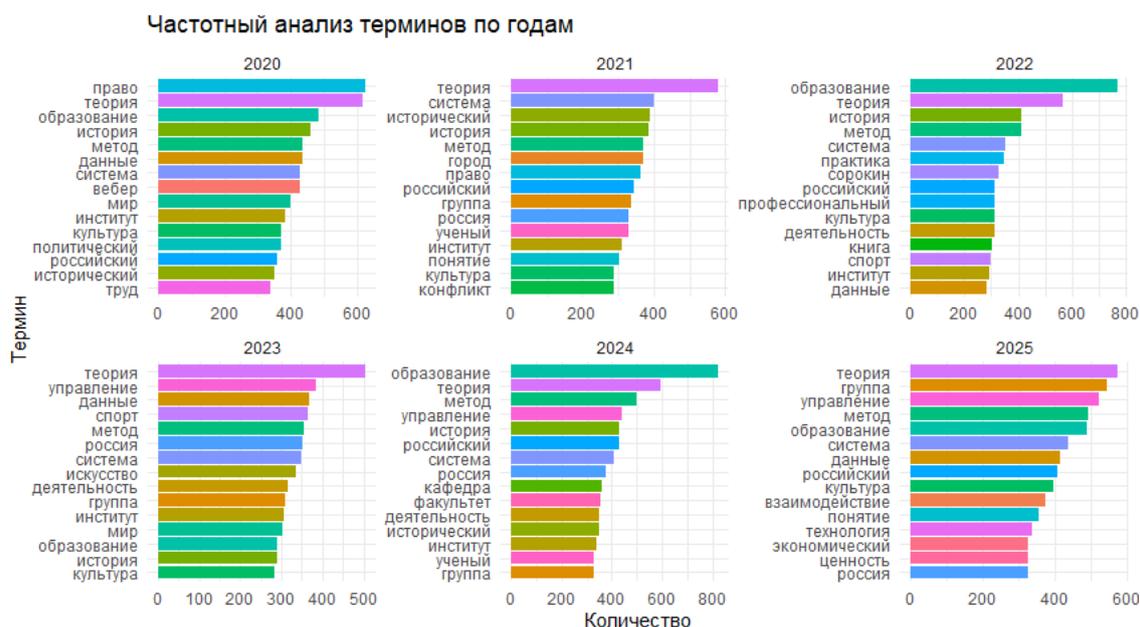


Рис. 2. Частотный анализ терминов по годам

На рис. 2 отображен частотный анализ терминов с разделением по годам. В 2020 г. в топе доминируют классические для гуманитарного знания термины: «право», «теории», «образования», «истории», «метод», а также «данные» и «системы». Присутствие фамилии «Вебер» указывает на активную работу с классическим теоретическим наследием. 2021 год сохраняет схожий набор: «образование», «теории», «истории», «метод», «система», но появляются «практика» и «книга», а «спорт» впервые входит в топ, что может быть ранним сигналом роста соответствующей темы. 2023 год знаменуется существенными изменениями: термин «управление» резко выходит на вторую позицию, появляется «спорт», сохраняются «данные» и «метод». Термин «искусство» также входит в список, что может отражать междисциплинарный поворот. 2024 год демонстрирует дальнейшую трансформацию: на первый план выходят «образование», «теории», «истории», но ключевым нововведением является появление термина «технологии», а также «экономической» и «цепи» (вероятно, в контексте «цепочки» или «сети»). 2025 год консолидирует тренды: лидируют «теория», «управление», «метод», «образования», «системы», «данные». Упрочняют свои позиции «технология» и «экономической», а также появляется «взаимодействие» и «понятие». Отсутствие в топе 2023-2025 гг. таких терминов 2020 г., как «Вебер» и «право», может указывать на смещение фокуса с историко-теоретической и правовой проблематики на более прикладные и актуальные управленческие и технологические вопросы.

Динамика частотности терминов по годам представляет собой лексико-метрическое свидетельство эволюции исследовательских интересов и дискурсивных поворотов в российской социологии начала 2020-х годов. Стабильное присутствие ядра из «теории», «метода», «образования», «истории» подтверждает непрерывность базовых исследовательских программ. Однако стремительный рост значимости «управления» и появление «технологий», начиная с 2023 г., является ключевым трендом. Это можно интерпретировать как реакцию социологического сообщества на два мощных внешних импульса. Во-первых, на усиливающийся запрос со стороны государства и бизнеса на экспертизу в сфере социального управления, эффективности институтов и управления персоналом. Социология активно встраивается в логику «управления знаниями». Во-вторых, это ответ на технологическую трансформацию общества, что выражается в росте частотности «данных» и «технологии». Социология осваивает роль интерпретатора цифровой реальности.

Появление и рост термина «спорт» в 2021 и 2023 г. может быть связано с мега-событиями или с растущим признанием спорта как значимого социального и политического феномена. Исчезновение из топа «права» и «Вебер» к концу периода требует осторожной интерпретации: это не обязательно означает забвение данных тем, но может сигнализировать об их переходе в статус «фонового знания» или о перераспределении интереса в пользу более сиюминутных, заказных и финансируемых исследовательских направлений (управление, технологии). Таким образом, годовой частотный анализ ярко иллюстрирует, как дисциплина балансирует между автономией своего классического концептуального аппарата и гетерономным влиянием социально-политического и технологического заказа, актуализируя свой лексикон в соответствии с меняющимся контекстом.

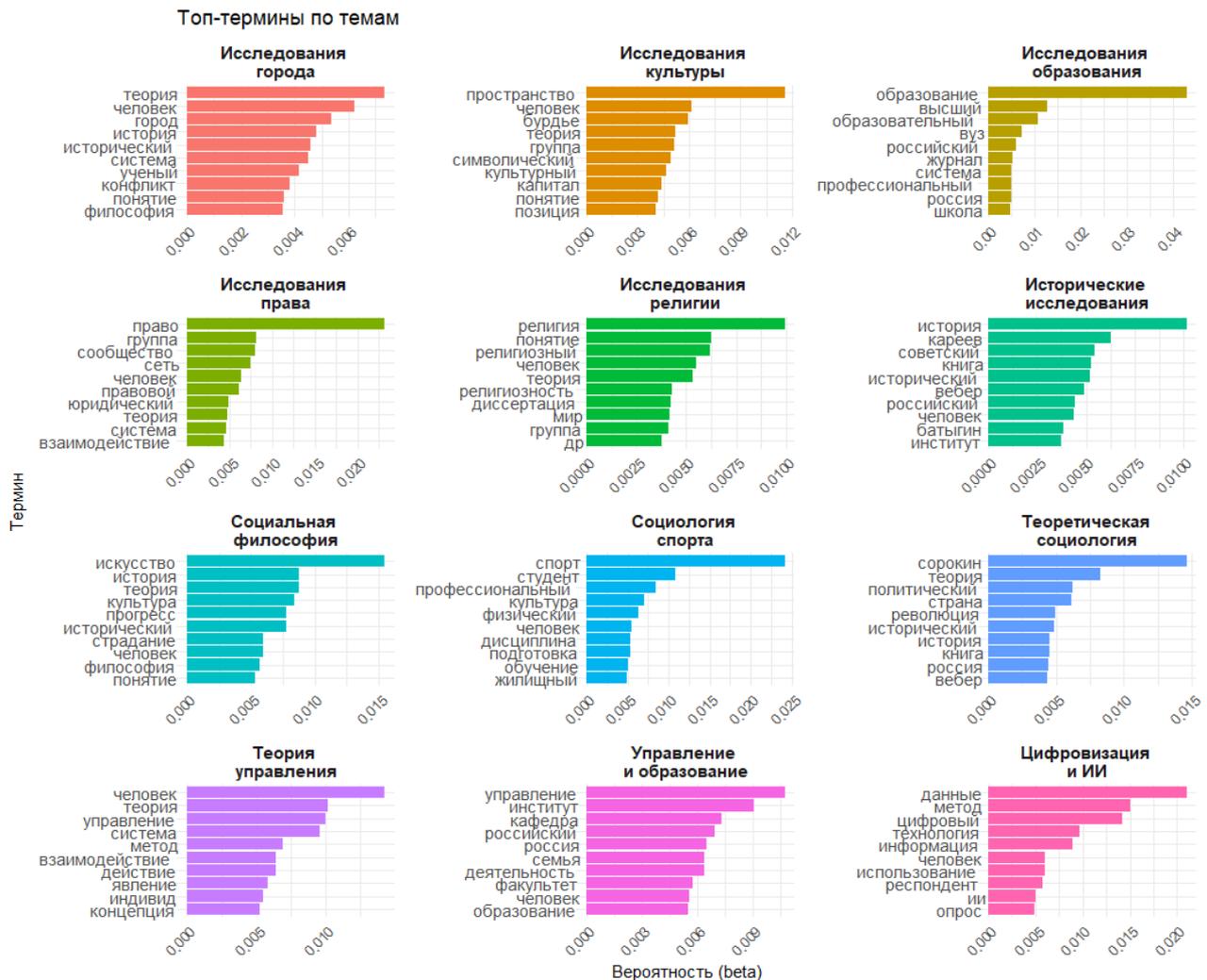


Рис. 3. Тематическое моделирование LDA, топ-10 терминов для 12 топиков

График на рис. 3 визуализирует результаты тематического моделирования (Topic Modeling) с использованием латентного размещения Дирихле (Latent Dirichlet Allocation, LDA). График представляет собой многопанельную диаграмму, где каждая панель соответствует одной теме. По оси ординат отложены термины в порядке убывания их вероятности (β) внутри темы, по оси абсцисс – сама вероятность. Анализ позволяет выделить содержательные кластеры. Например, тема «Исторические исследования» характеризуется терминами «история», «книга», «советский», «исторический», «вебер», «российский», что указывает на направленность на историко-социологический анализ, часто с фокусом на советский период и классическое наследие. Тема «Цифровизация и ИИ» концентрируется вокруг «данных», «метода», «цифровой», «технологии», «информации», «использования», отражая дискурс о новых технологиях в социальных исследованиях. Тема «Исследования

образования» объединяет «образование», «высший», «образовательный», «вуз», «российский», «системы», «профессиональный», что соответствует академическим исследованиям высшей школы. Важно отметить перекрестное присутствие некоторых терминов (например, «российский», «система») в нескольких темах, что демонстрирует их общеметодологический или общеконтекстуальный характер для всего поля.

Выделенные темы можно рассматривать как основные кластеры или исследовательские программы в российской социологии рассматриваемого периода. Их спектр отражает как традиционные для дисциплины области (теоретическая социология, социальная философия, исторические исследования), так и активно развивающиеся прикладные и междисциплинарные направления (управление, цифровизация, исследования спорта, города, права). Наличие тем «Исследования религии» и «Исследования культуры» подтверждает значимость социологии религии и культурологии. Тема «Управление и образование» является ярким примером гибридизации, где академический дискурс соединяется с практиками административного управления университетами. Содержание тем свидетельствует о преемственности по отношению к классическим теоретическим канонам (упоминание Вебера) и одновременно об активной рецепции современных трендов (цифровизация, data science). Структура тем в целом соответствует институциональной дифференциации социологического знания, существующей в виде специализированных секций, кафедр и журналов.

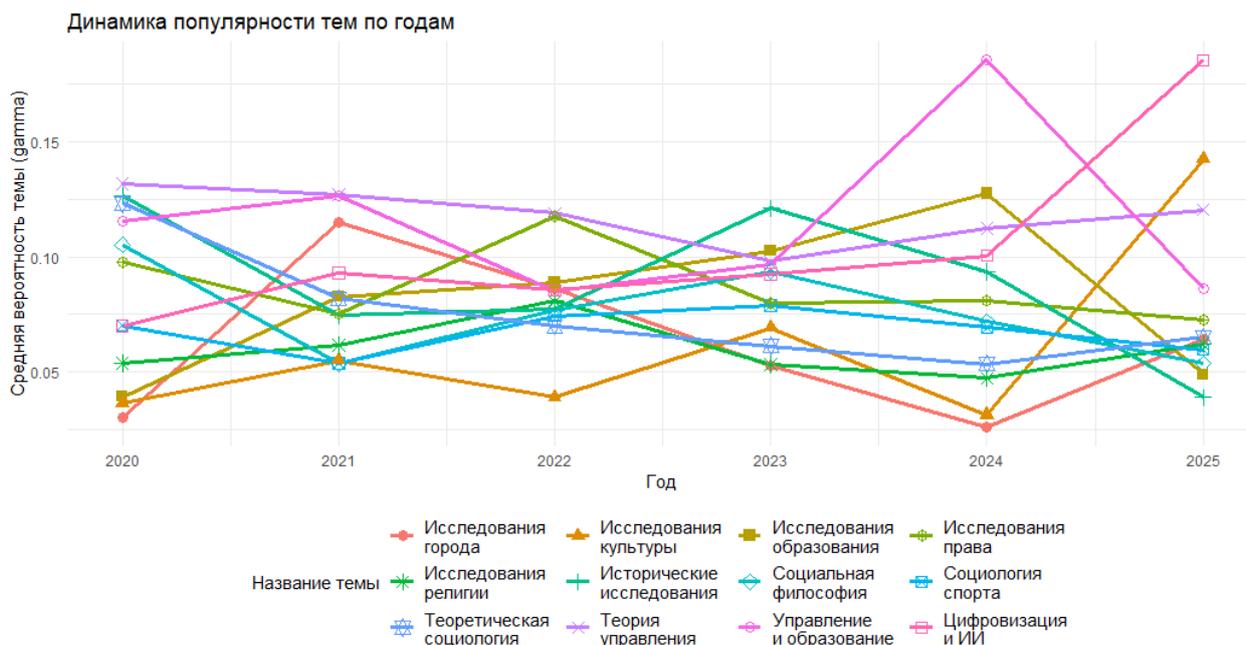


Рис. 4. Динамика распределения тем LDA по годам

Рис. 4 отображает динамику средней вероятности (γ) каждой из 12 тем по годам с 2020 по 2025 г. Данные получены из той же LDA-модели: для каждого документа вычислена вероятность его принадлежности к каждой теме (γ), затем эти значения усреднены по всем документам, опубликованным в конкретный год. Метод позволяет отследить изменение относительной значимости или «популярности» тематических направлений во времени. Визуализация выполнена в виде линейного графика с отдельными линиями и точками для каждой темы, что облегчает сравнение траекторий.

График выявляет разнонаправленные тренды. Некоторые темы демонстрируют относительно стабильный уровень вероятности на протяжении всего периода (например, «Исторические исследования», «Социальная философия»). Другие – заметный рост. Наиболее выраженный восходящий тренд наблюдается у темы «Цифровизация и ИИ»: ее средняя вероятность начинает расти с 2022 г. и достигает пика к 2025 г. Аналогичный, но менее резкий

рост характерен для тем «Теория управления» и «Социология спорта». Тема «Исследования города» показывает всплеск в 2021-2022 гг. с последующим спадом. Тема «Исследования религии», напротив, демонстрирует спад после 2022 г. Такая динамика позволяет предположить влияние внешних по отношению к академическому полю факторов (социальный запрос, государственная политика, технологические изменения) на исследовательскую повестку.

Динамика популярности тем является индикатором реакции социологического сообщества на актуальные социальные вызовы. Экспоненциальный рост интереса к цифровизации и ИИ напрямую коррелирует с глобальным технологическим трендом и внутренней политикой РФ в сфере цифровой трансформации, что порождает спрос на соответствующие социологические исследования (изучение цифровых практик, социальных последствий ИИ, данных цифровых платформ). Рост «Теории управления» и «Социологии спорта» может быть связан с усилением внимания к эффективности государственного и корпоративного управления, а также с мега-событиями в спорте и их социально-политическим контекстом. Спад интереса к «Исследованиям религии» после 2022 г. требует отдельного изучения, но может быть гипотетически связан со смещением фокуса научного внимания на более острые социально-политические и экономические проблемы. Стабильность историко-социологических и социально-философских исследований указывает на их статус фундаментального, менее подверженного сиюминутным колебаниям ядра дисциплины. Таким образом, график динамики иллюстрирует сложное взаимодействие внутренней логики развития науки и внешнего социального контекста.

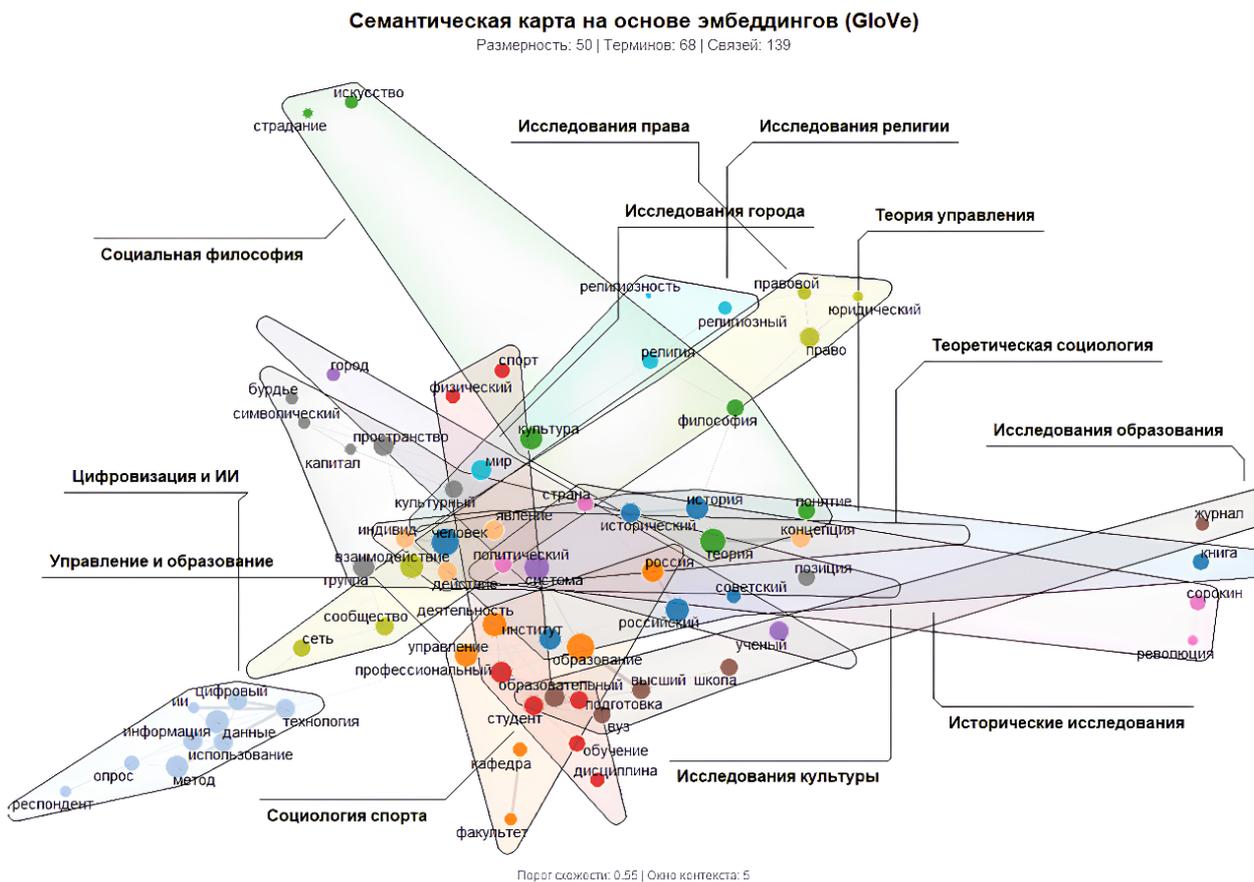


Рис. 5. Граф GloVe с группировкой по темам LDA

Рис. 5 представляет собой семантическую карту, построенную на основе векторных представлений слов (эмбедингов), обученных по алгоритму GloVe (Global Vectors for Word Representation). GloVe является методом обучения без учителя, который конструирует

векторные представления слов на основе статистики их совместной встречаемости в корпусе, стремясь к тому, чтобы скалярное произведение векторов двух слов было пропорционально логарифму вероятности их совместного появления [6]. В данном случае модель была обучена на лемматизированном корпусе с параметрами: размерность = 50, окно контекста = 5. Затем для топ-терминов из LDA-тем была рассчитана матрица косинусной близости их векторов. На основе этой матрицы, с применением порогового значения ($\text{sim_threshold} = 0.55$), был построен неориентированный взвешенный граф, где вершины – термины, а ребра – семантические связи между ними. Для визуализации графа использован алгоритм размещения Фрюхтермана-Рейнгольда ($\text{layout_algo} = \text{«fr»}$), который стремится расположить сильно связанные вершины ближе друг к другу. Кластеры (оболочки) на графе соответствуют темам LDA, что позволяет проверить согласованность двух разных методов: статистического (LDA) и дистрибутивно-семантического (GloVe).

На семантической карте вершины (термины) раскрашены в соответствии с их принадлежностью к LDA-темам, а размер вершины пропорционален ее частоте в корпусе и степени центральности в графе. Визуально можно наблюдать формирование кластеров, которые в целом соответствуют тематическим группам. Например, термины из тем «Цифровизация и ИИ» (данные, метод, цифровой) и «Теория управления» (управление, система, метод) образуют сближенные, но различимые области, что указывает на их семантическую родственность, но не идентичность. Термины из «Исторических исследований» (история, книга, советский) локализованы в иной части графа. Присутствуют и «мостовые» термины, находящиеся на стыке кластеров, что отражает их междисциплинарный характер. Структура графа демонстрирует, как различные концептуальные поля соотносятся друг с другом в едином семантическом пространстве социологического дискурса.

Данная визуализация предоставляет мощное свидетельство о структуре и разрывах внутри социологического знания в России. Изолированность кластера «Цифровизация и ИИ» является ключевым выводом. Это означает, что язык, концепты и проблематика цифровой социологии, анализа больших данных слабо интегрированы в основной понятийный аппарат дисциплины. Дискурс о цифровизации существует параллельно, не вступая в активный семантический диалог с дискурсами об истории, образовании, культуре или теории. Это подтверждает тезис о том, что российская социология в своем основном потоке остается по преимуществу теоретико-гуманитарной и качественно-ориентированной дисциплиной, в которой количественные и цифровые методы не стали «мейнстримом» или общим языком.

Плотное ядро из исторических, теоретических, философских и образовательных тем указывает на глубокую взаимосвязь этих областей. Социология в России осмысливает общество через призму истории, питается философской рефлексией и часто фокусируется на образовательных институтах как на ключевом объекте и метафоре. Это ядро отражает сильную традицию классического социального мышления. Относительная дистанцированность кластера «Социальная философия» и «Теоретическая социология» подчеркивает их особый, метатеоретический статус. Графовый анализ, таким образом, дополняет LDA, предоставляя не только набор тем, но и карту их взаимного расположения и связности, что является ценным инструментом для изучения структуры научного знания, его фрагментации или интеграции. Данный подход соответствует принципам наукометрии и анализа дискурса, позволяя объективно изучать концептуальные ландшафты академических дисциплин.

Таким образом, анализ корпуса статей по запросу «социология» выявляет дисциплину в состоянии напряжённого динамического равновесия и внутренней дифференциации. Российская социология, представленная на платформе CyberLeninka, демонстрирует:

- 1) рефлексивный и теоретический характер – язык дисциплины насыщен метапонятиями (теория, метод, система, понятие), что говорит о высокой степени самоосмысления;
- 2) доминирование образовательной и управленческой проблематики;

3) сильный историко-философский вектор (классическое наследие, исторический анализ и социально-философская рефлексия образуют смысловое ядро дисциплины, особенно усилившееся в последние годы, что можно трактовать как реакцию на современные вызовы);

4) субдисциплинарную специализацию – чётко выделяются области по религии, праву, городу, спорту;

5) критический разрыв между традиционным ядром и цифровой периферией.

Наиболее значимый вывод – слабая интеграция цифровых методов и тематики в основной дискурс. Цифровизация существует как обособленная, хоть и растущая, «планета» в семантической вселенной российской социологии. Это указывает на возможное методологическое и поколенческое отставание мейнстрима от современных трендов цифровой социологии, где компьютерные методы становятся неотъемлемой частью исследовательского инструментария многих субдисциплин. Тем не менее мы можем увидеть нарастающую связь темы «Цифровизация и ИИ» с темами «Исследования права» и «Управление и образование», что может говорить о потенциале сближения этих тем и более тесном внедрении цифровых методов в основной дискурс гуманитарной науки.

Данная картина отражает более широкий контекст развития социальных наук в России, где сильные традиции теоретической и гуманистической мысли сосуществуют с фрагментарным и затруднённым внедрением новых исследовательских технологий и междисциплинарных подходов, характерных для «цифрового поворота».

Литература

1. Медведев М.М. Казачьи колыбельные песни: опыт классификации и обобщения этнографического материала методами интеллектуального анализа текста / М.М. Медведев, А.А. Сергеев // *Голос минувшего*. – 2024. – № 3. – С. 30–40. – EDN PBTAVZ. – URL: <https://www.elibrary.ru/item.asp?id=PBTAVZ> (дата обращения: 26.01.2026).
2. Николина Н.В. Дискурс-анализ как рефлексивная практика в социологии научного знания / Н.В. Николина // *Философия науки*. – 2022. – № 3. – С. 12-20. – URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/koха:000999554> (дата обращения: 26.01.2026).
3. Рунаев Т.А. Интеллектуальный анализ текста в социальных науках: учебное пособие / Т.А. Рунаев; Министерство науки и высшего образования Российской Федерации, Кубанский государственный университет. – Краснодар: Кубанский гос. ун-т, 2024. – 127 с.
4. Blei D.M. Latent dirichlet allocation / D.M. Blei, A.Y. Ng, M.I. Jordan // *Journal of Machine Learning Research*. – 2003. – Vol. 3. – P. 993–1022. – URL: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf> (дата обращения: 26.01.2026).
5. Grimmer J. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts / J. Grimmer, B.M. Stewart // *Political Analysis*. – 2013. – Vol. 21, № 3. – P. 267–297. – DOI: <https://doi.org/10.1093/pan/mps028>. – URL: <https://web.stanford.edu/~jgrimmer/tad2.pdf> (дата обращения: 26.01.2026).
6. Pennington J. GloVe: Global Vectors for Word Representation / J. Pennington, R. Socher, C.D. Manning // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. – 2014. – P. 1532–1543. – URL: <https://aclanthology.org/D14-1162/> (дата обращения: 26.01.2026).
7. Straka M. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing / M. Straka, J. Hajic, J. Strakova // *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. – Portoroz, Slovenia, 2016. – P. 4290–4297. – URL: <https://ufal.mff.cuni.cz/biblio/attachments/2016-straka-p4247732638960608255.pdf> (дата обращения: 26.01.2026).