

ПРИКЛАДНОЙ СЕТЕВОЙ АНАЛИЗ ДЛЯ РЕШЕНИЯ СОВРЕМЕННЫХ ЗАДАЧ ГОСУДАРСТВА, БИЗНЕСА И ОБЩЕСТВА

МЕТОДОЛОГИЧЕСКИЕ РАЗРАБОТКИ И ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

November 27, 2023

Содержание

| | |
|---|-----------|
| 1 Основные термины | 2 |
| 1.1 Социальный капитал | 2 |
| 1.2 Социальное доверие | 2 |
| 1.3 Актуальность | 2 |
| 1.4 Литература | 3 |
| 1.5 Сети международной торговли | 4 |
| 1.6 Experiments with model's robustity | 5 |
| 2 1.1 Современные статистические методы для сетевого анализа: возможности и потенциал применения байесовской статистики | 5 |
| 3 1.2 Акторно-ориентированные стохастические модели для изучения сетевой динамики и социального влияния | 14 |
| 3.1 Введение | 14 |
| 3.2 Ограничения | 16 |
| 3.3 Вырождение | 19 |
| 3.4 1.5 Применение современных методов машинного обучения для предсказания связей в социальных сетях | 20 |
| 4 Глава 2 | 28 |
| 4.1 Графики | 28 |
| Заключение | 29 |
| 5 Изучение гендерной специфики деструкторов руководителей крупных российских компаний | 30 |
| 5.1 Введение | 30 |
| 5.2 Цели и задачи исследования | 31 |
| 5.3 Практическая значимость исследования | 31 |
| 5.4 Обзор литературы | 31 |
| 5.5 Методы сбора и обработки данных | 34 |
| 5.6 Обсуждение | 37 |
| 5.7 Результаты исследования | 38 |
| 5.8 3.1.1 Программа «Bib-eLib» для сбора и обработки библиографических данных на русском языке из электронной библиотеки eLibrary | 39 |
| 6 Методологические особенности предобработки данных по российским авторам в Web of Science | 43 |
| 6.1 4.2.1 Библиометрический сетевой анализ коллaborаций российских социологов на материалах Web of Science | 48 |
| 6.2 4.2.3 Картирование научного поля: применение VOSviewer и Biblioshiny на материалах Web of Science | 59 |
| 6.3 5.11.1 Модели управления благотворительными фондами – бенефициарными собственниками бизнес-компаний | 68 |

| | | |
|----------|--|-----------|
| 6.4 | 5.11.3 Изучение гендерной специфики деструкторов руководителей крупных российских компаний | 74 |
| 7 | Профессиональные роли журналистов: об исследовательском проекте | 84 |
| 7.1 | Введение | 84 |
| 7.2 | Методы сбора и обработки данных | 85 |
| 8 | Применение ERGM для анализа конференций | 97 |
| 8.1 | Введение | 97 |
| 8.2 | Данные и методы | 98 |
| 8.3 | Результаты | 99 |
| 8.4 | Заключение | 104 |
| 8.5 | 5.9 SNA для анализа пользовательского контента: взгляд через призму маркетинга | 105 |
| 8.6 | 3. Процесс применения SNA для анализа пользовательских отзывов | 111 |
| 8.7 | 3.2 Анализ основных метрик SNA | 111 |

1 Основные термины

1.1 Социальный капитал

Вот цитирование (*larranaga2013?*) и без скобок (*larranaga2013?*) — и сноска¹. А вот пример ссылки на часть текста sec. 4.1.

Попробуем написать что-то про социальный капитал. Но никто из моих цитирований про это не писал. У меня есть ссылка на две статьи по блокмоделингу: такую (Batagelj et al., 1999) и такую Marcot & Penman (2019). Но эти статьи не подходят и для следующего раздела sec. 1.2. Еще что-то написали. ### Bridging
Bonding

1.2 Социальное доверие

Также хотелось сделать картинку, но пока оставляю вот эту:

Модульный шрифт Джозефа Альберса

Figure 1: Модульный шрифт Джозефа Альберса

Цитата - это важно. *A еще лучше писать ее курсивом.*

1.3 Актуальность

Сетевой анализ как консистентная исследовательская методология сформировался в 1970-80-е гг., объединив ряд наработок в области социальной психологии, социометрии, социологии, антропологии, экономики, политологии, социальной географии, математики (теории графов) и статистики, а с 2000-х гг. стал разрабатываться также в естественных науках, что привело к появлению науки о сетях (Network science). Сетевой анализ относится к системному уровню анализа и рассматривает эмпирически обозримые отношения в виде сети, состоящей из узлов, связанных направленными или ненаправленными связями различной интенсивности. Предметом исследования выступают глубинные социальные структуры, оказывающие ограничивающее влияние на акторов с разным положением в социальной структуре и неравным доступом к ресурсам.

¹Внутренняя сноска

К настоящему времени в прикладном сетевом анализе разработано большое количество продвинутых методов для анализа различных типов сетевых данных. Использование этих методов позволяет отвечать на множество важных вопросов и решать современные задачи, стоящие перед государством, бизнесом и обществом.

Данный текстовый отчет содержит описание основных результатов работы Международной лаборатории за 2022 год по основным тематическим направлениям деятельности сотрудников. В главе 1 описана история появления и развития сетевого анализа, теоретические положения, определения основных понятий и принципы проведения сетевых исследований, а также представлены основные методологические разработки в сетевом анализе, сделанные в рамках реализации проекта. В главе 2 приведено описание проектов, реализованных сотрудниками лаборатории в 2022 году, где применялись методы, модели и инструменты для сбора, очистки и анализа социальных сетей. Для каждого проекта описаны их цели и задачи, методы сбора и обработки данных, полученные результаты, область их применения и степень внедрения. А вот пример цитирования — ([larranaga2013?](#)).

1.4 Литература

1.5 Сети международной торговли

1.5.1 Введение

Уолтер Рэли (1552–1618), английский мореплаватель, историк, поэт. *Кто владеет мировой торговлей, владеет богатствами мира, а следовательно – и самим миром.*

Рассказываем про сети международной торговли, как они образуются, что влияет на их развитие и т.п.

1.5.2 Графики

Как отмечается в (**bhattacharyaInternationalTradeNetwork2008a?**) торговые сети характеризуются высокой плотностью.² Вместе с тем, в работе (**deandradeUseNodesAttributes2018?**) отмечается сильная взаимозависимость.

Торговая сеть в 2010 году по наиболее крупному партнеру

Figure 2: Торговая сеть в 2010 году по наиболее крупному партнеру

Заключение

²Данный индекс был рассчитан в программе Pajek, данные по количеству публикаций топ-авторов могут отличаться от данных, полученных в VOSviewer, так как файлы для Pajek создаются с помощью программы WoS2Pajek, которая использует встроенные алгоритмы статистической обработки данных. В целом количественно данные по топ-авторам отличаются несущественно, что позволяет проводить сравнения разных метрик. В таблице жирным шрифтом отмечены авторы с самым высоким индексом коллаборативности.

1.6 Experiments with model's robustity

Я читала такую статью: (Zhou et al., 2021), она написана Zhou et al. (2021). Надеюсь этот файлик загрузится ^ ^

2 1.1 Современные статистические методы для сетевого анализа: возможности и потенциал применения байесовской статистики

Обнаружение сообществ в сетях широко изучалось в сетевой науке. Оно привлекло внимание в 1970-х годах, когда Лоррейн и Уайт (Lorrain & White, 1971) ввели функциональный мэппинг для получения глобальных сетевых паттернов, таких как структурная эквивалентность. Было предложено множество алгоритмов и методов, направленных на понимание структуры сложных систем. Однако, даже после недавнего оживления этой области, связанного с методом, основанным на спектральной модулярности (M. E. Newman, 2004), в исследовательском сообществе уже давно существует понимание того, что различия в простой концептуализации того, что такое сообщество, не являются основанием для существования единого метода для каждого приложения. Хотя не существует единого метода для каждого приложения (Fortunato & Hric, 2016), особенно при различной концептуализации сообществ, важно изучить различные сети и подходы. В данном разделе представлены современные статистические методы на основе байесовской статистики для сетевого анализа, применяемые для обнаружения сообществ в статических и динамических сетях. В разделе описаны особенности изучения развивающихся (динамических) больших разреженных сетей с использованием методов сетевого анализа. Описаны особенности и выявлены преимущества использования байесовского подхода в сетевом анализе для обнаружения сообществ в динамических сетях. Приведены примеры применения байесовского подхода к обнаружению сообществ при разработке больших разреженных сетей в области изучения различных социальных процессов.

Интерес к изучению социальных групп является фундаментальным для социологии. Классики социологии исследовали природу и эволюцию сообществ, сравнивая традиционные и современные формы, концепции Тенниса о *Gesellschaft* и *Gemeinschaft*, Дюркгейма о механической и органической солидарности продолжают оказывать влияние на социологические теории и в настоящее время. Однако также существует набор теорий для более сфокусированного анализа взаимодействий индивидов и, в последствии, анализа социальных сетей. Георг Зиммель (Simmel, 2010), Чикагская школа социологии (Cooley, 1909), и теоретики социального обмена (Homans, 2017) изучали характеристики социальных групп. Одним из важных понятий является первичная группа - небольшое неформальное сообщество с личным контактом и позитивным отношением (Freeman, 1992), и важность этих групп заключается в формировании паттернов социального поведения и ценностных установок по отношению к участникам группы и индивидам за ее пределами (Friedkin, 1984). Структура подобных групп была впервые описана Дэвисом (Davis et al., 2009) в 1941 году на примере женских сообществ в южных штатах США, где он выделил первичных и вторичных участников групп на основе частоты их участия.

К операционализации социальных групп можно подходить по-разному, в зависимости от того, как исследователи рассматривают групповую сплоченность. Существует несколько методов: 1) предполагая

полную взаимность связей, когда каждый участник выбирает другого; 2) наличие доступных участников, подключенных через других участников; 3) подчеркивание частых контактов или большого количества соседних узлов в сетевом анализе; 4) изучение внутригрупповой связанности по сравнению с внегрупповой связанностью (Wasserman & Faust, 1994a). Каждый метод опирается на классические концепции и методы анализа социальных сетей, некоторые из которых являются частью современных инструментов обнаружения сообществ. Одним математическим понятием, соответствующим полной взаимности, является клика, определяемая как полный подграф взаимно смежных узлов, при этом никакие другие узлы не имеют такой же смежности (Luce & Perry, 1949). Однако клики имеют ограничения в реальных социальных сетях, которые часто могут быть разреженными, что затрудняет поиск плотных групп. Кроме того, клики не дают представления о внутренней структуре группы, поскольку все участники имеют одинаковое количество связей и фактически идентичны. Таким образом, клики обладают ограниченным исследовательским потенциалом на практике.

Математические определения сплоченных социальных групп основаны на концепции клики, но с ослабленными свойствами. Одной из таких концепций является n -клика, члены которой могут относительно легко связываться друг с другом через посредников. N -клика - это максимальный подграф, в котором наибольшее геодезическое расстояние между любыми двумя узлами не превышает n (Alba, 1973). По мере увеличения n все больше узлов распознаются как часть n -клики, включая те, которые подключены к остальным максимум через n посредников. Однако, в отличие от клики, n -клика менее сплочена. Он может не иметь диаметра размера n , и даже при $n=2$ наибольшее геодезическое расстояние внутри n -клики может быть больше n . Кроме того, n -клика может быть не подключен, так как путь между узлами может проходить в обход элементов n -клика. Для решения этих проблем есть два решения: n -клан и n -клуб (Mokken, 1979), но они не получили широкого применения в исследованиях (Stokman & Sprenger, 1989).

Сплоченная подгруппа, построенная вокруг частых (но не обязательно постоянных) контактов ее членов, отражена в концепциях k -plex и k -core. Они были созданы потому, что n -клики подвержены уязвимости, т.е. исчезновению из-за удаления одного из узлов (Seidman & Foster, 1978), поэтому было необходимо разработать что-то, что повысило бы общую сплоченность подгрупп. В этом отношении k -plex определяется как максимальный подграф S с n узлами, которые привязаны к минимуму узлов n_s-k . Иными словами, у каждого узла в S могут отсутствовать связи максимум с k членами S . Степень (количество смежных узлов) для каждого узла в S равна не менее n_s-k и не может быть больше n_s-1 . При $k = 1$ это дает нам клику, и по мере увеличения k увеличивается и количество недостающих звеньев. Но поскольку связи охватывают более широкий диапазон, это делает подграф более стабильным. Однако крайне важно установить правильный размер подграфа (количество узлов), чтобы k -множества не были очень разреженными. k -ядро — это максимальный подграф, где каждый узел имеет степень не менее k (Seidman, 1983b). K -core является противоположностью k -plex, потому что в нем указано, сколько связей узел должен иметь с другими элементами своего подграфа, а не сколько у него может не быть. Вот почему k -ядра также часто являются вложенными, т.е. узлы могут находиться как в k -ядре, так и в $(k+1)$ -ядре. Хотя этот подход не приводит к созданию плотных группировок, он служит показателем для общей кластеризации и определения мест потенциальных группировок (Seidman, 1983b).

Понимание того, как устроено сообщество, включает в себя связи, близость и частоту контактов между членами группы. Множества LS и λ отражают идею многочисленных и прочных связей внутри

сообщества. LS имеет больше связей внутри своих подграфов, чем за их пределами (Seidman, 1983a), в то время как λ использует линейную связность для определения минимального количества связей для разъединения узлов. Множества λ являются более общими и отражают уникальное понятие связности. Эти концепции являются основополагающими для изучения сообществ в рамках анализа социальных сетей. Понятие клики было расширено, чтобы охватить близость (n -клик) и частоту контактов (k -сплетений и k -ядер). Комплекты LS и λ сочетают в себе внутреннюю относительную прочность и малое количество внешних соединений. Эти концепции необходимы для передовых методов обнаружения сообществ в разреженных сетях.

В области выявления сообществ проводится множество исследований по новым и систематизированным методам, и хотя согласованной таксономии не существует, есть некоторая основа классификации. Статичные сети далее изучаются с помощью методов секционирования (partitioning) или традиционной кластеризации, а также методов статистического вывода (statistical inference). Некоторые исследователи рассматривают спектральную кластеризацию отдельно, в то время как другие классифицируют ее как основанную на модульной оптимизации или традиционную кластеризацию. Были предложены различные подходы к классификации, включая нулевые модели, блочные модели, потоковые модели и такие функции, как внутренняя плотность, структурное сходство, динамическое сходство и возможность разделения разделов. Вместо генеалогических классификаций мы фокусируемся на общих исследовательских задачах и признаем примеры, на которых основаны методы. В одном разделе мы обсуждаем статические методы обнаружения сообществ, а в другом - динамические методы, описывая подходы, мотивацию, механику и ситуации для их использования

Сначала определим группу подходов, основной целью которых является разделение сети на подходящее количество сообществ. Подходы, основанные на разрезании (cut-based), направлены на разделение сети на сообщества путем минимизации внутренних связей и максимизации внешних. Это достигается с помощью таких методов, как обрезка дуг (edge cut) и пропорциональная нарезка (ratio cut), обеспечивающая сбалансированное и оптимальное разделение. Спектральная кластеризация - популярный метод, который преобразует сеть в точки в многомерном пространстве с использованием собственных векторов, позволяя идентифицировать сообщества в сложных данных. Однако он не масштабируем для больших сетей или разреженных данных. Также используются методы разделения, такие как k -средние значения и алгоритмы иерархической кластеризации (HCA). HCA работает сверху вниз или снизу-вверх для обнаружения иерархических сообществ, но полагается на правильное расположение узлов. Алгоритмы Гирвана-Ньюмана (GN) устраниют границы с высокой степенью связанности для поиска сообществ, но могут быть дорогостоящими с точки зрения вычислений. Эти подходы практичны для различных данных, особенно для исследовательских целей, но могут плохо работать с большими или разреженными наборами данных. Кроме того, они не обладают высокой масштабируемостью.

Поиск среза минимальной дуги сам по себе может не дать удовлетворительных результатов, поскольку он игнорирует внутренние связи подграфов. Чтобы решить эту проблему, проводимость (conductance) была введена как мера среза дуг (edge cut) относительно объема подграфа (Kannan et al., 2004). Проводимость обеспечивает локальное развертывание и снижает вычислительную сложность (Rosvall et al., 2019). Другой важной концепцией является модульность, которая количественно определяет количество дуг внутри групп по сравнению со случайной сетью (M. E. Newman & Girvan, 2004).

Максимизация модулярности позволяет идентифицировать сообщества в сети. Спектральная и жадная (greedy) оптимизации — это обычно используемые методы оптимизации модулярности. Спектральная оптимизация использует собственные векторы и собственные значения матрицы модулярности. Жадная оптимизация, такая как алгоритм Louvain, постепенно увеличивает модулярность за счет объединения кластеров. Однако эти алгоритмы могут создавать несвязанные сообщества (Javed et al., 2018) и в попытке создать не пересекающиеся кластеры, произвести неинтерпретируемые результаты в случае, если в сетях пересекающиеся сообщества, что свойственно реальным сетям (Traag et al., 2019a). Алгоритм Leiden решает эти проблемы с помощью процедуры быстрого локального перемещения (Traag et al., 2019a). Несмотря на популярность, методы оптимизации модулярности имеют ограничения, и исследователи ставят под сомнение их применимость к реальным сетям Fortunato & Barthelemy (2007). В целом, методы, основанные на кластеризации, такие как Louvain и Leiden, широко используются из-за их эффективности в крупномасштабных сетях, но они могут давать неубедительные результаты.

Существуют подходы к выявлению сообществ, основанные на статистическом выводе и проверке гипотез, которые позволяют исследователям тестировать различные типы сообществ в сети. Традиционные методы предполагают плотные и эквивалентные группы, но при анализе социальных сетей бывают случаи, когда узлы могут не группироваться на основе сходства. Могут существовать различные типы сетевых структур, такие как диссортативные и структуры ядро-периферия. Стохастический блокмоделинг (SBM) — это метод, используемый для тестирования конкретных структур сообщества в сети (Wasserman & Faust, 1994a), (Holland et al., 1983). Он перестраивает связи на основе идеальных блоков, представляющих различные типы сообществ. Цель состоит в том, чтобы свести к минимуму разницу между эмпирической структурой сети и идеальной моделью при оценке соответствия модели. SBM широко используется в SNA и в качестве эталона для методов выявления сообществ (Fortunato & Hric, 2016), (Javed et al., 2018). Однако у классического SBM был недостаток, поскольку он не учитывал степень неоднородности в реальных сетях. Для решения этой проблемы был разработан SBM с поправкой на степень (DCSBM) (Karrer & Newman, 2011), но для такого алгоритма требуется заранее указать количество сообществ, чтобы избежать переобучения модели (Fortunato & Hric, 2016). Методы выбора модели, такие как иерархическая вложенность SBMS, были предложены для улучшения сжатия данных (Peixoto, 2014). Методы, основанные на SBM, являются мощными для выявления сообществ, выступая в качестве генеративных моделей и ориентиров (Dao et al., 2020). Однако они сопряжены с более высокими вычислительными затратами и более длительным временем выполнения. Подробности SBM и его байесовского варианта будут рассмотрены позже.

Сообщества обычно рассматриваются как статичные структуры, но также важно учитывать и их динамику. Динамические модели помогают понять функции и временные сообщества внутри сложных систем (Rosvall et al., 2019). Модели, основанные на случайном блуждании (Pons & Latapy, 2005), идентифицируют сообщества, в которых «случайное блуждание» (random walker) попадает в ловушку, обходя все узлы в этом регионе, и затем эти сообщества группируются с использованием традиционных методов или методов, основанных на разрезе (cut-based) (Xu & Wunsch, 2008). Однако в первоначальной формулировке этот метод очень требователен к вычислениям и не может быть использован в больших сетях (Fortunato & Hric, 2016). Подходы, основанные на теории информации, заменяют случайного ходока кодовыми словами для описания структур сообщества и минимизации длины описания. Алгоритм Infomap (Rosvall & Bergstrom, 2008) вычисляет минимальную длину описания

бесконечного случайного блуждания и обладает способностью обнаруживать иерархические (Rosvall & Bergstrom, 2011) и перекрывающиеся сообщества (Esquivel & Rosvall, 2011) и даже воздействовать на сохраненную память о предыдущих состояниях (Persson et al., 2016). Существует еще один метод - spin glass, название которого отсылает к моделированию поведения неупорядоченных материалов, таких как стекло, полимеры и сверхпроводники, где большое количество частиц, называемых “спинами”, случайнym образом взаимодействуют. В модели spin glass (Mydosh, 1993) структура сообщества представлена в виде конфигурации спинов, и цель состоит в том, чтобы минимизировать энергию путем согласования спинов со структурой сообщества. Тем не менее алгоритмы случайного блуждания показали лучшую производительность, чем модели spin glass (Reichardt & Bornholdt, 2006). Динамическое обнаружение сообществ относительно игнорируется (Rosvall et al., 2019), с акцентом на статические подходы и фиксацию динамики с помощью моментальных снимков структуры сети. Однако для обнаружения перекрывающихся сообществ и поддержания согласованности с течением времени был разработан алгоритм динамического байесовского детектора перекрывающихся сообществ (DBOCD).

Для исследования социальных сетей актуальным является вопрос изучения больших разреженных динамических сетей, что предполагает наличие двух качеств сети – «разреженных» и «динамических». Сеть называется разреженной, если в ней гораздо меньше связей, чем возможный максимум (Barabási, 2013). Динамическая сеть рассматривается как функция времени (Lotker, 2021), что подразумевает, что для каждого момента в заданном периоде существует граф и изучается множество графов. Следовательно, большая разреженная динамическая сеть может быть определена как подмножество сетей временного масштаба с большим числом узлов и гораздо меньшим количеством связей, чем максимально возможное число.

Причина изучения таких объектов заключается в том, что разреженные сети на самом деле чрезвычайно распространены в реальной жизни, в отличие от плотных модели (Barabási, 2013), например, в нейробиологии нейронная система содержит огромное количество нейронов, но связи между ними немногочисленны, что затрудняет анализ. Динамическая характеристика сети также важна, поскольку, изучая эволюцию взаимоотношений узлов и изменения в структуре, исследователи смогли бы отслеживать распространение информации и выявлять особенности построения сообщества.

Ключевой работой, которая обеспечивает основу для математических расчетов в области изучения динамических разреженных сетей, является “Моделирование динамических сетей с разреженными и слабыми связями во временном масштабе” (Chow & Kokotovic, 1987). В книге рассматриваются такие темы, как методология моделирования в масштабе времени, разделение во временном масштабе, преобразование траектории и многое другое. Книга полезна для анализа крупномасштабных систем, таких как энергосистемы, и выявления важных узлов в сети.

Существует несколько методов анализа больших разреженных динамических сетей. В первую очередь стоит рассмотреть алгоритмы анализа графов. Эти алгоритмы используются для анализа структуры сети и выявления закономерностей и взаимосвязей между узлами. Примеры алгоритмов анализа графов включают поиск в ширину, поиск в глубину, кратчайший путь, обнаружение цикла, минимальное связующее дерево и раскраску графа (Ghoorjian, 2019).

Теория динамических графов. Этот метод моделирует и анализирует сети как двухкратные разреженные динамические графовые сети с кластерами, представляющими связную подсистему (Meyer-Bäse et al., 2017). Применение теории графов позволяет формализовать определения макросостояния системы,

микросостояния на микроуровне и динамических структурных изменений (Gignoux et al., 2017). Для классической и динамической теорий графов на первый план выходят различные проблемы. Более конкретно, в классической теории графов основной задачей оптимизации является поиск подграфа или связующего дерева с определенными характеристиками, такими как минимальный вес. В теории динамических графов ключевой проблемой является установление взаимосвязи между оптимизационными решениями на разных графах. Это позволяет нам определить наследование в классе динамических графов с общими правилами перехода. Затем мы можем связать наследственное свойство с операциями перехода в траектории динамического графа. Если мы установим эту связь, мы сможем достичь программируемой самоорганизации и получить гарантированные унаследованные структурные свойства и характеристики динамических графов (Kochkarov et al., 2015).

Спектральные алгоритмы (Singh & Humphries, 2015), основанные на матричных представлениях сетей, часто используются для обнаружения сообществ, но классические спектральные методы, основанные на матрице смежности и ее вариантах, терпят неудачу в разреженных сетях. Недавно были внедрены новые спектральные методы, основанные на случайных блужданиях (random walks) без обратного отслеживания, которые успешно обнаруживают сообщества во многих разреженных сетях. Разбиение спектрального графа на самом деле представляет собой семейство методов. Эти методы зависят от собственных векторов матрицы Лапласа или ее родственников в графе. В зависимости от способа разбиения графика спектральные методы можно разделить на два класса. Первый класс использует ведущий собственный вектор графа Лапласиана для двойного разбиения графа. Второй класс подходов вычисляет k-образное разбиение графа с использованием нескольких собственных векторов (Ruan & Zhang, 2007). Однако стандартные методы, основанные на матрице смежности и связанных с ней матрицах, не работают для очень разреженных сетей, которые включают в себя множество сетей, представляющих практический интерес. В качестве решения этой проблемы недавно было предложено вместо этого сосредоточиться на спектре матрицы без обратного отслеживания - альтернативном матричном представлении сети, которая демонстрирует лучшее поведение в разреженном пределе с несколько иным определением, чтобы обладать желаемыми свойствами, особенно в общем случае сетей с широкой степенью распределения (M. E. J. Newman, 2013).

Вычисления с разреженной матрицей для динамической централизации сети могут быть использованы для получения нового алгоритма вычисления зависящей от времени централизации, который работает с разреженной версией матрицы динамической коммуникальности. Таким образом, требования к вычислениям и хранилищу сводятся к требованиям разреженной статической сети в каждый момент времени (Arrigo & Higham, 2017). «Поскольку зависящая от времени пограничная структура обычно позволяет информации широко распространяться по сети, естественная сводка разреженных, но динамичных парных взаимодействий, как правило, принимает форму большой плотной матрицы. По этой причине вычислительные узловые центры для зависящей от времени сети могут быть чрезвычайно дорогостоящими как с точки зрения вычислений, так и с точки зрения хранения; гораздо более дорогостоящими, чем для отдельной статической сети» (Arrigo & Higham, 2017).

Методы пертурбации используются для выделения функциональных аспектов созданной сети, таких как динамика в данной сети. В случае, если исследователи предполагают наличие случайного шума в данных, при выполнении статистического анализа, они многократно пертурбируют и группируют данные, а затем агрегируют результаты (Mirshahvalad et al., 2012).

Подходы группы Лассо (Lasso-type approach), зачастую используются в методах сетевого анализа, таких как байесовские подходы и спектральные алгоритмы. Байесовская регрессия Лассо — это тип регрессионного анализа, который выполняет как выбор переменных, так и регуляризацию с целью повышения точности прогнозирования и интерпретируемости результирующей статистической модели (Tibshirani, 1996).

Подходы группы Лассо также используются при оценке и выборе переменных в рамках модели единого индекса (Chiuso & Pillonetto, 2012). Спектральные алгоритмы, основанные на матричных представлениях сетей, часто используются для обнаружения сообществ, но классические спектральные методы, основанные на матрице смежности и ее вариантах, терпят неудачу в разреженных сетях. Метод «Сетевого Лассо» недавно был адаптирован для наборов данных с сетевой структурой, и было показано, что он может быть точным при определенных условиях, зависящих от базовой структуры сети и набора выборок.

Обнаружение сообществ в больших разреженных сетях является сложной задачей, и традиционные методы зачастую не справляются с такими данными. Байесовский подход приобрел известность благодаря своей способности решать эти проблемы, предоставляя вероятностную структуру, моделирующую неопределенность и включающую предварительную информацию. Он использует теорему Байеса для обновления убеждений, основанных на новых доказательствах. Байесовские методы широко используются в статистике, сетевом анализе, машинном обучении и искусственном интеллекте. По своей сути байесовский подход основан на теореме Байеса, которая представляет собой математическую формулу, описывающую, как обновлять наши убеждения или вероятности относительно события, когда становятся доступны новые доказательства или информация.

Байесовская вероятность - это математическая основа для обновляющихся убеждений перед лицом новых доказательств. Он объединяет предварительные знания, представленные априорными вероятностями, с наблюдаемыми данными, измеряемыми с помощью вероятностей, для вычисления апостериорных вероятностей. Предварительные данные выражают первоначальные убеждения относительно гипотезы, в то время как вероятности количественно определяют поддержку, предоставляемую данными. Обновляя априорные значения с использованием теоремы Байеса, мы можем получить апостериорные вероятности, которые включают в себя как предшествующие знания, так и новую информацию, что позволяет более точно представлять убеждения (Kruschke, 2010). Байесовский подход отлично подходит для обработки неопределенной или неполной информации. Он позволяет систематически обновлять убеждения по мере появления новых свидетельств. Он широко используется в статистике, машинном обучении и сетях. При обучении с учетом разреженности мы фокусируемся на байесовском подходе (Theodoridis, 2015). Это помогает сделать вывод о структуре сети и динамике с течением времени, учитывая разреженность. Обнаружение сообщества включает в себя поиск групп узлов со схожими схемами подключения. Байесовские методы справляются со сложностями динамических и разреженных сетей. Они дают ценную информацию, учитывая неопределенность и разреженность данных. Байесовский вывод рассматривает вероятности как меру неопределенности и обновляет их новыми данными. Это позволяет принимать рациональные решения и делать прогнозы, сочетая предварительные знания с наблюдаемыми фактическими данными.

Во многих исследованиях изучался байесовский вывод для выявления сообществ. Они используют вероятностные модели для обработки заданий сообщества, структуры сети и временной эволюции.

Байесовские априорные значения кодируют убеждения о размерах сообщества, вероятностях связей и динамике. Статья «Байесовское обнаружение сообществ» (van der Pas & van der Vaart, 2018) посвящена применению байесовского вывода к стохастическому блокмоделингу (SBM) для сетевого анализа, в частности для обнаружения сообществ. Байесовские методы восстанавливают метки классов в SBM, предлагая вероятностный подход. В статье приведены теоретические результаты, подтверждающие высокую согласованность байесовского апостериорного метода при обнаружении сообществ, подчеркивая его надежность. Это предполагает проведение будущих исследований по включению предварительных знаний и оценке числа сообществ с использованием байесовских методов. Байесовский вывод улучшает обнаружение сообщества в SBM (van der Pas & van der Vaart, 2018).

В статье «Последовательное байесовское обнаружение сообществ» представлен байесовский подход к обнаружению сообществ в развивающихся больших разреженных сетях, включающий ковариаты. Это решает проблему использования сетевой структуры и ковариационной информации для обнаружения. Предлагаемый байесовский SBM включает ковариаты посредством предварительного случайного разбиения, зависящего от ковариат, выражая их влияние на членство в кластере. Примечательно, что он узнает количество сообществ на основе данных без предварительного знания (Hoffmann et al., 2020).

Байесовские методы превосходят остальные в выявлении структурных свойств динамических сетей и моделировании развивающихся взаимосвязей. В статье “Обнаружение сообществ в сетях без наблюдаемых связей” (Hoffmann et al., 2020) представлена байесовская иерархическая модель для обнаружения сообществ в данных временных рядов без наблюдаемых связей. Этот целостный подход включает в себя байесовский вывод для сравнения моделей и выбора оптимального масштаба сообщества. Он решает вычислительные задачи, проблемы неопределенности и многомасштабного обнаружения. Иерархические байесовские модели интегрируют различные уровни структуры, временные зависимости и свойства сети в рамках единой структуры (Glynn et al., 2015).

В статье «Байесовский подход к идентификации разреженной динамической сети» описывается использование динамических байесовских сетей (DBNs) для идентификации разреженных динамических сетей. DBNs — это вероятностные графические модели, которые фиксируют взаимосвязи между переменными с течением времени (Chiuso & Pillonetto, 2012). DBNs расширяют байесовские сети для работы с динамическими данными и вероятностными зависимостями. Они находят применение в различных областях, включая обработку языковых данных и финансовое моделирование, для понимания временных взаимосвязей. Более того, байесовские методы могут включать в себя априорные данные, вызывающие разреженность, для идентификации как динамических, так и отсутствующих связей в разреженных сетях. Это предпочтение более простых структур помогает обнаружить значимые взаимосвязи и избежать переобучения (Chiuso & Pillonetto, 2012).

Более того, на основе байесовских методов было предложено множество новых подходов и решений. В статье «Байесовский подход к идентификации разреженной динамической сети» авторы представляют два байесовских подхода, Stable-Spline GLAR (SSGLAR) и Stable-Spline Exponential Hyperprior (SSEH), оба из которых способствуют разреженности при выборе модели и оценке импульсного отклика. Принимая байесовскую точку зрения, можно получить ту же формулировку, моделируя компоненты φ как независимые гауссовские случайные величины. Авторы предлагают байесовскую модель для разреженной идентификации, где распределения вероятностей представляют неопределенность.

Байесовские методы предлагают способ оценки неопределенности с помощью апостериорных

распределений. Это особенно важно при выявлении отсутствия динамических связей, поскольку позволяет исследователям не только заявить об отсутствии доказательств наличия связи, но и количественно оценить свою уверенность в этом утверждении. Эта информация ценна для принятия решений и проверки гипотез.

В статье «Обнаружение сообществ и их эволюция в динамических социальных сетях — байесовский подход» (T. Yang et al., 2011) предлагается динамическая стохастическая блочная модель для анализа сообществ и их эволюции в единой вероятностной структуре. В методологию интегрирован байесовский подход, использующий байесовский вывод для оценки апостериорных распределений параметров. Это повышает устойчивость модели к помехам в данных и фиксирует неопределенность в значениях параметров. Использование байесовской обработки является ключевой особенностью, которая обеспечивает вероятностную основу для выявления сообществ и анализа эволюции (T. Yang et al., 2011).

Динамические сети также имеют переменные прибытия и выбытия узлов. Байесовское моделирование обрабатывает это с помощью вероятностных моделей для этих событий. Предыдущие распределения используются для определения вероятности присоединения новых узлов к сети или выхода из нее на каждом временном шаге. Байесовский вывод обновляет назначения сообществ и параметры модели по мере изменения узлов (T. Yang et al., 2011).

Байесовская структура обычно используется для оценки гиперпараметров. Распределения правдоподобия, основанные на модели, включают гауссовские распределения для измерений с шумом (Y. Wang et al., 2022).

Гауссовые процессы - распространенный инструмент байесовского моделирования - дают оценки и количественную оценку неопределенностей в сценариях с шумом. Они используются для задач регрессии и оценки. В статье «Байесовская оценка импульсных откликов» наиболее вероятный импульсный отклик определяется по наблюдаемым данным. Байесовские подходы включают регуляризацию, чтобы сбалансировать сложность и подгонку данных. В гауссовых процессах выбор ковариационных функций и параметров регуляризации управляет плавностью и регулярностью оцениваемых функций (импульсных откликов). Байесовские методы включают неопределенность и регуляризацию в идентификацию системы, что соответствует подходу, описанному в статье (Chiuso & Pillonetto, 2012).

На практике исследователи часто сталкиваются с тем, что узлы сети группируются в определенные тесные сообщества с преобладанием внутренних связей внутри кластеров над связями между кластерами. Исследователи отмечают, что в сетевом анализе важно эксплицитно моделировать структуру сообществ рассматриваемых сетей. Так, исследуя 16 эмпирических сетей, начиная от классических «Клуба каратэ Захарии» и дельфинов-афалин из «Doubtful Sound» до футбольных и энергосетей, Мёруп и Шмидт показывают, что структура сообществ возникает из всех данных эмпирическим путем и предлагают алгоритм для ее моделирования, опирающийся на байесовскую статистику (Møgup & Schmidt, 2012).

Байесовские методы к обнаружению сообществ также могут опираться на классическое понятие ассортативности в формировании связей между узлами сети. В частности, исходя из этого определения, был разработан новый алгоритм стохастического блокмоделинга, который использует логистическую регрессию с поправкой на характеристики узлов. С помощью данного алгоритма авторы выявили сообщества в публикациях политической блогосферы США, а также обнаружили сообщества исходя из политической литературы, которую одни и те же пользователи чаще всего приобретают на Amazon

(Hopkins & Steurer, 2017).

Развитием байесовского подхода к обнаружению сообществ также является алгоритм неотрицательного матричного разложения (non-negative matrix factorization, NMF). Данный алгоритм позволяет получить вероятностные оценки принадлежности узлов к тому или иному сообществу, а также протестировать конкретное, смоделированное разложение на определенных бенчмарках. Алгоритм применялся, в числе прочего, для анализа сетей взаимодействия между нематодами *Caenorhabditis elegans*; джазовыми музыкантами; студентами одного из университетов в Facebook (Psorakis et al., 2010).

Подводя итог, мы можем выделить следующие преимущества Байесовского подхода к обнаружению сообществ:

1. байесовские методы эффективно обрабатывают разреженные данные, обеспечивая значимые результаты при ограниченных наблюдаемых взаимодействиях;
2. байесовские методы автоматически уравновешивают соответствие модели и ее сложность, предотвращая переобучение;
3. байесовский подход предлагает более широкие интерпретации и количественную оценку неопределенности с помощью целых апостериорных распределений;
4. байесовские методы превосходно отражают эволюцию сообщества в динамических сетях;
5. благодаря включению нескольких типов данных в байесовскую структуру повышается точность обнаружения сообщества, что дает всестороннее представление о поведении сети;
6. байесовский подход обеспечивает непротиворечивые результаты, согласованные с лежащими в их основе допущениями.

В заключение, применение байесовских методов в сетевом анализе позволяет лучше понять структуру сообщества и его эволюцию с течением времени, решая проблемы неопределенности, разреженности и временной динамики. Они являются ценными инструментами в сетевом анализе.

3 1.2 Акторно-ориентированные стохастические модели для изучения сетевой динамики и социального влияния

3.1 Введение

В контексте развития сетевого анализа исследования сетей в динамике становится все более значимыми для понимания сложных взаимосвязей. Осознавая эту потребность, исследователи прибегают к разработке новых методологий для анализа и построения сетей. В этом контексте нельзя не вспомнить про акторно-ориентированные стохастические модели (Stochastic Actor-Oriented Models, SAOMs), представляющие собой одно из наиболее развивающихся и перспективных средств анализа механизмов социального развития, взаимосвязей и эволюции различных сетей. В связи с этим ученые из различных областей, таких как социология, экономика, эпидемиология и коммуникационные исследования, первоочередно прибегают к использованию данного аналитического инструмента для понимания сложных взаимосвязей между акторами.

По сути, SAOM выступает в роли линзы, через которую исследователи могут расшифровать сетевую динамику, раскрывая глубинные процессы, определяющие эволюцию сети. Преодолевая

разрыв между наблюдаемым и ненаблюдаемым, модель предоставляет ценный инструмент для сетевых аналитиков, стремящихся выявить скрытые закономерности и механизмы, управляющие социальными взаимодействиями и сетевыми структурами.

В данной работе мы рассматриваем основные принципы и области применения этой методологии. Мы стремимся выяснить отличительные особенности SAOM и их значимость в области стохастического анализа сетей. Проведя сравнительный анализ SAOM и временных экспоненциальных моделей случайных графов (TERGM), мы подчеркнули сильные стороны и уникальный вклад SAOM в раскрытие динамики сетей. Наконец, мы также рассматриваем процесс работы алгоритмов, которые позволяют оценить качество подобных моделей, включая такие статистические показатели как оценка адекватности модели (Goodness of Fit). Кроме того, мы приводим релевантные примеры эмпирических работ, которые позволяют напрямую увидеть практическую значимость методологии в современных исследованиях и дальнейший потенциал для разработки сложных сетевых феноменов.

Цель: Провести сравнительный анализ методологий применения Акторно-ориентированных стохастических моделей (SAOMs) для изучения сетевых динамик и социального влияния.

Задачи: 1. Дать характеристики применения акторно-ориентированных стохастических моделей (Stochastic Actor Oriented Models, SAOM) как одного из направлений развития подходов к анализу динамических сетей. 2. Сравнить акторно-ориентированные стохастические модели (SAOMs) и темпоральные экспоненциальные модели случайных графов (TERGMs) для изучения динамических сетей. 3. Описать алгоритмы для оценки качества и адекватности акторно-ориентированных стохастических моделей. ## Изучение сетевой динамики: перспективы, ограничения и применения SAOM

Хотя SAOM все еще является развивающимся методом сетевого анализа, его уже успешно применили в различных областях: от анализа небольших сетей дружбы подростков до политологического анализа транснациональных союзов. Мы сделаем обзор некоторых из наиболее цитируемых работ в нескольких научных областях, а также предложим способы применения этого подхода в работе ANR-Lab. ### Применения

Дружба и влияние сверстников Применение SAM породило довольно много исследований на более маленьких выборках сетей дружбы, в которых исследовались реципрокность и гомофилия как каналы социального влияния на курение, потребление веществ, ожирение и т. д. в группах сверстников. Среди подобных исследований: эффекты пола на распределение индивидуальных характеристик в сетях дружбы (Van De Bunt et al., 1999); актуальные проблемы курения и употребления алкоголя через призму гомофилии: влияния сверстников (De La Haye et al., 2019) (Kiuru et al., 2010) (G. C. Huang et al., 2014) (Kiuru et al., 2010) (Kiuru et al., 2010) (Schaefer et al., 2012) и родительского примера (Mercken et al., 2013).

Кроме того, на пересечении медицинской социологии и сетевого анализа, существуют также исследования, применяющие SAOM, для выявления социального влияния на вероятность заболеваний, например, СДВГ (Aronson, 2016), подростковой депрессии (Zalk et al., 2010), образа жизни и ожирения (De La Haye et al., 2011). В этой области также существуют работы о социальном влиянии этнического самоопределения (Jugert et al., 2018), лидерских динамик (Mehra et al., 2009), религии (Kretschmer & Leszczensky, 2022) и владения оружием (Dijkstra et al., 2012) на дружбу и процесс выбора друзей.

Библиометрический анализ Другим перспективным направлением сетевого анализа социальных

сетей, в котором применяется модель Siena, является библиометрический анализ. Наши коллеги по ANR-Lab А. Ферлигой и Л. Кроннегер, в соавторстве с создателем моделей Т. Снейдером, провели впечатляющий анализ научного сообщества и динамику соавторства в Словении с 1996 по 2010 год (Ferligoj et al., 2015), а также провели дополнительную работу на лучших данных, указав на важность институциональных контекстов на среду работы ученых, его не-механическую природу (Kronegger et al., 2012).

Иновации в научных исследований, проанализированные с помощью SAOM, уделяют внимание гендерной гомофилии (Lungeanu & Contractor, 2015), а также сетевым динамикам более молодых областей науки (Vermond et al., 2022), роли административных ресурсов университета на паттерны коллaborации (Roesler & Broekel, 2017), коллаборациям между университетом и индустрией (H. Chen et al., 2022) и диффузии инноваций (Liang & Liu, 2018).

Политические науки Одной из наиболее заметных областей исследований, в которых применяется SAOM, является политология и исследования законодательства. Политические акторы и связи между ними оказались исключительно подходящими для этой модели и позволили провести широкий спектр исследований. Во-первых, это исследования по определению и изменению паттернов коллaborации среди законодателей (Ingold & Fischer, 2014), использующие влияния рисков/ресурсов на принятие решений (Berardo & Scholz, 2010).

В поле исследований также входят статьи по анализу дипломатических связей между странами (Kinne, 2014); международной кооперацией и работой международных союзов в связи с проблемами координации (Kinne, 2013) и предсказания будущего Европы как структуры транснациональной сети (Thurner & Binder, 2009).

Лонгитюдный подход к сетям также полезен при анализе распространения правил и законов через институты, регионы и страны (например, законы вокруг международной торговли (Mohrenberg, 2017), а также коэволюции доступа к диджитал-инструментам, демократии и торговых связей (Rhue & Sundararajan, 2014). Наконец, SOAM может быть применен для анализа политического действия (Redhead & Von Rueden, 2021) и исторического моделинга событий (Box-Steffensmeier & Jones, 2004).

3.2 Ограничения

SAOM обладает большими теоретическими основаниями, чем TERGM, и модель сочетает как социологические, так и статистические методы (Leifeld & Cranmer, 2019). В то же время теоретическая укорененность модели может выступать и ограничением. Для начала, идея о том, что любые изменения в сети акторы выполняют только последовательно, а не одновременно не позволяет использовать данные из имейл сообщений, электрических систем, очень больших сетей и других ситуаций, когда “социологическая рамка не соответствует реальности” или она не может быть полностью фальсифицирована (Leifeld & Cranmer, 2022). Из-за этого, наложение новых теоретических предположений на сеть невозможно без предварительной проверки базовых предположений SAOM.

Кроме того, этот метод предполагает, что каждый актор размышляет о своих действиях в одинаковой логике, что удобно для статистического упрощения расчетов, но не всегда соответствует модели с акторами, которые обладают разными классами и задачами (Ceoldo et al., 2023). При этом, нельзя отрицать, что хотя процесс и ускоряется, работа с моделью требует больших временных затрат (Т.

A. Snijders, 1996), а получение необходимых данных высокого качества более финансово требовательно (Steglich et al., 2010). ## Перспективы

Перед исследователями стоят еще много методологических проблем, в том числе работа в модели с коррелирующими рандомизированными эффектами (Ceoldo et al., 2023), и разработкой моделей с необнаруженной гетеронормативностью между акторами, которые позволят применить подход к более крупным сетям (T. A. Snijders, 2017).

Наконец, мы считаем, что SAOMs могут быть применены к нескольким направлениям исследований российского общества. До сих пор только в образовательной сфере были созданы литературные обзоры, направленные на выявления факторов академического успеха и поведения учеников с помощью SAOM (Докука, 2021),(Докука & Валеева, 2015). Мы предлагаем несколько потенциальных направлений развития.

Библиометрический анализ российского научного сообщества, которым занимается одна из исследовательских групп ANR-Lab уже использует лонгитюдные данные (A. Kim & Maltseva, 2021), (Matveeva & Ferligoj, 2020). Более того, в нескольких проектах мы также обладаем крупными датасетами библиометрических данных из Web of Science, которые могут быть в разрезе по отдельными направлениям или институциям проанализированы с помощью SAOM. Кроме того, кажется перспективным анализ открытых судебных данных, а также коллaborаций в законодательных инициативах и торговых международных договорах. В данном случае, наиболее сложным этапом работы был бы сбор и предобработка данных.## Оптимизация и оценка Для оценки SAOM используются различные статистические методы, наиболее распространенным из которых является метод моментов (MoM) или метод максимального правдоподобия (MLE).

Метод моментов (MoM) – это метод имитационного моделирования, используемый для оценки параметров в SAOM, подробно описанный Снайдерсом (Snijders, 2001). Он заключается в сравнении описательных статистик наблюдаемой сети со значениями, полученными в результате моделирования при разных значениях параметров. Целью является обнаружение значений гиперпараметров, минимизирующих разницу между наблюдаемой и моделируемой статистикой сети.

Сначала оценки параметров часто задаются произвольно, а затем, путем итерационного моделирования SAOM с различными наборами значений параметров, рассчитываются сводные статистики и сравниваются с соответствующими характеристиками наблюдаемых данных на основе функции расхождения, которая количественно оценивает разницу между наблюдаемой и моделируемой статистиками. Затем оценки параметров обновляются таким образом, чтобы минимизировать функцию расхождения. Функции расхождения, используемые в SAOM, могут оценивать такие свойства, как количество связей, транзитивность, распределение мер центральности, паттерны образования и распада связей, вклад характеристик акторов, специфические параметры диад и временная динамика. Выбор функции расхождения зависит от вопроса исследования, конкретной используемой SAOM и характеристик наблюдаемых сетевых данных.

Несмотря на концептуальную простоту MoM и возможность работы со сложными SAOM, а также гибкую спецификацию модели, подходящую для различных типов сетевых данных, ее использование сопряжено с определенными трудностями. Для больших сетей или сложных моделей MoM может быть вычислительно трудоемким, а также требует тщательной настройки алгоритмов оптимизации. Предпринимаются попытки повысить эффективность и расширить спектр использования MoM.

Например, развитие обобщенного метода моментов (GMoM) позволяет обогатить оценку временными данными, вводя в нее в качестве параметров статистику из различных временных моментов (Amati et al., n.d.). Однако этот метод, как предполагают авторы, не может стабильно превосходить традиционный MoM, в частности, из-за избыточности признаков, что препятствуют сходимости.

При оценке по методу максимального правдоподобия (MLE) необходимо найти такие значения параметров, при которых наблюдаемые данные наиболее вероятны в рамках данной модели. Параметры обновляются таким образом, чтобы максимизировать логарифм функции правдоподобия, выбранной исходя из характера сетевых данных и предположений об их эволюции. Функции правдоподобия разнообразны и подходят для различных типов данных (бинарных, непрерывных, мультиномиальных, событийных и т.д.). MLE дает оценки, которые асимптотически эффективны: увеличение размера выборки связано с ростом точности и уменьшением погрешности. Этот метод широко используется в статистике и может работать с различными спецификациями SAOM. Однако MLE может требовать больших вычислительных затрат, особенно для сложных SAOM. Сходимость к глобальному максимуму функции правдоподобия может быть не гарантирована, а процесс оценки может потребовать тщательной инициализации.

Более того, определение полной функции правдоподобия может стать сложной задачей из-за зависимостей между диадами в процессе эволюции сети. Другими словами, одна диада может влиять на поведение других диад в сети. Дальнейшая адаптация MLE – оценка максимального псевдоправдоподобия (MPLE) – решает эту проблему путем максимизации функции псевдоправдоподобия: вероятность для каждой диады вычисляется на основе наблюдаемого состояния этой диады и состояний соседних диад, заданных SAOM. MPLE менее требователен к вычислениям по сравнению с заданием полного совместного правдоподобия для сложных сетей, поскольку требует моделирования только условных связей между диадами, однако, несмотря на вычислительные преимущества, он не всегда может давать асимптотически эффективные оценки. Бесаг утверждает, что максимальная оценка псевдовероятности отражает “локальную” (пространственную) информацию о соседях, в отличие от оценки максимального правдоподобия, которая отражает “глобальную” информацию о соседях (Besag, 1986). Более того, Снайдерс утверждает, что результаты исследований показывают, что обычно используемые модели случайных графов имеют скорее глобальную, чем локальную структуру, что в конечном итоге приводит к плохим статистическим свойствам MPLE-оценок (Snijders, 2001). Далее он предполагает, что адаптация спецификации модели, например, подходы, основанные на соседстве, с ограничениями на возможные связи между соседями (Pattison & Robins, 2002), подходы, основанные на латентном пространстве [Nowicki & Snijders (2001)](Hoff et al., 2002)(Schweinberger & Snijders, 2003), обладают большими возможностями для решения этой проблемы.

Дальнейшая валидация модели, а также сравнение SAOM с различными характеристиками осуществляется с помощью таких тестов качества, как:

1. *Goodness-of-fit (GOF)* тесты оценивают, насколько хорошо SAOM воспроизводит наблюдаемые сетевые данные, сравнивая статистики сетей. Тесты GOF могут использовать имитационное тестирование или методы бутстрепа, но страдают от переобучения, плохой генерализации и чувствительности к размеру выборки. Lospinoso и Snijders (Lospinoso & Snijders, 2019) предлагают в качестве решения этой проблемы вспомогательные статистики (например, характеристики триад, транзитивность), не включенные в модель в явном виде. Они моделируют расстояние Махalanобиса между вектором вспомогательной статистики

и оценкой модели с помощью симуляций Монте-Карло, повторно используя их из вычислений МОМ в SAOM. Вводя собственный принцип минимального описания модели (MMD), они анализируют влияние вспомогательных статистик на GOF, добиваясь баланса между сложностью модели и ее описательной способностью.

2. *Критерии отбора моделей*, такие как информационный критерий Акаике (AIC) или Байесовский информационный критерий (BIC), предлагают количественную сравнительную меру для SAOM: чем меньше значения, тем лучше модель подходит под данные. AIC совмещает оценку соответствия модели данным и штраф за сложность модели:

$$AIC = -2 * \log(likelihood) + 2 * \text{number of model parameters}$$

BIC штрафует сложность модель сильнее, чем AIC. Этот критерий рассчитывается как:

$$BIC = -2 * \log(likelihood) + \log(samplesize) * \text{number of model parameters} *$$

3. *Тесты на сходимость* позволяют определить, сходится ли алгоритм оценки, используемый для подгонки SAOM, к стабильным оценкам параметров. Визуальное изучение графиков параметров модели может помочь выявить проблемы сходимости. В идеале графики должны стабилизироваться по мере выполнения оценки. В отношении несождшихся моделей от интерпретации следует отказаться.

- t-ratio является количественной мерой степени отклонения смоделированной статистики от целевой в среднем.
- Чем меньше t-ratio, тем лучше сходимость. Как правило, t-ratio менее 0,1 считается показателем хорошей сходимости.
- Чтобы считать модель сходящейся, общее максимальное t-ratio сходимости не должно превышать 0,25.
- В тех случаях, когда модель не сходится, рекомендуется повторно провести анализ с использованием опции “prevAns”.

3.3 Вырождение

Другой проблемой, возникающей при оценке SAOM, является вырождение. В работах Штрауса (Strauss, 1986), Снайдерса (Snijders, n.d.) и Хэндкока (Handcock, n.d.) показано, что экспоненциальные модели случайных графов могут быть почти вырожденными, и то же самое может иметь место для SAOM в перспективе отсутствующих временных лимитов (хотя на практике время обычно ограничено). Вырожденность в SAOM возникает, когда несколько наборов значений параметров приводят к одной и той же наблюдаемой структуре сети. Это может затруднить оценку “истинных” или наиболее точных значений параметров и точное определение механизма, управляющего эволюцией социальной сети. Проблема вырождения представляется особенно опасной в сетевом анализе, поскольку сходимость к целевому распределению становится еще более медленной и менее устойчивой в мультиаддитивных сетях, где типичные алгоритмы, обновляющие отдельные связи или структурные элементы, имеют чрезвычайно малую вероятность перемещения между модальными областями (Snijders, n.d.).

Для решения проблемы вырождения SAOM исследователи обычно используют различные стратегии

(Handcock, n.d.), такие как проверка робастности, сравнение различных инициализаций модели, предоставление дополнительных данных для обучения.

В этих практиках также отдается предпочтение байесовскому фреймворку (Nowicki & Snijders, 2001). Помимо уменьшения вырождения модели, он облегчает распространение неопределенности параметров на окончательный вывод и позволяет учитывать предварительные знания экспертов, если они существуют (Handcock, n.d.). Кроме того, Лоспиносо и др. (Lospinosa & Snijders, 2019) предполагают, что введение в модель временной неоднородности может снять проблему вырождения. Временная неоднородность добавляет временное измерение в модель, делая ее более способной улавливать и различать различные состояния сети в разные моменты времени, что, в свою очередь, приводит к улучшению предсказательной силы и качества подгонки, а также позволяет вводить временные ограничения и включать внешние события в качестве параметров модели. Проблема вырождения в бимодальных сетях может быть минимизирована путем адаптации методов оценки, как это было предложено в работе (Snijders, n.d.).

3.4 1.5 Применение современных методов машинного обучения для предсказания связей в социальных сетях

С каждым годом компьютерные технологии все глубже интегрируются в различные научные дисциплины. Использование искусственного интеллекта, нейронных сетей в таких далеких от математики областях, как психология, филология, литературоведение, растениеводство и т.д. становится обыденностью. В данном параграфе описаны возможности использования контролируемого (supervised) и неконтролируемого (un-supervised) машинного обучения (ML) для предсказания связей (link prediction) в социальных сетях.

В разделе представлен сравнительный анализ литературы в парадигме экспертной методологии, которая предполагает качественную стратегию экспертного отбора источников на основе анализа литературы по теме исследования, что позволяет преодолеть барьеры, связанные с усложнением, специализацией и фрагментацией научных областей, а также с ростом публикационной активности исследователей, характерным для развития современной науки. Использовалась литература по применению машинного обучения в области социального сетевого анализа (social network analysis – SNA), причем более подробно рассматривалось использование контролируемых и неконтролируемых методов машинного обучения для прогнозирования связей в социальных сетях.

В начале каждого исследования ученым необходимо собрать данные для своего проекта. Но редко данные поступают в виде удобном для обработки. Часто они не структурированы, загрязнены шумом и ненужной информацией. А если сложных данных слишком много, то на их ручную обработку уходит много ресурсов и времени, в течение которого исследование может стать неактуальным. Поэтому уже на этом этапе машинное обучение очень помогает ученым и значительно ускоряет их работу.

Рассмотрим пример, когда нейронные сети предварительно обработали данные в исследовании (Haupt et al., 2021). Ученые собрали 34672 твита с 1 по 20 апреля 2020 года с необходимыми темами и ключевыми словами. Для дальнейшего анализа исследователям необходимо было отсортировать сообщения, относящиеся к теме исследования, и те, которые просто содержали подходящий набор слов, а затем определить, какие из отсортированных сообщений отражают позицию противников движения, а какие – последователей. Для этого они создали программу, использующую неконтролируемый подход к машинному обучению, в котором применяются тематическое моделирование и обработка

естественного языка (natural language processing – NLP). Эта технология предназначена для выявления закономерностей в данных и обобщения содержания твитов в отдельные темы с высокой степенью корреляции. Исследователи использовали модель Biterm Topic Model (BTM), которая выявляет закономерности в коротких текстах. Этот метод кластеризации тем моделирует совпадение слов, что повышает производительность для документов с небольшим разреженным текстом, таких как твиты. После проведенных манипуляций ученые получили актуальные данные, рассортированные по кластерам, что позволило им в дальнейшем провести сетевой анализ и обнаружить интересные закономерности структурирования и влияния общественного мнения.

В нашем мире существует множество вещей и аспектов жизни, по которым людей можно разделить на условные группы, например: интересы и увлечения, сообщества, взгляды и принадлежность. Таким образом, каждый человек в социальной сети может быть охарактеризован набором меток. Однако в реальной работе маркировка занимает много времени и является дорогостоящей, поэтому люди маркируются либо частично, либо достаточно редко. Задача классификации узлов состоит в том, чтобы с учетом структуры сети предсказать метки немаркированных узлов, используя их связи с меченными узлами. Как утверждают Tan и соавторы (Tan et al., 2019): «Существующие методы можно разделить на две категории, например, методы, основанные на случайному прохождении, и методы, основанные на извлечении признаков». Метод random walk направлен на распространение меток, а механизм второго метода – на извлечение характеристик узла с использованием информации и статистики, окружающей его.

Ранее работа по классификации сетей строилась следующим образом: сначала извлекались характеристики узлов сети с помощью методов обучения представлению, а затем использовались классификаторы машинного обучения (например, машина опорных векторов, наивный байесовский алгоритмический классификатор и логистическая регрессия для прогнозирования). Сейчас ученые отходят от разделения этапов и разрабатывают структуру, позволяющую объединить эти две задачи таким образом, чтобы отличительная информация, полученная из меток, способствовала обучению встраиванию сети.

Кластеризация узлов подразумевает разбиение сети на кластеры или подграфы таким образом, что узлы одного кластера более похожи друг на друга, чем узлы других кластеров. В социальных сетях кластеры можно широко наблюдать в виде групп людей с общими интересами или общих сообществ. Ранее основные работы по кластеризации были направлены на кластеризацию сетей с различными показателями близости или силы связи между узлами. Например, на минимизацию количества связей между кластерами с учетом максимизации количества связей внутри кластера. В настоящее время ученые пытаются использовать методы представления сетей для кластеризации узлов. Из этих методов можно выделить те, в которых Tan и соавторы (Tan et al., 2019) рассматривают «встраивание и кластеризацию как несвязанные задачи, где они сначала встраивают узлы в низкоразмерные векторы, а затем применяют традиционные алгоритмы кластеризации для создания кластеров».

Прогнозирование событий в структурах социальных сетей остается важной исследовательской задачей для SNA. Как утверждает Molokwu (Molokwu, 2021), «это предполагает понимание внутренних закономерностей связей, сохраняющих заданную структуру социальной сети, на основе изучения ряда структурных свойств, вычисляемых для составляющих ее социальных единиц в пространстве и времени». Часто проблема прогнозирования осложняется тем, что данные о действиях узлов социальной сети

скудны или недостаточны.

Еще одним применением машинного обучения в SNA является технология Trend and Pattern Analysis. Эта технология представляет собой модель, обученную на проверенных данных и применяемую к целевым данным. Она позволяет отслеживать и прогнозировать различные результаты действий. Особенно широкое распространение этот метод получил во время пандемии COVID-19. Многие группы ученых изучали это явление с разных сторон, в том числе с помощью Trend and Pattern Analysis. Например, анализ последствий пандемии в нескольких канадских штатах позволил выявить закономерность и получить возможность прогнозировать потребление средств индивидуальной защиты и спрос на них в других географических точках.

В социальных сетях часто встречается недостающая информация: между людьми (узлами) в сети нет связей несмотря на то, что они существуют в реальной жизни. Такие сети являются неполными. Предсказание связей позволяет на основе имеющихся данных об эволюции и структуре сети сделать выводы о ее дальнейшей динамике, а также предсказать будущие связи между узлами. Эта задача очень популярна в настоящее время. Поэтому существует множество способов решения этой задачи с помощью машинного обучения:

- предсказание связей с помощью подхода Strength of Ties,
- предсказание связей с помощью подхода Graph Embeddings,
- предсказание связей с помощью подхода Graph Embeddings на основе матричной факторизации,
- предсказание связей с помощью подхода Graph Embeddings на основе Random Walk(s),
- предсказание связей с помощью подхода Graph Embeddings на основе нейронных сетей (Molokwu, 2021).

В современном мире процесс установления отношений между социальными субъектами глубоко укоренен в существующих социальных сетях. Социальные сети стали движущим фактором изменения способа построения социальных взаимодействий, позволяя ускорить создание связей между социальными акторами и сделать поток информации практически безграничным. Вовлеченность социальных субъектов в различные коммуникативные процессы несет в себе ценные данные, которые могут быть использованы для достижения целей в самых разных сферах. Использование зависит от свойств связей между акторами – их прочности, взаимности, возможности будущих связей (P. Wang et al., 2014) и т.д.

Например, некоторые интернет-сервисы используют алгоритмы рекомендательных систем, основанные на взаимодействии пользователя с объектом и пользователем с пользователем, чтобы улучшить пользовательский опыт и предложить лучшие контентные решения (Z. Huang et al., 2005). Эффективное применение анализа сетевых связей также часто встречается в научных работах, посвященных сетям соавторства (Cho & Yu, 2018; Liben-Nowell & Kleinberg, 2007), в результате чего эта область является одной из наиболее процветающих в последние годы, поскольку способствует развитию эффективной системы совместной работы (Chuan et al., 2018; Z. Huang et al., 2005). К настоящему времени как в научных, так и в практических кругах предсказание связей рассматривается как перспективная область исследований, поскольку несет в себе огромную многоцелевую ценность.

Предсказание связей – это область исследований, которая занимается вопросами прогнозирования социального поведения акторов в социальных сетях (Daud et al., 2020). Социальные сети – это динамические образования (Leguia et al., 2019), которые развиваются и изменяются с течением времени, при этом связи исчезают и возникают в силу определенных свойств, связанных с узлами (акторами), структурой группы и т.д. (Hasan & Zaki, 2011). В предыдущие годы появилось множество научных работ, посвященных острым вопросам предсказания связей, касающихся метрик, используемых для целей предсказания (Cho & Yu, 2018; Mohan et al., 2017), обсуждающих задачу предсказания в различных типах сетей (Gou & Wu, 2021; Nasiri et al., 2022). Наконец, появились обзоры (Daud et al., 2020; Hasan & Zaki, 2011; P. Wang et al., 2014), в которых рассматривается вопрос о различных подходах к предсказанию связей. Цель данной части обзора – погрузиться в различные подходы к предсказанию связей и сосредоточиться в первую очередь на алгоритмических подходах машинного обучения. Мотивация данного обзора кроется в стремительном развитии технологий и социальных сетей. Мы также воспользуемся таксономией, предложенной для лучшей систематизации (Daud et al., 2020), которая помогает структурировать обзор.

Одной из наиболее простых в применении и традиционных групп подходов являются подходы, основанные на сходстве (Daud et al., 2020; Moradabadi & Meybodi, 2018). Методы, основанные на сходстве, исследуют структурную эквивалентность пары узлов, которая затем используется для оценки вероятности будущих связей между этой парой. Высокая степень сходства, соответственно, приводит к повышению вероятности возникновения связей в будущем. При расчетах на основе сходства используются в основном две точки зрения, касающиеся структурного уровня анализа. В подходах, основанных на сходстве, принято использовать локальные и глобальные индексы, а также их современную модификацию, называемую квазилокальными индексами. В основном разница заключается в расстоянии пути до ближайшего узла: при подходе с использованием локальных индексов узлы считаются соседними, если расстояние пути меньше двух (Daud et al., 2020; Lü & Zhou, 2011), а при подходе с использованием глобальных индексов, наоборот, интересны случаи, когда расстояние пути больше двух, что является необходимым условием для того, чтобы узел считался соседним. Квазилокальные подходы, напротив, используют дополнительную топологическую информацию, но в большей степени на локальном уровне (Daud et al., 2020; X. Liu et al., 2018), учитывают больше информации о соседних узлах (Lü & Zhou, 2011) и предсказывают другие возможности для связей. Квазилокальный подход развивается и по сей день, при этом вносятся изменения и улучшения в вычислительный алгоритм и метрики, которые часто являются модификациями традиционных метрик локальных индексов (X. Liu et al., 2018; Özcan & Öğüdücü, 2016; Srilatha & Manjula, 2016).

Часто для экономии вычислительного времени и эффективности используются методы, основанные на сходстве, и в этом случае индексы локального сходства оказываются как нельзя кстати, поскольку позволяют одновременно эффективно использовать ресурсы и иметь высокие прогностические характеристики. Однако из-за того, что метрики локального подобия анализируют только пути ближайших соседей, возникает дефицит информации (Hasan & Zaki, 2011), так как упускаются потенциальные связи. Основные работы в области локально-индексных подходов ведутся в области совершенствования метрик вычисления сходства и алгоритмов вычисления сходства (Daud et al., 2020; Wu et al., 2018). Подходы на основе глобальных индексов используют больше информации и раскрывают больше структуры, однако они крайне неэффективны с точки зрения затрат времени и энергии,

поскольку анализируют высокоразмерные связи сетей, что делает их не лучшим выбором для задач предсказания связей. Решение проблемы заключается в снижении размерности и взвешивании сетей, что позволит сократить время вычислений и повысить эффективность прогнозирования (Coskun & Koyutürk, 2015; Muniz et al., 2018). Возможности для совершенствования есть и у квазилокальных подходов, которые сильно зависят от особенностей данных и метрик расчета, выполняемых исследовательской группой. Поэтому существует множество исследований, посвященных устойчивости и робастности квазилокальных инструментов (Y. Liu et al., 2016; Özcan & Öğüdücü, 2016; S. Wang et al., 2017). В целом методы, основанные на подобии, эффективны и зависят от конкретного случая, поэтому исследователям следует внимательно относиться к условиям и задачам исследования и использовать описанные выше подходы.

Следующая группа методов, которые обычно используются для предсказания связей, называется вероятностными. Вероятностные методы используют статистическое моделирование вероятности в соответствии со структурой и размерностью существующей сети. Каждая пара узлов, еще не имеющих связи, включается в модель, которая вычисляет математическую статистическую меру в соответствии с параметрами сети. После вычислений используются гипотезы, которые измеряют степень вероятности того, что эти два узла будут иметь связи в будущем. Такая модель позволяет вписать в предсказание большинство параметров наблюдаемых данных, что делает эту группу подходов более гибкой и универсальной (Farasat et al., 2015). В этой группе методов исключительно важно обращать внимание на тип сети, так как в марковских и байесовских сетях используются разные методы и расчеты вероятностей (Farasat et al., 2015). Следует внимательно относиться к типу переменных, взаимности связей, типу взаимных связей и т.д. (Koller & Friedman, 2010). Вероятностные модели также имеют возможность работать с множеством измерений и проводить многомерный анализ, однако процесс вычисления структуры и параметров может оказаться непомерно сложным. Используя разграничение, приведенное в работе (Daud et al., 2020), мы также можем структурировать наш обзор, опираясь на четыре типа подходов в группе вероятностных методов:

1. Модель тензорной факторизации вероятностей (Cheng et al., 2012; J. Wang et al., 2021). Как указано в (Daud et al., 2020), они являются логическими расширениями моделей Probability Matrix Factorization, используемых для решения задачи тензорной факторизации. Развитие этого подхода можно увидеть в работах (X. Chen et al., 2019; Rettinger et al., 2012; W. Zhang et al., 2019).
2. Модель вероятностных латентных переменных (Hoff, 2009; Li et al., 2011). Эти модели развивают идею низкоранговых аппроксимаций для повышения точности предсказания и анализа блоков сетей со схожими свойствами.
3. Марковская модель. Эта группа стохастических моделей показала свою эффективность в применении к динамическим сетям, поскольку позволяет визуализировать эволюцию сети как процесс (Daud et al., 2020). Она также показала более высокую точность предсказания по сравнению с существующими моделями предсказания динамических связей. Из-за большого количества параметров вычисление этой модели занимает много времени, а набор параметров создает дополнительные препятствия для создания модели.
4. Моделирование меток связей (Agrawal et al., 2013). Эта группа моделей применима только к сетям с

подписями и решает задачу предсказания меток связей. Подписанные связи несут дополнительную информацию, уникальную для наблюдаемой связи, но в основном связи можно разделить на две группы по характеру их связи – положительные или отрицательные связи. Положительные связи представляют собой отношения взаимного доверия, близости и общего одобрения. Напротив, негативные связи обозначают антагонистические отношения, характеризующиеся высокой степенью неодобрения и холдности. Это повышает объяснительную силу модели, но одновременно увеличивает время вычислений, поскольку характер связей добавляет в сеть еще одно измерение.

В целом, вероятностные модели обладают более высокой предсказательной точностью и способны нести гораздо больше полезной для анализа информации, однако следует быть осторожным, поскольку дополнительные параметры делают модель более тяжелой и менее устойчивой, поэтому экономия вычислительного времени и ресурсов крайне необходима.

Еще одним подходом, обеспечивающим высокую эффективность прогнозирования, является набор алгоритмических подходов. Они широко представлены в литературе и продолжают развиваться по сей день, поскольку обеспечивают скорость и эффективность. Одним из их преимуществ по сравнению с подходами подобия и вероятностными подходами является возможность использования дополнительной информации из сети, а также использование дополнительной информации, которая может как-то повлиять на формирование связей (Daud et al., 2020).

Существуют исследования, в которых используются данные о структуре сообществ, что значительно повышает точность предсказания; одной из наиболее востребованных для получения результатов информации являются данные о поведении пользователей, которые могут нести в себе многоцелевую информацию. Чтобы более четко организовать обзор этой группы алгоритмов и сфокусироваться на одном конкретном подходе (а именно на машинном обучении), сначала целесообразно привести современные условия и методы, помимо машинного обучения. Подход машинного обучения будет рассмотрен позже и более подробно. Мы также используем разграничение, приведенное в работе (Daud et al., 2020).

Существует три группы подходов – метаэвристические, матричной факторизации и машинного обучения. Метаэвристическая группа методов содержит рекомендации и приемы, которые помогут исследователю применить наилучший вариант эвристического метода оптимизации. Эта группа была широко использована в исследовании (Daud et al., 2020), показав свою высокую эффективность в задачах, где анализировались большие сети, поскольку требовала меньшего времени и вычислительной эффективности по сравнению с другими алгоритмами. Исследования предполагают дальнейшее развитие этих подходов, поскольку они тестируются на большем количестве информации и характеристик сети с целью повышения их точности. Факторизация матриц – группа алгоритмов, использующих коллаборативную фильтрацию (Daud et al., 2020), которая принимает в качестве результата предсказания произведение, образующееся после слияния двух матриц меньшей размерности. Однако эта методика не столь надежна и стабильна, поскольку зависит от данных – если есть шум, выбросы, смещение или экстремальная дисперсия, то предсказание не будет стабильным и точным. Решить эту проблему достаточно просто, так как данная группа подходов является универсальной и гибкой, поэтому во многих работах уже рекомендуется использовать методы мешков, которые хотя и увеличивают время вычислений, но при этом значительно повышают точность прогнозирования. Данный подход также

применяется, когда ставится задача анализа неявных признаков динамической сети. Машинное обучение сочетает в себе достижения предыдущих алгоритмов, а также позволяет сэкономить гораздо больше вычислительного времени и усилий.

Контролируемое обучение – это одна из ветвей машинного обучения, известная как Supervised Machine Learning (SML). Этот метод отличается от других тем, что для обучения алгоритмов классификации данных или точного прогнозирования результатов используются полностью помеченные наборы данных. Наличие полностью маркированного набора данных означает, что каждый пример в обучающем наборе имеет правильный ответ, и цель алгоритма – получить этот ответ. Таким образом, помеченный набор данных с фотографиями фруктов позволит обучить нейронную сеть с фотографиями яблок, груш, бананов и т.д. Когда сеть получает новую фотографию фрукта, она сравнивает ее с примерами из обучающего набора данных, чтобы предсказать ответ.

Существует множество алгоритмов и вычислительных методик контролируемого обучения. Можно выделить несколько часто используемых методов: нейронные сети, Naive Bayes, линейная и логистическая регрессия, support vector machines (SVM), K-nearest neighbor, Random forest.

Контролируемое обучение имеет как преимущества, так и недостатки. К числу преимуществ контролируемого обучения относятся:

- простота обучения: поскольку алгоритм обучается на помеченных данных, обучать модель гораздо проще. Кроме того, этот процесс прост в случае реализации и понимания процессов;
- ясность данных: Каждый алгоритм контролируемого обучения использует помеченные данные, поэтому входные данные должны быть отнесены к определенным категориям, что позволяет уменьшить количество ошибок при работе с этими данными;
- прогнозы, как правило, более точны и надежны, если имеется достаточное количество соответствующих данных.

Несмотря на то, что контролируемое обучение имеет ряд преимуществ, при построении моделей такого типа возникают определенные трудности:

- при работе с большими и сложными наборами данных алгоритмы контролируемого обучения могут быть сравнительно более трудоемкими и вычислительно дорогими;
- при работе с большими наборами данных возрастает вероятность человеческой ошибки, приводящей к неправильному обучению алгоритмов;
- для контролируемого обучения необходимы помеченные данные, то есть данные должны быть классифицированы по определенным категориям, прежде чем алгоритм сможет на них обучаться.

Неконтролируемое обучение как метод автоматизированной обработки данных берет свое начало с перцептрона, построенного в 1958 году Фрэнком Розенблаттом. Перцептрон классифицировал примитивные изображения, используя солнечные (фотоэлектрические) элементы. Однако, пройдя значительный путь эволюции, сегодня алгоритмы неконтролируемого обучения стали действительно мощным инструментом. Например, неконтролируемое обучение может быть использовано для выравнивания графов знаний или построения вкраплений графов.

Основное преимущество бесконтрольного обучения заключается в том, что обучающий набор данных не обязательно должен быть помечен, т.е. для обучения сети не нужно давать правильные ответы или решения. В случае отсутствия помеченных данных бесподчиненное обучение часто оказывается более дешевым и быстрым решением, чем создание помеченного обучающего набора данных.

Для предсказания связей в социальных сетях обучение без контроля впервые было использовано в 2007 году Либен-Ноуэллом и Клейнбергом. Они изучали временные соавторские сети и использовали граф, соответствующий более раннему состоянию сети, для предсказания новых связей в сети, соответствующих более позднему периоду времени.

Большинство алгоритмов предсказания связей без наблюдения используют сходство между узлами для предсказания того, должна ли быть сформирована связь. Ниже приведены некоторые популярные методы вычисления этого сходства:

- общие соседи (Common Neighbors, CN) - более высокая вероятность предсказания ребер между узлами с большим числом общих соседей;
- алгоритм Jaccard (Jac) - зависит от количества общих и разных соседей у двух узлов;
- алгоритм Лейхта-Холма-Ньюмана (LHN) - сравнивает реальное количество общих соседей с ожидаемым количеством общих соседей;
- алгоритм Адамика-Адара (AA) - также дает более высокую вероятность предсказания ребер между узлами с большим числом общих соседей;
- алгоритм Local Path (LP) - этот алгоритм также учитывает 2- и 3-хоповых соседей.

Статистика является широко используемым инструментом для предсказания связей, например, вероятностная мягкая логика (PSL) (Getoor et al., 2002) и марковские логические сети (MLN) (Richardson & Domingos, 2006). Неподконтрольный механизм, использующий статистику, был реализован Куо и др (Kuo et al., 2013). Они работали с социальными онлайновыми сетями и анонимными отзывами пользователей. Вопрос, который задают авторы, заключается в следующем: «можем ли мы предсказать носителя мнения в гетерогенной социальной сети без каких-либо помеченных данных?».

Их алгоритм получил название «Factor Graph Model with Aggregative Statistics (FGM-AS)». В его основе лежат три слоя: “Кандидат”, “Атрибут” и “Счет”. Слой “Кандидат” - это слой с парами случайных вершин, которые потенциально могут иметь общее ребро. Слой “Атрибуты” содержит атрибутивную информацию о кандидатах. Слой “Count” кодирует агрегированную статистику кандидатов. Таким образом, исследователи используют три типа функций: Функции “атрибут-кандидат”, “кандидат-кандидат” и “кандидат-счет”.

Предсказание связей в социальных сетях часто является задачей прогнозирования временных изменений в графе. Соавторские сети, с которых началась история ненаблюданного предсказания связей (Liben-Nowell & Kleinberg, 2007), также изучались Мунисом и др (Muniz et al., 2018) во временной перспективе. Они объединили контекстную, временную и топологическую информацию для предсказания связей в соавторской сети.

Многие из современных механизмов предсказания связей без наблюдения используют ту же идею, что и Либен-Ноуэлл и Клейнберг (Liben-Nowell & Kleinberg, 2007): использование старых ссылок в качестве

обучающего множества и новых ссылок в качестве тестирующего множества. Однако при этом важно, какой тип вкраплений графа используется. Поэтому данная идея может быть очень удобно реализована с помощью CTDNE - Continuous-Time Dynamic Network Embeddings (Nguyen et al., 2018). Этот алгоритм обучает сеть динамически, внедряя временную информацию во вкрапления графов. Это делает его идеальным для предсказания временных связей в социальных сетях.

Одной из ключевых идей CTDNE является временное случайное блуждание: временно близкие ребра имеют больше шансов быть связанными. Игуен с соавторами (Nguyen et al., 2018) сообщают о среднем выигрыше в качестве предсказания связей в темпоральных графах на 11,9% по сравнению с другими алгоритмами встраивания в сеть DeepWalk, Node2Vec и LINE. Однако эти результаты могут быть сомнительными: DeepWalk, Node2Vec и LINE были представлены более чем за 3 года до CTDNE.

В работе «Towards Fast Evaluation of Unsupervised Link Prediction by Random Sampling Unobserved Links» Ванг и соавторы (J. Wang et al., 2021) рассматривают проблему оценки предсказания связей без наблюдения. Основная проблема оценки ненаблюдавшегося предсказания связей заключается в том, что ненаблюдавшихся потенциально возможных связей гораздо больше, чем наблюдаемых. Этот дисбаланс создает трудности для оценки, так как «нереально количественно оценить вероятность существования».

В качестве решения они предлагают выбирать для тестирования только некоторые из ненаблюдавшихся ребер, а не тестировать модели на всех связях сети. Этот метод позволяет быстрее оценивать модели и приводит к значительной стабильности. Однако авторы предостерегают читателей от использования этого метода на небольших сетях, поскольку он может быть рискованным и приводить к худшим результатам.

В рамках задачи изучения контролируемого и неконтролируемого обучения для предсказания связей в социальных сетях рассмотрены возможности использования машинного обучения для решения задач в области сетевого анализа. Описаны особенности задачи предсказания образования связей (предсказания связей) в сетевом анализе для понимания и объяснения процессов, управляющих социальными взаимодействиями. Проведено сравнение и выявлены особенности использования контролируемых и неконтролируемых методов машинного обучения для предсказания связей в социальных сетях. Приведены примеры применения контролируемого и неконтролируемого машинного обучения для предсказания связей в социальных сетях в области социальных наук. Машинное обучение может применяться на любом этапе анализа социальных сетей: предварительная обработка и обработка данных, классификация узлов, кластеризация, анализ на основе событий, анализ тенденций и предсказание связей.

4 Глава 2

See readings here Batagelj & Mrvar (n.d.)

Вот цитирование ([larranaga2013?](#)) и без скобок ([larranaga2013?](#)) — и сноска³. А вот пример ссылки на часть текста sec. 4.1.

4.1 Графики

Как отмечает Aoki (2007, p. 131):

³Внутренняя сноска

Пример цитаты. *A это пример курсива.*

4.1.1 Название подсекции

Заключение

5 Изучение гендерной специфики деструкторов руководителей крупных российских компаний

Марина Буракова (КСА, ФСН), Иван Климов

5.1 Введение

Стандартом социологических исследований считается проведение «пилотажа» - тестирования инструментария перед запуском «полевого» этапа. Однако на практике эта процедура реализуется нечасто: редко когда в публикациях и отчетах можно встретить хотя бы упоминание о действиях на предварительном этапе. Да и если посмотреть издающиеся учебники, пилотаж описывается не очень подробно.

«В многоплановых аналитических исследованиях, где применяется сложный методический инструментарий, может возникнуть потребность в двух или трех пилотажах. Их проведение целесообразно поручить разработчикам опросника, которые способны уловить мельчайшие недостатки тех или иных вопросов, обнаружить психологические барьеры их восприятия. В процессе пилотажа фиксируются реакция респондента на каждый вопрос, его замечания. Если он затрудняется ответить, нужно попытаться выяснить причину. В случае непонимания какого-то вопроса, последний формулируется по-новому» (Горшков & Шереги, 2011).

«Пилотажные исследования могут проводиться в два этапа: предварительная проба и генеральный пилотаж. На первом проверяются отдельные элементы методики. На втором проверяется вся процедура сбора информации. Это «разведка боем», цель которой — проверка всей полевой процедуры» (Таворкин Е. П., 2021) [Таворкин, 2009, С.11].

Существует множество способов предварительного тестирования вопросов анкеты: экспертные оценки, опрос интервьюеров, фокус-группы, эксперименты с альтернативными вариантами формулировок (Mohorko & Hlebec, 2016). Начиная с 1980 годов одними из самых широко распространенных становятся методы когнитивного интервью. Когнитивное интервьюирование выступает в качестве одной из ключевых тем Конференции разработки, оценки и тестирования дизайна анкеты в 2002 и 2016 годах (*International Conference on Questionnaire Design, Development, Evaluation and Testing (QDET2)*, n.d.). Бюро переписи населения США (Census Bureau), Национальный центр статистики здравоохранения США (National Center for Health Statistics), в России - ФОМ, ВЦИОМ и многие другие организации включают процедуру когнитивного интервьюирования в стандартный процесс разработки опросного инструментария (Beatty & Willis, 2007). Проблема, тем не менее, заключается в том, что в отчетах по исследованиям эти процедуры и их результаты описываются редко, что приводит к стагнации в развитии метода и подхода в секторе русскоязычной исследовательской литературы. Список работ о практике проведения когнитивного интервью немногочисленный. Основным источником, на который ссылаются большинство авторов, является работа Д. М. Рогозина (Рогозин, Д. М., 2002). Очевидна необходимость обзора основных теоретических парадигм когнитивного интервью, сравнения методологических преимуществ и ограничений каждого подхода, оценка возможностей КИ для задач не только тестирования опросного инструмента, но также и для других распространенных исследовательских задач, а также ознакомление российских исследователей с практикой проведения КИ.

5.2 Цели и задачи исследования

Цель: тематизировать основные направления в использовании и развитии методики КИ, оценить эвристические возможности КИ в некоторых актуальных направлениях методических исследований.

Задачи исследования: 1. На основе анализа отечественной и зарубежной литературы определить возможности и методические ограничения основных теоретических подходов проведения когнитивного интервью. 2. Определить возможности и ограничения методологии КИ для задачи культурной адаптации измерительных методик, определить возможности для корректировки метода когнитивного интервью применительно к задаче культурной адаптации опросного инструментария. 3. Описать ограничения и возможности работы с результатами когнитивного интервью на основе сетевого анализа и работы с неструктурированными массивами текстовых данных.

5.3 Практическая значимость исследования

Практическая значимость работы – в ее направленности на практические потребности исследователей, занимающихся разработкой опросного инструмента и работающих с переводными методиками. Мы не только провели систематизацию подходов и методических разработок в области когнитивного интервью. Также мы провели серию экспериментов по использованию этого подхода для задач кросс-культурной адаптации опросных методик (на примере одного психометрического теста). Наш следующий шаг – разобраться в вопросе о том, как анализировать результаты КИ и провести апробацию модели анализа неструктурированной текстовой информации на основе сетевого анализа.

5.4 Обзор литературы

Описание research gap: Использование в исследовательской практике переводного инструментария – довольно распространенная практика. Хорошо известны приемы, которыми пользуются исследователи для валидизации перевода (прямой и обратный перевод, перевод несколькими переводчиками с разной областью экспертизы, обсуждение разногласий и др.). Также существует обширная традиция кросс-культурной адаптации инструментария. Однако целенаправленное использование методики когнитивного интервью для этих задач встречается довольно редко. На наш взгляд, эвристические возможности КИ для задач кросс-культурной адаптации опросников остаются недооцененными и не проработанными. Для решения этой задачи мы выбрали одну из методик (психометрический тест), не существующих пока на российском рынке, но к которой у отечественных исследователей есть и интерес, и готовность ее перевести.

Степень научной разработанности проблемы. Использованные в нашей работе источники можно разделить на три группы: психометрические тесты и характеристики их эффективности; методы адаптации психометрических тестов, а также методы когнитивного интервьюирования.

1. Работы из первой группы послужили основой для составления общей картины исследования и разработки психометрических тестов. Наиболее значимыми в области психометрического тестирования являются работа Анастази “Evolving concepts of test validation” (Anastasi, 1986) и Американской психологической ассоциации (APA) “Standards for educational and psychological testing” (Association et al., 1985).

Анастази принадлежит наиболее распространенное определение понятия психометрического теста, которое согласуется с фундаментальной в образовательном и психологическом тестировании работой “Стандарты образовательного и психологического тестирования” и включает в себя его ключевые компоненты. Анастази определяет психометрический тест как объективное и стандартизированное измерение образцов (проб) поведения.

Кроме того, изучением особенностей психометрических тестов, разработанных для задачи измерения компетенций занимались Кронбах и Мил; Хэмблтон, Сваминатан и Роджерс; Радж и Элис; Гизелли и Браун; Гион; Моссхолдер и Арви; Бранник и Левин; Щербаум; Маккорник, Де Ниси и Шоу; Левинджер; Кларк и Уотсон; Кэмпбелл; Мэссик; Лорд и Новик.

Ключевой работой, посвященной изучению моделей компетенций личности, выступает работа Бояцис “The competent manager: A model for effective performance” (Boyatzis, 1991). В рамках данной работы было сформулировано определение компетенций, выступающее фундаментом изучаемого нами психометрического теста.

2. Разработка и тестирование подходов к адаптации психометрических тестов были описаны в руководстве Международной тестовой комиссии, работах Битон, Бомбардиер, Гуллемин и Ферраз, Солано-Флореса, Бекхоффа, Эпстайна, Санто, Фокс-Рушби, Хердмана. Ключевой работой в данном списке является руководство, разработанное Битоном, Бомбардиером, Гуллемином и Ферразом, в рамках которой нами были рассмотрены основные возможности и ограничения применения каждого из описанных авторами методов адаптации.

Одним из первых официальных документов, в котором были собраны рекомендации по переводу и адаптации психометрических тестов является сборник Международной тестовой комиссии (International Test Commission), работа над первым изданием которого была начата в 1992 году (Commission, 2017). На сегодняшний день последнее издание сборника, которое разрабатывалось в период с 2005 по 2015 год, датировано 2017 годом и включает в себя шесть разделов (Commission, 2017). В рамках последнего издания Международной тестовой комиссии адаптация определяется как совокупность таких процессов, как принятие решения о возможности измерения конструкта в целевой культуре, к которой адаптируется тест; подбор переводчиков; разработка схемы оценивания работы переводчиков; изменение формата проведения теста при необходимости; перевод; проверка эквивалентности переведенного варианта теста исходному (Commission, 2017)p.6-7.

Битон, Бомбардиер, Гуллемин и Ферраз определяют кросс-культурную адаптацию как процесс, в ходе которого исследователи преследуют цель обеспечения эквивалентности между адаптируемым и адаптированным вариантами психометрического теста на основе его содержания (Beaton et al., 2000). Исследователи разделяют понятия перевода, адаптации и кросс-культурной валидации. Перевод в интерпретации Эпштейна, Рут Миюки Сантоб и Гуллемина определяется как «единый процесс создания анкеты из исходной версии на целевом языке», при этом под адаптацией подразумевается «процесс рассмотрения смысловых и культурных различий между исходной и целевой версиями опросника». Процесс кросс-культурной валидации направлен на проверку сохранения тестом исходных психометрических свойств (Epstein et al., 2015), (Commission, 2017)p.12.

Таким образом, несмотря на существование разных подходов к определению понятия адаптации психометрического теста, все они выделяют схожие элементы процесса адаптации: проверка

существования конструкта, для измерения которого разработан тест, в культуре, к которой он будет адаптироваться; выбор переводчиков и критериев оценки их деятельности; непосредственно перевод; объединение разных версий перевода и проверка эквивалентности исходной версии теста переведенной. Необходимо подчеркнуть, что в нашей работе мы сосредоточились именно на последнем из аспектов процесса адаптации.

3. Ключевыми работами, которые посвящены изучению метода когнитивного интервью для задачи адаптации психометрических тестов, выступает работа американских исследователей Пан и Фонд, а также работа Виллиса. Пан и Фонд разработали классификацию проблем, которые могут быть выявлены с помощью метода когнитивного интервью в рамках адаптации психометрических тестов: культурные, социальные и лингвистические проблемы. Виллису принадлежит определение когнитивного интервью как одного из инструментов адаптации тестов: когнитивное интервью – это метод, нацеленный изучить то, как целевая аудитория понимает, мысленно обрабатывает и реагирует на материалы, предоставленные исследователем, с особым вниманием на смещения и сбои в этом процессе. Теоретико-методологическим изучением когнитивного интервью также занимались ДиМайо и Ротгеб; Мохорко и Глебец; Джоуб, Туранжо и Смит; Уилсон; Стритт и Смит; Конверс и Прессер; Оксенберг и Калтон; Блэр; Штраус и Корбин; Гербер и Велленс; Бикарт и Фелчер; Канеел и Фовлер; Д.М. Рогозин; А. Ипатова, К. Мануильская.

Когнитивное интервью относится группе методов, именуемых когнитивным анализом. Предмет когнитивного подхода заключается в выявлении “мыслительных процессов, которые активизируются при восприятии вопроса и ответе на него, а назначение – в том, чтобы «установить, что человек думает, отвечая на вопросы интервьюера» (Рогозин, Д. М., 2002). Чаще всего основной целью применения когнитивного подхода является тестирование чернового опросного инструментария.

Когнитивный подход включает в себя две группы методов: когнитивное кодирование и когнитивное интервью. Преимущество когнитивного кодирования заключается в возможности зафиксировать характеристики прохождения интервью и выявить те параметры, которые могут оказывать влияние на его прохождение. Недостатком данной группы методов является дороговизна и высокая степень продолжительности процедуры кодирования и анализа результатов, а также тот факт, что когнитивное кодирование не позволяет проанализировать соответствие интерпретаций вопроса исследователем и респондентом, а также выявить недостатки в формулировках вопросов [Cannel, Fowler, 1996. P.21]. То есть, этот подход не годится для задач адаптации уже созданного теста к новому культурному контексту.

Несмотря на многообразие классификацией методов когнитивного интервью, чаще всего исследователи выделяют две «парадигмы»: «проговаривание мыслей вслух» (think-aloud) и «метод вопросов-проб» (probes) (Beatty & Willis, 2007). Применительно к задаче адаптации психометрического теста парадигма «мышление вслух» является менее затратным инструментом, так как не требует предварительного обучения интервьюеров, подробного ознакомления с содержанием анкеты, а также тщательной проработки гайда интервью. Ограничением «проговаривания мыслей вслух» для задачи адаптации является тот факт, что данный подход не всегда может предоставить информацию об интерпретации содержания вопросов, так как информант не сталкивается с данным вопросом напрямую. Исследователь, перед которым стоит цель адаптации психометрического теста, никогда не может быть точно уверен, получит ли он необходимую информацию или нет. Таким образом, несмотря на снижение

вероятности возникновения эффекта интервьюера в парадигме «мышление вслух», мы не можем с уверенностью утверждать, что данный подход будет эффективным для достижения задачи адаптации теста.

Параллельно с мышлением вслух развивалась парадигма «интенсивного интервью» («intensive interviewing») или метод «вопросов-проб» (probes). Эспосито, Ротгеб, Поливка, Хесс и Кампанелли в качестве основных возможностей применения парадигмы «вопросов-проб» выделяют оценку того, насколько представления респондентов расходятся с определениями, представленными в словарях и справочниках, выявление неоднозначно воспринимаемых респондентом понятий, а также вопросов, которые не отражают цель исследования, установление вопросов, которые в большей степени отражают цель исследования, выявление того, как различные версии вопросов влияют на ответы респондентов [Esposito, 1993. p. 18-19]. Джоуб, Туранжо и Смит среди преимуществ применения метода когнитивного интервью называют получение более детальной информации о когнитивных процессах, которые происходят во время формирования ответа (Jobe et al., 1993). Гербер и Велленс считают, что данный метод позволяет оценить сензитивность вопросов по отношению к определенным социальным группам (Gerber & Wellens, 1997). Таким образом, парадигма «вопросов-проб» может предоставить более полную, исходя из целей исследования информацию, по сравнению с «мышлением вслух», так как данный подход позволяет контролировать процесс прохождения интервью информантом.

Американские исследователи из Бюро переписи населения США Пан и Фонд, опираясь на мысль о неотделимости языка, культуры и общества, разработали следующую классификацию проблем, которые могут быть выявлены с помощью когнитивного интервью в рамках адаптации инструментария: лингвистические, культурные и социальные проблемы. Лингвистические проблемы относятся к затруднениям информантов, вызванными использованием устаревших в языке понятий, грамматическими ошибками, особенностью порядка слов или структуры вопроса в целом. Культурные проблемы возникают из-за разных способов выражения одного и того же понятия в культурах. Социальные проблемы связаны с существующими в обществе практиками и социальными институтами (Pan & Fond, 2014). Таким образом, разработанная Пан и Фонд классификация соотносится к трем видам эквивалентности, для достижения которой необходимо привлечение носителей культурных практик, а также с подходом социолингвистики.

Основным ограничением работы Пан и Фонд является тот факт, что в своем исследовании они используют существующие пробы, не нацеленные на выявление описанных ими проблем. Таким образом, для преодоления данного ограничения когнитивного интервью применительно к задаче адаптации психометрического теста, мы предполагаем, что целесообразным будет развитие данного метода.

5.5 Методы сбора и обработки данных

В качестве подхода была выбрана качественная методология, а именно когнитивный анализ. Основной целью нашей работы является выявление ограничений и возможностей развития методики когнитивного интервью для задачи культурной адаптации психометрического теста как инструмента измерения компетенций. Основное назначение культурной адаптации психометрических тестов заключается в выявлении смысловых и культурных различий между его исходной и адаптированной версиями, в то время как предмет когнитивного подхода заключается в анализе семантического соответствия вопросов

изучаемому предмету. Таким образом, методология когнитивного анализа может выступить в качестве инструмента выявления смысловых и культурных особенностей восприятия информантами вопросов психометрического теста, что является необходимым условием для полноценной адаптации инструментов (Рогозин, 2000)р.18. В этом отношении цель нашей работы – развитие методики КИ.

В качестве метода исследования было выбрано когнитивное интервью в «парадигме» «вопросов проб», так как именно данный метод (по сравнению с «мышлением вслух») позволяет целенаправленно выявлять смысловые и культурные особенности восприятия вопросов информантами, а также контролировать процесс интервью.

В качестве определения понятия «когнитивное интервью», мы будем использовать подход Виллиса: когнитивное интервью – это метод, нацеленный изучить то, как целевая аудитория понимает, мысленно обрабатывает и реагирует на материалы, предоставленные исследователем, с особым вниманием на смещения и сбои в этом процессе [Willis, 2005]. Данный подход к определению когнитивного интервью согласуется с понятием адаптации в нашем исследовании как процесса, нацеленного на получение информации о смысловой интерпретации вопросов информантами.

Под развитием когнитивного интервью мы будем понимать использование данной методики для задачи новой проблематизации адаптации психометрического теста, а не только апробации чернового инструментария; разработку зондов, релевантных для задачи культурной адаптации психометрического теста; выявление возможностей и ограничений модифицированной методики когнитивного интервью для задачи культурной адаптации психометрического теста.

Дизайн исследования. За основу работы был взят психометрический текст Р*А, используемый для оценки профессиональных компетенций сотрудников при приеме на работу. В зарубежной практике этот тест используется на протяжении последних 20 лет. Также эта методика удовлетворяла нашим требованиям. При выборе инструмента для его дальнейшей адаптации мы руководствовались такими принципами, как надежность и валидность инструмента, тест не должен был ранее переводиться на русский язык и адаптироваться к российскому контексту, разработчики инструмента согласны предоставить содержание психометрического теста. От компании, которая занималась переводом теста и методических рекомендаций к нему, мы получили ограниченное количество вопросов для когнитивного тестирования (21 шт.), в отношении формулировок которых эксперты не могли прийти к соглашению. Условием передачи вопросов компанией являлось подписание договора о неразглашении коммерческой тайны.

В качестве основного метода сбора данных выступили полуструктурированные когнитивные интервью, проведенные в рамках парадигмы вопросов-проб.

Полевой этап реализовывался в два этапа. На первом нами было проведено 20 интервью, 10 интервью с “программистами” и 10 интервью с “филологами”. Каждая целевая группа была разделена на две подгруппы, каждой подгруппе был предоставлен свой вариант анкеты с разными альтернативами перевода вопросов. В качестве генеральной совокупности в нашем исследовании выступали студенты 3 и 4 курса бакалавриата, а также 1 и 2 курса магистратуры НИУ “ВШЭ” в г. Москва, так как на данном уровне обучение (незаконченное высшее или высшее образование) целесообразно проводить измерение компетенций. Нами рассматривались только информанты, обладающие опытом работы, так как анкета подразумевает прохождение информантами с опытом работы. В рамках выборочной совокупности нами были выделены 2 контрастные целевые группы: “программисты” (студенты, обучающиеся по

направлениям подготовки «Прикладная математика и информатика»; «Информатика и вычислительная техника»; «Информатика и вычислительная техника»; «Программная инженерия»; «Информационная безопасность»; «Информационная безопасность») и “филологи” (студенты, обучающиеся по направлениям подготовки «Филология»).

После окончания этого этапа мы проанализировали адекватность подхода, корректность работы зондов и всей методологии исследования в целом. После внесения корректировок в подход, были проведены 40 интервью. Из них:

- Интервью в формате видеоконференций (20 интервью, в том числе 10 интервью с респондентами моложе 25 лет и 10 интервью с респондентами старше 50 лет; дополнительно варьировался уровень образования – высшее / среднее),
- Интервью в формате «веб-пробирования» (20 интервью, в том числе 10 интервью с респондентами моложе 25 лет и 10 интервью с респондентами старше 50 лет; дополнительно варьировался уровень образования – высшее / среднее).

Для выявления “проблемных” вопросов мы использовали несколько типов «зондов»:

1. “Уверенность в ответе”, когда информант оценивал свой ответ по шкале от 1 до 5, где 1 - абсолютно уверен в ответе и 5 - абсолютно не уверен в ответе. Данный зонд использовался для индикации наиболее проблемных формулировок. Интервьюер задавал дополнительные зонды к вопросам, в ответе на которые информант был не уверен.
2. Метод парадигма (информант самостоятельно пытается переформулировать вопрос при помощи собственных лексических конструкций).
3. Уточняющий зонд (информанта просят объяснить причины выбора своего ответа).

На основе общего анализа ответов информантов после применения метода парадигма и уточняющего зонда мы оценивали семантическую адекватность для каждого вопроса по шкале от 1 до 3: 1 - адекватный ответ (уверенный ответ, без искажения смысла заданного вопроса), 2 - респондент сомневается (есть некоторые отклонения ответа от конструкции вопроса, респондент предлагает свои варианты ответа), 3 - неадекватный ответ (респондент отказывается делать выбор из предложенных вариантов, искажает смысл вопроса, дает свой ответ) (Рогозин, Д. М., 2002).

3. «Трехчастный зонд». Для некоторых, заранее отобранных вопросов, мы использовали “смыслоориентированный зонд», состоящий из трех частей, чтобы выявить несоответствие между опытом респондента и его определением обсуждаемого в вопросе концепта. Мы задавали уточняющие вопросы о том, как понимается концепт («светский разговор»), есть ли у него личный опыт и в каком контексте происходило подобное действие.

Кроме того, в конце интервью мы использовали еще два зонда:

- «сложные вопросы». У информанта спрашивали, какие вопросы ему показались сложными, на какие вопросы трудно ответить.
- «зонд на сензитивность». У информанта спрашивали, о каких вопросах ему неловко говорить, какие вопросы было бы некомфортно обсуждать с незнакомым человеком.

5.6 Обсуждение

Возможности и ограничения применения «трехчастного зонда». Основные возможности применения «трехчастного зонда» для задачи адаптации психометрического теста как инструмента измерения компетенций заключаются в выявлении смысловых интерпретаций вопросов и смещений в этих интерпретациях с помощью соотнесения личного опыта информанта и его определений ключевых концептов в обсуждаемых вопросах. Сбор данной информации является необходимым условием для принятия решения о достижении эквивалентности между исходной и адаптированной версиями тестов, что может повысить качество адаптации психометрических тестов. Результаты проведенного нами исследования продемонстрировали тот факт, что «трехчастный зонд» может выявлять потенциальные проблемы в вопросах именно с точки зрения процесса адаптации. Несмотря на то, что при разработке «трехчастного зонда» предполагалось, что данный инструмент будет направлен преимущественно на выявление социальных проблем, результаты анализа интервью продемонстрировали, что «трехчастный зонд» способен также выявлять и лингвистические затруднения участников.

В качестве основных ограничений применения «трехчастного зонда» необходимо выделить невозможность выявления культурных проблем о разных способах выражения одного и того же понятия в исходной и целевой культуре без наличия результатов когнитивных интервью или смысловых интерпретаций вопросов в исходной культуре.

Применение «трехчастного зонда» к большому количеству вопросов затруднительно из-за объемной конструкции данного инструмента, а также из-за факта обучения информантов: в нашем исследовании к концу каждого интервью информанты отвечали на все три части зонда, не дожидаясь вопросов интервьюера.

Специфика исследуемого нами психометрического теста заключается в краткости формулировок вопросов, из которых он состоит. Этот факт облегчает выделение ключевых понятий в вопросах для дальнейшего применения «трехчастного зонда» и составления ожидаемых проб «трехчастного зонда» перед проведением интервью. Применительно к опросному инструментарию, в рамках которого встречаются громоздкие конструкции вопросов, в которых могут быть несколько ключевых понятий, процедура подготовки к интервью осложняется составлением условных проб для сохранения стандартизации, либо появляется необходимость использования эмерджентных проб в зависимости от течения разговора интервьюера и участника, что понизит стандартизированность процедуры проведения интервью.

Использование когнитивного интервью как одного из методов адаптации и «трехчастного зонда» как основного инструмента когнитивного интервью повысит временные и денежные затраты для адаптации теста.

Основными ограничениями нашей работы является апробация развитой методики только на двух контрастных социально-демографических группах, отсутствие возможности сравнения полученных смысловых интерпретаций со смысловыми интерпретациями вопросов носителей культурных практик в англоязычной культуре для апробации «трехчастного зонда» как инструмента выявления культурных проблем, а также субъективность при выделении потенциальных проблем в вопросах теста. Использование зонда «понятность вопросов» в качестве индикатора для применения «трехчастного зонда» к вопросам, которые были малопонятны информантам, требовало введения уточняющих вопросов

о причинах низких оценок «понятности» на этапе применения данного зонда.

5.7 Результаты исследования

В результате проведения исследования была выявлена эффективность разработанного нами «трехчастного зонда» по трем критериям: выявление информации, соответствующей функциям зонда, количество и качество обнаруженных проблем.

Интервью в формате видеоконференций (средний показатель 2,8) более эффективные с точки зрения предоставления информации о смысловых интерпретациях вопросов информантами и личном опыте участников взаимодействия с обсуждаемым в вопросе конструктом, по сравнению с «веб-пробированием» (средний показатель – 2,4).

Интервью в формате видеоконференций также более эффективны, по сравнению с «веб-пробированием» с точки зрения выявленных проблем качественно и количественно (в видеоконференциях было выявлено 58 проблем, из них 50% социальных и 50% лингвистических, при «веб-пробировании» - 19 проблем, из них 47% социальных и 53% лингвистических).

Таким образом, наше предположение о меньшей эффективности «веб-пробирования» по сравнению с проведением интервью в формате видеоконференций на базе платформы Zoom подтвердилось.

Исходя из количества и качества выявленных проблем, наше предположение о преимущественном выявление «трехчастным зондом» социальным проблем подтвердилось.

В разрезе наличия опыта участия у информантов в социологических и маркетинговых исследованиях, значительных отличий в показателях выявления необходимой информации «трехчастным зондом» нет. С точки зрения выявления проблем количественно и качественно, и в видеоконференциях и при проведении «веб-пробирования» интервью с участниками без наличия соответствующего опыта выявило больше лингвистических и социальных проблем (в видеоконференциях: 32 проблемы выявлены в интервью с участниками без опыта, 26 – с опытом; при «веб-пробировании»: 12 проблем выявлено в интервью с участниками без опыта, 7 – с опытом).

Таким образом, наше предположение о большей эффективности «трехчастного зонда» при проведении интервью с информантами, у которых есть опыт участия в социологических исследованиях и присутствовали определенные ожидания по поводу интервью, не подтвердилось. Специфика взаимосвязи наличия опыта участия в социологических и маркетинговых исследованиях и эффективности «трехчастного зонда» требует проведения дополнительных исследований.

«Веб-пробирование» позволило избежать возникновение эффекта интервьюера и снизило временные и денежные затраты, необходимые для проведения интервью и составления транскриптов.

Один из выявленных нами недостатков «веб-пробирования» состоит в ограничении применения зонда «понятность вопроса» из-за невозможности использования дополнительных вопросов для уточнения причин выставления низких оценок по данной пробе информантами. Формат «веб-пробирования» также ограничивает использование зонда «понятность вопроса» в качестве индикатора для дальнейшего применения «трехчастного зонда» к вопросам, которые были малопонятны участникам. Невозможность использования дополнительных и уточняющих вопросов относится также в целом ко всей процедуре «веб-пробирования», что снижает информативность полученных ответов, по сравнению с форматом видеоконференций.

Среднее время, затраченное участниками на заполнение анкеты при участии в «веб-пробировании» больше, чем для видеоконференций. Участники исследования в качестве недостатков процедуры «веб-пробирование» отметили длительность процедуры, однообразие вопросов и трудность набора большого количества текста на клавиатуре.

Несмотря на вероятность возникновения эффекта интервьюера, применение «трехчастного зонда» в формате видеоконференций открыло перед нами возможность задавать уточняющие вопросы информантам, а также контролировать прохождение интервью и анализировать невербальные реакции участников. Некоторые информанты, принявшие участие в «веб-пробировании» в качестве комментариев, отметили тот факт, что им была не понятна цель исследования, в то время как при проведении видеоконференций такой проблемы не возникало. Возможно, данный факт связан с тем, что при прохождении «веб-пробирования» информанты могли не читать приветственное слово, в то время как в видеоконференциях они были вынуждены выслушать вступление информанта с объяснением целей исследования «принудительно».

Ограничением использования видеоконференций в нашем исследовании выступили плохое качество Интернет-соединения и динамиков у некоторых информантов, в связи с чем проанализировать некоторые проведенные интервью не удалось.

Основные направления будущих исследований могут быть связаны с кросс-культурным использованием «трехчастного зонда» в качестве инструмента выявления не только социальных и лингвистических, но и культурных проблем в рамках задачи адаптации опросного инструмента; проведение опросного эксперимента для выявления влияния выявленных нами проблем на психометрические свойства теста: валидность и надежность; проведение исследования на большей выборке испытуемых для изучения особенностей использования «трехчастного зонда» в разных социально-демографических группах; измерение когнитивной усталости информантов, к которым применяется «трехчастный зонд»; уточнение требований администрирования процесса применения «трехчастного зонда» с точки зрения поведения интервьюера, апробация использования «трехчастного зонда» в других форматах проведения интервью, а не только в формате видеоконференций и «веб-пробирования». Отдельным направлением стоит выделить задачу использовать методы сетевого анализа текстовой информации, чтобы развивать аналитические возможности работы с материалами, полученным с помощью когнитивного интервью.

5.8 3.1.1 Программа «Bib-eLib» для сбора и обработки библиографических данных на русском языке из электронной библиотеки eLibrary

5.8.1 Введение

В рамках реализации научного проекта лаборатории научным коллективом МЛ прикладного сетевого анализа разработан служебный результат интеллектуальной деятельности – программа для ЭВМ «Программа «Bib-eLib» для сбора и обработки библиографических данных на русском языке из электронной библиотеки eLibrary». Программа «Bib-eLib» разработана Д. В. Мальцевой, В. А. Ващенко, Л. В. Капустиной, А. В. Ким, Т. Е. Щегловой, Л.Г. Ципесом в НИУ Высшая школа экономики, являющейся ее правообладателем (свидетельство о государственной регистрации программы для ЭВМ № 2023684182, регистрация в реестре программ для ЭВМ 14.11.2023) (Рис.1).

Рис. 1. Свидетельство о регистрации программы «Bib-eLib» для сбора и обработки

библиографических данных на русском языке из электронной библиотеки eLibrary Рис. 1.

Свидетельство о регистрации программы «Bib-eLib» для сбора и обработки библиографических данных на русском языке из электронной библиотеки eLibrary

Цель работы программы «Bib-eLib» – сбор и обработка библиографических данных на русском языке из электронной библиотеки eLibrary для применения в области научометрических и библиометрических исследований, в особенности в области сетевого анализа библиографических сетей.

5.8.2 Описание работы программы

Программа «Bib-eLib» характеризуется следующими функциональными характеристиками:

- Программа осуществляет выгрузку массива данных о научных публикациях через API электронной библиотеки eLibrary при наличии предварительно заключенного контракта с Национальной электронной библиотекой НЭБ;
- Программа проводит предварительную обработку массива данных о научных публикациях, выгруженных через API электронной библиотеки eLibrary, в части имен авторов и их аффилиаций (организаций);
- Программа реализовывает алгоритм дизамбигуации авторов публикаций в предварительно обработанном массиве данных путем создания новых универсальных идентификаторов авторов;
- Программа проводит анализ итогового массива данных по основным статистическим метрикам в области библиометрических исследований и визуализировать распределение подсчитанных метрик;
- Программа создает двумодальную сеть связей между научными публикациями и их авторами в формате с расширением .net, пригодном для дальнейшей обработки в программе для анализа и визуализации больших сетей Pajek.

Программа подходит для реализации на Windows, MacOS, Linux на базе интегрированных с ОС графических пользовательских интерфейсов (GUI). Программа написана на языке программирования Python и включает архив файлов формата Jupyter Notebook (.ipynb) размером 266 КБ. Описание файлов .ipynb-files приводится ниже. В качестве примера для сбора взят массив данных о публикациях российских социологов, который был собран в рамках проекта “Паттерны коллаборации в Российском социологическом сообществе” (грант РНФ, руководитель Д.В. Мальцева, 2021-2023 гг.).

1. Файл “0. Выгрузка данных о статьях авторов.ipynb” - в этом файле производится выгрузка данных о статьях конкретных авторов на eLibrary через сервис API Elibrary 011 (не используется напрямую в программе, но может быть полезным). В качестве примера, в этом файле Jupyter Notebook производится выгрузка данных eLibrary по публикациям российских авторов в области социологии. Осуществляется импорт необходимых библиотек, загружаются id авторов, данные о которых необходимо выгрузить, производится выгрузка кратких данных обо всех статьях этих авторов из API Elibrary. Основной код выгрузки к API-011 Elibrary был подготовлен Ликой Капустиной в сентябре-октябре 2022 года. Цель выгрузки - выгрузить краткие справочные данные о статьях всех уникальных авторов, встречающихся в основном массиве данных (но не все данные по статьям). Собранные данные не использовались для дообновления финальных данных проекта, но код в этом ноутбуке может быть полезен для других исследователей. Результат выгрузки – файл

all_final_authors_with_papers.csv, содержащий в себе информацию о всех статьях всех авторов, присутствовавших в 8 колонках наших изначальных данных.

2. Файл “1. Выгрузка данных по статьям.ipynb” – представленный в этом файле код используется для выгрузки данных о статьях через API Elibrary. В этом файле Jupyter Notebook производится парсинг (выгрузка данных) данных eLibrary по статьям через API eLibrary для проекта “Паттерны коллаборации в Российском социологическом сообществе”. Осуществляется импорт необходимых библиотек, загружаются id необходимых к выгрузке статьи, производится сама выгрузка основного массива данных. Основной код выгрузки был подготовлен Львом Ципесом, в доработке кода и выгрузке данных участвовали Василиса Ващенко и Лика Капустина. Выгрузка производилась в июне-июле 2022 года. Эта версия была закончена и прокомментирована Ликой Капустиной. Представленный код можно использовать для выгрузки данных о статьях через API Elibrary. Предварительно вам нужно заключить договор с Elibrary, и тогда вы получите доступ к одному из сервисов API по IP-адресу, указанному в договоре. Для работы нужен файл формата .txt с id статей, информацию о которых нужно выгрузить через API Elibrary. Процесс выгрузки данных о статьях через API Elibrary происходит в несколько шагов:

- Шаг 1: подготовка к работе;
- Шаг 2: выгрузка данных
- Шаг 3: сохранение данных В результате получаются следующие файлы:
 - Промежуточные файлы с данными в формате .csv и .xlsx, которые сохраняются локально в рабочую директорию каждую 1000 итераций (с названием в формате df_текущая дата_число строк_rows.формат файла);
 - final_data.csv - pandas.DataFrame с информацией о всех статьях, находящихся в изначальном списке, в формате .csv;
 - final_data.xlsx - pandas.DataFrame с информацией о всех статьях, находящихся в изначальном списке, в формате .xlsx;

3. Файл “2.1. Предобработка данных и получение идентифицирующих признаков.ipynb” – в этом файле проводится предобработка данных, выгруженных на этапе 1, с точки зрения работы с именами авторов и получения идентифицирующих признаков. Часть 1. Предобработка данных На данном этапе стоит задача предобработать данные, выгруженные через API Elibrary ранее (см. файл «1. Выгрузка данных по статьям»). На этом этапе стоит несколько задач:

- Предобрбатать данные по русскоязычным фамилиям и инициалам;
- Предобрбатать данные по англоязычным фамилиям и инициалам;
- Предобрбатать данные по русско- и англоязычным аффилиациям;
- Провести процедуру присвоения собственных id авторов; В части 1 проводятся две первые процедуры - предобработка данных по русско- и англоязычным фамилиям и инициалам. Такая необходимость возникает в связи с тем, что в данных достаточно много авторов без id во внутренней системе eLibrary. Этот факт осложняет дальнейшую работу с данными и их анализ – нельзя построить сети на основании id авторов, если они не достаточно корректные. Поэтому было принято решение подготовить собственные id авторов, основываясь на ФИО и данных об

аффилиациях авторов. Однако на этом этапе возникла проблема – eLibrary не контролирует единобразность записи фамилий, имен и инициалов; туда можно вписать разные знаки препинания; иногда в фамилии вписываются сразу и фамилии, и инициалы, и так далее. В этом ноутбуке демонстрируются разные проблемы и способы их решения. Получаемые в результате первой части предобработки файлы:

- problem1.xlsx - строчки с информацией о статьях, для которых отсутствует информация даже о первом авторе (то есть, нет информации об авторах вообще). Необходимо заполнение вручную.
- problem2.xlsx - строчки с информацией о статьях, в фамилиях и инициалах авторов которых встречаются вопросительные знаки. Необходимо заполнение вручную.
- problem3.xlsx - строчки с информацией о статьях, среди авторов которых встречаются авторы с не-русскими фамилиями если в разделе с инициалами у них не указаны инициалы (то есть, в ячейку с фамилией отнесена и фамилия, и инициалы). Необходимо заполнение вручную.

4. Файл «2.2. Обработка аффилиаций.ipynb» – в этом файле проводится предобработка данных, выгруженных на этапе 1, с точки зрения работы с аффилиациями авторов. В этом файле проводится работа по дедупликации аффилиаций авторов путем их чистки и присуждения id. Вначале составляется общий список всех аффилиаций, затем рассматриваются варианты их сокращения до аффилаций и создается словарь сокращений. Работа подразумевает следующие шаги:

- Шаг 0: Предварительная очистка данных
- Шаг 1: Работа с ID аффилиаций
- Шаг 2: Работа с английским переводом аффилиаций
- Шаг 3: Создаем новые id для неидентифицированных аффилиаций
- Шаг 4: Создание новых ID для авторов
- Шаг 5: Дедупликация авторов по новым ID

5. Файл «2.3. Пост-обработка_новых_ID.ipynb» – в этом файле проводится пост-обработка данных – создание новых универсальных ID авторов на основе почищенных имен авторов и аффилиаций. Работа по дедупликации авторов осуществляется на основании ID авторов и аффилиаций, присужденных на предыдущих этапах 2.1 и 2.2. Для новых ID авторов были рассчитаны попарные расстояния Дамерау-Левенштайна (см. п. 5 в файле 2.2), и в этом файле новые ID анализируются на предмет пересечения важных идентифицирующих признаков для финализации дедупликации. Работа включает следующие шаги:

- Шаг 1: Обработка авторов с наличествующим идентификатором
- Шаг 2: Обработка авторов с отсутствующим идентификатором
- Шаг 3: Сбор результатов обработки воедино
- Шаг 4: Изменение рабочего файла с данными

6. Файл «“3. Анализ данных и визуализация.ipynb”» – в этом файле анализируются предварительно собранные и обработанные файлы, а также создаются графики. Работа подразумевает следующие шаги:

- Пункт 1. Распределение статей по числу авторов

- Пункт 2. Распределение цитирований
- Пункт 3. Абстракты и ключевые слова
- Пункт 4. Статистики по авторам

7. Файл «“4. Создание сетевых файлов для Pajek.ipynb”» – в этом файле создаются сетевые файлы для работы с графами в формате Pajek. В настоящее время прописанный код подразумевает создание двумодальной сети формата «Работа – Автор», на основе которой могут строиться сети коллaborаций (согласно цели проекта “Паттерны коллаборации в Российском социологическом сообществе”), однако в дальнейшем планируется написание кода для других видов сетей («Работа – Журнал», «Работа – Ключевое слово», сети соприсутствия слов в аннотациях / названиях статей и др.).

Согласно разработанной в лаборатории методологии анализа библиографических сетевых данных на русском языке дальний анализ сетевых данных осуществляется в программе для анализа и визуализации больших сетей Pajek.

Файлы формата Jupyter Notebook расположены в репозитории на GitHub по ссылке: <https://github.com/Daria-Maltseva/Collaboration>. Коллективом лаборатории планируется работа по продолжению развития программ ЭВМ для анализа русскоязычных данных и разработка полноценного отчуждаемого пакета для библиометрического анализа русскоязычных публикаций на языке программирования Python.

6 Методологические особенности предобработки данных по российским авторам в Web of Science

Первоначальный массив данных, который лег в основу этого исследования, состоял из более чем 1.38 миллиона публикаций российских исследователей за 1990-2022 гг., индексированных в престижной международной базе данных Web of Science (WoS)(Мальцева & Fiala, 2023). Уникальность массива состоит в отсутствии любых ограничений на тип записей, количество цитирований, научную область, регион и т.д. Благодаря этому можно говорить о том, что эти данные отражают реальную картину представленности российской науки в «Web of Science» на май 2022 года, рис. 6. Исходный набор данных данного исследования включает публикации WoS, выгруженные со спецификацией поля данных «CU=(Russia)» в режиме full record (полное библиографическое описание публикаций, включающих пристатейные списки литературы). Всего исходный набор данных включал 1383996 библиографических записей о российских публикациях, проиндексированных в WoS Core Collection до мая 2022 г. Топ-20 типов документов и их количество, доля в общем объеме, цитируемость, доля в общем количестве цитирований и количество цитат на статью (CPP)

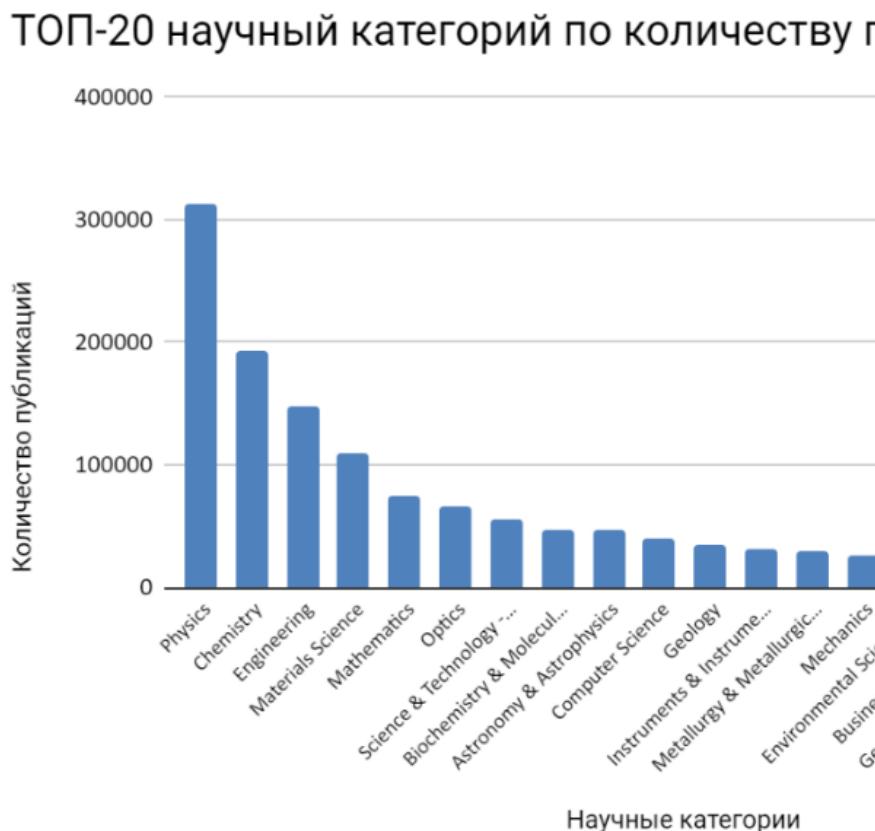
Ввиду отсутствия ограничений на этапе отбора данных и их большого размера, прежде чем приступить к анализу, мы были обязаны провести довольно крупный объем работ по предобработке данных. Сначала мы опишем общую структуру изначального массива данных, далее перейдем к процессу создания подсетов по отдельным научным категориям. Затем будут перечислены основные проблемы, которые встали перед исследовательской группой в обработке данных. После этого мы перейдем к представлению стратегии по решению конкретных проблем в именах авторов, состоящей из нескольких этапов. Наконец, будут описаны сложности, встречающиеся в записи организаций, и процесс их преодоления.

Полученные данные разделены по годам публикации, в каждой папке содержатся файлы с полными библиометрическими записями материалов в формате .txt, каждый из которых содержит максимум 500 записей. Каждая запись отделяется с помощью текстовых маркеров «PT» и «ER», как продемонстрировано на рис. 3. Основные библиометрические данные, необходимые для анализа включают в себя: название, имена авторов и их аффилиации, тип материала (статья, отчет по итогам конференции и т.д.), процитированных в работе авторов, страны, научное направление, дату публикации и уникальный идентификационный номер, присвоенный WoS.

Пример стандартной структуры полных библиометрических данных для 1 публикации

Figure 3: Пример стандартной структуры полных библиометрических данных для 1 публикации

Отбор данных по научной категории возможен с помощью параметров WC (Web of Science Categories) или SC (Research Areas): первый автоматически присваивается на основе журнала публикации, второй - определяется самими авторами, а потому обладает более точным определением научных направлений. Всего по SC в массиве определяется около 156 научных направлений, из них на первые 20 категорий



приходится 68% всех публикаций (рис. ??).

Подмножество по социологии было выделено на основе категории SC как категории исследования WoS (выполнение условия «SC = Sociology») и состоит из 7915 публикаций (менее 0,01% от всего массива публикаций). Для большинства публикаций в Web of Science характерна принадлежность нескольким категориям исследования. В исходном наборе данных у некоторых публикаций насчитывалось до 9 таких категорий, в среднем для публикации характерно 3-4 категории исследования. В нашем случае в качестве материалов исследования были отобраны все публикации, у которых как минимум одна категория исследования была указана как Sociology. Подмножество по социологии включает в себя публикации всех представленных типов – статьи, главы в монографиях, конференционные материалы и

т.д. – за период с 1992 до мая 2022 года.

Предобработка данных была реализована в Python с дополнительной ручной проверкой и корректировкой. Особое внимание уделялось именно авторам и организациям для приведения к единому виду и объединению разных вариантов написания фамилий и имен авторов, а также разных вариантов применения названий организаций. В целом на данном этапе было выявлено достаточно большое количество сложностей; некоторые из них приведены ниже как категории особенностей предобработки данных:

1. транслитерация ФИО авторов с кириллицы на латиницу (французский и английский варианты написания, сложные звуки (шипящие)). Пример: loukov – lukov, toshchenko – toshenko, alikhadzhieva – alikhadjieva и т.д.;
2. использование у российских авторов отчества, которое в библиографическом описании публикации может как присутствовать, так и отсутствовать. Отчество может стоять на первом месте вместо фамилии (valerevich, radaev vadim) или отсутствовать в принципе (radaev, vadim), что мультилинирует количество авторов. В этом случае усложняется обработка данных, так как для одного канонического написания ФИО автора (radaev, v. v.) необходимо выявить все разнообразные варианты имени этого автора в наборе данных. Например, для канонического ФИО автора как toshchenko, i. z. было выявлено 13 разных вариантов написания;
3. разные написания названий организаций – название организации с советского периода (Томский государственный университет им. В.В.Куйбышева – в настоящее время Национальный исследовательский Томский государственный университет), множественные варианты аффилиаций (Высшая школы экономики, НИУ Высшая школы экономики, Национальный исследовательский университет и т.д.);
4. сложная организационная структура (несколько кампусов, множество институтов и т.д.), когда указанное подразделение кодируется WoS как самостоятельная организация;
5. в библиографических описаниях публикаций в WoS категория независимого исследователя (independent researcher) никаким образом не учитывается, рассматривается как пропущенное значение. Также по непонятным причинам, в WoS не были указаны ряд аффилиаций коммерческих организаций, названия которых в публикациях были написаны по всем правилам (Yandex, GetBrand, Издательский дом «Коммерсант» и др.);
6. технические ошибки, опечатки и пропущенные данные (особенно у старых публикаций).

Предобработка данных об авторах включала два основных процесса: (1) предобработка данные по определенным правилам; (2) поиск схожих авторов, чье ФИО отличается на несколько символов с использованием технологии fuzzy matching. Главным приоритетом в работе, помимо максимального снижения числа «универсальных» ФИО авторов (в данных о российских социологах встречается до 8 вариаций ФИО одних и тех же авторов) являлось также минимальное число некорректных случаев соответствия ФИО авторов. Здесь и далее “ФИО авторов” и “имена авторов” являются синонимичными понятиями.

Одной из главных проблем, решенных на первом этапе работы с авторами стала проблема не унифицированности записей об авторах: хотя большинство записей имело вид “,<отчество или первая буква отчества, если указано>”, но встречались и другие формы записи, когда на месте фамилии

находилось отчество или имя. Нам был реализован алгоритм поиска имен и отчеств на некорректном месте и изменения порядка записей в сторону унифицированных (*andrey, kinyakin* -> *kinyakin, andrey; sergey, stepanov* -> *stepanov, sergey*). Помимо этого, мы заменили мало популярные формы имен на популярные (“*nadejda*” -> “*nadezhda*”), убрали лишние символы из имен (например, символ штриха), чтобы облегчить поиск сильно совпадающих имен авторов на следующем этапе.

После первого этапа предобработки мы имели предварительно обработанные имена авторов, приведенные к единому формату записи: “,<отчество или первая буква отчества, если указано>”. После этого мы решили реализовать процесс мэтчинга авторов на основании метрики расстояния Левенштейна: эта метрика позволяет получить число символов, на которые отличаются определенные строки. При простом поиске совпадающих строк с именами авторов существует риск случайно слить в один разных в реальности авторов; например, хотя для пары “*barsukova, s y*” и “*barsyukova, s y*” расстояние Левенштейна будет равно единице и этот мэтч может быть обработан далее корректно (это действительно один и тот же автор), то при отсутствии дополнительных правил мы могли бы привести к единой форме в реальности разных авторов, например, расстояние Левенштейна для строк “*ivanova, a a*” и “*ivanov, a a*” также равно единице. Поэтому мы реализовали поиск совпадающих авторов при наличии дополнительных правил: совпадение последнего звука фамилии (чтобы исключить мэтчинг мужчин и женщин-авторок), совпадение первого звука фамилии, совпадение всех или хотя бы части инициалов в том случае, если отчество автора не было указано. Далее производилось деление случаев на категории по типам (гласные/согласные звуки) для первых отличающихся звуков в фамилиях и совпадении этих типов:

| Тип | Уверенность в корректности дальнейшего мэтчинга фамилий Пример |
|---|--|
| :—: :—: :—: | |
| Фамилии отличаются на гласные звуки Высокая alekseeva, t. a. – alekseyeva, t. a. | |
| Фамилии отличаются на согласные звуки Средняя sebentsov, a. b. – sebentzov, a. b. | |
| Фамилии, отличающиеся лишь из-за одного дублирующегося звука Высокая isaev, l. m. – issaev, l. m. | |
| Фамилии отличаются на гласный и согласный звук Низкая lapin, v. s. –apkin, v. ** | |

Далее, во время формулирования разных категорий потенциальных мэтчей, возможно также посмотреть содержание всех списков и далее вручную удалить некорректные пары, однако, можно использовать и целый объект, полученный на этом шаге. Таким образом, на данных по российским социологам после проведения процедур первого типа удалось сократить число уникальных авторов на 11,2%, после проведения мэтчинга на основе поиска совпадающих фамилий еще на 4,8% от изначального числа уникальных авторов, а итоговый результат составил 16%-ное снижение уникальных авторов. По ощущениям, появившимся при первичном просмотре файла по социологам, примерно 20-25% всех записей являлись дублями и могли бы быть приведены к другой, более популярной форме имени автора. Полученная нами цифра, с одной стороны, не очень маленькая – что демонстрирует, что данный, даже очень аккуратный подход к мэтчингу авторов, способен снижать число уникальных авторов; с другой стороны, это относительно невысокий показатель, демонстрирующий аккуратность подхода авторов к поиску совпадающих авторов. Описанный процесс выше необходим для дальнейшего построения сетевых моделей на основе данных об авторстве и соавторстве, так как при наличии большого количества дублирующихся записей об авторах выводы любого исследования будут некорректны. Дальнейшие планы

по развитию проекта связаны с намерением повысить точность мэтчей и составить новые процедуры поиска совпадающих имен авторов.

Деятельность по предобработке данных об организациях, с которыми аффилированы авторы исследуемых публикаций, включала в себя два ключевых составных блока: итеративный fuzzy matching и следующий за ним keyword matching.

На предварительном этапе обработки данных организаций были определены проблемные аффилиации, для которых вместо названия указан адрес, и исключены из последующего анализа. К таковым были отнесены строки, содержащие цифры или маркеры адреса, такие как слова «lane» (переулок), «str» (ул.) и др. Эти аффилиации были размечены вручную при помощи обращения к публикациям, к которым они относятся. Прочие аффилиации были очищены от специальных символов и отдельно стоящих букв, а затем разделены на слова-токены.

После предобработки, мы приступили к итеративному мэтчингу токенов на основании метрики близость Дамерау-Левенштайна, реализованной в библиотеке jellyfish для языка программирования Python. Сначала мэтчинг применялся к отдельным токенам: для всех пар токенов длиннее 3 букв рассчитывалось расстояние Дамерау-Левенштайна и, после ручной проверки, пары с расстоянием менее 3 (токены отличаются друг от друга менее, чем 3 символами) были объединены. Подобная операция была произведена три раза последовательно с сокращением порога объединения до 1 отличающегося символа. Это позволило объединить слова с разным написанием при транслитерации на английский (e.g. ‘altay’ и ‘altaï’), альтернативные сокращения (e.g. ‘federal’ и ‘federat’), опечатки (e.g. ‘novasibirsk’ и ‘novosibirsk’), имена (e.g. ‘peter’ и ‘petr’), а также слова, написание которых варьируется между языками написания (e.g. ‘milan’ и ‘milano’ или ‘labor’ и ‘labour’). Затем дедуплицированные токены были снова объединены в полные названия. К измененным строкам названий также был применен мэтчинг: объединялись строки, отличающиеся не более чем на 2 символа.

Дедупликация токенов и мэтчинг строк позволяют избавиться лишь от части вариативности в написании названий организаций ввиду того, что наименования могут включать в себя слова в разной последовательности, неоднородный перевод, сокращения, разную степень детализации аффилиации (например, до уровня факультета). Дальнейшее удаление дубликатов производилось при помощи подхода, основанного на выделении ключевых слов для идентификации ряда крупных организаций и присвоении стандартизованных названий всем наблюдениям, содержащим указанные ключевые слова. Полный список использованных ключевых слов представлен в таблице ниже. При выделении ключевых слов мы ориентировались на задачу обнаружения последовательности минимальной длины, которая позволяет обнаружить как можно большее количество строк, относящихся к искомой организации.

| Организация | Ключевые слова |
|--------------------------|-------------------------------------|
| НИУ ВШЭ | hse, higher_sch, higher_econ |
| МГУ им. Ломоносова | lomonosov, msu |
| МГТУ им. Баумана | bauman |
| Российская Академия Наук | russian_acad_sci, ras |
| РУДН | friend, rudn |
| РАНХиГС | ranepa, russian_acad_natl_econ_publ |
| РГГУ им. Плеханова | plek |

| Организация | Ключевые слова |
|-------------|-----------------------|
| МГИМО | mgimo, inst_int_relat |

На этом этапе достигается наибольшее падение в количестве уникальных аффилиаций в базе данных. Финальным штрихом в автоматизированной обработке организаций стало приведение к однородному написанию всех государственных организаций и, в частности, министерств. После обработки все аффилиации с министерствами записываются в однородном формате: строки начинаются с ‘russian_minist’. Это также позволило идентифицировать и устраниить ряд дубликатов.

Дальнейшая работа с аффилиациями требовала экспертного вмешательства. Так, были идентифицированы «подозрительные» аффилиации, которые были затем проанализированы вручную. К «подозрительным» были отнесены аффилиации, длина которых не превышает 5 символов, а также содержащие слова «faculty», «fac», «dept», «school», «inst» с целью обнаружения случаев, в которых в аффилиации сохранилось только подразделение, а не основная организация. Аналогично, ручной обработки требовало сопоставление не англоязычных аффилиаций с англоязычными: в нашей базе данных присутствуют названия организаций не только на английском, но и на испанском, итальянском и немецком.

По итогам обработки удалось сократить количество уникальных аффилиаций в выборке с 1644 до 1309 (на 21%).

6.1 4.2.1 Библиометрический сетевой анализ коллабораций российских социологов на материалах Web of Science

Включенность в международное исследовательское сообщество является важной предпосылкой становления и развития исследовательских институтов и научных школ. Конкурентоспособность научных коллективов, работающих в рамках любой научной дисциплины, во многом зависит от сотрудничества как внутри страны, так и на международном уровне.

Библиометрические исследования на протяжении всей истории их существования фиксируют тенденцию к увеличению научной коллаборации (Shrum et al., 2007), в том числе коллаборации международной. Такой анализ не только позволяет узнать, каким будет облик науки будущего, но и оценить рост влияния совместной работы ученых на перспективы того или иного научного сообщества. Все это делает изучение коллабораций ученых в контексте влияния их публикаций крайне актуальным. Наиболее распространенный и эффективный метод анализа научных коллабораций – сетевой анализ сетей соавторства (Lundberg et al., 2006), данные для которых могут быть получены из наукометрических баз данных (Pike, 2010) (Scopus, WoS, eLibrary и др.).

И.Н. Трофимова проанализировала на основе базы данных Web of Science публикации российских ученых с 2018 по 2022 годы, фокусируясь на международной коллаборации российского научного сообщества. В частности, она отмечает рост числа публикаций, написанных российскими учеными в соавторстве с иностранными коллегами, происходящий на фоне снижения влияния российских публикаций в мировых масштабах (Трофимова, 2023). Также данное исследование подтверждает положительную связь между международным соавторством и цитируемостью публикаций (треть цитируемых публикаций российских ученых написаны в соавторстве с иностранцами, эти публикации

чаще выходили в журналах Q1). География международного соавторства российских ученых в большей степени определяется исторической развитостью научных центров и объемами финансирования, чем территориальным расположением государств и культурной близостью (наибольшее число иностранных соавторов российских ученых из США, стран Европы, Китая и Японии, а научное сотрудничество со странами СНГ менее продуктивно).

Х.Ф. Моэд, В.А. Марсукова и М.А. Акоев провели сравнительное исследование трендов публикационной активности российских ученых на основе баз библиометрических данных Web of Science и Scopus (Moed et al., 2018). Анализ показал сильную разницу в оценках роста числа российских публикаций и их влияния. Авторы исследования пришли к выводу, что на положительную динамику российского научного вклада “наложились” изменения числа русскоязычных изданий, включенных в базы Web of Science и Scopus, что вызывает трудности в ее оценке и говорит о необходимости учитывать особенности каждой из баз данных для построения валидных выводов при работе с ними.

В практике оценки продуктивности научного сообщества растет важность как самих методов наукометрического и библиометрического анализа, так и баз данных (БД), которые являются основными поставщиками библиографических метаданных о публикационной активности исследователей. Как правило, библиографические базы данных Web of Science (WoS) и Scopus определяются в качестве наиболее полных источников данных для различных аналитических целей (Zhu & Liu, 2020). Несмотря на то, что две основные специализированные БД – WoS и Scopus – по-прежнему считаются наиболее надежными источниками библиографических данных, именно WoS все же рассматривается как «золотой стандарт» библиометрического использования (Pranckuté, 2021).

По данным российской наукометрической базы eLibrary, средняя цитируемость статей, опубликованных в WoS, включенных в ядро Российского индекса научного цитирования (РИНЦ), отличается от аналогичных статей из Scopus – в WoS она в 1,25 раз выше, чем у статей в Scopus, в 9,3 раза выше, чем в ESCI (Emerging Sources Citation Index), в 6,7 раз выше, чем в RSCI (Russian Science Citation Index). При распределении на квантили средняя цитируемость статей в WoS Q1 в 1,36 раз выше Scopus Q1, в 28,2 и 20,3 раза выше, чем в ESCI и RSCI соответственно. Средняя цитируемость российских статей в WoS Q4 в 2,7 раз выше, чем в Scopus Q4, в 4,4 и 3,1 раза выше, чем в ESCI и RSCI соответственно (eLibrary.ru, n.d.). Преимущества базы данных WoS как с точки зрения публикующихся авторов, так и с точки зрения качества метаданных делают ее наиболее подходящим источником для получения валидных результатов библиометрического анализа.

Настоящее исследование посвящено библиометрическому анализу международного измерения публикаций российского социологического сообщества на основе материалов базы Web of Science за 1992-2022 гг. Основной акцент в работе сделан на сетях соавторства в интересах выявления уникальных паттернов коллабораций российских социологов через публикации, индексируемые в наукометрической базе WoS. Подмножество по социологии было выделено на основе категории SC (SC = Sociology, категория исследования/research area в WoS) и состоит из 7915 публикаций со спецификацией поля данных «CU=(Russia)» в режиме full record.

Всего было проанализировано 7915 публикаций всех типов (статьи, монографии, конференционные публикации и др.) из 172 изданий, опубликованных за период с 1992 по 2022 (май). Для целей данного исследования авторы не исключали никакие публикации из набора данных для того, чтобы все публикации, проиндексированные в WoS, попали в анализ. Более 40% публикаций не имели соавторов.

Такое значительное количество работ без соавторов предполагает, что в российском социологическом сообществе довольно значительное число авторов работает индивидуально, не вступая ни в какие коллаборации. На каждую публикацию в среднем приходилось 1,57 соавтора, 1,419 цитирований, а возраст всех публикаций в среднем составил 12,7 лет. Доля международного соавторства составила 4,611%.

Зачастую для библиометрических исследований сетей характерно применение комбинаций программных продуктов, которые можно использовать алгоритмически комплементарно друг другу. Выбор комбинаций программ для анализа часто также зависит от исследовательского вопроса. В данной работе построение библиометрических сетей и проведение библиометрического анализа осуществлялись при помощи нескольких программных продуктов – VOSviewer, Pajek и R (библиотека bibliometrix/biblioshiny).

В анализе цитирования представлены как зарубежные, так и российские исследователи. Это предполагает, что в российском социологическом сообществе сформировались отечественные научные школы, заметные международному исследовательскому сообществу. Однако список топ-журналов, где публикуются российские социологи, довольно ограничен. На Рис. 4 представлена сеть цитирований с источниками (изданиями), где публикуют работы эти авторы. На первом месте с большим отрывом стоит журнал «Социологические исследования».

Сеть цитирований по источникам (изданиям) публикаций

Figure 4: Сеть цитирований по источникам (изданиям) публикаций

В Таблице ?? приведены данные по топ-10 изданиям по показателям общего количества публикаций, цитированию и общей силы связей (total link strength). Обращает на себя внимание, что в топ-10 источников по количеству опубликованных документов вошли международные конференции, которые готовили публикационные материалы как журналы (по факту те же самые сборники конференционных материалов) – Social and Cultural Transformations in the Context of Modern Globalism (European Proceedings of Social and Behavioural Sciences) или International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM 2016 (SGEM Conference Proceedings). Можно сделать вывод, что гонка 2010-х за продуктивностью международных публикаций, с которой пришлось столкнуться российскому академическому сообществу, нашла свое отражение в списке источников публикаций, где среди топ-источников оказались представленными сборники материалов конференций, которые не имеют эквивалентного reputационного веса по сравнению с академическими журналами. В этом контексте будет полезным сравнение не только по количеству опубликованных документов, но также по количеству цитирований и метрике «Общая сила связей» (total link strength). В VOSviewer для элемента сети учитывается количество связей элемента с другими элементами (links) и общую силу связей элемента с другими элементами (total link strength). Например, в случае сетей соавторства, авторы, имеющие одинаковое число соавторов, будут иметь один и тот же показатель связей (буквально, сколько у них было коллокаций). Если же один из них будет чаще публиковаться совместно с кем-либо, но число его соавторов будет неизменным, то показатель общей силы связи у него будет выше, чем у другого исследователя. Таким образом, показатель общей силы связи учитывает не только наличие совместных публикаций, но и интенсивность соавторства, что позволяет получить более точные выводы относительно статуса ученых в сети.

| НПП | Публикации, | | Цитирования, | | Общая сила связей | |
|-----|--|------|---|------|--|-----|
| | Журнал | шт. | Журнал | шт. | Журнал | |
| 1 | Социологический исследования | 4922 | Социологический исследования | 5525 | Социологический | 260 |
| 2 | Вестник Российского университета дружбы народов. Серия: Социология | 679 | Social Indicators Research | 690 | Социологический обозрение / Экономическая социология | 129 |
| 3 | Экономическая социология | 540 | International Journal of Intercultural Relations | 557 | Вестник Российского университета дружбы народов. Серия: Социология | 47 |
| 4 | Социологический обозрение | 517 | Социологический обозрение | 396 | Социология науки и технологий | 38 |
| 5 | Социальные и культурные трансформации в контексте современного глобализма (конференция, European Proceedings of Social and Behavioural Sciences) | 322 | Экономическая социология | 394 | Current Sociology | 19 |

| НПП | Журнал | Публикации, | | Цитирования, | | Общая сила связей |
|-----|---|-------------|---|--------------|--|-------------------|
| | | шт. | Журнал | шт. | Журнал | |
| 6 | Социология науки и технологий | 195 | Вестник Российской Федерации по дружбы народов. Серия: Социология | 270 | Comparative Sociology | 16 |
| 7 | Changing Societies & Personalities | 85 | Population and Development Review | 181 | International Journal of Sociology and Social Policy | 12 |
| 8 | International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM 2016 (Psychology and Psychiatry) | 78 | Ethics and Racial Studies | 115 | American Sociologist | 11 |
| 9 | International Journal of Sociology and Social Policy | 41 | Annals of Tourism Research | 106 | Critical Sociology | 9 |
| 10 | Comparative Sociology | 29 | European Societies | 103 | Filosofija. Sociologija | 7 |

На Рисунке 5 представлен топ списка самых цитируемых публикаций исследователей (локальное цитирование и глобальное цитирование). Глобальное цитирование означает общее количество цитирований, которое статья, включенная в коллекцию, получила из документов, проиндексированных в библиографической базе данных в целом. Среди глобальных цитирований в основном представлены статьи, опубликованные в международных журналах. Локальные цитирования получены по публикации «внутри коллекции» (массива данных). Среди локально цитируемых публикаций в основном

представлены статьи из журнала «Социологические исследования».

Топ цитируемых публикаций (вверху – локальные цитирования, внизу – глобальные цитирования)

Figure 5: Топ цитируемых публикаций (вверху – локальные цитирования, внизу – глобальные цитирования)

Библиометрические сети соавторства являются одними из основных видов сетей в библиографическом анализе и выражают определенный тип взаимосвязей между элементами изучаемого нами пространства. Сеть соавторства, как гласит название, отражает связи совместного участия агентов (исследователей, организаций, стран) в производстве академических публикаций. В этой сети узлами выступают авторы, а связь между ними отражает частоту, с которой они совместно публиковали статьи. Благодаря рассмотрению сетей соавторства мы можем оценить структуру научной коллaborации, выявить ключевых и периферийных акторов этого процесса. Сеть соприсутствия ключевых слов позволяет нам картировать тематический ландшафт академического поля, выявить приоритетные и популярные темы исследований, а также то, на что исследователи обращают меньше внимания, либо не обращали его вовсе. Технически это осуществляется благодаря подсчету частоты, с которой термины одновременно встречаются в обозначенном поле библиографических данных.

В случае обоих сетей связь между элементами оценивается благодаря совместному подсчету авторства либо ключевых слов. Этот подсчет может быть полным либо фракционным. При полном подсчете мы считаем, что каждая связь между узлами сети имеет вес, равный числу документов, которое они опубликовали вместе (сеть соавторства) либо где они встречались вместе (сеть соприсутствия ключевых слов). Например, если 5 авторов выпустили одну публикацию, вес каждой их связи друг с другом равен 1; либо если в одной публикации встречаются 5 ключевых слов, вес связей между ними также будет 1. При фракционном подсчете вес связей будет определяться обратно числу узлов. Теперь каждый из соавторов нашей публикации будет связан друг с другом с весом $\frac{1}{n}$, как и каждый термин будет связан с другим весом $\frac{1}{n}$. Разница кажется небольшой, однако использование полного подсчета, который делается во многих исследованиях по умолчанию, приводит к значительному (практически квадратичному) увеличению числа связей с ростом числа авторов, что существенно искажает реальную картину научных коллабораций (Perianes-Rodriguez et al., 2016). В связи с этим, в нашем анализе мы используем фракционный подсчет для построения сетей соавторства.

Наш массив данных состоит из 6765 авторов и 1664 организаций. За весь изучаемый период только 28% ученых опубликовали 2 и более работ (3 и более – 14%, 4 и более – 8%). Далее представлен топ-10 авторов по числу публикаций, числу цитирований и метрике общей силы связей (Таблица ??). Активно публикавшиеся авторы также, в основном, имеют сильные связи с другими авторами. Среди наиболее цитируемых авторов, однако, немного исследователей из России. Это обстоятельство объясняется международной кооперацией – многие из представленных ниже активно цитирующихся международных авторов работали с российскими коллегами и публиковались в российских журналах, индексируемых Web of Science.

| НПП | Автор | Публикации, | | Цитирования, | | Общая сила связей |
|-----|--------------------|-------------|----------------------|--------------|--------------------|-------------------|
| | | шт. | Автор | шт. | Автор | |
| 1 | Троцук И.В. | 69 | Инглахрт Рональд | 575 | Троцук И.В. | 33 |
| 2 | Тощенко Ж.Т. | 53 | Вельцель Кристиан | 532 | Зборовский Г.Е. | 21 |
| 3 | Кравченко С.А. | 44 | Делхи Ян | 407 | Голенкова З.Т. | 21 |
| 4 | Радаев В.В. | 41 | Ньютон Кеннет | 397 | Пузанова Ж.В. | 21 |
| 5 | Зборовский Г.Е. | 36 | Шмит Петер | 393 | Нарбут Н.П. | 20 |
| 6 | Пузанова Ж.В. | 34 | Давидов Эльдад | 318 | Тощенко Ж.Т. | 17 |
| 7 | Барсукова С.Ю. | 34 | Берри Джон | 214 | Игитханян Е.Д. | 14 |
| 8 | Лапин Н.И. | 33 | Барсукова С.Ю. | 154 | Коротаев А.В. | 13 |
| 9 | Горшков М.К. | 30 | ван де Вийер Фонс | 125 | Ларина Т.И. | 13 |
| 10 | Голенкова З.Т. | 28 | Кравченко С.А. | 124 | Иванов В.Н. | 13 |

Паттерн представленности организаций примерно соответствует представленности ученых. Из 1664 институций лишь 33% выпустили 2 и более публикации. Наиболее активно выпускавшая публикации РАН тесно соседствует с ВШЭ, тогда как следующий за ними по числу публикаций вуз, РУДН, имеет на 77% меньше публикаций, чем в среднем выпустили ВШЭ и РАН (Таблица ??). Из числа региональных ВУЗов лишь УрФУ им. Ельцина попал в топ списка организаций. Также обратим внимание на то, что в топе присутствует ЕУСПб, который по своим размерам значительно уступает всем остальным.

В разрезе цитирований из топа пропадают УрФУ, МГИМО и РУДН. ВШЭ поднимается на первое место, опережая РАН на 30%. ЕУСПб также практически вплотную соседствует с МГУ и РГГУ в середине списка, а СПбГУ соперничает с Бременским университетом Якобса.

Наиболее сильными академическими связями обладает РАН. На 40% меньшую силу связей имеет ВШЭ, остальные близко находящиеся к ним организации (МГУ, РУДН, СПбГУ, РАНХиГС) имеют примерно на 80% менее сильные связи. Обращает на себя внимание то, что в топе присутствует два чеченских университета (ЧГУ и ЧГПУ), причем один из них (ЧГУ) имеет более сильные связи, чем РГГУ и УрФУ. В рамках институционального ландшафта коллaborации ЧГУ и ЧГПУ хронологически являются довольно молодыми.

| НПП | Организация | Публикации, | | Цитирования, | | Общая сила связей |
|-----|-------------|-------------|--|--------------|---|-------------------|
| | | шт. | Организация | шт. | Организация | |
| 1 | РАН | 1943 | ВШЭ | 3653 | РАН | 404 |
| 2 | ВШЭ | 1081 | РАН | 2563 | ВШЭ | 247 |
| 3 | РУДН | 354 | СПбГУ | 468 | МГУ | 81 |
| 4 | СПбГУ | 337 | Бременский университет Якобса | 412 | РУДН | 73 |
| 5 | МГУ | 302 | МГУ | 330 | СПбГУ | 63 |
| 6 | УрФУ | 164 | ЕУСПб | 306 | РАНХиГС | 61 |
| 7 | РГГУ | 140 | РГГУ | 242 | Чеченский государственный университет (ЧГУ) | 55 |
| 8 | РАНХиГС | 128 | Университет Куинс | 228 | РГГУ | 39 |
| 9 | ЕУСПб | 122 | Институт демографических исследований им. Макса Планка | 195 | Чеченский государственный педагогический университет (ЧГПУ) | 30 |
| 10 | МГИМО | 93 | РАНХиГС | 187 | УрФУ | 27 |

При рассмотрении сети соавторства всего 394 автора соответствуют критерию в минимум 5 публикаций (Рис. 6). В этой сети один значительный компонент (связанный подграф, 28% выборки), а также несколько не связанных с данным компонентом более мелких. Этот компонент представляет из себя ядро сети, и далее мы разберем его более подробно.

Сети соавторства коллабораций российских социологов за период 1992-2022 (фракционный счет, слева барьер – 5 и более публикаций, справа – 15 и более публикаций)

Figure 6: Сети соавторства коллабораций российских социологов за период 1992-2022 (фракционный счет, слева барьер – 5 и более публикаций, справа – 15 и более публикаций)

При минимальном ограничении в 5 работ, опубликованных за 30 лет индексирования в WoS (5% выборки), основной компонент сети представляет из себя сеть с одним основным ядром и разветвленной периферией, которая либо находится близко к центру сети, либо слабо связана с ним единственным «маршрутом» (Рис. 7). В центре основного ядра присутствуют наиболее известные и цитируемые социологи (Ж.Т. Тощенко, З.Т. Голенкова, Иванов В.Н., Рукавишников В.О., Игитханян Е.Д., Горшков М.К. и др.), которые находятся друг от друга на определенном отдалении и замыкают на себя слабее связанных авторов.

Если же мы строим сеть соавторства только для тех, кто выпустил за 30 лет как минимум 15 работ, в сеть попадают лишь 49 авторов (0,7% выборки), а основной компонент состоит лишь из 12 человек (Рис.

Сети соавторства российских социологов - наибольший связанный компонент сети при наличии у авторов минимум 5 публикаций за период 1992-2022

Figure 7: Сети соавторства российских социологов - наибольший связанный компонент сети при наличии у авторов минимум 5 публикаций за период 1992-2022

8). Сюда попадают социологи из ядра предыдущей сети. Ядро представляют авторы, соединяемые Ж.Т. Тощенко, у которого, опять же, самая разветвленная сеть. Однако в отличие от предыдущей сети, здесь сила связи исследователей примерно одинаково низкая, за исключением Голенковой З.Т. и Игитханян Е.Д. Они являются наиболее интенсивно кооперирующими друг с другом социологами, к тому же З.Т. Голенкова связывает между собой два участка данной сети. Социально-экономические исследователи из ФНИСЦ РАН и ВШЭ в данном случае находятся на периферии главного связанного компонента сети соавторов.

Сети соавторства российских социологов - наибольший связанный компонент сети с барьером отсечения в 15 публикаций за период 1992-2022

Figure 8: Сети соавторства российских социологов - наибольший связанный компонент сети с барьером отсечения в 15 публикаций за период 1992-2022

Таким образом, при анализе сети соавторства мы можем четко выделить ядро научной коллаборации, которое в свою очередь представлено отдельными центрами притяжения. Эти группы могут быть объединены либо вокруг конкретных персоналий, организаций, либо тематик исследований. Многие из выделенных центров притяжения сохраняются при отсечении менее продуктивных (в смысле международно рецензируемых в WoS публикаций) исследователей. Это говорит об отчетливой поликентричности такой сети и сниженной кооперации (опять же, исключительно в смысле соавторства) между более “плодовитыми” социологами. Отметим, что в целом количество продуктивных авторов и коллокации является не очень большим – для сети соавторства скорее характерны индивидуальные работы, что может быть признаком специфических паттернов исследовательской работы с сфере социологии. Сеть достаточно фрагментарна и представлена относительно небольшим ядром коллокации соавторов, среди которых международных участников нет.

Исследователи в целом отмечают рост публикационной активности в российской науке (Трофимова, 2023), при этом соотношение долей коллокационной активности характеризуется перераспределением долей коллокаций – доля национальных коллокаций растет, в то время как доля международных снижается (Moed et al., 2018). Для социологического сообщества также характерен рост публикационной продуктивности. Однако характеристикой коллокаций социологов является ориентация на внутренние, российские коллокации или индивидуальную работу. В нашем случае встают два вопроса: (1) какое критериальное количество публикаций может демонстрировать индивидуальную научную продуктивность для представленного набора данных публикаций по социологии в WoS; (2) какое количество публикаций в соавторстве можно рассматривать как продуктивную научную коллокацию, устойчивую во времени.

Индекс коллокативности авторов, посчитанный на основе построенных сетей соавторства в программе Rajeck, подтверждает структурную разрозненность и относительно невысокую склонность к выстраиванию коллокаций⁴. Индекс рассчитывается как единица минус отношение общего

⁴Данный индекс был рассчитан в программе Рајек, данные по количеству публикаций топ-авторов могут отличаться от данных,

фракционного вклада автора в свои работы к общему количеству публикаций и показывает тенденцию автора к работе с другими авторами (Таблица ??). Полученные результаты соотносятся с анализом сетей соавторства, представленных графически на Рисунках 7 и 8, – даже высокопродуктивные авторы могут иметь низкий уровень коллаборационной активности. У автора может быть значительное количество публикаций, но он может работать индивидуально или с очень узким кругом соавторов.

| Автор | Общий вклад автора | Количество публикаций | Индекс коллаборативности |
|-------|--------------------|-----------------------|--------------------------|
| 1 | TROTSUK_I | 49,08 | 68 |
| 2 | TOSHCHEN_Z | 36,97 | 49 |
| 3 | KRAVCHEN_S | 35,50 | 38 |
| 4 | #RADAEV_V | 31,58 | 36 |
| 5 | #YANITSKI_O | 35,00 | 35 |
| 6 | ZBOROVSK_G | 24,17 | 35 |
| 7 | LAPIN_N | 26,28 | 33 |
| 8 | PUZANOVA_Z | 16,53 | 33 |
| 9 | IVANOV_V | 21,19 | 32 |
| 10 | ROMANOVS_N | 25,28 | 29 |
| 11 | GORSHKOV_M | 23,48 | 29 |
| 12 | GOLENKOV_Z | 13,51 | 27 |
| 13 | BARSUKOV_S | 22,33 | 25 |
| 14 | LEVASHOV_V | 21,21 | 25 |
| 15 | TIKHONOV_N | 21,17 | 25 |
| 16 | NARBUT_N | 12,53 | 25 |
| 17 | FILIPPOV_A | 20,00 | 22 |
| 18 | SOKOLOV_M | 18,33 | 22 |
| 19 | TESLYA_A | 21,00 | 21 |
| 20 | STEPANOV_E | 15,67 | 21 |

Коллaborации научных организаций (Рис. 9) однозначно показывают два центра притяжения – РАН и ВШЭ. Особенность положения РАН заключается в том, что она не только является центром в и так довольно связанном ядре сети организационной коллаборации (т.е. соединяет сильно связанные институции), но и открывает путь к этим коллаборациям со стороны слабо связанных организаций (справа сверху), которые, к тому же, практически не связаны друг с другом. Специфика ВШЭ состоит в коллаборации с иностранными вузами (например, Университетом Мичигана, Тильбургским университетом и др.). Между этих двух больших организаций находятся более мелкие, однако относительно ближе интегрированные с другими ВУЗы, РАНХиГС и Университет им. Г.В. Плеханова. Мы также можем отчетливо наблюдать географические группировки вокруг СПбГУ и довольно крупный кластер чеченских университетов, которые также соединяются с другими ВУЗами южных регионов

полученных в VOSviewer, так как файлы для Pajek создаются с помощью программы WoS2Pajek, которая использует встроенные алгоритмы статистической обработки данных. В целом количественно данные по топ-авторам отличаются несущественно, что позволяет проводить сравнения разных метрик. В таблице жирным шрифтом отмечены авторы с самым высоким индексом коллаборативности.

России.

Сети соавторства организаций

Figure 9: Сети соавторства организаций

Динамически картина организационных коллабораций характеризуется преобладанием сначала РАН, потом ВШЭ, а затем региональных вузов в пространстве публикационных коллабораций. В начале 2010-х было характерно преобладание традиционно крупных московских организаций (РАН, МГУ). Затем к ним (из крупных) добавились РУДН, СПбГУ, РГГУ, Плехановский университет, после 2015 г. – ВШЭ и иностранные университеты, и уже после 2017 г. РАНХиГС. Совершенно новые организации на академическом ландшафте, которые появились в районе 2020 г. и позже – это чеченские ВУЗы, хотя отдельные южные университеты начали свою активную деятельность гораздо раньше даже крупных московских организаций, обозначенных выше. В анализе публикационной активности в хронологическом аспекте также обращает внимание на себя тот факт, что РАН (Институт социологии) показывал положительную динамику роста по публикациям за весь период с 1992 года, в то время как начало роста публикаций по университетам-лидерам приходится на конец 2000-х – начало 2010-х гг.

В общей сложности выделены 63 страны, с которыми сотрудничают российские коллективы социологов, однако только 27 стран удовлетворяют требованию наличия в выборке минимум 5 публикаций. В топ-5 входят США, Германия, Великобритания, Италия, Нидерланды, но при этом 90% соавторства документов принадлежат России. Эти выделенные топ-5 стран представляют «традиционную» географию сотрудничества. Условная «новая» география сотрудничества включает Китай, Швейцарию, Австралию, Швецию, Испанию и другие страны.

При построении сети соавторства из всех стран за весь временной период (30 лет) обращает на себя внимание следующая особенность. Коллаборации по странам можно отнести к 2 категориям: двусторонние отношения (правая часть Рис. 6.1) и многосторонние коллаборации (левая часть Рис. 6.1), куда относятся как раз страны «традиционной» географии. Такие многосторонние коллаборации, безусловно, имеют больший потенциал охвата научного пространства, больше возможностей для привлечения новых участников коллабораций и более высокую публичность.

Сети коллабораций российских социологов по страновой принадлежности Анализ соавторства публикаций показал, что социологическое сообщество достаточно неоднородно, большое количество авторов не входит в ядро коллабораций. Также публикационная активность авторов весьма невысокая – критерию порога в 5 и более статей в выборке (за 1992-2022 гг., все типы публикаций) соответствуют только 394 из 6765 авторов. Значительное количество авторов работает индивидуально, а имеющиеся научные коллаборации ограничены устоявшимися коллективами из ведущих научно-образовательных организаций.

В общей сложности выделены 63 страны, с которыми сотрудничали российские коллективы социологов за период 1992-2022 гг. Первой особенностью международных коллабораций российских социологов является декомпозиция на «традиционную» и «новую» географии сотрудничества. Вторая особенность коллабораций – это количество стран-участников. С рядом стран выстраиваются только двусторонние коллаборации, а с другими российские социологи участвуют в многосторонних коллаборациях.

Анализ соавторства организаций продемонстрировал модель сотрудничества «ядро-периферия», где

ядро представлено коллаборациями двух доминирующих организаций – Российской академии наук и Высшей школы экономики. Также данная модель сотрудничества характеризуется наличием группы традиционно представленных в коллаборациях институтов в силу своей истории, репутации и географии (в основном Москва и Санкт-Петербург), а также присутствием относительно новых участников, что может отражать институциональную трансформацию научного ландшафта в связи с изменениями в национальной образовательной и исследовательской политике.

Международные коллаборации российских социологов малочисленны и в основном представлены российскими авторами с незначительным участием зарубежных ученых. Первым фактором, ограничивающим включенность в международные коллаборации, является языковой фактор (84,37 % публикаций представлены на русском языке). Вторым важным фактором является особенность выстраивания коллабораций – либо склонность к индивидуальной работе, либо сотрудничество с отечественными исследователями.

6.2 4.2.3 Картирование научного поля: применение VOSviewer и Biblioshiny на материалах Web of Science

Анализ соприсутствия ключевых слов (keywords co-occurrence) позволяет картировать тематические кластеры ключевых слов публикаций – построить карты (сети) ключевых слов. Соприсутствие ключевых слов показывает, как соотносятся друг с другом библиометрические объекты (ключевые слова) на основе документов, в которых они одновременно присутствуют (соприсутствуют). Если ключевые слова не указаны автором публикации, то они могут быть присвоены журналом, базой данных или автоматически извлечены из заголовка, что позволяет обозначить тематическую направленность на основе метаданных академической работы (Maltseva & Batagelj, 2020). В библиометрических исследованиях анализ соприсутствия ключевых слов является весьма популярным самостоятельным подходом, часто определяемым как картирование структуры знаний по соответствующему научному направлению (Павлова, 2023).

В нашем случае картирование сетей соприсутствия ключевых слов производилось для отдельных научных областей из собранных данных Web of Science. Цель данного этапа, как и этапа обработки данных, в первую очередь была связана с определением наиболее удобного и наглядного отображения существующей тематической структуры разных дисциплин. Во вторую очередь мы попытались выделить содержательные категории, в которые объединяются встречающиеся в публикациях термины, проанализировать эволюцию популярности тех или иных терминов, выявить основные тематические тренды, а также оценить статус тех или иных тематик в ракурсе (бес)перспективности их разработке в текущем научном дискурсе. Обозначим заранее, что данная предварительная работа не является междисциплинарным анализом в полном смысле слова. Она проводилась в рамках дедуктивно определенных дисциплин (в частности, политологии и социологии) и не включает в себя перекрестные тематические совпадения между дисциплинами (например, такие точно есть между политологией и социологией).

С технической точки зрения картирование научного ландшафта осуществлялось с помощью программного обеспечения VOSviewer и biblioshiny (часть пакета bibliometrix на языке R). VOSviewer (www.vosviewer.com) – программа, разработанная ученым в Лейденском университете (Королевство

Нидерланды) специально для построения библиометрических сетей. Разработчики программы предложили метод VOS (visualization of similarities) – визуализации сходств между объектами при построении библиометрических карт (сетей) на основе расстояний между этими объектами, которые отражают силу связи между элементами (Van Eck & Waltman, 2010). В нашем случае для работы в этой программе была сделана предобработка данных в Python с доработкой в ручном режиме, подробно освещенные в третьей части отчета. Корректировки уникальных имен (ФИО) ученых, а также названий организаций вносились через тезаурусы (списки с правилами замены встречающихся имен в метаданных на пользовательские). С помощью VOSviewer удалось произвести качественные визуализации сетей соприсутствия ключевых слов, а также выделить кластеры терминов, тематически связанных друг с другом.

В дополнение к VOSviewer мы также использовали программу biblioshiny из пакета bibliometrix. Эта программа, разработанная учеными Неаполитанского университета (Италия), предназначена для систематического анализа как ключевых слов, так и связей между учеными, институциями, и в целом не ограничивается одним лишь сетевым анализом (Aria & Cuccurullo, 2017). С помощью biblioshiny нам удалось произвести общий дескриптивный анализ публикационной активности российских ученых и университетов, а также осуществить довольно подробное первичное картографирование научного ландшафта как в разрезе «иерархии» тех или иных тем (популярность-нишевость), так и проследить изменение статуса данных тематик во времени.

Для начала представим самые общие сведения о тематиках исследований. С помощью VOSviewer были составлены топ-10 самых часто упоминаемых и наиболее важных с точки зрения соприсутствия ключевых слов. Данные списки не совпадают до конца, поскольку одного лишь упоминания в разделе «ключевых слов» недостаточно, чтобы можно было считать ту или иную тематику популярной – соответствующее ключевое слово должно не просто встречаться в большом числе публикаций, но и активно присутствовать во взаимосвязи со многими другими ключевыми словами. Чем больше терминов, с которыми соприсутствует определенное ключевое слово, и чем чаще оно с ними соприсутствует, тем более достоверно можно говорить, что данная тематика встроена в активный поток научных разработок, причем во многих областях. Чтобы учесть такое положение ключевых слов, в VOSviewer применяется показатель «общей силы связи» (total link strength). Мы ранжировали наши ключевые слова как по нему, так и по «сырой» встречаемости.

В таблицах ниже представлены топ-10 ключевых слов из социологии (табл. 1.1) и политологии (табл. 1.2). Ключевые слова выбирались таким образом, чтобы они присутствовали минимум в 15 публикациях – это позволило исключить из анализа редко упоминаемые темы, а также различные варианты написания одного и того же слова. Можно заметить, что термины, которые часто упоминаются, также имеют сопоставимо высокую общую силу связи, однако это верно не для всех ключевых слов. Хотя при ранжировке новых слов не добавляется, можно сказать, что исходя из общей частоты соприсутствия, основные тематики исследований сконцентрированы скорее вокруг молодежи и образования, тогда как исходя из общей силы связей ценности являются более популярной темой, которая широко и интенсивно присутствует в сети.

Таблица 1.1 Топ-10 терминов в анализе соприсутствия ключевых слов (социология)

| НПП | Ключевое слово | Соприсутствие в наборе данных (количество раз) | Ключевое слово | Общая сила связей |
|-----|----------------|--|----------------|-------------------|
| 1 | russia | 178 | russia | 185 |
| 2 | youth | 93 | values | 119 |
| 3 | education | 78 | youth | 115 |
| 4 | sociology | 78 | identity | 110 |
| 5 | identity | 77 | education | 108 |
| 6 | values | 72 | culture | 106 |
| 7 | culture | 71 | gender | 97 |
| 8 | trust | 61 | trust | 82 |
| 9 | gender | 58 | society | 82 |
| 10 | migration | 58 | inequality | 80 |

Похожая ситуация, но в меньшем масштабе характерна и для политологии. Полное совпадение относительно 7 приоритетных тем – «Russia», «democracy», «China», «politics», «state», «elections», «international relations» – сочетается с некоторой вариацией популярности тем «identity», «authoritarianism» и «power», которая, однако, несущественна.

Таблица 1.2 Топ-10 терминов в анализе ключевых слов (политология)

| НПП | Ключевое слово | Соприсутствие в наборе данных (количество раз) | Ключевое слово | Общая сила связей |
|-----|-------------------------|--|-------------------------|-------------------|
| 1 | russia | 385 | russia | 326 |
| 2 | democracy | 142 | democracy | 129 |
| 3 | china | 139 | china | 113 |
| 4 | politics | 103 | politics | 88 |
| 5 | state | 100 | state | 81 |
| 6 | elections | 88 | elections | 74 |
| 7 | power | 68 | identity | 56 |
| 8 | identity | 67 | authoritarianism | 54 |
| 9 | authoritarianism | 58 | power | 54 |
| 10 | international relations | 54 | international relations | 44 |

На рисунках 1.1 и 1.2 представлены визуализации сетей соприсутствия ключевых слов с разбиением на кластеры для социологии и политологии соответственно. Приведем нашу интерпретацию тематических кластеров в социологии:

- *Историко-теоретический* (красный кластер): capitalism, evolution, revolution, ideology, crisis, corruption, democracy, sociology (в различных вариациях), discourse, politics, economy, society, state, Russia, China, Max Weber и т.д.
- *Социально-демографическая_ политика* (зеленый кластер): age, children, health, family, gender, life, women, а также inequality, justice, solidarity, Europe и т.д.

- *Социальные технологии* (синий кластер): mobility, risk, behavior, identity, culture, management, modernization, globalization, innovation, integration, adaptation, tolerance и т.д.
- *Социально-экономический* (желтый кластер): labor, labor market, social structure, work, education, higher education, employment, human capital, precariat, patriotism и т.д.
- *Цифровые технологии* (фиолетовый кластер): internet, (social) media, digitalization, communication, information society, а также social networks, social capital, impact, consumption, civil society, social inequality, trust и т.д.

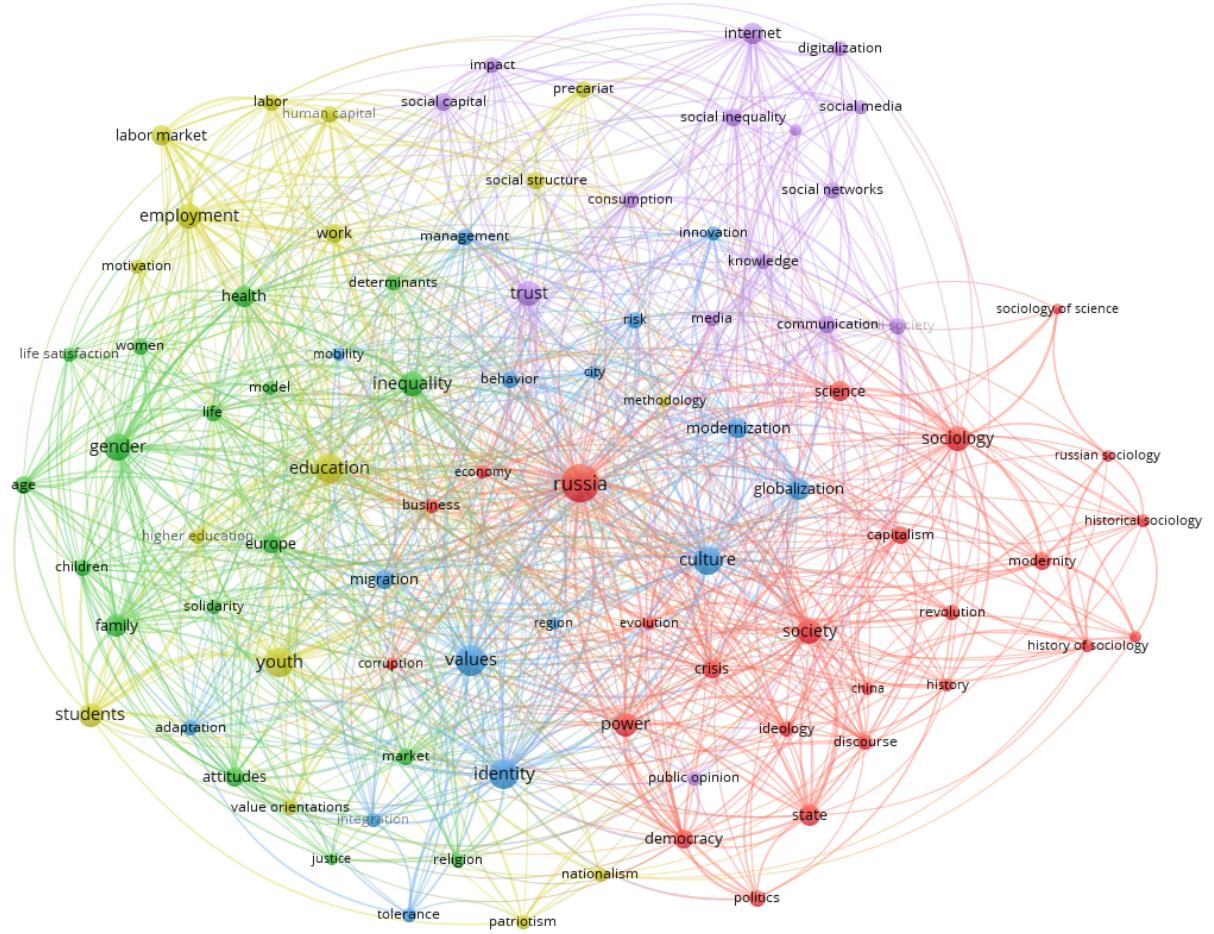


Рис. 1.1. Сеть соприсутствия ключевых слов (социология)

Теперь приведем нашу интерпретацию получившихся тематических кластеров в области политологии:

- *Геополитика и международные отношения* (зеленый кластер): Russia, China, European Union, globalization, geopolitics, regionalism, international relations, BRICS и т.д.
- *Фундаментальные политические явления* (красный кластер): politics, democracy, authoritarianism, institutions, power, governance, corruption, parties, protest и т.д.
- *Социальные институты и процессы* (синий кластер): civil society, state, ideology, values, education, modernization, law, culture, society, cooperation, migration и т.д.
- *Область социально-политической напряженности* (желтый кластер): nationalism, terrorism, war, conflict, Ukraine, crisis, identity, ethnicity, islam, trust, media, public opinion и т.д.
- *Политологический инструментарий* (серый кластер): comparative analysis, methodology, elites.

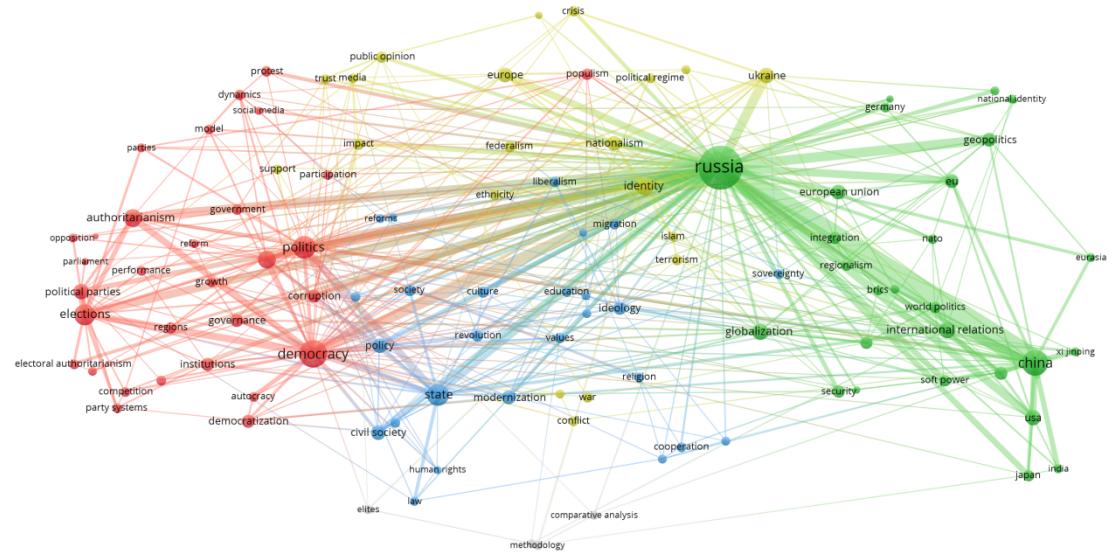


Рис. 1.2. Сеть соприсутствия ключевых слов (политология)

Представленные визуализации позволяют составить первое впечатление о том, в какой контексте встроена та или иная тема, а также увидеть потенциальные области, связанные между которыми пока еще не проработаны в литературе. Например, в случае публикаций по социологии можно увидеть сравнительное отсутствие связей между темами из социально-экономического (желтого) и историко-теоретического (красного кластера), за исключением тем «России» и «образования». Или в случае политологии, достаточно заметно разграничение между областью фундаментальных политических явлений (красный кластер) и геополитикой (зеленый кластер). В случае политологии эти области, помимо России как объекта исследования, связываются через области социально-политической напряженности (желтый кластер) и, в несколько меньшей степени, через изучение социальных институтов и процессов (синий кластер). Для социологии, в данном случае, сравнительно труднее выявить связующие области и разъединенные области, однако сама идея использования сетей соприсутствия для составления впечатления о состоянии той или иной области, согласно результатам нашего анализа, выглядит продуктивной.

Анализ тематических трендов мы выполнили в нескольких видах. Во-первых, это такой же дескриптивный анализ частотности употребления тех или иных терминов (в нашем случае, заголовков, в силу малонаполненности области ключевых слов). Во-вторых, мы обратились к параметрам сетевой центральности и степени для тематических кластеров (так же составленных из заголовков), чтобы количественно оценить степень популярности и «укорененности» тематик публикаций. В-третьих, мы выполнили анализ центральности-степени для публикаций из разных хронологических периодов отдельно, а также визуализировали общую схему эволюции публикационной активности по тематикам (также выделенных с помощью кластерного анализа).

Описание трендов по частотности тех или иных слов в заголовках можно провести как с хронологической точки зрения, так и с точки зрения длительности обращения к той или иной тематике. Так, на примере публикаций по социологии (рис. 2.1) можно выделить (хотя и с оговорками) некоторые тренды конкретного временного периода: например, появление «пандемии» и «ковида» в

публикациях 2020 г., «этнического» в публикациях 2010 г. на волне беспорядков на Манежной площади в Москве, «федерализм» в связи с переустройством федерального устройства России в 2000 г. и т.д. Однако в общем виде нельзя проследить однозначного тематического тренда, т.к. практически все представленные слова из заголовков упоминаются в публикациях практически за все годы, являются долгоиграющими.

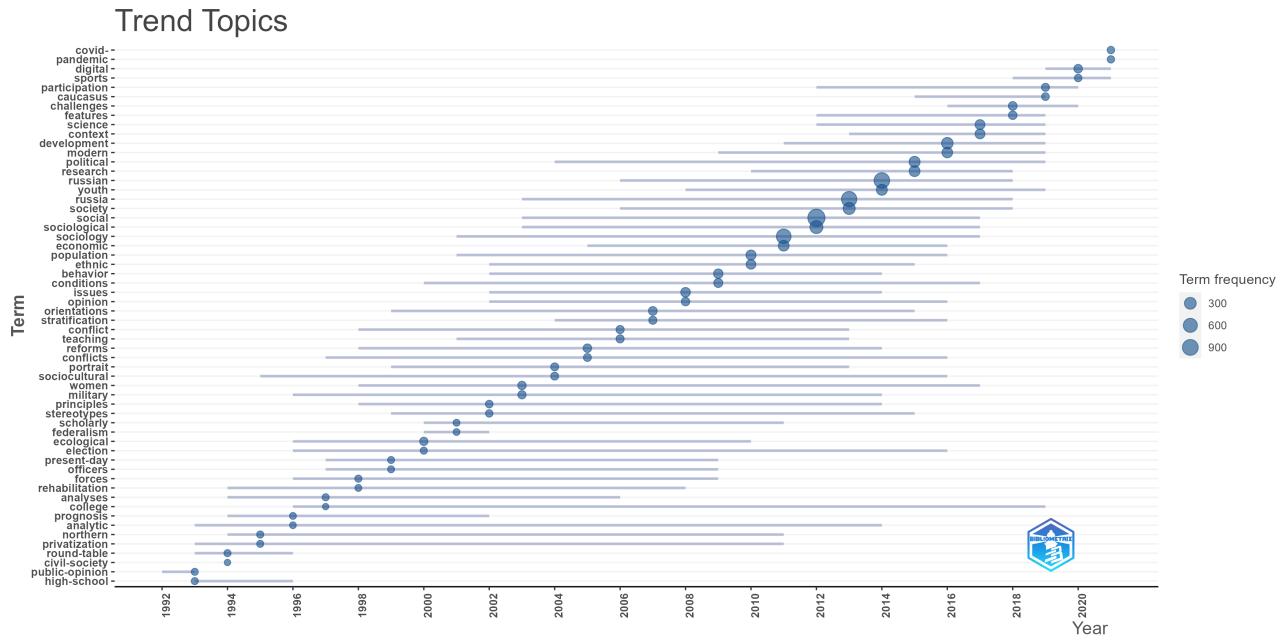


Рис. 2.1 Топ-2 популярных заголовка за каждый год (социология)

Ситуация с трендами в политологии заметно отличается от того, что происходило в социологии. Отчетливо заметно, что порядок частотности слов в целом меньше, чем в социологии, равно как и общее число выделенных терминов, несмотря на то, что для анализа отбирались 3 (а не 2, как для социологии) самых популярных слова из заголовков за каждый год. Отметим также, что использование большей части слов ограничивается серединой 2000-х годов – практически нет примеров появления одной и той же темы, начиная с 1990-х, вплоть до текущего момента. Общая тематическая эволюция показывает, что в 1990-х – 2000-х фокус внимания был сосредоточен на рыночных реформах, глобализации и модернизации России. С 2013 г. отчетливо появляется тренд на национализацию исследовательских тематик (пик частотности отдельных слов максимален в 2014 г.). После этого можно отследить расширение «географического» фокуса в исследованиях, а также фокус на конкретные сферы практической политической деятельности и смежных сферах. Как и в социологии, наиболее популярные слова из заголовков 2020 г. касаются пандемии и ковида.

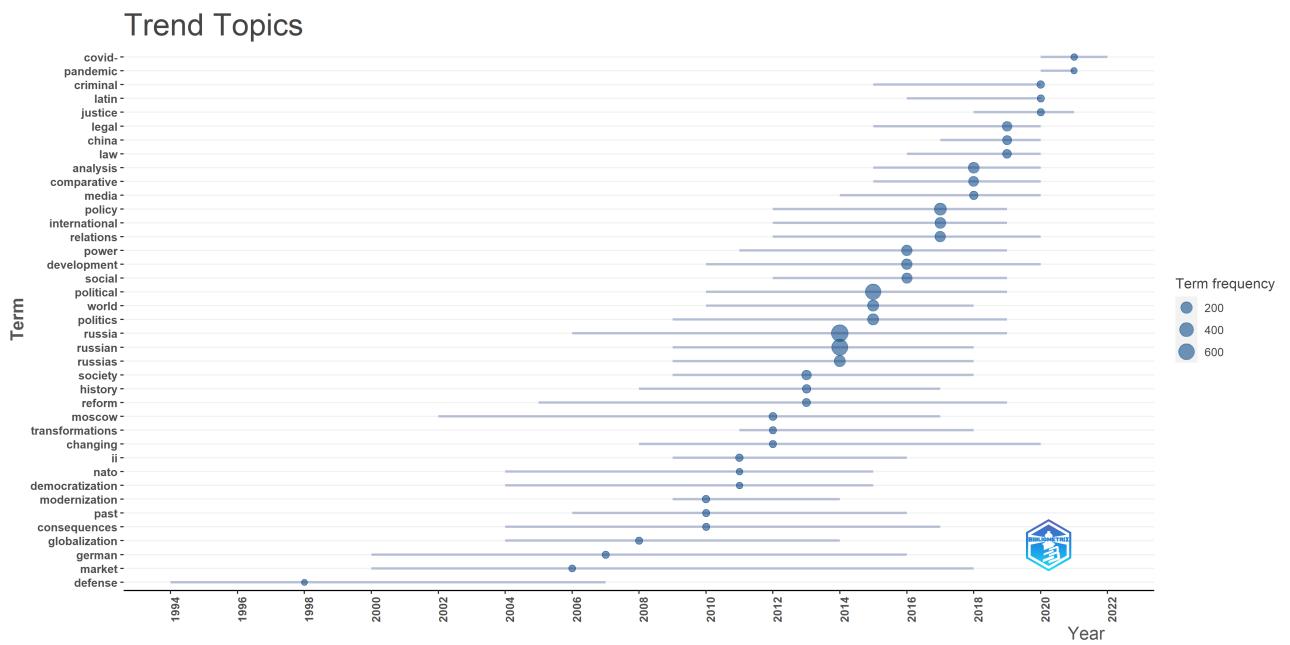


Рис. 2.2 Топ-3 популярных заголовка за каждый год (политология)

Выводы, полученные с помощью анализа частотности упоминания тех или иных слов в заголовках публикаций, являются практическим инструментом для анализа наиболее общих тематических паттернов библиографических записей. Такой анализ дает возможность получить общее представление о языке дисциплины, а также выделить некоторые исторические тренды, связанные с ее развитием. Например, в нашем случае для социологических публикаций характерно продолжительная встречаемость определенных слов на всем протяжении анализируемого периода, что может говорить о систематической роли этих понятий в языке науки. Напротив, в политологии как вариация терминов, так и их встречаемость во времени более ограничены, термины чаще отражают конкретные образования/процессы/акторов, нежели фундаментальные понятия, что также может говорить о специфике развития российской политической науки.

Тем не менее, стоит крайне осторожно относиться к этим выводам ввиду того, что встречаемость слова в заголовках – не прямой результат мотивированных действий авторов, агрегируя которые можно получить общее впечатление о мнениях и вопросах ученых, которые они озвучивают в публикациях. Для подлинно тематического анализа в идеале стоит обращаться к ключевым словам, потому что именно через них авторы определяют смысл своей публикации. Тем не менее, в наших условиях мы не могли провести анализ частотности ключевых слов, поскольку упоминания о них отсутствовали более чем в 50% библиографических записей (как политологических, так и социологических). Причины данного обстоятельства видятся в не проработанности базы данных Web of Science, однако более точный анализ может показать иные результаты.

Далее представим результаты тематической эволюции в публикациях российских социологов (рис. 3.1) и политологов (рис. 3.2). Для проведения данного вида анализа также применялся кластерный анализ, который сгруппировал публикации с тематически схожими заголовками. Темпоральные изменения в кластерах определялись с помощью других сетевых алгоритмов (Cobo et al., 2011). Временные срезы были заданы исходя из динамики публикационной активности в соответствующих дисциплинах как ориентира для общей динамики развития дисциплины.

В российских социологических публикациях стабильно выделилось меньше связанных тематических

групп, чем в политологии. Исторически тематический фокус в социологических публикациях сначала задавался политической сферой (реформы, федерализация и т.д.), затем перешел в область прикладных исследований социальной сферы (название кластера «жизнь»), а также глобальной динамики. После 2014 в исследованиях стабильно присутствовал кластер национально-ориентированных тематик, а также публикаций, сконцентрированных на фундаментальных социологических тематиках.

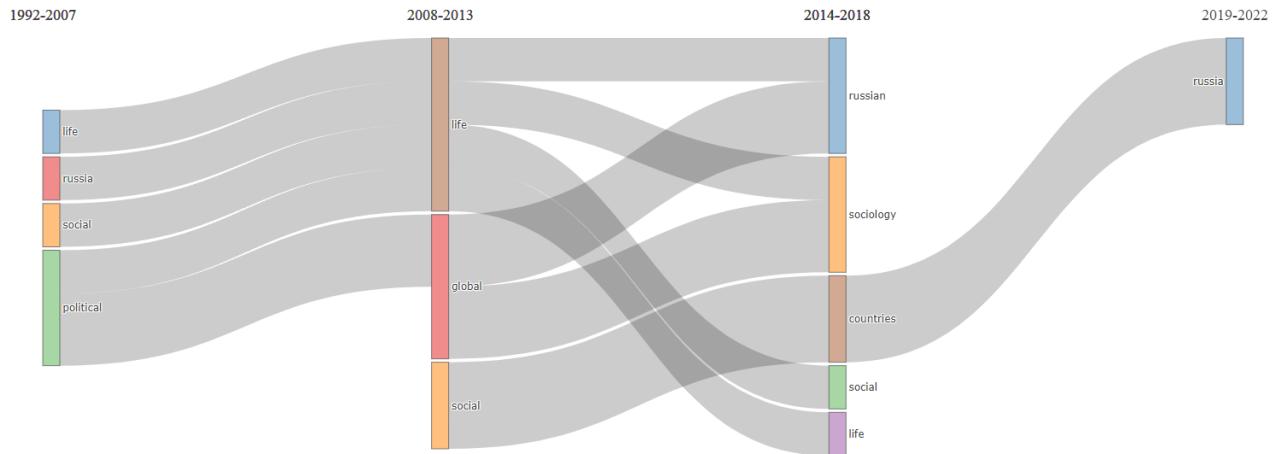


Рис. 3.1 Тематическая эволюция (социология)

В политологии вектор тематической эволюции задали работы по геополитике, а также советскому прошлому и текущим (на тот момент) конфликтам. Период 2005-2013 характеризуется широким разнообразием тематических направлений, начиная от узко-региональных (например, «Кавказ») и заканчивая фундаментальными вопросами политологической теории и политической практики. Интересно, что после 2014 г. многие темы смещаются в электоральную (или, вероятно, прикладную) область, а некоторые из национально-ориентированных тем перетекают в общую категорию «идентичности». Период с 2019 по 2022 гг. характеризуется преобладанием изучения России, а также законодательной сферы (как в России, так и в других странах).

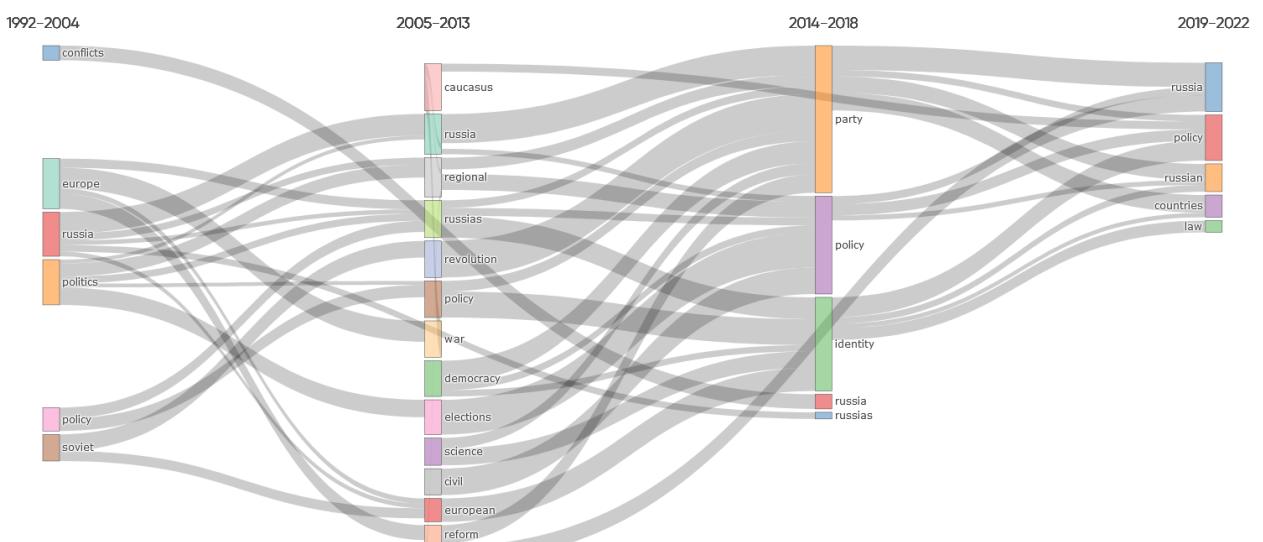


Рис. 3.2 Тематическая эволюция (политология)

Анализ тематической эволюции, в отличие от сравнительно менее изощренного дескриптивного анализа частотных трендов позволяет более глубоко оценить, какие темы и как именно существовали/трансформировались

на протяжении времени. Благодаря использованию методов кластеризации и других сетевых алгоритмов обнаружения схожести между узлами сети, возможно построить схему преемственности публикаций, и, как следствие, отдельных тематик и тематических направлений. Более того, реализация данного подхода в biblioshiny позволяет получить доступ к полностью размеченному массиву данных, который затем можно анализировать с помощью более точных инструментов. Однако следует оговориться, что, несмотря на группировку схожих публикаций, эти кластеры остаются сугубо эвристическими, а потому нельзя делать окончательных содержательных выводов о дисциплинарной эволюции, не проведя критический анализ интерпретации алгоритмически сгенерированных кластерных решений.

Наконец, представим результаты структурного анализа центральности-степени тематик публикаций российских социологов (рис. 4.1) и политологов (рис. 4.2). Принципы построения тематических кластеров аналогичны описанным выше. Главное отличие – в расположении групп на двумерной оси координат, где ось абсцисс представляет степень «релевантности» той или иной тематики для остального научного поля (замеряется с помощью сетевой центральности), а ось ординат показывает степень «разработанности» тематики – количества работ в определенной области (операционализируется через сетевую степень).

Анализ схемы центральность-степень для социологии показывает, что на начальном этапе наиболее активно развивающимися и релевантными были публикации, касающиеся социологии как науки, а также студентов, образования и других социальных сфер и групп и культуры (в том числе ценностей). На следующем этапе данные темы в определенной степени укоренились, либо пролиферировались и частично перешли в область «нишевых» тем, с более узким фокусом и меньшей релевантностью для остального поля науки. Также с 2008 по 2018 гг. можно наблюдать появление области исследований, сконцентрированных на изучении глобализации, урбанизма и благополучия (они оставались «нишевыми»). В последний анализируемый период изучение образования встроилось в контекст общего социально-экономического развития, а также к этим темам добавилось изучение цифровых технологий и их последствий.



Рис. 4.1. Центральность-степень тематических кластеров во времени (социология)

Схожий анализ для сферы политологии показал следующие результаты. Изначальным двигателем

публикационной активности были темы, посвященные внутренним конфликтам, а также политическим и региональным трансформационным процессам. На следующем этапе к уже имеющимся группам тем добавляется много новых, а в качестве фундаментальных закрепляются изучение демократии, пост-советского пространства и региональной политики. Проявляются работы по дискурс-анализу и правам человека. После 2014 г. тематический спектр сильно сужается, на передний план выходят темы внутrerоссийской и глобальной политики, появляется небольшое, относительно незаметное число публикаций касательно украинского кризиса и электоральных систем. На последнем этапе фундаментальной темой обозначается изучение законодательной сферы России, а локомотивом выступает изучение внутrerоссийской политической трансформации. Глобальный сравнительный анализ сильно уступает как в плане проработанности, так и в плане релевантности для остального дисциплинарного контекста.



Рис. 4.2. Центральность-степень тематических кластеров во времени (политология)

В заключение отметим, что ранжирование тематик по степени их проработанности и релевантности – важный этап библиографического сетевого анализа. В отличие от предыдущих инструментов, этот позволяет увидеть структуру научного поля и понять, какое место занимает изучение того или иного вопроса, чего нельзя достичь сугубо сетевой визуализацией либо подсчетом частот тех или иных ключевых слов. Однако опять же встает вопрос содержательной консистентности результатов, полученных алгоритмическим путем, в связи с чем и к выводам данного анализа стоит относиться критически.

6.3 5.11.1 Модели управления благотворительными фондами – бенефициарными собственниками бизнес-компаний

6.3.1 Введение

Как мы себе представляем, что такое благотворительный фонд? Какой образ возникает в нашем воображении? Как правило возникает более-менее стандартный набор: «острая социальная

проблема», самоотверженные, активные и мобилизованные проблемой люди и постоянный поиск денег. Соответственно выстраивается образ человека, его личных и профессиональных качеств – сотрудника или руководителя благотворительного фонда или НКО.

А если задуматься о другой модели, когда благотворительный фонд владеет бизнесом или группой компаний? В этой модели он сам получает доходы и сам распоряжается ими. Кроме того, в зоне его ответственности оказываются очень разноплановые задачи – не только развивать благотворительные проекты, но также и заниматься устойчивостью бизнеса и его стратегическим развитием.

В европейских странах такая модель – фонды-собственники компаний, ФСК - оказывается довольно распространенной. В числе фондов-лидеров по размеру капитала — Novo Nordisk Foundation (Дания, группа компаний Novo), INGKA Foundation (Нидерланды, группа компаний INGKA/IKEA), Fondazione Cariplo (Италия, финансовая группа Intesa Sanpaolo).

Именно эта практика и стала предметом исследования, которое мы будем обсуждать.

Один из важнейших вопросов – кто и как управляет такими фондами? Какими компетенциями обладают люди, входящие в их руководство, какие карьерные треки ими реализуются? По каким принципам формируются команды руководителей? Поэтому мы будем обсуждать не только уже полученные результаты работы. Мы попробуем определить принципы следующего исследования и наметить шаги, направленные на изучение профессиональных сетей и связей руководителей таких фондов, их профессиональный бэкграунд и «композицию» советов директоров, необходимую для успешной реализации непростых и довольно разноплановых задач развития.

В российской практике такая модель управления практически не представлена. Только в последний год наметилось движение в эту сторону у некоторых фондов и компаний. Вместе с тем, определенное развитие получила смежная форма управления и финансирования социальных и благотворительных программ – эндаументы, фонды целевых капиталов. Их главное отличие от «industrial foundations» в том, что сфера деятельности эндаументов ограничена финансовым капиталом, тогда как ФСК также владеют акциями компаний, участвуют в получении дивидендов и в той или иной степени и объеме участвуют в управлении компаниями.

6.3.2 Цели и задачи исследования

Цель работы — сформировать углубленное представление о моделях управления ФСК в странах, где подобные практики широко применяются и имеют существенный вес в экономике, а также определить возможности и стратегии для развития рассматриваемых сценариев в российских реалиях.

Для достижения цели в исследовании формируются следующие задачи: - изучить, что является мотиватором собственника при принятии решения о выстраивании управления активами через создание фонда; - сформировать представление о том, что представляет собой структура управления фондом и как оценивается ее эффективность; - проанализировать, как выстроено управление с бизнес-активами и взаимодействие с другими значимыми участниками (основателями, наследниками, советом директоров и др.), выявить экономическую обоснованность управления активами посредством фонда; - описать структуру и приоритеты благотворительных программ и программ социального развития, оценить факторы их устойчивости и эффективности, - описать страновые регуляторные особенности, влияющие на деятельность IF - определить другие преимущества такой моделью управления бизнесом, понять, для

кого и как результат такого управления является или может являться полезным и выгодным.

6.3.3 Практическая значимость исследования

Модель, когда благотворительный фонд выступает собственником успешного бизнеса, оказывается распространенной в Европе, в первую очередь – в странах Северной Европы и в германоязычных странах. В долгосрочной перспективе именно такие решения оказываются наиболее эффективными и демонстрируют стабильность в осуществлении социальных программ и проектов.

В России в настоящее время активно растет сектор целевых капиталов: законодательно эндаументы имеют возможность становиться собственниками больших активов. Однако в российской практике таких примеров пока не много.

В исследовании наглядно продемонстрирована важная роль страновых особенностей при выборе модели управления фондами. Практическая значимость работы – в задаче наметить направления развития в масштабах российского сектора благотворительности и филантропии. Результаты могут быть использованы для анализа достоинств и недостатков такого рода решений, сравнительного анализа структуры собственности, формирования более комплексного взгляда на инфраструктуру благотворительности, для расширения нашего понимания собственности и способов управлять ею.

По итогам работы были сформулированы рекомендации для собственников компаний, задумывающихся о стратегической устойчивости корпоративных социальных программ, для благотворительных фондов, управляющих эндаументами, для государства – по развитию регулирования в этой сфере.

6.3.4 Методы сбора и обработки данных

Исследование построено в формате кейс-стади. Оно опирается на анализ и обобщение существующих практических примеров десяти фондов-собственников компаний и предлагает выводы о применимости этого подхода в российских реалиях. В каждом из кейсов рассматривается перечень вопросов, помогающих выполнить задачи исследования.

Авторами были разобраны и проанализированы десять фондов – бенефициарных владельцев компаний: Novo Nordisk, Bosh, IKEA, Карл Цейс, Hempel, Ramboll, Pierre Fabre, индийская Tata, итальянские Intesa Sanpaolo и La Caixa.

Тематическая структура case-study (Климов et al., 2023): 1. История возникновения фонда, компании: основные моменты 2. Миссия и цели фонда 3. Структура владения и управления компанией 4. Владение 1. Управление 2. Процесс и ступени принятия решений в фонде относительно компании 4. Финансовые показатели фонда и компании 5. Регуляторный режим

На этом этапе методология работы полностью ориентировалась на принципы разработки «делового кейса». Однако следующий аналитический шаг нацелен на изучение практики формирования советов директоров ФСК с использованием сетевого подхода. Изучение состава СД фонда Ново Нордиск за десять лет показало, что его члены не только перемещаются по значимым позициям внутри всей холдинговой структуры, но также включены в управляющие органы других важных стейкхолдеров. Таким образом в СД фонда аккумулируется необходимая экспертиза и социальный капитал каждого из участников. Сетевой подход представляется очень уместным для понимания сути и принципов одной из ключевых практик функционирования ФСК.

6.3.5 Обзор литературы

Практика, когда БФ являются бенефициарными владельцами бизнес-активов, не развита, она только начинает формироваться. В российских исследованиях все больше появляется работ про межпоколенческий трансфер российских бизнесов и частных благотворительных проектов (Мисютина et al., 2015). Много работ посвящены барьерам, сдерживающим развитие ФЦК, эндаументов: неразвитость фондового рынка, недостатки в управлении налоговыми льготами, трудности взаимоотношений с донорами, с управляющими компаниями и другие (Цель Под Ключ - PBWM.RU, n.d.) [Барковец (2013)](Соколова, 2011)[Савченко et al. (2015)](Подольская & Харламова, 2016)

«Фонды – собственники компаний» (industrial foundations) – это организации, владеющие акциями или же долей одной или нескольких бизнес-образований напрямую или же через холдинговую структуру (Thomsen, 1999). Формы ФСК зависят от специфики законодательства в конкретном государстве. Это могут быть фонды, учреждения, трасты с образованием или без образования юридического лица. Но их все объединяет параметр – у них нет собственников или акционеров.

Зачастую ФСК создается основателем и акционером компании и такой выбор является альтернативой передаче прав владения компанией наследникам или внешним инвесторам. Другой вариант их возникновения, когда бизнес изначально создается со стратегической социальной целью (как Novo Nordisc), под влиянием государственного регулирования (Фонд Карипло, Италия) или же по решению наследников (трасты Tata). В большинстве случаев это некоммерческие организации, они владеют или контрольным пакетом акций или особыми классами акций (Hansmann & Thomsen, 2021). Существует практика, когда акции передаются в фонд на принципе безотзывности (Hansmann & Thomsen, 2021).

Модели управления, принципы формирования совета директоров, приоритеты деятельности, структура социальных и благотворительных программ, а также традиции целеполагания в значительной мере определяются страновой спецификой.

Выделяют четыре параметра вариативности форм и решений (Schröder & Thomsen, 2021): - традиции филантропии как в стране, так и в семье основателя; - социально-экономические и культурные особенности страны; - деловые процессы в компании и динамика в «ее» секторе экономики; - специфика и изменения в законодательстве.

Модель ФСК принципиально отличен от грантодающих и благотворительных НКО. Набор задач в них более широкий и требует более широкого набора компетенций и профессионального опыта. Однако главная цель ФСК – реализация акционерных прав согласно с принятыми положениями и ценностями.

В литературе и в исследованиях ФСК выделяются несколько важных для темы направлений.

Управленческая дистанция (Hansmann & Thomsen, 2021). Предметом анализа становятся полномочия советов директоров, распределение функций контроля в Фонде и компании, в управлении операционной деятельностью бизнеса, состав и принципы формирования советов директоров компаний и фонда, в какой мере распространено пересекающееся членство. Оценивается, насколько ФСК оказываются вовлечеными и активными собственниками. В качестве акционера фонд участвует в назначении своих представителей в органах управления компании и участвует в утверждении основных решений, контролирует финансовое положение и др. (Børsting & Thomsen, 2017) (Hilt, 2008) (Schmidt, 2015) (Burkart et al., 2021)

Исследования показывают, что грамотно выстроенное управление ФСК и корпоративного управления оказывают положительное влияние на эффективность компаний – независимо от системы финансового

стимулирования директоров (Kaya & Thomsen, 2022). Тем не менее, здесь фиксируется нелинейная зависимость между управленческой дистанцией и эффективностью компании: позитивное влияние возрастающей дистанции после определенного момента демонстрирует обратный эффект (Hansmann & Thomsen, 2021).

Назначение директоров и работа СД. Исследования фокусируются на большом спектре вопросов и проблем: принципы формирования, оценки рисков для стратегического управления в зависимости от принятых принципов формирования СД, роль законодательного регулирования общественная подотчетность благотворительных фондов [Thomsen (2016b)](Hansmann & Thomsen, 2021). Еще одна большая область исследований - система мотивации и финансовые условия участия в СД в ФСК, анализ KPI и принципов его работы, проблематика нефинансовой мотивации и т.д. (Hansmann & Thomsen, 2021)

Страновая специфика. Большой сегмент исследований посвящены страновой специфике и изучению социо-культурных и социально-экономических детерминант в развитии практики ФСК. Наиболее распространение ФСК получают в Северной Европе и германоговорящих странах, хотя они встречаются и в других странах мира и бывают в различных формах [Thomsen & Kavadis (2022)](Kaya & Thomsen, 2022)(Schröder & Thomsen, 2021). Проводится оценка их вклада в ВВП стран и в объем благотворительного сектора (Thomsen & Degn, 2014). Систематический обзор страновых отличий проведен исследователем Центра корпоративного управления БШ Копенгагена С. Томсеном (Thomsen, 2018).

Синдром «сиюминутного мышления». В работах, посвященных корпоративной социальной ответственности и стратегическому развитию институциональных условий для сектора благотворительности эта проблематика обозначается концептом «short-termism»: стремясь в моменте максимизировать возможные выгоды, «временщик» не просто теряет выгоды в стратегической перспективе. Такие действия и такой стиль мышления проецируются в общество и формируют специфический тип культуры и тип социальных отношений. Исследования нацелены на анализ механизмов, которые помогают бороться с «сиюминутным мышлением» и как корпоративной практикой, и как социальным феноменом (Thomsen et al., 2018). Исследователи показывают, что ФСК как специфическая форма управления и сами компании, принадлежащие благотворительным фондам, оказываются хорошим инструментом для преодоления синдрома, т.к. с ними связано стабилизирующее влияние на уровень и качество занятости, на развитие местной экономики. Кроме того, фирмы, принадлежащие фондам, лучше справляются с экономическими кризисами, показывают более высокую выживаемость в длинной временной перспективе (Thomsen et al., 2018). Например, для них выше вероятность выживания в 40-летней перспективе (выживают 30% от их числа – против 10% для компаний с другой формой управления). Причем эти же исследователи показали, что такая форма управления компанией показывает свою эффективность и в кризисное время, когда требуются быстрые решения и решительные действия.

Семейный бизнес. В работах показано, что ФСК имеют много общего с практиками семейного бизнеса (Thomsen et al., 2018). Однако некоторые из базовых параметров ФСК делают его еще более долгосрочным предприятием, чем только лишь семейное владение, поскольку иначе решается вопрос преемственности и собственности, вопросы управленческих и предпринимательских компетенций преемника, иное отношение к дивидендам, иные горизонты планирования и основания для построения долгосрочных планов [Thomsen (1999)](Bertrand & Schoar, 2006)[Souder et al. (2016)](Thomsen & Degn, 2014).

6.3.6 Обсуждение

Главный вопрос и миссия нашей работы – поиск ответа на вопрос: почему нам и нашему государству необходимо интересоваться этой темой?

Ответ на этот вопрос лежит в плоскости экономического развития, драйверов устойчивости корпоративного управления, детерминант появления долгосрочных бизнес-стратегий и развития инфраструктуры благотворительности. Такая аргументация важна в контексте осмысления 30-ти летнего опыта формирования рыночной экономики в нашей стране. По мере ее усложнения и развития появляются возможности создавать все более сложные и долговременные институты.

Мы видим несколько аргументов, которые позволяют отнести к Фондам предприятий как к стратегической задаче экономического и социального развития (Климов et al., 2023). 1. Постепенно менять отношение россиян к «богатству», к «богатым людям», и в том числе – к занятию предпринимательством (SKOLKOVO, n.d.). ФСК как модель владения и управления оказывается противопоставленной личному владению бизнес-активами, что потенциально может сформировать иные представление и новое понимание «частной собственности». Можно предполагать, что практика коллективного принятия решений по вопросам управления компаниями (особенно крупными) будет с большим пониманием воспринято нашими согражданами, чем «эгоистичное владение» (Климов et al., 2023). 2. Государству, чтобы реализовывать долгосрочные экономические стратегии, необходим мандат легитимности относительно базовых инструментов, с помощью которых оно осуществляет регулирование социально-экономических процессов (Климов et al., 2023). ФСК и принадлежащие им компании со временем могут стать устойчивым и эффективным институтом, создающим долговременный эффект и не требующим непосредственного управляющего участия государственных структур в его деятельности. 3. Описываемая модель корпоративного управления снимает с государства необходимость регулировать частные капиталы и определять их судьбу – со всеми сопутствующими этой активности негативными последствиями: конфликтами элит, коррупцией, разрушением устойчиво действующих компаний, компрадорскими установками элит и т.д. (Климов et al., 2023). Кроме того, ФСК оказываются одним из наиболее эффективных инструментов, которые заставляют деньги оставаться и работать в стране. Причем делают это не принуждением, а через создание возможностей нового типа. 4. Владение фондом, по-видимому, является примером частного предприятия другого типа, которое не подвержено недостаткам финансализации – сведения роли бизнеса только лишь к зарабатыванию денег и формированию частного капитала (Dzansi, 2011). Владение фондом также, по-видимому, позволяет избежать проблем наследования, семейных конфликтов и кумовства, которые преследуют семейный бизнес (Børsting & Thomsen, 2017). Более того, фонды являются частными организациями, которые не подвержены знакомым проблемам государственных предприятий, таким как политическое вмешательство или мягкие бюджетные ограничения (Thomsen, 2018). 5. Что делает владение фондами особенно перспективной формой корпоративного управления, так это сопряжение благотворительности с бизнесом (Климов et al., 2023). Фонды должны и могут служить неэгоистичным целям и ценностям, разделяемым в обществе. В отличие от англо-американской традиции, регулирующие органы в странах Северной Европы рассматривают владение компанией как служение полезной социальной цели. Поэтому их законодательство изначально ориентировалось на создание инфраструктуры для деятельности социально-ответственных акторов (Thomsen, 2018). 6. Многие исследователи рассматривают Фонды

предприятий как способ борьбы с «парадоксом временщика» (Климов et al., 2023). Фонды являются долгосрочными владельцами, что может дать их компаниям, конкурентное преимущество в некоторых сферах бизнеса, например, в фирмах, интенсивно занимающихся НИОКР, с длительным жизненным циклом продукции (Børsting & Thomsen, 2017). Похоже, что преимущества долгосрочного подхода более выражены в крупных фирмах, чем в небольших стартапах.

Несколько факторов, по-видимому, усиливают практику долгосрочного планирования в компаниях, принадлежащих фондам (Климов et al., 2023). Например, консервативная долгосрочная ориентация может быть заложена базовыми правовыми положениями. Например, в ряде стран есть положения, которые обязывают Фонды сохранять полученные активы. В отличие от компаний, они не могут (за исключением случаев банкротства) быть распущены до тех пор, пока их бизнес-актив остается действующим [Schröder & Thomsen (2021)](Thomsen, 2016a)

6.3.7 Результаты

Основной результат работы – систематическое описание десяти отобранных кейсов фондов – бенефициарных владельцев компаний: Novo Nordisk, Bosh, IKEA, Карл Цейс, Hempel, Ramboll, Pierre Fabre, индийская Tata, итальянские банки Intesa Sanpaolo и La Caixa. Для каждого из примеров были описаны (Климов et al., 2023): 1. История возникновения фонда, компании: основные моменты 2. Миссия и цели фонда 3. Структура владения и управления компанией 1. Владение 2. Управление 3. Процесс и ступени принятия решений в фонде относительно компании 4. Финансовые показатели фонда и компании 5. Регуляторный режим

Кроме того, были приведены аргументы в пользу подобной практики развития корпоративного управления с точки зрения собственников и акционеров компаний, с точки зрения государства и в контексте развития социальных и экономических процессов.

6.4 5.11.3 Изучение гендерной специфики деструкторов руководителей крупных российских компаний

6.4.1 Введение

В научной литературе и в эмпирических исследованиях существует множество работ, посвященных изучению гендерной специфики лидерства. Существуют «свои» традиции в самых разных дисциплинах – в социологии, политической науке, в психологии и социальной психологии. Заметная доля публикаций приходится на междисциплинарные области, такие как феминистические и гендерные исследования, изучение организаций и менеджмента, сочетающие в себе самые разные жанры: публицистика, научно-популярно изложение теорий и концепций, описание лабораторных экспериментов и полевых исследований, а также обобщающие работы, использующие методологию мета-анализа. Однако увеличение эмпирических данных не приводит к целостному пониманию лидерства как феномена и его гендерной специфики. Можно согласиться с исследователями, которые давно говорят о накопленных знаниях как о множестве несвязанных, противоречивых, фрагментарных сведений и фактов, о череде одновременно амбициозных и тривиальных теорий и исследований (Scheidlinger, 1980).

На фоне большого числа публикаций о лидерстве в целом и женском лидерстве – в частности, заметно меньшее число публикаций посвящено «деструктивному лидерству». Многие авторы утверждают, что

этого рода публикации фрагментированы, основаны на разношерстной методологии, их совокупность плохо осмыслена (Terperg, 2007). Тем не менее, на основании проведенного анализа литературы и исследований мы можем предположить, что существует гендерная специфика в проявлении деструкторов у руководителей, а ряд исследователей и подходов дают сильные объяснительные модели для интерпретации проявляющихся гендерных отличий. Именно этот сюжет стал предметом нашего исследования. Источником данных послужила база данных результатов оценки руководителей российских компаний по методике Хогана, собранная за период с 2019 по 2021 год.

Мы не можем утверждать, что имеющийся массивreprезентативен для всех российских руководителей: он формировался в результате коммерческой деятельности нескольких компаний, предоставляющих услуги по оценке руководителей. Тем не менее, для задач исследования важны другие характеристики: точное попадание в целевую группу и структура массива, позволяющая возможности сравнивать разные категории руководителей.

Массив данных, который использовался для статистической обработки, был полностью анонимизирован. У исследователей не было доступа ни к персональным данным респондентов, ни к данным о конкретных компаниях.

6.4.1.1 Цели и задачи исследования Научная проблема: проанализировать гендерные различия в деструкторах топ-менеджерах российских компаний, опираясь на эмпирические данные. В основе эмпирики – датасет на основе психометрического теста Хогана (Hogan & Hogan, 2009). Проанализированная структура деструкторов должна быть соотнесена с традицией психоаналитических исследований лидерства и токсичного лидерства.

Задачи исследования:

1. Разобраться с понятием «деструкторы» в контексте проблематики «токсичного лидерства»; углубиться в специфику их измерения в методике Хогана;
2. Провести анализ имеющегося дата-сета с прицелом на гендерной специфики деструкторов; найти ключевые деструкторы, где есть дифференциация по гендерным группам; изучить возможности анализа гендерных отличий деструкторов по различным отраслям.
3. Оценить эвристические возможности психоаналитического подхода к изучению деструкторов руководителей; увидеть их связь со спецификой женского и мужского лидерства.

С точки зрения авторов исследования, именно психоаналитическая традиция системно ведет разработку природы деструкторов и рассматривает бессознательное в качестве их природы. Традиция формируется работами таких ученых, как Фрейд, Лакан, Кляйн, Баум, Клули, Кетс де Врис, Миллер, Штейн, Залезник, Костас, Тахери, Драйвер, Обхольцер, Маккоби, Кохут, Габриели, Мак-Вильямс и другие. Целый ряд направлений изучает деструкторы как связь между невротическими дисфункциями лидера и организационной патологией, взаимозависимость и взаимное влияние лидера и его последователей, различие между лидерами и менеджерами, аутентичное лидерство, опираясь на клиническую парадигму, объясняют природу деструкторов и связанное с ними иррациональное поведение.

Также можно сказать, что для психоаналитической школы больший интерес представляет влияние ранних периодов на лидерский потенциал и не гендер как таковой, а скорее роли отца и матери в развитии

и формировании не только лидерской позиции, но и лидерских деструкторов. Традиционно лидерская тема, включая гендерный аспект, рассматривается с точки зрения прохождения нарциссической стадии и формирования «нормального» или «патологического» нарциссизма у обоих полов. Можно резюмировать результаты анализа следующим образом: в вопросе формирования деструкторов лидеров основная роль принадлежит не гендеру, а специфике прохождения стадий развития и влиянию родителей. При этом, похоже существуют отдельные социальные и культурные нормы, которые определяют различия в том, как матери относятся к сыновьям и дочерям, что формирует специфику в проявлении лидерства.

6.4.1.2 Практическая значимость исследования Новизна работы состоит в использовании психоаналитической традиции для понимания природы гендерных отличий в деструкторах, обнаруженных у руководителей российских компаний. Фокус исследования конструируется пересечением четырех тем и подходов: лидерство, деструкторы, гендер, психоаналитический подход. Проблематика в этих направлениях «приземляется» на эмпирические данные выбранной целевой группы: реально действующие руководители российских компаний.

Практическая значимость работы состоит в комплексности подхода и в применимости полученных результатов для задач бизнес-консультирования, персональной работы с теми, кто претендует на лидерские позиции, для оптимизации командной работы и гармонизации внутреннего взаимодействия команд, для решения задачи управлять деструктивными характеристиками лидера (лидеров).

6.4.2 Обзор литературы

«Темная сторона» лидерства была предметом научного изучения, по крайней мере, в последнее десятилетие (Scheffler & Brunzel, 2020). Исследователи приписывают этой теме такие термины, как деструктивное лидерство (Einarsen et al., 2007), нарциссическое лидерство (Rosenthal & Pittinsky, 2006) или токсичное лидерство (Lipman-Blumen, 2010), и это лишь некоторые из них.

Первое впечатление, которое ожидает исследователя темы деструкторов и «темной стороны» лидерства, связано с отсутствием общепринятого определения данного понятия.

Так Хиггс (Higgs, 2009) указывает, что в настоящее время не существует общепринятых определений и разграничений для термина «деструктивное лидерство». Уместнее говорить о том, что «деструктивное лидерство» является зонтичным понятием. Также Хиггс утверждает, что ключевыми словами, которые определяют этот термин, являются «дирейлеры лидерства», «токсичное лидерство», «лидерство темной стороны», « злоупотребление лидерством» и «деструктивное лидерство». Аналогичным образом, Типпер (Terper, 2007) утверждает, что литература по данной тематике «фрагментирована, плохо интегрирована и использует различные методологии». С этой точки зрения интересны работы, структурирующие данное понятие. Так исследование Чжан, Лесли и Ханну (Y. Zhang et al., 2013) помогло определить несколько слабых областей лидерства, обобщенных позднее Коутом (Cote, 2018). По данному исследованию деструкторы лидера с поведенческой точки зрения можно определить как проблемы с производительностью, отношениями, изменениями, созданием и управлением командами, опытом.

Липман-Блюмен (Lipman-Blumen, 2010) относят токсичное лидерство к процессу, в котором лидеры в силу своего разрушительного поведения и/или дисфункциональных личностных характеристик наносят серьезный и устойчивый вред своим последователям, организациям и другим людям. Автор делает

упор на политических лидеров, хотя его выводы распространяются и на организационный контекст. Он так же, как и Терпер, включает в свой подход разных акторов (лидер, подчиненный, последователь и т.д.) и влияние контекста, истории, культуры. Он указывает, что определение токсичных лидеров - нетривиальный процесс, так как токсичный лидер для одного человека может считаться героем для другого.

Липман-Блюмен один из немногих авторов, кто разделяет поведенческие и личностные характеристики токсичной или «темной стороны» лидерства, а также обращает внимание на разнообразие токсичного лидерства из-за различной степени его проявления и уровня осознанности при демонстрации такого поведения. Чтобы проработать концепт токсичного лидерства, Липман-Блюмен предлагает использовать многомерный фреймворк: учитывать интенциональность или целенаправленность поведения токсичных лидеров, уровень интенсивности их токсичности, типы деструктивного поведения, типы дисфункциональных личностных качеств, которые управляют их решениями и действиями, и последствия самих решений и действий.

Можно обратить внимание, что деструкторы лидерства чаще всего изучаются с поведенческой точки зрения, через описание того, как ведут себя токсичные руководители и какое влияние они оказывают. Однако нас в первую очередь интересует не только проявление деструкторов, но их природа. Многие авторы описывают деструкторы как иррациональное поведение, иногда не поддающееся объяснению ((Terper, 2007), (Ashforth, 1997), (Einarsen et al., 2007), (Lipman-Blumen, 2010)). Значительная часть нашей мотивации и поведения происходит вне осознания, и мы не имеем полного контроля над нашими процессами восприятия ((Kets de Vries, 1989), (Kets de Vries & Miller, 1984)).

Считается, что большая часть психоаналитических подходов к изучению деструкторов опирается на концепции Фрейда и Кляйн (Dashtipour & Vidaillet, 2016). К авторам этого направления принято относить Баума, Клули, Кетс де Вриса, Миллера, Штейна, Залезника, а меньшее количество публикаций также используют теорию Лакана (Dashtipour & Vidaillet, 2016). Например, Костас и Тахери, Драйвер. Психоаналитический взгляд на лидерство основан на том, что невозможно иметь представление о вопросах развития авторитета и лидерства без учета внутреннего мира индивида (Obholzer, 1996). При этом сам предмет изучения лидерства в психоаналитическом подходе достаточно комплексный и можно выделить следующие направления в области исследований деструкторов лидеров:

1. Исследования, раскрывающие связь между невротическими патологиями, такими как нарциссизм, и организационными дисфункциями ((Kets de Vries & Miller, 1984), (Kets de Vries, 1991); (Maccoby, 2000); (Rosenthal & Pittinsky, 2006); (Schwartz, 1990)).
2. Исследования лидерства, фокусирующиеся на вопросе связи лидера и последователей. Так Кохут (Кохут, 2003) популяризовал идею о том, что последователи часто рассматривают лидеров и относятся к ним так, как это отражает их более ранние детские отношения с родительскими фигурами. Габриэль (Gabriel, 2011) изучал различные фантазии с помощью историй, которые последователи рассказывают о своих лидерах. Некоторые также изучали эмоциональную связь между лидерами и последователями ((Kets de Vries, 1989), (Kets de Vries & Miller, 1984)). Кетс де Врис показывает, как эта связь между лидером и последователями часто включает в себя защитные механизмы, процесс переноса, а также управление тревогами, что в комплексе может приводить к дисфункциональному поведению.

3. Исследования, фокусирующиеся на различиях между менеджерами и лидерами. Например, подход Залезника (Kets de Vries, 1991), в котором он выступает за развитие лидеров, а не менеджеров в организациях. Аргументируя это тем, что в то время, как менеджеры сосредоточены на рациональном, порядке и контроле, лидеры более интуитивны, эмоциональны и креативны, и у них есть страсть и видение. Лидеры более эмоциональны и могут лучше справляться с неопределенностью. Именно менеджеры, по мнению Залезника, ставят процесс, политику и личные интересы выше содержания работы (Kets de Vries, 1991). Что, в свою очередь, делает их очень манипулятивными, хитрыми и расчетливыми.
4. Исследования, которые опираются на подход Лакана. Драйвер (Driver, 2013), фокусируется на изучении лидерских идентичностей, которые в конечном итоге приводят к неудаче. Костас и Тахери (Costas & Taheri, 2012) обсуждают популярный в настоящее время подход «аутентичного лидерства» и его акцент на положительных эмоциях.

Поскольку понятие деструкторов или «темных сторон» лидерства оказывается сложносоставным, комплексным, всегда существует потребность в некоторой систематизации и синтезе этих понятий для практического использования. В психоаналитической школе таким исследователем можно считать Мак-Вильямс, которая предложила выделить девять типов организации личности, опираясь на два измерения: определение личностной структуры, которое основано на возрастном периоде переживания травмирующего события, и сочетание защитных механизмов. Объединяя указанные выше понятия, Мак-Вильямс определила типы организации личности или треки, изучая и диагностируя которые, можно определить стиль лидерства и специфическое проявление его «темной стороны» или деструктора.

Хорошим примером практической реализации теоретических подходов оказываются психометрические инструменты. В частности – это методика Хогана и ее возможности для оценки деструкторов руководителей. По мнению Хогана, — это не просто отсутствие навыков, а скорее, дисфункциональные наклонности и связанное с ними поведение (Hogan & Hogan, 2009). Специфика таких личностных характеристик или деструкторов заключается в том, что они могут ухудшить илинейтрализовать любые навыки и компетенции, которые принято относить к «светлой стороне» лидерства. Хоган указывает, что феномен «дисфункциональных диспозиций» характеризуется сочетанием технической компетентности и межличностной неадекватности. В худшем случае лидеры, которые транслируют деструкторы, могут восприниматься своими подчиненными и организацией в целом как «абьюзивные» (Terper, 2007) или «тираннические» (Ashforth, 1997). Так мы видим, что подход Хогана сочетается с другими моделями деструктивного лидерства и скорее представляет собой изучение причин в виде дисфункциональных личностных характеристик или деструкторов, которые на уровне

поведения приводят к негативным проявлениям и последствиям.

Хоган считает себя приверженцем социоаналитической теории (J. Hogan & Holland, 2003), которая, по его мнению, отражает динамику, связанную с успехом в любом групповом процессе. Социоаналитическая теория утверждает, что, как животные, живущие в группах, люди разработали стратегии для максимального индивидуального и группового выживания (J. Hogan & Holland, 2003). Все группы организованы по статусной иерархии, и взаимодействие внутри групп включает в себя две темы: о принятии и о статусе. Соответственно, люди мотивированы как “ладить с другими” (максимизировать популярность), так и “продвигаться вперед” (максимизировать статус по отношению к другим членам

группы).

Такой подход предполагает, что организационный контекст тесно связан с социальным взаимодействием и требует реализации этих мотивов: ладить с другими и продвигаться вперед (Hogan & Hogan, 2007). Чтобы реализовать первый мотив – «ладить с другими» - люди должны сотрудничать, а также восприниматься дружелюбными и позитивными. Чтобы «продвинуться вперед», необходимо проявлять инициативу, стремиться к ответственности и стараться быть признанными.

Природу деструкторов как дисфункциональных наклонностей Хоган раскрывает через теории Адлер, Хорни, Эриксон, Салливан (Hogan & Hogan, 2009), которые утверждали, что проблемы людей основаны на том, как они взаимодействуют с другими. При этом, ранний опыт (в семье, в школе, в группе сверстников) гарантирует, что почти каждый чувствует себя неадекватным в отдельных аспектах. То есть детство и юность почти неизбежно сопряжены со стрессом, и у большинства людей развиваются ожидания, что в определенных ситуациях их будут критиковать и/или они будут чувствовать себя неадекватными или беспомощными.

Ключевой теорией для определения природы деструкторов Хоган считает работу Хорни, которая выделила 10 «невротических потребностей». Позже она обобщила эти потребности в терминах трех тем (Хорни, 2019):

1. Движение навстречу людям — т.е. управление своей неуверенностью путем создания союзов, в которых угроза критики может быть сведена к минимуму.
2. Отдаление от людей — т.е. управление своим чувством неадекватности, избегая истинной связи с другими.
3. Движение против людей — т.е. управление сомнениями в себе, через доминирование и запугивание других.

Опираясь на приведенные выше понятия и теории, в 1997 году Хоган разработал опросник The Hogan Development Survey ((Hogan & Hogan, 2009)) для измерения и оценки вероятности проявления деструкторов в профиле человека. HDS оценивает личность по 11 шкалам, которые соответствуют дисфункциональным личностным характеристикам.

При формировании шкал для опросника HDS Хоган взял за основу 10 расстройств личности из Диагностического и статистического руководства по психическим расстройствам, 4-е издание, пересмотренное, а также добавил шкалу для описания пассивно-аггрессивной личности, которая включена в DSM (Hogan & Hogan, 2009).

При этом, Хоган подчеркивает, что HDS не предназначен для измерения расстройств личности, а оценивает саморазрушительные или деструктивные проявления нормальной личности (Hogan & Hogan, 2009).

Также Хоган обращает внимание, что шкалы опросника HDS представляют собой своего рода спектр дисфункциональных наклонностей, возникающих в межличностных отношениях. И в зависимости от того, на какой части спектра находятся результаты конкретного человека, возможны различные поведенческие проявления: от эффективных навыков и уверенности в межличностном взаимодействии до некомпетентности и поведения, связанного с расстройством личности. Так более высокие баллы указывают на большую вероятность возникновения дисфункционального поведения в любом межличностном контексте. Хоган отмечает, что большинство людей находятся в середине этого

распределения, и любой отдельный человек может иметь высокие или низкие баллы по любой из шкал (Hogan & Hogan, 2009).

6.4.3 Описание research gap

На основе обзора публикаций и исследований о гендерной специфике лидерства, можно сделать несколько выводов:

Исследования различий между мужчинами и женщинами - реальными лидерами и руководителями недостаточны и фрагментарны. По-видимому, получаемые данные зачастую оказываются противоречивы, поэтому метаанализ эмпирических работ не показывает значимых отличий. Есть ряд характеристик, по которым гендерные различия у мужчин и женщин- лидеров установлены как эмпирическими исследованиями, так и подтверждаются метаанализом. В частности - гендерная идентичность, мотивация к достижениям, эффективность лидерства в однородных и смешанных группах, конкурентность/кооперативность, отказ от лидерства, самооценка.

Тема гендерной специфики деструкторов у руководителей изучена мало, этот сюжет не выделяется в тематическом репертуаре исследований, посвященных изучению гендерной психологии лидерства в целом и женского лидерства - в частности. Это связано с объективной труднодоступностью этой целевой категории и невозможностью построить сколь-нибудь масштабное (для достижения статистической значимости) и методологически обоснованное исследование. Одновременно с этим психоаналитическая традиция в изучении гендерной специфики лидерства и лидеров есть, но в контексте дисциплинарных исследований она представлена мало, теряется на фоне массы психологических работ.

Кроме того, источник затруднений мы видим также и в том, что большинство исследований используют методики, рассчитанные на когнитивную работу, или же выстраивают модели экспериментов, рассчитанных на проявление поведенческих реакций. Психоаналитическая традиция в изучении гендерной специфики лидерства и лидеров есть, но в контексте дисциплинарных исследований она представлена мало, теряется на фоне массы психологических работ.

Учитывая сказанное, мы считаем уникальной возможность изучить гендерную специфику в проявлении деструкторов у руководителей российских компаний на реальных эмпирических данных, собранных целевым образом у интересующей нас группы с помощью методики, хорошо фундированной как с точки зрения теоретической традиции, так и с точки зрения методологии личностной оценки

6.4.4 Методы сбора и обработки данных

Источником данных стала база данных результатов оценки руководителей российских компаний, собранных по методике Хогана. Для целей исследования важно, что это именно состоявшиеся руководители, занимающие достаточно высокие посты в структуре компаний, представители топ-менеджмента, которые прошли опросник HDS за период с 2019 по 2021 год. Исходный массив составлял 1743 респондента. Однако в 47% случаев наблюдались пропуски ответов респондентов по ряду ключевых характеристик исследования, включая как вопросы измерения лидерских качеств, так и социально-демографические характеристики. Всего после исключения невалидных или нерелевантных данных база содержала сведения о 922 людях из 37 компаний 9 секторов экономики.

Методическая рамка исследования базируется на нескольких принципах. Во-первых, это акцент на

предварительном анализе объекта исследования на основе ряда социально-экономических переменных и переменных методики Хогана. Структура данных сформирована таким образом, что значительная часть интересующих нас переменных имеют в среднем только 2 возможные опции (мы говорим о бинарных дихотомических шкалах, о которых речь пойдет далее). Это накладывает определенные ограничения в свободе выбора методологии статистического анализа. Поскольку нас интересует идентификация и описание возможной статистической связи между компонентами деструкторов (измеряемых по методике Хогана) и рядом социально-экономических характеристик, выбор был сделан в пользу бинарной логистической регрессии, наиболее оптимальной как с точки зрения требований, так и возможностей по верификации эффектов. Для построения серии логистических моделей предполагается выполнения ряда принципиальных шагов:

1. Разведочный анализ (EDA). На данном этапе проводится анализ используемых в модели логистической регрессии переменных, в первую очередь на предмет вероятных отклонений значений деструктивных шкал от общей тенденции. Дополнительно рассматриваемые шкалы сравниваются признаку и отрасли деятельности руководителей.
2. Формирование уравнения логистической регрессии. В первую очередь проводится оценка распределения зависимой переменной (пол респондента, отрасль деятельности) и независимых переменных (деструкторов по методике Хогана). Далее эти переменные преобразуются в бинарные (дихотомические) по следующей схеме:
 - Для параметра «пол» зависимая переменная получила значения: «0» – мужчины, «1» – женщины.
 - Каждая из 11 шкал деструкторов были перекодированы по схеме: от 0 до 70 – «0», а 70 и более – «1», что соответствует уровням проявления деструкторов у респондента. Значения выше 70 попадают в область умеренно высокого риска, значения выше 90 – в область высокого риска. Шкала считается деструктором в профиле конкретного респондента, если значение по ней составило 70 и более процентилей.
3. Оценка статистической значимости гендера, отраслевой принадлежности руководителя и диагностируемых у него деструкторов на основе продвинутых методов логистического регрессионного моделирования. Сюда входит построение простых бинарных логистических регрессий, а также аналогичных решений с применением алгоритмов сэмплинга с использованием Марковских цепей Монте-Карло (МСМС-подход). Анализ статистической значимости подразумевает интерпретацию специализированных коэффициентов — информационный критерий Акаике (AIC), Байсовский информационный критерий (BIC), в том числе оценку статистической значимости. Далее проводится поиск и формирование системы аргументации на основе полученных результатов.

6.4.5 Обсуждение

Происходящие в стране и в мире события усиливают актуальность и значимость как нашей работы, так и любых исследований в этой области. Кризис, стресс, ситуация неопределенности провоцируют актуализацию защитных механизмов, которые проявляются как деструкторы. Первой на линии удара оказывается категория руководителей, лидеров, вынужденных действовать в этих условиях, принимать

решения, поддерживать и мобилизовывать своих последователей. Деструкторы руководителей имеют потенциальную опасность для благополучия команды и эффективности ее действий, предопределяют форму регресса группы, когда она оказывается неспособной адекватно реагировать на ситуацию, анализировать ее, принимать верные решения и реализовывать их. Наше исследование позволяет углубиться в понимание природы деструкторов и их гендерной специфики.

6.4.6 Результаты исследования

Результаты исследования были использованы для подготовки и защиты магистерской диссертации, легли в основу научной статьи, находящейся на завершающей стадии подготовки. Кроме того, материалы работы используются в программах компании Kontakt Intersearch.

Частотный анализ деструкторов показал:

1. В кластере «Отдаление от людей» по Хорни лидирует деструктор Скептичный.
2. В кластере «Движение против людей» по Хорни очень близкими оказываются значения сразу двух деструкторов: Самоуверенный и Увлекающийся.
3. В кластере «Движение навстречу людям» по Хорни лидирует деструктор «Исполненный сознания долга», и он же оказывается безусловным лидером среди всех остальных деструкторов.
4. Самым частотным оказывается кластер «Движение навстречу людям» по Хорни: деструкторы из этого кластера встречаются в среднем в два раза чаще, чем деструкторы по шкалам из других кластеров.

Изучение результатов регрессионного анализа позволяет зафиксировать несколько моментов:

1. Шкалы деструкторов Скептичный, Осторожный, Сам в себе, Самоуверенный, С богатым воображением и Исполненный сознания долга оказываются наиболее важными и дифференциирующими для гендерных групп.
2. Мужчинам более всего свойственны деструкторы Скептичный и С богатым воображением.
3. Для женщин характерны три деструктора: Осторожный, Сам в себе, Самоуверенный и Исполненный сознания долга.

6.4.7 Заключение

Использование бинарной логистической регрессии позволило подтвердить **основную гипотезу исследования**: гендерная специфика в проявлении деструкторов у руководителей российских компаний существует. **Ключевой результат работы** в том, что мы выделили шкалы, по которым наблюдалась устойчивая связь с гендером. Ими стали: «Скептичный - Параноидный» и «С богатым воображением - Шизотипический» у мужчин; и «Осторожный - Уклоняющийся», «Сам в себе - Шизоидный», «Самоуверенный - Нарциссический», «Исполненный сознания долга - Зависимый» - у женщин.

Научная новизна заключается в выборе психоаналитической традиции в качестве модели объяснения и теоретической проработки темы. Мы продемонстрировали, что анализ эмпирических данных позволяет увидеть гендерную специфику деструкторов, выявленных у руководителей российских компаний. Это, в свою очередь, позволяет нам говорить о подтвержденной исходной гипотезе нашего исследования.

Одновременно с этим опора на целый ряд классических исследований и теорий позволяет увидеть в эмпирических результатах психоаналитические основы деструкторов, проявляющихся у мужчин и женщин - руководителей.

Почему мы считаем психоаналитический подход полезным и эвристичным? Если говорить в целом, эта традиция дает возможность прикоснуться к первоосновам деструктивных моделей лидерства. Это особенно важно на фоне большого числа публикаций как о «токсичном лидерстве», так и о гендерной дискриминации («стеклянный потолок» для женщин, претендующих на руководящие позиции), стилистика и характер которых говорит о выраженной «моральной панике». По большому счету, приоритет в таких работах отдается социальному контексту, который рассматривается в качестве ключевого фактора как деструктивного лидерства, так и связанных с ним патологий организаций.

Психоаналитический взгляд на проблему нам представляется более аккуратным и точным в своей базовой аналитической стратегии. Во-первых, он демонстрирует связь лидерских качеств (в том числе и деструкторов) с личностными характеристиками, сформировавшимися у человека в детстве под влиянием особенностей его взаимодействия с матерью и отцом. Область бессознательного становится, таким образом, еще одним объясняющим фактором - наряду с социо-культурным контекстом.

Во-вторых, он естественным образом вводит идею гендерной специфики лидерства, поскольку опирается на описанные различия в процессе формирования личностных характеристик у мальчиков и у девочек. На наш взгляд такое более глубокое и гендерно дифференцированное понимание феномена позволяет избежать примитивизации как наших аналитических стратегий, так и предложений, что следует делать, чтобы найти «быстрое и эффективное» решение как поддержать развитие женского лидерства. Например, идея «позитивной дискриминации» женщин — руководителей.

В-третьих, он позволяет увидеть природу деструкторов сразу же в их гендерной несопоставимости, асимметричности для мужчин и женщин. Генезис «мужских» деструкторов скорее предопределен ролью матери и приводит к параноидальному или шизотипическому типу их проявления в лидерстве. На формирование женских деструкторов сильное влияние оказывает роль отца, если она была «токсичной». Это чревато переносом модели отношений отца и дочери на отношения с авторитетными и властными «значимыми мужчинами».

В-четвертых, психоаналитический подход предлагает эвристичные модели объяснения эмпирически наблюдаемым фактам (наблюдению, экспериментальным и статистическим данным, материалам консультаций и коучинговых сессий), предлагает рабочий понятийный аппарат: «защитные механизмы», «системы мотивационных потребностей», «главная тема конфликтных отношений», «типы организации личности» и др.

В-пятых, психоаналитическая традиция укрепляет наше понимание, почему деструкторы проявляются именно в стрессовой ситуации, объясняет механизмы такого регресса относительно «нормальной», рабочей ситуации как на индивидуальном, так и групповом уровне.

7 Профессиональные роли журналистов: об исследовательском проекте

7.1 Введение

Содержание различных медиа-платформ и новостей меняется в разных культурах. Целью этой работы является лучшее понимание факторов, объясняющих различные модели журналистики во всем мире, а также разрыв между нормами, профессиональными идеалами и позициями, представленными в новостных публикациях.

Анализ основан на результатах второй волны проекта «Реализация журналистских ролей» (Journalistic Role Performance, JRP) (www.journalisticperformance.org) - международного исследования, основанного на данных 365 средств массовой информации в 37 странах. Руководитель исследования – профессор Клаудия Мелладо (Университет Вальпараисо, Чили). С российской стороны руководителем национальной исследовательской команды выступает аналитик МЛ ПСА С.Г. Давыдов. Основной задачей второй волны исследования, которая проводилась в 2020 – 2021 гг., стал систематический анализ того, как различные профессиональные роли анализируются в новостях. Анализ журналистских ролей в печатных СМИ и проектах национальных газетных новостей был проведен в рамках первой волны исследования, которая проводилась в период с 2013 по 2017.

Исследование базируется на комплексном подходе. Основываясь на стандартизированной операционализации ролей «сторожевого пса», интервенциониста, лояльного посредника, гражданской, информационно-развлекательной и обслуживающей ролей в журналистике, сначала было проведено измерение уровня представленности той или иной журналистской роли в новостных текстах посредством контент-анализа. Для выявления связи между оценочным и перформативным уровнями журналистской культуры далее был проведен опрос о концепции роли, ее восприятии и воплощении среди журналистов, работающих в новостных медиакомпаниях, включенных в нашу выборку. Это позволило сравнить оценки журналистов со средними показателями работы их новостных СМИ по странам. Каждая национальная команда собирала информацию на организационном/институциональном уровне своих средств массовой информации и на уровне стран. В исследовании используется методическая схема, основанная на сравнительном изучении демократических, переходных и недемократических стран. Журналистская практика встроена в рутину и осуществляется в рамках социальной системы, которая служит основой для создания медиаконтента. Таким образом, возникает возможность глубоко изучить профессиональные роли, как различные социальные и культурные контексты объясняют различия в воплощении журналистских ролей для разных медиаплатформ и тем медийной повестки дня.

Полученные данные были проанализированы посредством методологии контент-анализа. Затем к полученным результатам была применена методология сетевого анализа для построения сети со-встречаемости журналистских ролей.

Полученные результаты показывают, что осуществляемые в практической деятельности журналистов роли в значительной степени отличаются от «идеальных типов», сформулированных и описанных в теоретических работах по медиа и журналистике. Результаты сетевого анализа показывают связи как между ролями, так и между отдельными индикаторами конкретных ролей, что позволяет делать выводы об их целостности и пересечению (наслоению) друг с другом.

7.2 Методы сбора и обработки данных

Сбор данных. Стремясь получить гетерогенную выборку, авторы исследования выбрали страны, которые представляют различные политические режимы, географические регионы и классификации медиасистем. В исследование вошли государства Северной Америки, Латинской Америки, Западной Европы, Восточной Европы, Азии, Африки, Ближнего Востока и Океании. Следуя моделям западных медиасистем Д. Халлина и П. Манчини, в выборку вошли страны, которые представляют либеральную, демократическую корпоративистскую и поляризованную плюралистическую модели. Также были использованы индексы демократии и отчеты о свободе прессы для классификации демократических стран с переходной экономикой и недемократических стран из разных частей мира.

Данные были собраны соответствующими национальными исследовательскими коллективами в 37 странах – Аргентине, Австралии, Австрии, Бельгии, Бразилии, Канаде, Чили, Колумбии, Кубе, Эквадоре, Египте, Великобритании, Эстонии, Эфиопии, Франции, Германии, Венгрии, Ирландии, Израиле, Италии, Японии, Кувейте, Ливане, Мексике, Парагвае, Польше, Катаре, России, Руанде, Сербии, Испании, Швейцарии, Тайване, Объединенных Арабских Эмиратах, США и Венесуэле.

Полевые работы по контент-анализу, онлайн-опросу и сбору данных на организационном и социальном уровне проводились в период с 2020 по 2021 год. Все данные контент-анализа, опроса и измерения структурного контекста были собраны независимо. Это позволило организовать работу на разных уровнях анализа: новости, журналисты, средства массовой информации и страны. **Обработка данных контент-анализа.** В выборку контент-анализа вошли новости, опубликованные в крупнейших газетах, на сайтах, в новостных радио- и телепрограммах стран-участниц исследования.

Хотя для отбора конкретных текстов внутри каждой страны использовались разные процедуры в зависимости от типа анализируемой медиаплатформы, все они были связаны общими техническими аспектами, чтобы гарантировать возможность использования глобальной выборки для проведения сравнений. Эти аспекты связаны с общим периодом времени, для которого были взяты тексты для анализа, одинаковыми днями анализа и с одними и теми же единицами анализа для всех стран. Была использована следующая модель построения выборки исследования. Исследователи в каждой из 37 стран-участниц выбрали от двух до четырех новостных СМИ на каждую платформу. Критериями выбора средств массовой информации были размер аудитории, охват и уровень влияния на формирование национальной повестки дня. Исследователям было предложено отобрать наиболее популярные СМИ (наиболее читаемые, просматриваемые или прослушиваемые) в своем классе на основе рейтингов или аналогичных параметров, и предпочтение отдавалось медиа национального масштаба, хотя в исследование были включены также региональные и местные медиа площадки.

Учитывая, что структура и формат медиасистем в разных странах во многом различаются, включая размер, ориентацию аудитории, собственность в сфере медиа, политические взгляды и наличие более чем одного языка на территории, исследователей попросили убедиться, что выбранные средства массовой информации представляют разнообразие медиа каждой страны в максимально возможной степени. Членам исследовательских команд пришлось принять во внимание тот факт, что количество включенных средств массовой информации может варьироваться от страны к стране и что большая гетерогенность в системе средств массовой информации приведет к тому, что исследователи будут включать в выборку больше средств массовой информации, и наоборот.

Чтобы контролировать потенциальную перепредставленность и/или недопредставленность конкретных типов СМИ в выборке, возникающую из-за того, что некоторые средства массовой информации включают в выборку больше материалов, чем другие, было принято решение взвесить данные по средствам массовой информации для каждой страны. Это гарантировало, что каждый тип СМИ – телевидение, радио, онлайн-новости и газеты – в каждой стране будут иметь одинаковый вес в обобщенных результатах. Количество проанализированных медиа и новостей в каждой стране представлено в Таблице ??.

Количество новостных текстов и СМИ в выборке исследования по странам

| СТРАНА | НОВОСТНЫЕ ТЕКСТЫ | КОЛИЧЕСТВО СМИ |
|----------------|-------------------------|-----------------------|
| Аргентина | 5368 | 10 |
| Австралия | 1965 | 8 |
| Австрия | 4821 | 10 |
| Бельгия | 2411 | 7 |
| Бразилия | 3679 | 9 |
| Канада | 3727 | 12 |
| Чили | 7512 | 11 |
| Колумбия | 5138 | 8 |
| Эквадор | 2892 | 8 |
| Египет | 3484 | 11 |
| Великобритания | 4185 | 15 |
| Эстония | 2409 | 11 |
| Эфиопия | 1400 | 10 |
| Франция | 4661 | 9 |
| Германия | 4777 | 6 |
| Венгрия | 3358 | 13 |
| Ирландия | 2421 | 8 |
| Израиль | 2448 | 8 |
| Италия | 4494 | 11 |
| Япония | 3757 | 9 |
| Кувейт | 1868 | 9 |
| Ливан | 3665 | 14 |
| Мексика | 7905 | 12 |
| Польша | 6230 | 14 |
| Катар | 1559 | 6 |
| Россия | 6955 | 10 |
| Руанда | 2644 | 7 |
| Сербия | 6067 | 11 |
| Южная Корея | 3959 | 8 |
| Испания | 6089 | 12 |

Количество новостных текстов и СМИ в выборке исследования по странам

| | | |
|-----------|---------|-----|
| Швейцария | 3555 | 10 |
| Тайвань | 6790 | 10 |
| ОАЭ | 2726 | 9 |
| США | 3992 | 11 |
| Венесуэла | 2443 | 12 |
| Куба | 2834 | 8 |
| Парагвай | 4286 | 8 |
| ВСЕГО | 148,474 | 365 |

С помощью метода «сконструированной недели» была сформирована стратифицированно-систематическая выборка размерностью две недели для каждого средства массовой информации в каждой стране со 2 января по 31 декабря 2020 года. Во всех странах, включенных в исследование, были проанализированы одни и те же дни. Поскольку ежедневные и ежемесячные колебания публикационной активности являются важными факторами, которые следует учитывать при проведении анализа контента новостей, мы разделили год на два периода (полугодия) по шесть месяцев: с января по июнь и с июля по декабрь. Затем внутри этих периодов по определенному алгоритму были отобраны дни. Алгоритм выбора дней подробно представлен на сайте проекта <https://www.journalisticperformance.org/>. В результате в каждом из двух периодов (полугодий) была сконструирована неделя из неповторяющихся дней для проведения измерений, и итоговая выборочная совокупность в исследовании составила 14 дней одного года. Использование такого алгоритма позволило достигнуть презентации по месяцам (в каждом месяце выбран хотя бы один день для измерения) и медиа-источникам (в каждый из дней хотя бы одна из программ/выпусков/изданий) для их более равномерного представления.

Исследовательская команда в каждой из стран-участниц проекта использовала такие критерии для отбора единиц наблюдения в отобранных СМИ: - для новостей на выбранных телеканалах: выпуски с наибольшим количеством просмотров; - для новостей на выбранных радиоканалах: программы с наибольшим количеством прослушиваний; - для отобранных печатных (газет): полный газетный выпуск; - для отобранных онлайн средств массовой информации (сайтов): полная заглавная страница и все идущие с нее ссылки.

Поскольку онлайн средства массовой информации имеют возможность публиковать информацию не только в заранее определенное время (как другие СМИ), а по мере ее поступления, замеры на сайтах проводились в две определенные исследователями точки времени: в 11:00 и 23:00. Этот 12-часовой перерыв между двумя снимками с высокой долей вероятности обеспечивает наибольшую вариативность контента. Домашние страницы и соответствующие ссылки открывались и сохранялись в режиме реального времени.

Единицей анализа была новость. Новость определялась как совокупность связанных элементов из слов, аудио-записей и визуального материала, аффилированных с определенным происшествием/вопросом/персоной. Для каждой из отобранных единиц анализа собирались все новости по широкому кругу тем (правительство и законодательные органы, выборы и предвыборные кампании, протестная активность, экономика и бизнес, полиция и преступность, судебная деятельность, оборона и безопасность, образование,

здоровье, проблемы социальной сферы, окружающая среда, энергетика, транспорт, жилье, происшествия и несчастные случаи, религия и церковная деятельность, занятость и трудовая деятельность, медиа-среда, спорт, наука и технологии, образ жизни, культурный досуг и развлечения, жизнь знаменитых людей). Затем тематические позиции были перекодированы в широкие темы.

Об одном событии, проблеме или явлении средство массовой информации может сообщать более чем в одной новости в один и тот же день. Если одно и то же событие, явление или проблема освещались более чем в одном материале, они считались отдельными историями и кодировались отдельно.

Редакционные статьи, колонки мнений, прогнозы погоды, гороскопы, обзоры фильмов (или других культурных событий), головоломки, страницы в социальных сетях и аналогичный контент на радио и телевидении не были включены в исследование. Авторы проекта исключили приложения/журналы/специальные программы и заголовки на первых полосах газет и в начале теле- и радиовыпусков. В целях сравнения данных опроса о ролевых концепциях журналистов с содержанием их новостных организаций также был исключен контент, который не был создан сотрудниками соответствующих редакций — например, статьи новостных агентств или статьи не журналистов, размещенные на новостных сайтах. В случае онлайн-СМИ кодировались только те новости, которые появлялись на расширенной домашней странице. Также были закодированы элементы, включающие встроенные видео или аудиоклипы.

В целом, выборка состоит из 148 474 новостей, опубликованных 365 СМИ (102 газет, 96 телевизионных выпусков новостей, 74 новостных радиопрограмм и 93 новостных веб-сайтов).

Построение сети со-встречаемости журналистских ролей. На основе имеющихся данных может быть создана сеть, основанная на связях между публикаций и журналистскими ролями, на основе которой может быть построена сеть связей между журналистскими ролями на основе их совместного присутствия в рассматриваемых и кодируемых в рамках исследования публикациях.

Сетевые данные были представлены в формате несимметричной матрицы, где по строкам были указаны все наблюдения, а по столбцам – индикаторы, соответствующие изучаемым показателям (журналистским ролям). Фрагмент такой матрицы на примере информации о тематиках, раскрывающихся в статьях, приведен на Рис. ???. С помощью кода, написанного на языке программирования Python, матрицы была конвертированы в сетевые файлы в виде двумодальных сетей с расширением .net. Пример конвертации данных в формат 2-модальной сети приведен на Рис. ???. На имеющихся данных была построена 2-модальная сеть (сети, состоящие из двух наборов узлов, где связи могут существовать только между узлами из разных наборов) публикаций (Publications) и ролей (Roles) журналистов PR.

Создание 1-модальных сетей из 2-модальных сетей производилось через сетевое перемножение. Эта процедура подразумевает, что двумодальная сеть, представленная в виде матрицы, перемножается на свою транспонированную версию, и один из наборов узлов становится посредником для формирования связей между узлами другого набора. В частности, в данном случае была построена сеть со-встречаемости журналистских ролей RR, где вес связей указывал на количество раз, когда эти сущности встречались вместе в одном закодированном материале (статье). Анализ полученных сетевых данных проводился в программе для анализа и визуализации больших сетей Rajeck.

| Номер узла | 148475 | 148476 | 148477 | 148478 | 148479 | 148480 | 148481 | 148482 | 148483 | 148484 | 148485 | 148486 | 148487 | 148488 | 148489 | 148490 | 148491 | 148492 | 148493 | 148494 | 148495 | 148496 | 148497 | 148498 |
|--------------------------------------|------------|-----------|--------|--------|---------|---------|-----------|-------------|--------|----------------|---------|----------|--------|----------|--------|---------|-------------|--------|--------|---------|-----------|---------|---------------|--------|
| Номер узла/ID наблюдения и категория | GOVERNMENT | CAMPAIGNS | POLICE | COURT | DEFENSE | ECONOMY | EDUCATION | ENVIRONMENT | ENERGY | TRANSPORTATION | HOUSING | ACCIDENT | HEALTH | RELIGION | LABOR | PROTEST | SOCIALISSUE | MEDIA | SPORT | SCIENCE | LIFESTYLE | CULTURE | ENTERTAINMENT | OTHER |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 7 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 9 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 11 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 12 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 13 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 14 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 15 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 16 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 17 | 17 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 19 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 21 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 22 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 24 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 26 | 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27 | 27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28 | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29 | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30 | 30 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| Узлы в 1-mode (до 148,474) | | Узлы в 2-mode (с 148,475) | | Связи между узлами | |
|--|--|---|--|---|--|
| *Vertices 148498 148474 1 1 2 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9 10 10 11 11 12 12 13 13 14 14 15 15 16 16 17 17 18 18 19 19 20 20 21 21 22 22 23 23 24 24 25 25 | | 148473 148473 148474 148474 148475 "GOVERNMENT_LEGISLATURES" 148476 "CAMPAIGNS_ELECTIONS" 148477 "POLICE_CRIME" 148478 "COURT" 148479 "DEFENSE_SECURITY" 148480 "ECONOMY" 148481 "EDUCATION" 148482 "ENVIRONMENT" 148483 "ENERGY" 148484 "TRANSPORTATION" 148485 "HOUSING" 148486 "ACCIDENTS" 148487 "HEALTH" 148488 "RELIGION" 148489 "LABOR_EMPLOYEMENT" 148490 "PROTEST" 148491 "SOCIAL_ISSUES" 148492 "MEDIA" 148493 "SPORT" 148494 "SCIENCE_TECHNOLOGY" 148495 "LIFESTYLE" 148496 "CULTURE" 148497 "ENTERTAINMENT_CELEBRITY" 148498 "OTHER" | | *Arcs 1 148477 2 148486 3 148488 4 148485 5 148482 6 148491 7 148476 7 148488 7 148475 8 148489 8 148491 9 148477 10 148498 10 148495 11 148496 11 148497 12 148493 13 148477 13 148491 13 148490 14 148480 14 148475 14 148476 15 148475 15 148483 15 148480 | |

Результаты исследования Базовые характеристики полученных сетей. Двумодальная сеть публикаций и ролей PR состоит из 6 ролей и 67 индикаторов ролей. Количество ролевых индикаторов по каждой роли приведено в Табл. ???. Роль «сторожевого пса» содержит значительно большее количество индикаторов – 30 – по сравнению с другими ролями, куда согласно дизайну исследования включено от 5 до 10 индикаторов. Визуализация ролей и индикаторов представлена на Рис. 10.

Глобальные роли и ролевые индикаторы

| N | Global role | N of indicators | Freq% |
|-------|-------------------|-----------------|-------|
| 1 | WATCHDOG | 30 | 45 |
| 2 | LOYAL-FACILITATOR | 10 | 15 |
| 3 | CIVIC | 10 | 15 |
| 4 | SERVICE | 7 | 10 |
| 5 | INTERVENTIONIST | 5 | 7.5 |
| 6 | INFOTAINMENT | 5 | 7.5 |
| Всего | | 67 | 100 |

Некоторые ролевые индикаторы, которые относятся к различным журналистским ролям, чаще других встречаются в публикациях. Таблица ?? показывает, что наибольшее число упоминаний имеют

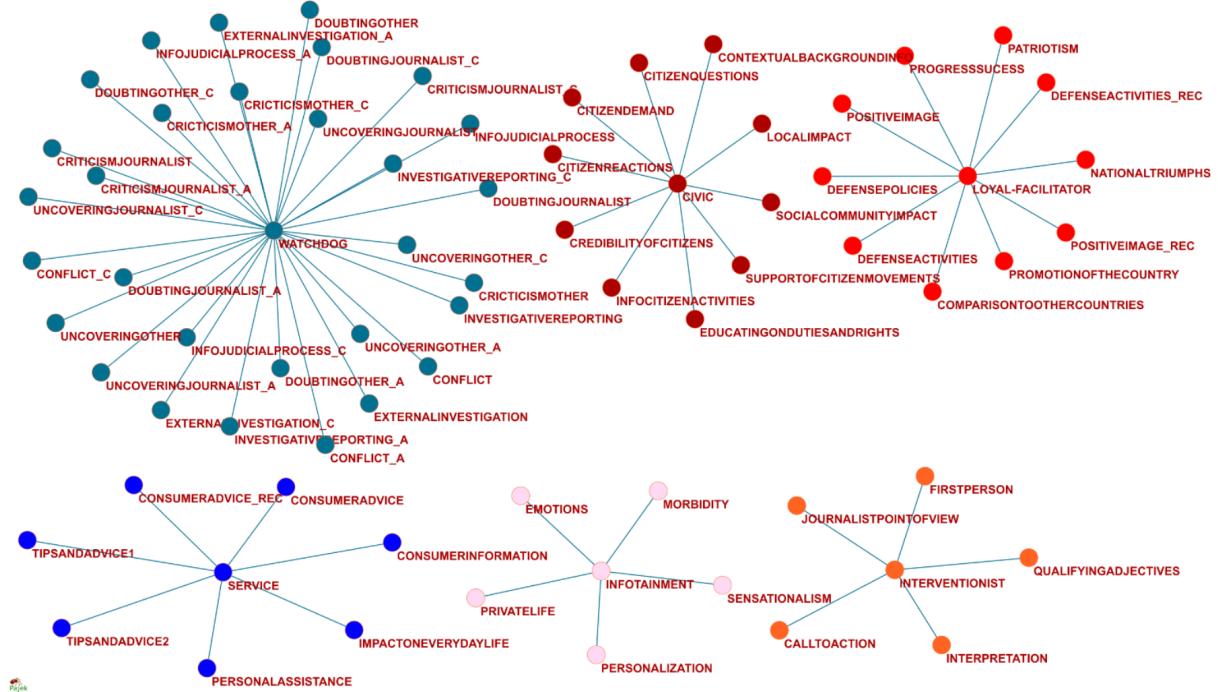


Figure 10: Визуализация сети публикаций и ролей PR

такие ролевые индикаторы как QUALIFYINGADJECTIVES (квалифицирующие прилагательные) и INTERPRETATION (интерпретация), которые встречаются заметно чаще остальных индикаторов (50,297 и 42,687 раз соответственно). Следующие за ними пять индикаторов встречаются в 20+ тыс. публикаций.

Ролевые индикаторы с
наибольшим количеством
упоминаний в публикациях

| # | Value | Id |
|----|-------|--------------------------|
| 1 | 50297 | QUALIFYINGADJECTIVES |
| 2 | 42687 | INTERPRETATION |
| 3 | 26850 | PERSONALIZATION |
| 4 | 25868 | CRITICISMOTHER |
| 5 | 24302 | JOURNALISTPOINTOFVIEW |
| 6 | 23330 | CRITICISMOTHER_A |
| 7 | 21853 | EMOTIONS |
| 8 | 19274 | SENSATIONALISM |
| 9 | 17113 | INFOJUDICIALPROCESS |
| 10 | 16387 | IMPACTONEVERYDAYLIFE |
| 11 | 15203 | LOCALIMPACT |
| 12 | 13347 | CONSUMERINFORMATION |
| 13 | 13248 | CONTEXTUALBACKGROUNDINFO |
| 14 | 12724 | SOCIALCOMMUNITYIMPACT |

Ролевые индикаторы с
наибольшим количеством
упоминаний в публикациях

| | | |
|----|-------|-----------------------|
| 15 | 12623 | CITIZENREACTIONS |
| 16 | 11939 | DOUBTINGOTHER |
| 17 | 11802 | PRIVATELIFE |
| 18 | 11331 | POSITIVEIMAGE_REC |
| 19 | 11331 | POSITIVEIMAGE |
| 20 | 11075 | INFOJUDICIALPROCESS_A |
| 21 | 10675 | DOUBTINGOTHER_A |

Одномодальная сеть совстречаемости журналистских ролей RR состоит из 67 узлов и 2,195 связей между ними; при этом сила связей между узлами варьируется от 1 до 23,458. Особенность этой сети состоит в том, что она является плотно связанный – практически все узлы связаны друг с другом, и плотность составляет 0.98. Стратегия анализа сетей такого рода заключается в обрезке наименее сильных связей ниже определенного порогового значения с помощью алгоритма line cut для того, чтобы в сети между узлами остались наиболее сильные связи. На Рис. 11 представлена визуализация полученной сети в двух вариациях: с пороговым значением веса связей в 1,000 и 1,500 (количество публикаций, в которых два индикатора ролей встречались совместно).

Для того, чтобы видеть связи внутри и между группами ролевых индикаторов, соответствующие им узлы в сети сгруппированы по ролям. Можно видеть многочисленные связи между группами (по цепочке) синих, красных, оранжевых и розовых узлов, то есть ролями WATCHDOG («Сторожевой пес»), CIVIC, INTERVENTIONALIST и INFOTAINMET, а также синих и оранжевых - WATCHDOG и INTERVENTIONALIST. Индикаторы внутри двух оставшихся ролей – LOYAL-FASILITATOR и SERVICE – тоже связаны с индикаторами других ролей, однако в менее сильной степени; обращает на себя внимание отсутствие связей между индикаторами из группы CIVIC и LOYAL-FASILITATOR.

Эти выводы становятся еще более видными после укрупнения этой сети до уровня ролей и связей внутри них (представленных в виде петель) и между ними (Рис. ??). Можно обратить внимание, что сеть представляет собой почти полную клику, за исключением отсутствующих связей между ролями CIVIC и LOYAL-FASILITATOR.

Сетевая кластеризация - блокмоделинг. Репрезентация сети в виде матрицы с удачным порядком расположения узлов уже может демонстрировать разделение сети на группы и показать пересечения между узлами, относящимися к разным категориям. Матричная репрезентация сети связей между индикаторами ролей по со-встречаемости RR представлена на Рис. 12. Матрица симметрична по столбцам и строкам, где представлены индикаторы журналистских ролей по 6 ролям (отмечены разным цветом). Темно-синие линии делят матрицу на группы (кластеры). Кластеры по диагонали включают индикаторы, относящиеся к одной роли, и показывают связи между ними. Кластеры вне диагонали показывают связи между индикаторами разных групп. Можно видеть наличие сильных связей внутри оранжевой и коричневой группы индикаторов, между ними также наблюдаются связи. Внутри большой группы индикаторов, относящихся к роли «Сторожевого пса», индикаторы связаны

друг с другом неравномерно – тогда как у некоторых есть много связей с другими, часть индикаторов занимают довольно изолированную позицию. Однако более точным для выделения групп и определения связей между ними является вариант сетевой кластеризации – блокмоделинг. В работе был применен непрямой подход к блокмоделингу, в ходе которого сначала рассчитывается мера близости / дальности между узлами на основании их структурного положения, на основе которой узлы затем относятся к тому или иному кластеру. В ходе анализа для расчета несходства элементов (dissimilarity) использовалось скорректированное Евклидово расстояние, которое учитывало различные веса связей в сети (вариант d5 в программе Pajek). На основе полученной дендрограммы (Рис. 13) было сделано предположение о наличии в сети трех кластеров, матрица с которыми показана на Рис. 8. Можно видеть, что наиболее плотно связанный кластер (справа внизу) состоит из восьми ролевых индикаторов. Второй кластер из 32 индикаторов также имеет достаточно много связей внутри себя, а также сильно связан с кластером 1. Индикаторы, вошедшие в кластер 3, не имеют значительного числа связей с индикаторами из своей группы, а также имеют выборочные связи с индикаторами из групп 1 и 2. Чтобы понять, к каким журналистским ролям относятся индикаторы, вошедшие в три кластера, для каждого лейбла был задан цвет, к которому он относится исходя из принадлежности к определенной роли (Рис. ??, справа). Полученные результаты позволяют говорить о том, что похожие друг на друга по структурным характеристикам кластеры (в смысле соотношения индикаторов друг с другом) на деле образованы индикаторами, которые относятся к разным журналистским ролям. Так, наиболее плотно связанный кластер 1, где имеется много связей между индикаторами, формируется индикаторами из ролей, обозначенных коричневым (гражданская журналистика), оранжевым (интервенционист) и синим («сторожевой пёс»).

Для визуализации веса связей нормализованы по максимальному значению. Размер узла приведен в соответствие с взвешенным показателем входящей центральности (weighted degree). Цвет узлов соответствует разделению на группы по 6 ролям.

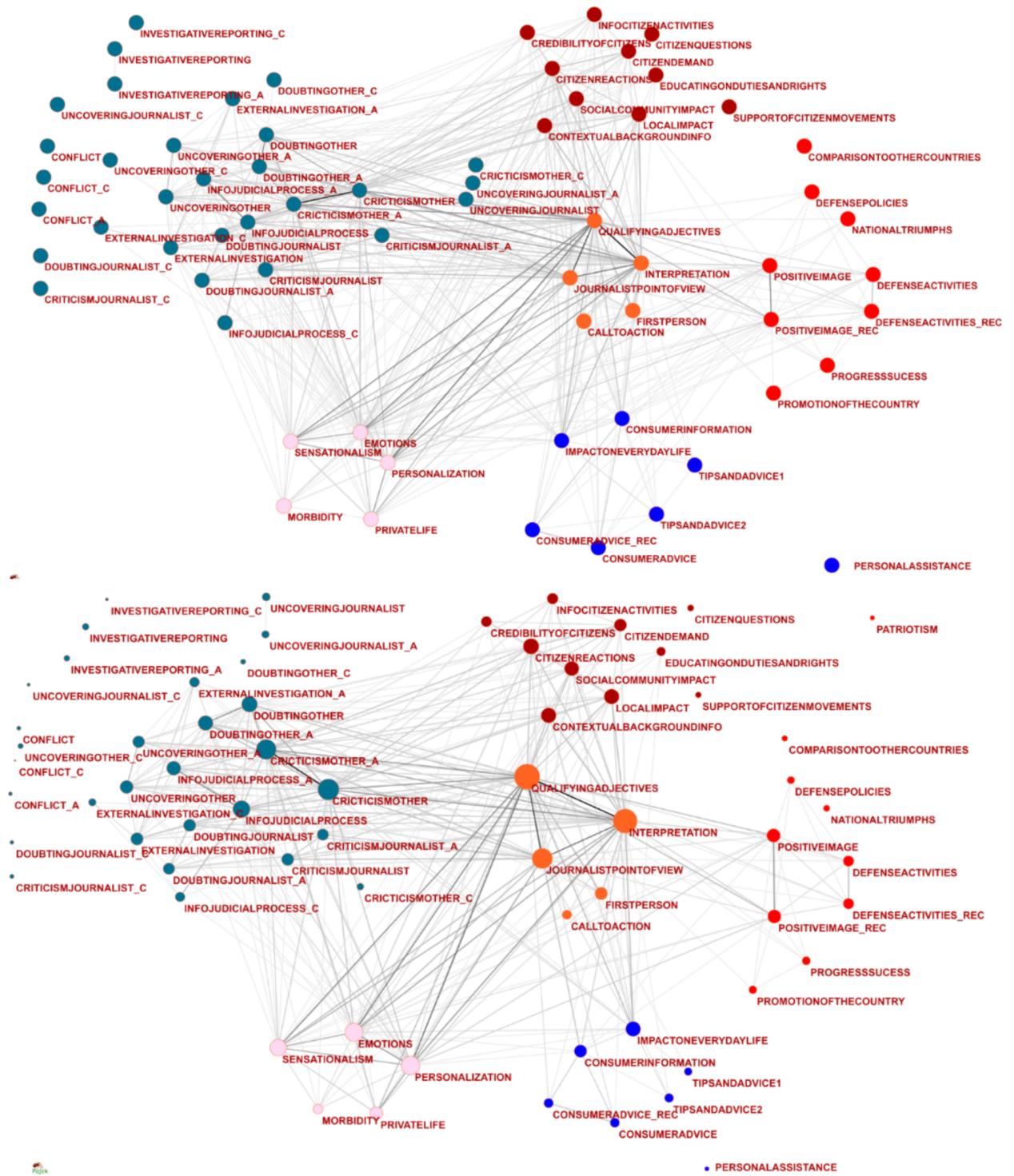
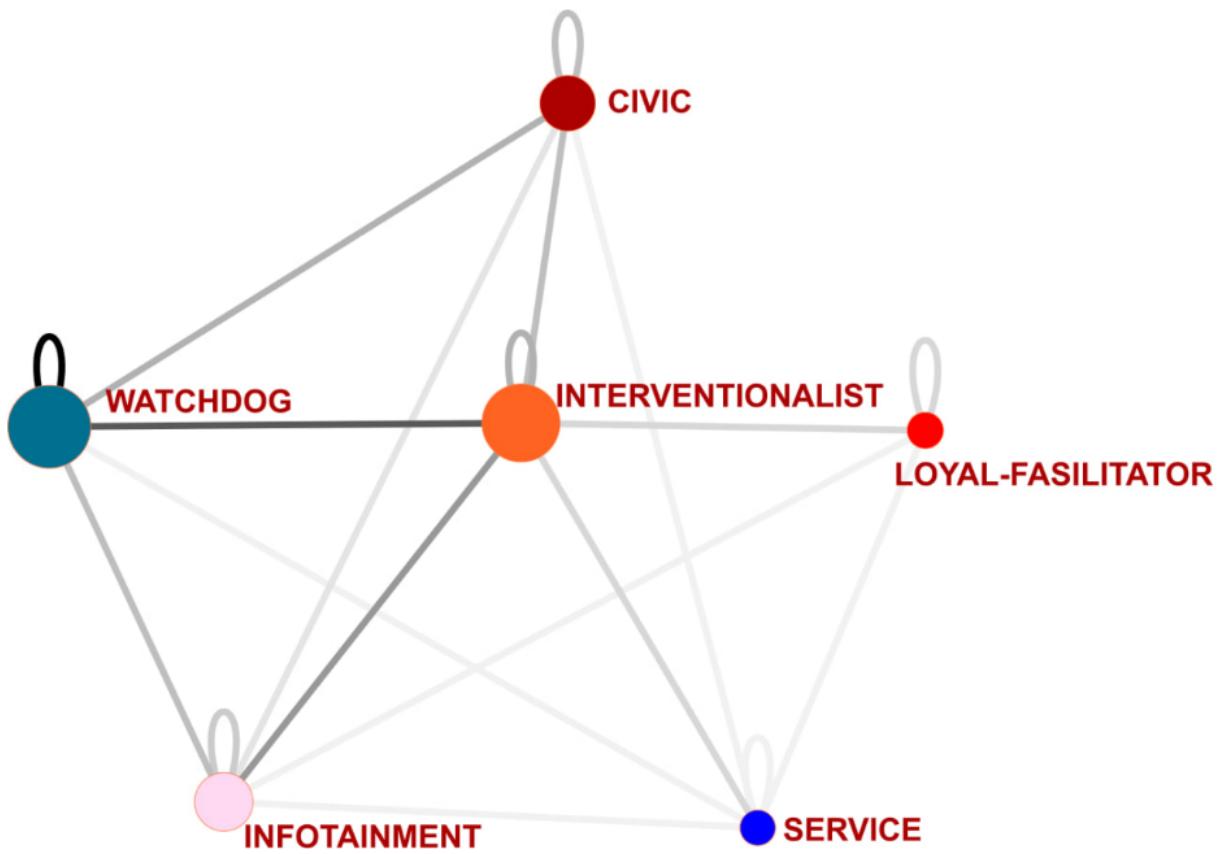
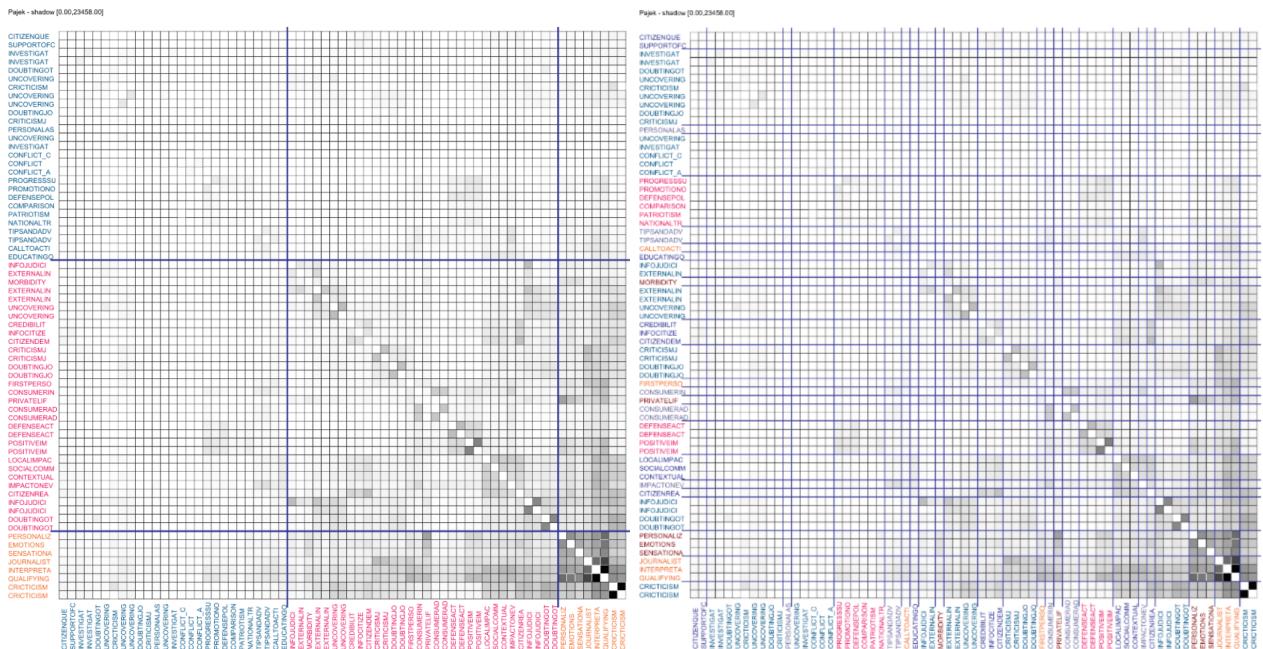


Figure 11: Визуализация сети связей между индикаторами ролей по со-встречаемости RR (пороговые значения по весу связи 1000, 1500)



Укрупненные узлы по ролям расположены на тех же местах графа, что и группы узлов на Рис. 11.



Обсуждение и выводы Полученные результаты показывают, что осуществляемые в практической деятельности журналистов роли в значительной степени отличаются от «идеальных типов», сформулированных и описанных в теоретических работах по медиа и журналистике. Оказалось, что роли интервенциониста, «сторожевого пса» и гражданского журналиста в значительной степени пересекаются по своим

Pajek - shadow [0.00,23458.00]

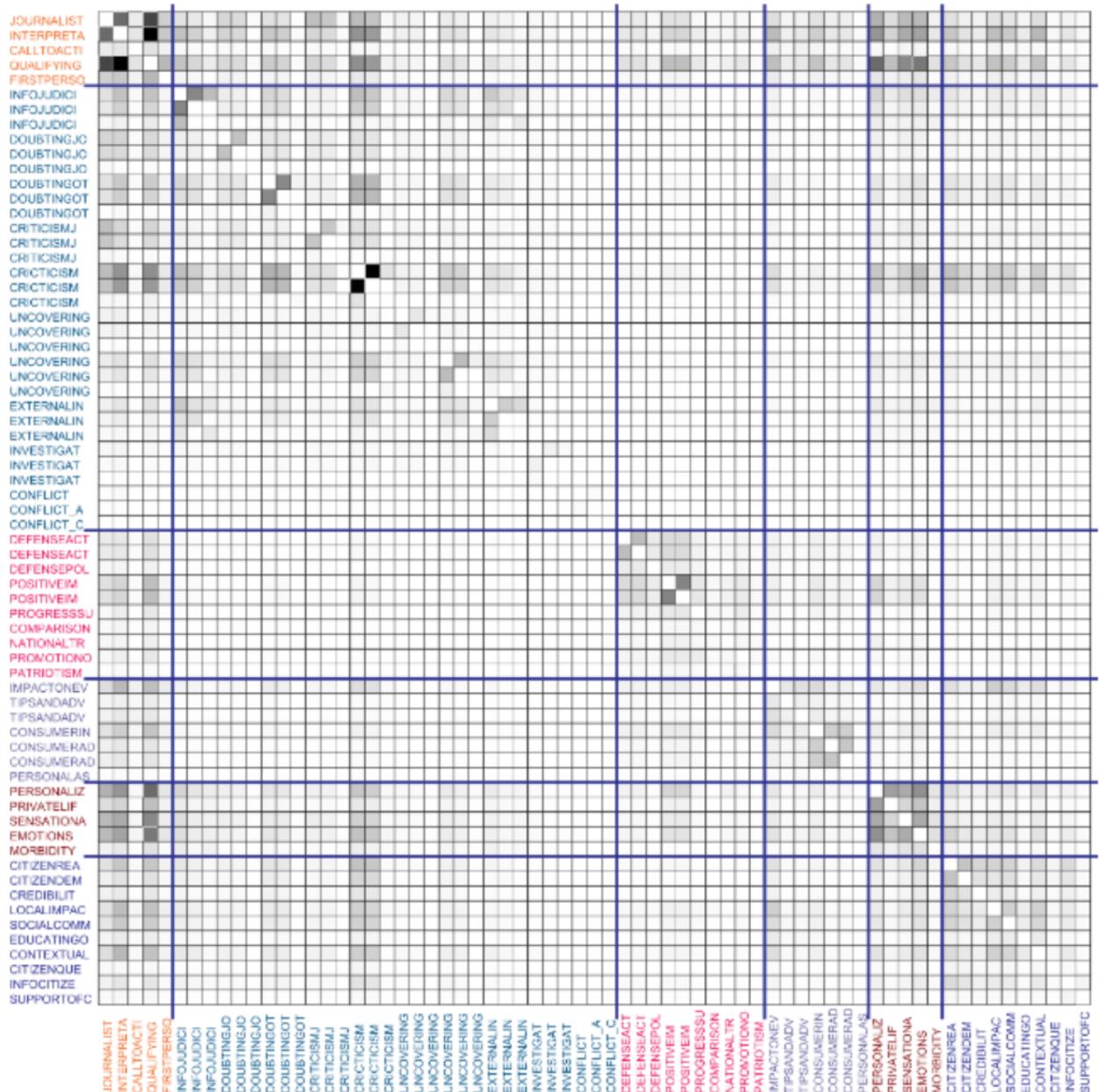


Figure 12: Визуализация сети связей между индикаторами ролей по со-встречаемости RR в виде матрицы

Pajek - Ward [0.00,295056.88]

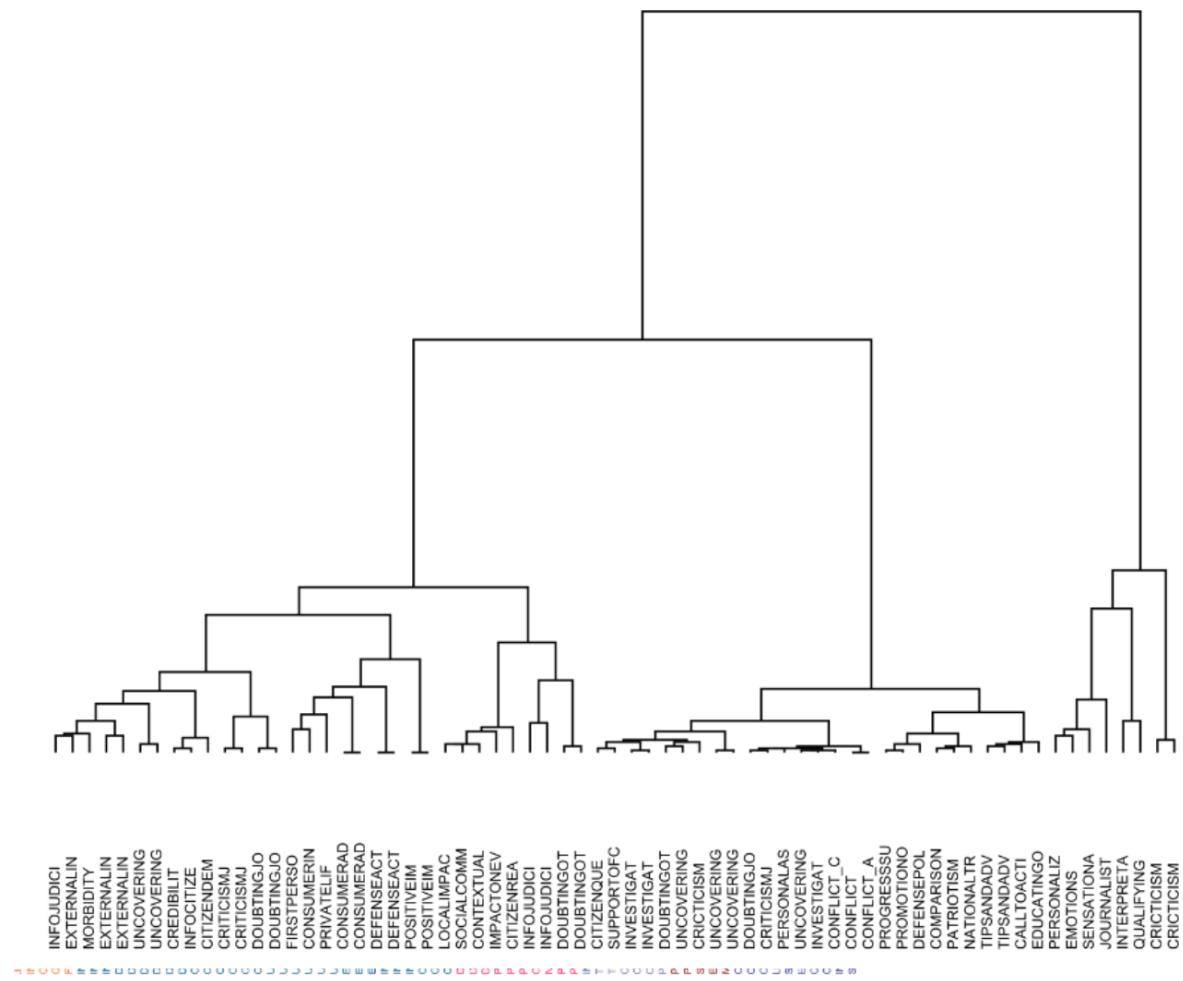


Figure 13: Дендрограмма по расчетам скорректированного Евклидова расстояние

индикаторам. В наименьшей степени пересекаются с другими ролями индикаторы лояльного фасилитатора.

Наиболее целостной, если судить по пересечению индикаторов внутри одной роли, является роль «сторожевого пса». У остальных ролей данный показатель заметно ниже, слабее всего внутренние связи выражены у сервисной роли. Самая слабая связь наблюдается между ролями гражданского журналиста и лояльного фасилитатора, что вполне ожидаемо, поскольку первый отстаивает интересы гражданских сообществ, тогда как второй в первую очередь поддерживает государственную бюрократию и находящиеся у власти элиты.

8 Применение ERGM для анализа конференций

8.1 Введение

В последние годы концепция “умного города” привлекает все большее внимание, поскольку городские центры сталкиваются с проблемами, связанными с быстрой урбанизацией, нехваткой ресурсов и необходимостью устойчивого развития (Hollands, 2015; Kirimtac et al., 2020). В умном городе используются передовые технологии и подходы, основанные на данных, для оптимизации городской инфраструктуры, повышения качества жизни горожан и обеспечения устойчивого экономического роста. По сути, эта концепция воплощает в себе видение технологически обусловленной городской утопии, направленной на преобразование и улучшение городской среды (Townsend, 2013). Хотя глобальное понимание умного города в первую очередь связано с внедрением интеллектуальных технологий и аспектов искусственного интеллекта, важно признать, что этот термин включает в себя не только цифровизацию. Т. Нам и Т. Прадо (Nam & Pardo, 2011) придерживаются комплексного подхода к этому понятию, охватывающего три аспекта: технологии, людей и сообщества/институты. Таким образом, каждый город ставит большие задачи перед людьми и институтами, ответственными за его прямое или косвенное управление, планирование и модификацию (Ricciardi & Za, 2015).

В силу своей междисциплинарной природы концепция охватывает различные области - городское планирование, внедрение инноваций, устойчивое развитие и менеджмент. Академические дискуссии вокруг темы “умного города” по своей природе характеризуются междисциплинарностью. Среди упоминаний “умного города” доминируют области ИКТ, экологии и энергетики, а также урбанизации (Ingwersen & Serrano-López, 2018). Сложность этих дискуссий обусловлена разнообразным и многообразным использованием термина, а также его связями со смежными понятиями (например, интеллектуальный, цифровой, информационный) (Hollands, 2015). Как отмечает А. Коккиа, “концепция умного города до сих пор не имеет единого определения и по-разному интерпретируется в зависимости от области внимания” (Cocchia, 2014).

Помимо присущей этому термину неоднозначности, практическое понимание умного города и целей, лежащих в основе программы применения концепции в России, также имеет свои особенности. Крупные российские агломерации активно участвуют в разработке инициатив, охватывающих различные сферы, таких как: здравоохранение, образование, утилизацию отходов, сбор данных и мониторинг. Москва, Санкт-Петербург и Казань достигли стадии “Умный город 3.0”, характеризующейся участием граждан в инновациях (Vishnivetskaya & Alexandrova, 2019). Россия представляет собой крайне особенный

пример реализации концепции в силу значительной концентрации городского населения и относительно небольшой доли населения, проживающего в сельской местности. Неосведомленность граждан является существенным препятствием для развития “умных городов” в России (Yuloskov et al., 2021).

Однако настоящее исследование отвечает на другой вопрос, не связанный с концептуализацией термина. Его цель - проанализировать, какие институты вносят вклад в развитие концепции “умного города” в российском академическом контексте. В исследовании рассматривается процесс сотрудничества в российском научном сообществе, выявляются его закономерности, ключевые участники и анализируется структура такого взаимодействия. В отличие от наукометрического подхода к рассмотрению термина “умный город”, использующего библиометрические методы для анализа сетей цитирования (Ingwersen & Serrano-López, 2018; Ricciardi & Za, 2015; Su et al., 2015), данное исследование сосредоточено на анализе компаний - участников научных мероприятий, проходивших в России с 2007 по 2022 год. Выборка была получена смешанным методом, как парсингом, так и ручным сбором данных с платформы *eLIBRARY.RU* - крупнейшей российской полнотекстовой базы данных научных журналов от ведущих академических, университетских, отраслевых и коммерческих издательств.

Мы работали с тремя исследовательскими вопросами:

1. Кого можно назвать ключевыми игроками, способствующими развитию концепции “умного города” в российском научном сообществе?
2. Какие сообщества можно выделить в сети сотрудничества между организациями?
3. Каковы ключевые характеристики сети сотрудничества среди организаций, участвующих в формировании повестки умного города?

8.2 Данные и методы

С платформы *eLIBRARY.RU* использовались документы типа “материалы конференций” с поисковым запросом “умный город”. Нас интересовали все материалы, в которых данная тема упоминалась в названии статьи или в ключевых словах публикации. Общее количество 652 научных мероприятий было собрано вручную, поскольку единого подхода к сбору такого рода данных из первоисточника не существует. Список уникальных участников лег в основу второй группы вершин сетевых данных - списка компаний. Всего было получена 2761 уникальная организация. После процедуры чистки данных количество было уменьшено до 2211 уникальных значений.

Итоговый набор данных состоит из мероприятий ($n = 652$) и компаний ($n = 2211$), которые принимали участие в этих событиях. Для формирования сети взаимодействия был создан список ребер (*edgelist*) (первый столбец назван *from*, второй - *to*). Под бимодальными данными понимаются данные, фиксирующие связи между двумя разобщенными группами, поэтому связи между этими группами отсутствуют. Кроме того, для обоих групп вершин были созданы файлы атрибутов. С помощью программы *txt2rajek 3* (Pfeffer et al., 2014) файл со списком ребер размером 6186×2 был преобразован в бимодальную сеть *EC*, состоящую из событий × компаний. Затем бимодальная сеть была разбита на две одномодальные сети. Сеть “Компании” (Companies) была получена умножением исходной сети *EC* на транспонированную *CE*, а сеть “Мероприятия” (Events) - умножением *CE* на *EC*. В итоге для

дальнейшего анализа мы сформировали 3 отдельные сети и присвоили их вершинам соответствующие атрибуты (табл. 1). Множественные петли и петли (loops) были удалены.

| | # Nodes (sum) | # Mode 1 | # Mode 2 | # Edges |
|-----------|---------------|----------|----------|---------|
| EC | 2,863 | 652 | 2,211 | 6,186 |
| Events | 652 | | | 22,573 |
| Companies | 2,211 | | | 52,824 |

Для изучения двудольных сетей традиционно применяются несколько подходов. Наиболее распространенный заключается в преобразовании бимодальных данных в одномодальные, что позволяет применять традиционные методы сетевого анализа (Borgatti, 2009). Такое преобразование достигается путем создания сети, в которой узлы представляют вершины из одной группы, а ребра - связи между вершинами из другой группы (например, компаниями и мероприятиями). Исследование сплоченных групп на одномодальных сетях позволит выявить основные группы участия в сети колабораций. Под сплоченными подгруппами понимаются кластеры или подмножества узлов в сети, которые демонстрируют более высокую степень взаимосвязанности между собой по сравнению с остальной частью сети (Wasserman & Faust, 1994b). Эти подгруппы характеризуются сильными внутренними связями и относительно более слабыми связями с узлами за пределами подгруппы. Рассматривая сплоченные подгруппы в сети взаимодействия компаний, мы остановимся на следующих: клики (cliques) (Wasserman & Faust, 1994b), k-ядра (Pittel et al., 1996), Лувенский алгоритм (Blondel et al., 2008) и Лейденский алгоритм (Traag et al., 2019b).

Несмотря на то что единый подход к анализу бимодальных данных в анализе социальных сетей основан на их преобразовании в одномодальные, в последние десятилетия исследователями была разработана отдельная группа методов, позволяющих анализировать обе группы одновременно (Borgatti, 2009; Borgatti & Everett, 1997). Для непосредственного анализа бимодальных данных будут использоваться “хабы и авторитеты” (hubs and authorities) Клейнберга (Kleinberg, 1999), бимодальные ядра (two-mode cores). Кроме того, для ответа на первый исследовательский вопрос будут использованы три меры центральности (degree centrality, betweenness centrality и eigenvector centrality).

Для ответа на третий исследовательский вопрос мы используем экспоненциальное моделирование случайных графов (Exponential Random Graph Modelling, ERGM) и сетевой автокорреляции для исследования структур зависимости в сетях сотрудничества. ERGM представляют собой подход к статистическому моделированию, который позволяет исследователям изучать закономерности связей или отношений в этих сетях. Применяя ERGM, мы выявим основные механизмы и процессы, управляющие формированием связей, и изучим факторы, определяющие формирование отношений сотрудничества.

8.3 Результаты

8.3.1 Исследовательский вопрос 1

Первоначально использовался бимодальный вариант узлов и авторитетов Клейнберга (Kleinberg, 1999). Используя программу Rajek, мы задали количество вершин из первой и второй группы равным 10. С помощью этой операции мы создаем подсеть. Здесь интересны несколько моментов. Во-первых,

большинство мероприятий являются международными научными конференциями, поскольку они потенциально привлекательны для политиков, исследователей и профессионалов отрасли. Во-вторых, все организации являются ведущими государственными университетами. Эти университеты можно рассматривать как ключевые движущие силы обмена знаниями.

Другим распространенным подходом при анализе бимодальной сети напрямую является применение бимодальных ядер (Cerinšek & Batagelj, 2015). Мы выбрали пороговые значения $p = 10$ и $q = 6$. Эти значения образуют разбиение, которое дает возможность выделить подсеть журналов, в которых не менее 10 компаний (в этой подсети) принимали участие в каком-либо мероприятии, и мероприятия, в которых участвуют не менее 10 компаний (в этой подсети). Большинство мероприятий - это международные конференции, которые проводились в Москве. Большинство компаний были крупными (размер > 3). Из этого можно сделать вывод, что как для участия в мероприятиях, так и для их организации необходимы весомые источники, и только крупные компании могут привлекать участников из влиятельных и центральных с точки зрения положения в сети компаний.

Далее к бимодальной сети были применены меры центральности. Под степенной мерой центральности общего количества связей мы будем понимать количество связей, которое имеет вершина. Она позволяет количественно оценить уровень вовлеченности и активности акторов в сети. Более высокое значение указывает на большее количество связей или более высокую вовлеченность компании в различные события. Анализируя сеть, можно отметить центральное положение российской компании "Ростелеком" (Comp1721) в сети и ее значительную вовлеченность в многочисленные события (degree centrality = 67). Высокая степень центральности говорит о том, что эта компания установила прочные связи, партнерские отношения и аффилированность с широким кругом мероприятий в области. Другая важная компания - Российская академия народного хозяйства и государственной службы при Президенте РФ (РАНХиГС) (Comp1648) - также может рассматриваться как ключевой игрок, имеющий большее число связей или вовлеченний компаний в различные мероприятия (degree centrality = 65). Эти компании можно рассматривать как локальные центры в промышленности и науке.

Центральность по посредничеству организации НИУ ВШЭ (Comp1383) характеризует ее как компанию с высоким потенциалом контроля над потоком информации между другими компаниями и конференциями (центральность по посредничеству = 572569,3). Показатель определяет степень, в которой компания выступает в качестве связующего звена или посредника, способствующего коммуникации и сотрудничеству в сети (Wasserman & Faust, 1994). Другим важным участником сети с точки зрения центральности по посредничеству является компания "Ростелеком", которая играет важнейшую роль в обеспечении связи между другими компаниями и событиями (центральность по посредничеству = 441545,6). Российская академия народного хозяйства и государственной службы при Президенте РФ (РАНХиГС) (Comp1648), опять же, может рассматриваться как ключевой игрок (центральность по посредничеству = 348313,1), а также Университет ИТМО (Comp990) (центральность по посредничеству = 236246,4).

С точки зрения центральности по собственному вектору наибольшее значение имеет Российская академия народного хозяйства и государственной службы при Президенте РФ (РАНХиГС) (Comp1648), что свидетельствует о ее высокой значимости и влиянии в домене "умный город" (центральность по собственному вектору = 1,0). Аналогичным образом в качестве важных вершин рассматриваются Российский экономический университет имени Г.В. Плеханова (Comp1747) (центральности по

собственному вектору = 0,96) и Финансовый университет при Правительстве РФ (Comp2116) (центральность по собственному вектору = 0,91). Эти компании связаны со значительным числом важных и влиятельных событий, что свидетельствует об их ключевой роли в содействии сотрудничеству, распространению знаний и инноваций в сети.

Таком образом, ключевыми игроками, судя по высоким показателям центральности, являются ОАО “Ростелеком”, НИУ ВШЭ, Президентская академия (РАНХиГС), Университет ИТМО, Российский экономический университет имени Г.В. Плеханова и Финансовый университет при Правительстве РФ. Эти организации можно рассматривать как ключевые движущие силы обмена знаниями, поскольку они занимают выгодное положение в сети коллaborаций.

8.3.2 Исследовательский вопрос 2

Сеть компаний имеет 2211 вершин, 52824 ребра. Сеть является взвешенной, так как представляет собой проекцию базовой бимодальной сети *EC*. Плотность сети равна 0,022, что означает наличие 2,2% всех возможных связей. Сеть имеет 21 (слабую) компоненту, размер самой большой компоненты равен 2170. Транзитивность (или коэффициент кластеризации) равна 0,354. Веса варьируются от 1 до 24. Диаметр (длина кратчайшего пути) равен 7.

Основная стратегия работы с такими проекциями заключается в том, чтобы выбрать пороговое значение веса ребер и выделить подсеть (Borgatti, 2009). Мы решили удалить все связи со значением меньше 7 и удалить все изолированные вершины. Редуцированная сеть содержит 69 вершин и 159 ребер. Плотность редуцированной сети равна 0,068, что означает наличие 6,8% всех возможных связей. По данным рис. 1 можно сказать, что визуализация редуцированной сети имеет некоторые интересные особенности (размер узла зависит от размера компании). Большинство компаний являются государственными, то есть они составляют ядро всей сети. Также примечательно, что сеть имеет разделение: одна часть состоит из университетов, а другой - из коммерческих компаний. Такое разделение позволяет говорить об очевидной дифференциации и разделении типов организаций в сфере “умного города”: университеты занимаются преимущественно исследованиями и академической деятельностью, а бизнес-компании - коммерческой деятельностью и практическим применением.

Визуализация редуцированной сети компаний с атрибутами

Figure 14: Визуализация редуцированной сети компаний с атрибутами

Транзитивность сети равна 0,355. Для оценки значения этого коэффициента был проведен CUG тест (Conditional Uniform Graph) (рис. 2). Учитывая структуру сети (обусловленную количеством ребер и диад), значение 3,355 выше по сравнению со случайными графами той же топологии, что свидетельствует о тенденции к кластеризации компаний. Таким образом, можно утверждать, что наша редуцированная сеть обладает высокой транзитивностью, имея в разы больше треугольников, чем симулированные сети ($n=1000$). Дальнейшее изучение характеристик, описывающих структуру сети, может способствовать более глубокому пониманию закономерностей взаимодействия, информационных потоков и обмена знаниями между различными заинтересованными сторонами в сфере умного города. Поэтому применение более современных методов представляется чрезвычайно актуальным.

Наибольшая клика редуцированной сети состоит из 7 вершин (Кубанский государственный аграрный университет, Донской государственный технический университет, Финансовый университет при

Результаты СUG теста

Figure 15: Результаты СUG теста

Правительстве РФ, Сибирский федеральный университет, Северо-Кавказский федеральный университет, Российский экономический университет имени Г.В. Плеханова, Президентская академия (РАНХиГС)). Несмотря на географическую удаленность друг от друга, наличие сплоченной и тесно взаимосвязанной подгруппы среди этих университетов свидетельствует о тесном взаимодействии и общности интересов. Их коллективное участие в сети свидетельствует о совпадении целей, направленности исследований и научной деятельности. Сплоченность этой группы свидетельствует о наличии устоявшейся сети научного взаимодействия и совместных инициатив, выходящих за пределы географических границ.

Затем мы провели анализ для выявления k-ядер в сети (рис. 3). Примечательно, что самыми крупными k-ядрами оказались 3-ядро(зеленый цвет) и 6-ядро (оранжевый цвет). В 3-ядро входят в основном государственные компании, что свидетельствует об активном участии государственных структур в совместной академической деятельности. Данное наблюдение подчеркивает стремление государства к развитию синергетических отношений и облегчению процесса сотрудничества в экосистеме умного города. Напротив, в 6-ядрах и 4-ядрах группах (желтые цвета) преобладают университеты, что свидетельствует об общей заинтересованности и вовлеченности этих учебных заведений в области. В 5-ядре (голубой цвет) представлены университеты с высшим рейтингом, что свидетельствует об их известности и признанной компетентности в данной области. Присутствие этих уважаемых учебных заведений в пятом ядре означает их лидерство и влияние на формирование дискурса, программы исследований и совместных инициатив в предметной области.

Визуализация редуцированной сети на основе 3-ядер

Figure 16: Визуализация редуцированной сети на основе 3-ядер

Далее мы применили алгоритмы выявления сообществ (Leiden и Louvain) для анализа подгрупп в сети (рис. 4). Применение алгоритма Louvain позволило разделить сеть на четыре различных сообщества. Светло-голубой кластер состоит в основном из университетов, что подчеркивает их коллективное участие и совместное взаимодействие в изучении концепции “умного города”. Зеленый кластер включает в себя как университеты, так и коммерческие компании, что подчеркивает взаимосвязь и потенциальный обмен знаниями между этими двумя структурами. В синем кластере преобладают государственные компании, что свидетельствует об их значительной роли и участии в совместной деятельности в контексте обсуждения умного города. Наконец, в розовом кластере представлены ведущие российские технические университеты, что подчеркивает их признанный опыт и влияние на формирование дискурса. Рассчитанная модульность равна 0,44. Это значение указывает на относительно высокую степень сплоченности выявленных сообществ, что говорит о том, что алгоритмы обнаружения сообществ эффективно отражают значимые и отчетливые модели взаимодействия в сети коллоквиций. Аналогичным образом алгоритм Лувена дал сопоставимые результаты, но с одним существенным отличием. Он объединил зеленый и синий кластеры, что свидетельствует о более высоком уровне взаимосвязи и совместного взаимодействия между университетами и государственными компаниями. Такое объединение кластеров отражает потенциально более сильную синергию и общие цели этих организаций.

Визуализация редуцированной сети на основе алгоритмов кластеризации Louvain (слева) и Leiden (справа)

Figure 17: Визуализация редуцированной сети на основе алгоритмов кластеризации Louvain (слева) и Leiden (справа)

8.3.3 Исследовательский вопрос 3

Для изучения зависимости между связями в сети взаимодействия компаний и размером компании мы провели сетевой автокорреляционный анализ. В ходе анализа были рассчитаны две широко используемые статистики автокорреляции: Индекс I Морана и Индекс С Гири. Расчетное значение I Морана составило 0,54, что указывает на умеренную положительную автокорреляцию между связями в сети и размером компаний. Это говорит о том, что компании со схожими размерами склонны к сотрудничеству друг с другом, образуя кластеры или группы в сети. Положительная автокорреляция говорит о том, что крупные компании чаще сотрудничают с другими крупными компаниями, а мелкие компании - с мелкими. Аналогичным образом был рассчитан индекс С Гири, значение которого составило 0,22, что также указывает на умеренную положительную автокорреляцию. Это еще раз подтверждает идею о том, что компании схожего размера склонны к сотрудничеству в сети.

Полученные результаты свидетельствуют о том, что размер компании играет определенную роль в формировании моделей сотрудничества в сети. Наблюдаемая умеренная положительная автокорреляция означает, что компании сопоставимого размера с большей вероятностью будут создавать партнерства и участвовать в совместной деятельности, что может быть обусловлено наличием общих ресурсов, опыта или взаимных интересов.

Экспоненциальные модели случайных графов (ERGM) - это статистические модели, используемые для анализа сетевых структур и позволяющие исследователям делать выводы о закономерностях сетевых связей. Результат применения ERGM представлен в табл. 2. Самая простая модель (Модель 1) не представляет особого интереса - она не дает нам ничего сверх простой вероятности образования ребра в существующей сети (вероятность = 0,067). С помощью этой компоненты мы ответим на вопрос: имеет ли сеть тенденцию к образованию взаимных связей? Эта мера равна плотности сети. Положительное значение коэффициента означает, что мы имеем больше ребер, чем в случайной модели; отрицательное - меньше, чем в случайной модели.

Таблица 2. Сводная статистика ERGM для сети взаимодействия

Figure 18: Таблица 2. Сводная статистика ERGM для сети взаимодействия

Далее мы добавляем параметры, основанные на атрибутах, генерируя случайные сети, которые по ним соответствуют наблюдаемой сети. Количество связей внутри компаний схожего размера - *nodematch('size')* - и схожих значений рейтинга - *nodematch('size')*, - согласно модели 2, говорит о том, что в данной сети нет статистически значимой вероятности этих параметров. *Nodematch* позволяет определить, склонны ли участники связываться с компаниями, имеющими схожий размер. Другими словами, гомофилия по этим признакам в сети отсутствует.

Далее мы включили *nodefactor('gov')* - количество раз, которое узлы с заданным уровнем категориального узлового атрибута встречаются в наборе ребер. Также мы рассматриваем атрибутивную переменную - абсолютное значение разницы в размере компаний - *absdiff('size')*. Следует помнить, что

коэффициенты, их стандартные ошибки и р-значения находятся в логарифмической шкале, поэтому мы интерпретируем их как коэффициенты logit.

Результаты *nodefactor* свидетельствуют о том, что в этой сети существует статистически значимая вероятность связи с компаниями, которые являются или не являются государственными. В нашем случае государственные компании связаны в 1,261 раза чаще по сравнению со случайной сетью. Аналогично, компании с размером 3 и компании с размером 4 связаны в 0,55 и 1,27 раза больше, чем случайно, соответственно.

Absdiff исследует, образуют ли два узла связь, потому что они оба имеют схожие значения (т.е. меньшую разницу) или потому что они не имеют схожих значений (т.е. большую разницу). Вероятность того, что узлы с одинаковыми оценками будут общаться, в 0,999 раза выше, чем вероятность случайности. Аналогично, вероятность того, что узлы с одинаковым размером будут общаться, в 1,106 раза выше, чем вероятность. Коэффициенты относительно невелики, однако они демонстрируют высокий уровень статистической значимости.

В заключительной части анализа мы оцениваем, насколько модель “подходит к данным”, т.е. насколько хорошо она воспроизводит некоторые глобальные свойства (общая статистика Goodness-of-Fit). Для этого мы выбираем базовую сетевую статистику, отсутствующую в модели, и сравниваем значение этой статистики, наблюдаемое в исходной сети, с распределением значений, полученных в симулированных сетях с помощью нашей модели. Можно с уверенностью сказать, что 3-я модель хорошо согласуется с данными. Выборочная статистика представлена на рисунке 5.

Выборочная статистика итоговой модели

Figure 19: Выборочная статистика итоговой модели

8.4 Заключение

Концепция “умного города” находит все большее применение в качестве ответа на различные городские проблемы, обусловленные такими факторами, как рост численности населения, ускоренное экономическое развитие и экологические вызовы. Умный город направлен на решение таких проблем, как преступность, пробки, недостаточное качество услуг и экономические ограничения, предлагая при этом перспективы инклюзивного процветания и повышения благосостояния.

Ключевыми игроками, судя по высоким показателям центральности, являются ОАО “Ростелеком”, НИУ ВШЭ, Президентская академия (РАНХиГС), Университет ИТМО, Российский экономический университет имени Г.В. Плеханова и Финансовый университет при Правительстве РФ. Эти институты можно рассматривать как ключевые движущие силы обмена знаниями, поскольку они обладают связями с влиятельными и хорошо известными событиями.

Редуцированная сеть обладает высокой транзитивностью. В качестве целостной и тесно взаимосвязанной подгруппы можно рассматривать 7 учебных заведений: Кубанский государственный аграрный университет, Донской государственный технический университет, Финансовый университет при Правительстве РФ, Сибирский федеральный университет, Северо-Кавказский федеральный университет, Российский экономический университет имени Г.В. Плеханова, Президентская академия (РАНХиГС). Результаты выявления сообществ показали, что в сети сотрудничества существует несколько отдельных

сообществ (университеты, университеты и частные предприятия, государственные компании, технические университеты высшего уровня).

В процессе анализа удалось установить, что размер компаний играет важную роль в формировании моделей сотрудничества в сети. Наблюдаемая умеренная положительная автокорреляция означает, что компании сопоставимого размера с большей вероятностью образуют партнерства и участвуют в совместной деятельности, что может быть обусловлено наличием общих ресурсов, опыта или взаимных интересов. Результаты моделирования с применением ERGM показали, что государственные компании связаны между собой в 1,261 раза чаще, чем случайно. Аналогично, средние и крупные компании связаны между собой в 0,55 и 1,27 раза больше, чем случайно, соответственно. Вероятность связи между университетами с одинаковым рейтингом в 0,999 раза выше случайности.

8.5 5.9 SNA для анализа пользовательского контента: взгляд через призму маркетинга

8.5.1 Рынок электронной коммерции: пользователи и отзывы

С каждым годом растёт доступность получения информации населениями разных стран. Так, в России по данным на 2021 год более 85 % населения страны имеют доступ к интернету (*Digital 2021: The Latest Insights into the “State of Digital” - We Are Social UK, 2021*). Одновременно с этим происходят и более глобальные тенденции, например развитие эры Web 4.0, которая является отражением развития интернет-технологий.

Так, всё большее количество процессов и взаимодействий переходят в онлайн среду, а экономика разных стран в свою очередь развивается под влиянием новых цифровых технологий, машинного обучения и искусственного интеллекта, инноваций и др. Электронная коммерция, являясь одним из секторов экономики России, тоже продолжает активно расти, достигнув в 2020 году 2,5 трлн. руб. (Ребязина et al., 2022) Пользователи же в свою очередь также начинают проводить всё большее количество времени в онлайн, совершая там разные действия, включая различные экономические транзакции (напр., онлайн-покупки) и поддерживая коммуникацию, как между пользователями, так и с брендами.

Таким образом, в эру онлайн-торговли пользователи совершают не только покупки онлайн, но и после их завершения оставляют отзывы о предоставленных услугах и купленных товарах. Например, онлайн-отзывы в академической литературе зачастую относят к таким понятиям, как пользовательский контент (UGC или user-generated content) и электронное сарафанное радио (e-WOM). В академической и бизнес литературе по-прежнему нет единого устоявшегося определения, поэтому в рамках данной работы под пользовательским контентом мы будем понимать медиаконтент, созданный или произведенный широкой публикой, а не оплачиваемыми профессионалами, и в основном распространяемый в Интернете (Daugherty et al., 2008). Второй термин, электронное сарафанное радио будем приравнивать к онлайн-отзывам, полученным от онлайн-пользователей (Hennig-Thurau et al., 2004).

Существует большое количество работ, написанных о пользе электронного сарафанного радио и в целом пользовательского контенте, поскольку они зачастую рассматриваются как фактор, влияющий на принятие решений о покупке. Помимо прочего, отзывы и комментарии, оставленные на онлайн-платформах, рассматриваются через призму такого психологического явления, как социальное доказательство. Поскольку процесс принятия решения потребителем о покупке онлайн становится

всё более комплексным, влияние мнение других людей может быть очень сильным при принятии решения о покупке. Потребители поэтому часто опираются на отзывы других пользователей онлайн-ресурсов, рейтинги товара, чтобы оценить качество, функциональность и ценность продукта. На основе проведённого анализа теоретических источников был выявлен перечень основных факторов электронного сарафанного радио, которые могут повлиять на принятие решений в онлайн. К ним относятся:

- Качество отзыва (Chu & Kamal, 2008):
 - Достоверность источника (Watts & Zhang, 2008)
 - Уровень субъективности, читабельность, орфографические ошибки (Ghose & Ipeirotis, 2010)
- Полезность отзыва и воспринимаемая полезность отзыва (S.-M. Kim et al., 2006):
 - Репутация авторов
 - Полнота (насыщенность) отзыва, простота понимания, краткое, сжатое резюме (C. C. Chen & Tseng, 2011)
 - Воспринимаемая информативность отзыва (Park & Lee, 2008)

Перечисленные выше факторы представляют особый интерес для изучения в области управления и маркетинга, поскольку помогут принимать более эффективные решения по управлению продуктом и брендом, а также позволяют лучше понимать потребителя. В настоящий момент было проведено разведывательное эмпирическое исследование по изучению покупателей в электронной коммерции и их отношения к написанию / чтению отзывов. На следующем шаге предпринимается попытка систематизации знания об использовании методов сетевого анализа для изучения отзывов, оставленных в сети Интернет.

8.5.2 Разведывательное эмпирическое исследование

Данное исследование ставило перед собой задачу более детального изучения покупателей молодого возраста (18-25 лет) г. Москвы, как одной из лидирующих социально-демографических групп, совершающих онлайн-покупки, и их поведения относительно написания отзывов о покупке в онлайне. В частности, в данном исследовании была предпринята попытка дать ответы на следующие исследовательские вопросы:

- Есть ли связь между частотой и длиной оставляемого отзыва?
- Как можно классифицировать потребителей в возрасте 18-25 лет, исходя из их особенностей онлайн-поведения?
- Что мотивирует молодых людей оставлять отзывы о продукте в интернете?

Исследование строится на смешанной методологии с использованием количественных данных, собранных с 2017 по 2022гг. среди слушателей онлайн-курса «Маркетинг», которое размещено на платформе «Открытое образование», а также качественных данных, полученных в 2022 году. Эмпирической базой количественного этапа исследования стали 1254 респондента (в возрасте от 18 до 25 лет, проживающих в городе Москва), а также 16 информантов, отобранных по невероятностной выборке с заданными критериями. Важно отметить, что в анализируемой эмпирической базе большая часть выборки приходится на 2018-2021 гг. (88%).

В целом, в проанализированной выборке 48% респондентов совсем не оставляют отзывы о компаниях/брэндах/продуктах в интернете (на форумах, в социальных сетях, на сайтах компаний, специализированных порталах и др.), 17% делают это один раз в год или реже. Однако 56% опрошенных молодых людей говорят, что они принимают решения о покупке на основе обзоров и отзывов. Одновременно с этим растёт доля тех, кто стал чаще оставлять отзывы в интернете (на основе самоопределения респондентов): так, в 2017 году 56% скорее не согласились, не согласились, полностью не согласились с высказыванием «За последний год я стал чаще оставлять отзывы в интернете», в то время как в 2021-2022 годах доля таких респондентов стала меньше 50% (см. рис. 1).

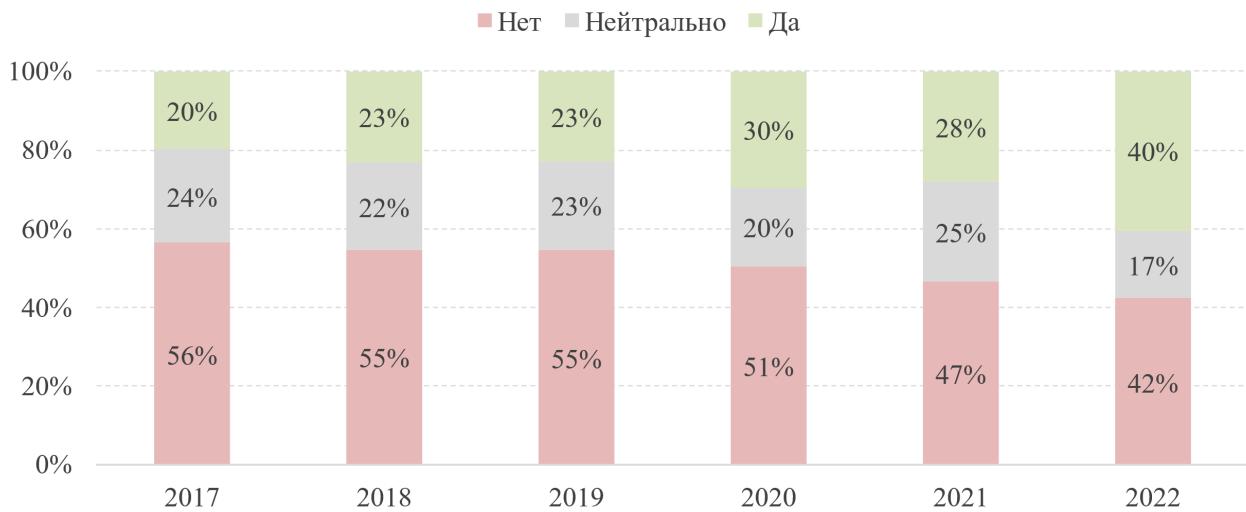


Рис 1. Укрупнённое распределение ответов на вопрос-высказывание «За последний год я стал чаще оставлять отзывы в интернете»

Далее опишем основные поведенческие характеристики респондентов в возрасте 18-25 лет и проживающих в г.Москве, кто оставляет отзывы в интернете ($n = 656$ респондентов). Среди тех, кто оставляет отзывы, 92% готовы написать как положительный, так и отрицательный отзыв. Это свидетельствует о том, что стереотип о написании только негативных отзывов среди молодой аудитории Москвы не распространён. Также большинство из них (66%) склонны доверять информации потребителей о компаниях, продуктах, товарах, оставленной в интернете.

В результате проведённого корреляционного анализа была обнаружена умеренная связь между частотой и длиной написанных отзывов. Так, если пользователи чаще оставляют отзывы, то они их пишут в более развернутом формате. Критерий Хи-квадрат, применённый к анализу таблиц сопряжённости, показал, что оценочность содержания отзыва (положительный или отрицательный характер отзыва) не связан с его длиной. Таким образом, негативный отзыв о товаре и, возможно, негативный опыт совершения покупки необязательно связан с тем, что пользователь оставит длинный отзыв. Таким образом, важно понимать, что мотивирует и демотивирует пользователей оставлять отзывы.

Перед тем, как классифицировать молодых людей на несколько групп на основе их покупательского поведения, доверию интернет-торговле и поведения, связанного с написанием отзывов, на первом этапе проводился факторный анализ, чтобы уменьшить количество анализируемых высказываний и отобрать латентные переменные. В результате было выявлено 5 факторов: (1) отношение к онлайн покупкам в целом (2) лояльность к интернет-магазинам (3) безопасность покупок в интернет-магазинах (4) частота и

длина отзывов (5) внимание к обзорам в онлайн-магазинах (см. табл. 1).

Табл. 1. Факторные нагрузки (5 компонент)

| Высказывание | 1 | 2 | 3 | 4 | 5 |
|--|------|---|---|---|---|
| Следующую покупку я совершу в интернет-магазине, где уже покупал раньше | .816 | | | | |
| Я совершу покупку в интернет-магазине, в котором уже покупал раньше, в течение следующего года | .809 | | | | |
| Я намерен продолжать совершать покупки в интернет-магазине, в котором уже покупал раньше, и не собираюсь от этого отказываться | .795 | | | | |
| Если бы мне пришлось снова купить тот же товар, я бы купил его в том же интернет-магазине | .784 | | | | |
| Я рекомендую своим друзьям / коллегам / родственникам интернет-магазины, в которых совершаю покупки | .750 | | | | |
| Я рассказываю своим друзьям про интернет-магазины, где совершаю покупки | .744 | | | | |
| Я бы хотел, чтобы мои друзья / коллеги / родственники покупали в тех же интернет-магазинах, где покупаю я | .628 | | | | |
| Мне легко взаимодействовать с интернет-магазином во время совершения покупки | .781 | | | | |
| В целом, совершая покупки в интернете легко | .779 | | | | |
| Мне легко совершать любые операции во время покупки в интернет-магазине | .764 | | | | |
| В целом, в интернете удобно покупать | .742 | | | | |
| Мне нравится совершать покупки в интернете | .733 | | | | |
| Процесс покупки в интернете имеет много преимуществ | .724 | | | | |
| Мне нравится искать товары в интернете | .657 | | | | |
| Я беспокоюсь, что мои финансовые данные могут быть переданы другим компаниям без моего согласия | .904 | | | | |
| Я беспокоюсь о сохранности персональных данных в интернете | .864 | | | | |
| Я беспокоюсь о безопасности финансовых операций в интернете | .854 | | | | |
| Я чувствую себя некомфортно, оставляя в интернете номер банковской карты | .849 | | | | |
| Сообщения, отправленные через интернет, могут быть прочтены посторонними людьми или компаниями без моего ведома | .677 | | | | |
| Я пишу длинные отзывы (более 1-2 предложений) | .896 | | | | |
| За последний год я стал чаще оставлять отзывы в интернете | .889 | | | | |
| Чем больше отзывов, тем выше вероятность, что я выберу товар и/или магазин для покупки | .866 | | | | |

Затем на основе результатов факторного анализа был проведён кластерный анализ. Было выявлено, что не все молодые люди г. Москвы в возрасте 18-25 лет имеют схожие паттерны поведения и отношения к онлайн-покупкам. В результате проведённого кластерного анализа были выявлены 5 кластеров. Так, например, кластер «Вдумчивые» склонны оставлять более длинные отзывы. Такие молодые люди в целом позитивно относятся к совершению покупок онлайн и доверяют отзывам, которые написаны в интернете. Одновременно с этим, среди кластера «Скептики» больше тех, кто не доверяет онлайн-транзакциям, в связи с чем реже совершают покупки в онлайне, реже оставляет отзывы и в целом более негативно относится к совершению онлайн-покупок.

Чтобы ответить на вопрос о том, как мотивировать молодёжь оставлять отзывы, были проведены глубинные интервью. В результате были выявлены такие драйверы, как:

1. недовольство товаром (в явной критической степени), в связи с чем покупатели хотят поделиться негативным опытом с другими потенциальными покупателями и предостеречь их;
2. полная удовлетворённость товаром (превосходящая ожидания);
3. получение материальных бонусов и стимулирующих действий от продавца. Среди основных барьеров были выявлены: отсутствие анонимности отзыва, отсутствие внутренней мотивации к оствалению отзыва, траты времени, а также отсутствие вознаграждение.

В результате проведённого эмпирического исследования, было изучено онлайн-поведение молодёжи г.Москвы в возрасте 18-25 лет. Были определены основные поведенческие паттерны по отношению к онлайн-покупкам и оствалению отзывов в онлайне. В рамках продолжения исследования планируется использовать такие методы как SNA и SEM для моделирования и определения влияния роли отзывов о совершении онлайн-покупок в интернете, а также более детального изучения текстов самих отзывов.

8.5.3 2. Применение методологии сетевого анализа для изучения пользовательского контента

Сильный рост развития платформ социальных сетей привел к всплеску пользовательского контента (UGC). Как исследователи, так и практики обращаются к анализу социальных сетей (SNA) как к надежному методу расшифровки сложных закономерностей и взаимосвязей в этих больших наборах данных. Анализ литературы показал, что есть пул исследований, которые демонстрируют применимость и эффективность SNA при анализе пользовательского контента.

Анализ совокупности текстов пользователей в Интернете с помощью (SNA) начинается с преобразованием текстовых данных в сетевое представление, где узлы (nodes) представляют пользователей или фрагменты текста, а связи (links) представляют отношения или взаимодействия (Z. Yang et al., 2019). Затем проводится анализ текстов или отношений, которые позволяют решить ряд управлеченческих задач в области бизнеса (в частности, в маркетинге). Именно о применении методологии сетевого анализа для таких задач пойдёт речь в данной главе.

8.5.4 2.1 Расширенное знание о пользовательском контенте вокруг продукта и бренда

Во-первых, анализ пользовательского контента с помощью SNA позволяет определять наиболее ключевых влиятельных лиц. Это может быть сеть вокруг бренда или продукта, которая включает как покупателей, так и в том числе бренды-конкуренты. Обнаруживание наиболее значимых лиц в сети позволит наладить сотрудничество, а взаимодействие с этими акторами может расширить охваты и повлиять на результаты маркетинговых кампаний.

Помимо обнаружения отдельных акторов, SNA помогает идентифицировать сообщества, что в свою очередь улучшает настройку таргетинга. Так, после выявления сообществ и групп, имеющих общие интересы или поведение, маркетологи могут адаптировать маркетинговые стратегии для конкретных сообществ, гарантируя, что контент и сообщения будут откликаться у отдельных членов этих групп.

8.5.5 2.2 Позиция на рынке и отношения с конкурентами

Когда компания только запускает новый продукт или услугу, особенно важно отслеживать информацию, которой делятся пользователи, поскольку это может повлиять на дальнейшую судьбу всего бренда (K. Kim et al., 2015). Чёткое понимание обсуждаемых тем, вычленение достоинств и недостатков продукта, позволяет маркетологам оптимизировать маркетинговые кампании через понимание влияния пользователей друг на друга.

Немаловажно оценивать позицию бренда и продукта относительно конкурентов. Если в одну сеть добавить в том числе названия брендов, то можно проанализировать связи и взаимодействие между пользователями, которые обсуждают бренд в сравнении с конкурентами (He et al., 2023). С помощью проведения такого анализа можно получить карту позиционирования бренда относительно конкурентов, увидеть конкурентные преимущества и скорректировать маркетинговые кампании.

8.5.6 Маркетинг и реклама

В маркетинге есть целый ряд метрик, которые чаще всего измеряются классическими онлайн-опросами или проведением интервью. Однако изучение текстов отзывов может воссоздать более объективно многие метрики у маркетологов может появиться более чёткое понимание и видение своего продукта. Например, можно измерять вовлечённость клиентов через анализ моделей взаимодействия пользователей с брендом или пользователей друг с другом через упоминание бренда (Isler et al., 2023). Эта позволяет оценить эффективность маркетинговых усилий для дальнейшего улучшения и усовершенствования стратегий для повышения вовлечённости клиентов.

Другой немаловажный аспект – это сарафанное радио (word-of-mouth marketing). Пользователи любят рекомендовать понравившийся товар (или проводить антирекламу непонравившемуся товару). Изучая потоки и обмен информации между пользователями, можно идентифицировать пользователей, наиболее часто разделяющие положительные оценки о бренде и поощрить дальнейшее развитие сарафанного радио с помощью маркетинговой поддержки (Hambrick & Pegoraro, 2014).

Изучение потребительского контента позволяет определить пользователей, которые с наибольшей вероятностью будут защищать бренд и в целом лояльны к нему (Wassouf et al., 2020). Разрабатывая и внедряя программы по промотированию бренда через таких авторов постов, позволить продвигать бренд в социальных сетях более эффективно и с меньшим вложением затрат.

8.5.7 2.4 Клиентский маркетинг

На основе отзывов в онлайне можно воссоздать как действующую базу клиентов, так и потенциальную. С помощью SNA можно проводить сегментацию клиентов наиболее естественным образом (Slaninová et al., 2010). Это позволяет адаптировать маркетинговые активности (через рекламные сообщения, акции и т.д.) к предпочтениям конкретных сегментов.

Помимо прочего, SNA может стать рабочей методологией в разработке стратегий удержания клиентов или улучшения сервиса и обслуживания. Так, с помощью текстового анализа отзывов можно понять негативные настроения и отрицательно окрашенные темы, что в дальнейшем может позволить улучшить обслуживание, а также решить проблемы, быстро реагируя на отзывы взаимодействуя с пользователями. Отдельно можно выделить и обнаружение «центральных» акторов сети, через которых можно не только удерживать клиентов, но и влиять на базу и в целом охватывать потенциальных клиентов.

8.6 3. Процесс применения SNA для анализа пользовательских отзывов

SNA может быть ценным инструментом для анализа отзывов потребителей, рассматривая отношения между пользователями, оставляющими отзывы (в дальнейшем будем называть их рецензентами), или связи между продуктами и рецензентами как единую сеть (Aiello et al., 2012). Ниже рассмотрим пошаговую методологию по применению SNA для анализа пользовательских отзывов.

8.6.1 3.1 Подготовка массива данных

На первом этапе необходимо собрать данные (data scrapping) пользовательских отзывов с интернет-платформ (это могут быть публикации в социальных сетях, отзывы на сайтах маркетплейсов, обсуждение товаров на форумах и т.д.). Для использования SNA будет необходимо собрать корпус текстовых данных на основе отзывов.

Далее необходима предварительная обработка этих данных. Будет произведена очистка данных (удаление ненужной информации, работа с пропущенными данными и т.д.), а также произведена стандартизация текстов в единый формат. Для того, чтобы язык поддавался изучению, будет необходимо токенизировать текст с помощью стемминга и лемматизации.

Последним шагом будет необходимо построить сеть на основе имеющихся данных, где узлы представляют собой пользователей (рецензентов) или фрагменты текста, а связями являются отношения или взаимодействия между ними. Связи могут также быть построены на ответах, упоминаниях или взаимных упоминаниях брендов / продуктов или рецензентов.

8.7 3.2 Анализ основных метрик SNA

Использование атрибутов позволяет более комплексно подойти к анализу текстов. Так, добавлений атрибутов узлам (node attributes) и связям (edge attributes) помогут провести сравнительный анализ текстов отзывов по частоте оставления публикаций определённым автором или более детально изучить динамические тренды при добавлении дат создания. Также можно добавить характеристики текста для сети (например, количество слов в отзыве), чтобы проверить гипотезу об эффективности оставления отзыва в зависимости от длины сообщения.

Метрики центральности (centrality measures) позволяют идентифицировать наиболее влиятельных пользователей (Romero et al., 2011) обнаружение инфлюенсеров, влияющих на продажи того или иного продукта), а также выявление наиболее значимых фрагментов текстов отзыва. Помимо обнаружения индивидуальных центральностей важно изучение сообществ (community detection). Так, с помощью применения алгоритмов возможно выявить наиболее и наименее влиятельные группы сообществ (Paranjape et al., 2017), а также выявить тематические группы текстов.

8.7.1 3.3 Методы анализа данных: более детальный взгляд

В качестве первого метода планируется использовать content analysis, который позволит с помощью методов изучения и обработки естественного языку изучить содержание текстов. В частности, можно использовать выделение основных тем (topic modelling) и определение ключевых слов для определения преобладающих тем в корпусе.

Далее можно использовать анализ настроений (sentiment analysis), чтобы понять тональность сообщения, оставленного пользователем в сети. В частности, используются техники машинного обучения для классификации текстов на положительные, нейтральные и отрицательные фрагменты текста и сообщения в целом.

Для изучения трендов во времени используется временной анализ (temporal analysis), который позволяет изучить в динамике сеть взаимодействий во времени, определить различия в обнаруженных темах, а также сравнить настроения пользователей в разный период времени.

8.7.2 3.4 Визуализация и выводы

Самой наглядной демонстрацией выводов при изучении сети является визуализация, которую можно нарисовать с помощью программного обеспечения R, Rajek или Gephi. Визуализация (в том числе интерактивная) позволит наглядно обозначить выводы, полученные в результате применения техник и методов, обозначенных в подпунктах 2 и 3. Наконец, интерпретация и анализ технических методологических показателей, рассмотренных в данной главе, позволит маркетологам и управленцам в разработке контент-стратегий, управлении сообществами, а также контролировать и влиять на настроения пользователей относительно продуктов и услуг.

- Agrawal, P., Garg, V. K., & Narayananam, R. (2013). *Link Label Prediction in Signed Social Networks*. 2591–2597.
<https://doi.org/10.25781/KAUST-TZW78>
- Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, 6(2), 1–33.
- Alba, R. D. (1973). A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology*, 3(1), 113–126.
- Amati, V., Schönenberger, F., & Snijders, T. A. B. (n.d.). *Estimation of Stochastic actor-oriented models for the evolution of networks by generalized method of moments*.
- Anastasi, A. (1986). Evolving Concepts of Test Validation. *Annual Review of Psychology*, 37(1), 1–16. <https://doi.org/10.1146/annurev.ps.37.020186.000245>
- Aoki, M. (2007). Endogenizing institutions and institutional changes*. *Journal of Institutional Economics*, 3(1), 1–31. <https://doi.org/10.1017/S1744137406000531>

- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
- Aronson, B. (2016). Peer influence as a potential magnifier of ADHD diagnosis. *Social Science & Medicine*, 168, 111–119. <https://doi.org/10.1016/j.socscimed.2016.09.010>
- Arrigo, F., & Higham, D. J. (2017). Sparse matrix computations for dynamic network centrality. *Applied Network Science*, 2(1), 17. <https://doi.org/10.1007/s41109-017-0038-z>
- Ashforth, B. E. (1997). Petty Tyranny in Organizations: A Preliminary Examination of Antecedents and Consequences. *Canadian Journal of Administrative Sciences / Revue Canadienne Des Sciences de l'Administration*, 14(2), 126–140. <https://doi.org/10.1111/j.1936-4490.1997.tb00124.x>
- Association, A. E. R., Association, A. P., & Education, N. C. on M. in. (1985). *Standards for educational and psychological testing*. <https://cir.nii.ac.jp/crid/1130000794987222016>
- Barabási, A.-L. (2013). Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987), 20120375. <https://doi.org/10.1098/rsta.2012.0375>
- Batagelj, V., Ferligoj, A., & Doreian, P. (1999). *Generalized Blockmodeling*.
- Batagelj, V., & Mrvar, A. (n.d.). *Pajek – Program for Large Network Analysis*.
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186–3191. https://journals.lww.com/spinejournal/Fulltext/2000/12150/Guidelines_for_the_Process_of_Cross_Cultural.14.aspx
- Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 71(2), 287–311. <https://academic.oup.com/poq/article-abstract/71/2/287/1928986>
- Berardo, R., & Scholz, J. T. (2010). Self-Organizing Policy Networks: Risk, Partner Selection, and Cooperation in Estuaries: SELF-ORGANIZING POLICY NETWORKS. *American Journal of Political Science*, 54(3), 632–649. <https://doi.org/10.1111/j.1540-5907.2010.00451.x>
- Bertrand, M., & Schoar, A. (2006). The Role of Family in Family Firms. *Journal of Economic Perspectives*, 20(2), 73–96. <https://doi.org/10.1257/jep.20.2.73>
- Besag, J. (1986). On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(3), 259–279. <https://doi.org/10.1111/j.2517-6161.1986.tb01412.x>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). *Fast unfolding of communities in large networks*. <https://doi.org/10.48550/ARXIV.0803.0476>
- Borgatti, S. P. (2009). Social Network Analysis, Two-Mode Concepts in. In R. A. Meyers (Ed.), *Encyclopedia of Complexity and Systems Science* (pp. 8279–8291). Springer New York. https://doi.org/10.1007/978-0-387-30440-3_491
- Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, 19(3), 243–269. [https://doi.org/10.1016/S0378-8733\(96\)00301-2](https://doi.org/10.1016/S0378-8733(96)00301-2)
- Børsting, C., & Thomsen, S. (2017). Foundation ownership, reputation, and labour. *Oxford Review of Economic Policy*, 33(2), 317–338. <https://doi.org/10.1093/oxrep/grx023>
- Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling: A guide for social scientists*. Cambridge University Press.
- Boyatzis, R. E. (1991). *The competent manager: A model for effective performance*. John Wiley & Sons. [https://www.google.com/books?hl=ru&lr=&id=KmFR7BnLdCoC&oi=fnd&pg=PR11&dq=Boyatzis,+R.+E.+\(1982\).+The+competent+manager:+A+model+for+effective+performance.+John+Wiley+%26+S](https://www.google.com/books?hl=ru&lr=&id=KmFR7BnLdCoC&oi=fnd&pg=PR11&dq=Boyatzis,+R.+E.+(1982).+The+competent+manager:+A+model+for+effective+performance.+John+Wiley+%26+S)

- ons.+21+p.&ots=wx2IRuNqTC&sig=WZq1yWhxkW6Zln175Ra1DO6pnUI
- Burkart, M., Miglietta, S., & Ostergaard, C. (2021). *Why Do Boards Exist? Governance Design in the Absence of Corporate Law* (SSRN Scholarly Paper No. 2902617). <https://doi.org/10.2139/ssrn.2902617>
- Ceoldo, G., Snijders, T. A., & Wit, E. C. (2023). Stochastic Actor Oriented Model with Random Effects. *arXiv Preprint arXiv:2304.07312*.
- Cerinšek, M., & Batagelj, V. (2015). Network analysis of Zentralblatt MATH data. *Scientometrics*, 102(1), 977–1001. <https://doi.org/10.1007/s11192-014-1419-z>
- Chen, C. C., & Tseng, Y.-D. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, H., Song, X., Jin, Q., & Wang, X. (2022). Network dynamics in university-industry collaboration: a collaboration-knowledge dual-layer network perspective. *Scientometrics*, 127(11), 6637–6660. <https://doi.org/10.1007/s11192-022-04330-9>
- Chen, X., He, Z., Chen, Y., Lu, Y., & Wang, J. (2019). Missing Traffic Data Imputation and Pattern Discovery with a Bayesian Augmented Tensor Factorization Model. *Transportation Research Part C: Emerging Technologies*, 104, 66–77. <https://doi.org/10.1016/J.TRC.2019.03.003>
- Cheng, C., Yang, H., King, I., & Lyu, M. (2012). *Fused Matrix Factorization with Geographical and Social Influence in Location-based Social Networks*. 26(1), 17–23. <https://ojs.aaai.org/index.php/AAAI/article/view/8100>
- Chiuso, A., & Pillonetto, G. (2012). A Bayesian approach to sparse dynamic network identification. *Automatica*, 48(8), 1553–1565. <https://doi.org/10.1016/j.automatica.2012.05.054>
- Cho, H., & Yu, Y. (2018). *Link prediction for Interdisciplinary Collaboration via Co-authorship Network*. 8(25). <https://doi.org/10.1007/s13278-018-0501-6>
- Chow, J. H., & Kokotovic, P. V. (1987). Time scale modeling of dynamic networks with sparse and weak connections. In P. V. Kokotovic, A. Bensoussan, & G. L. Blankenship (Eds.), *Singular Perturbations and Asymptotic Analysis in Control Systems* (Vol. 90, pp. 310–353). Springer-Verlag. <https://doi.org/10.1007/BFb0007181>
- Chu, S.-C., & Kamal, S. (2008). The effect of perceived blogger credibility and argument quality on message elaboration and brand attitudes: An exploratory study. *Journal of Interactive Advertising*, 8(2), 26–37.
- Chuan, P. M., Son, L. H., Ali, M., Khang, T. D., Huong, L. T., & Dey, N. (2018). Link prediction in co-authorship Networks Based on Hybrid Content Similarity Metric. *Applied Intelligence*, 48(8), 2470–2486. <https://doi.org/10.1007/s10489-017-1086-x>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Cocchia, A. (2014). Smart and Digital City: A Systematic Literature Review. In R. P. Dameri & C. Rosenthal-Sabroux (Eds.), *Smart City* (pp. 13–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-06160-3_2
- Commission, I. T. (2017). *The ITC guidelines for translating and adapting tests*. Second edition.
- Cooley, C. H. (1909). *Social organization: A study of the larger mind*. Charles Scribner's Sons.
- Coskun, M., & Koyutürk, M. (2015). *Link Prediction in Large Networks by Comparing the Global View of Nodes in the Network*. 485–492. <https://doi.org/10.1109/ICDMW.2015.195>

- Costas, J., & Taheri, A. (2012). "The Return of the Primal Father" in Postmodernity? A Lacanian Analysis of Authentic Leadership. *Organization Studies*, 33(9), 1195–1216. <https://doi.org/10.1177/0170840612448157>
- Cote, R. (2018). Dark Side Leaders: Are their Intentions Benign or Toxic? *Journal of Leadership, Accountability and Ethics*, 15(2). <https://doi.org/10.33423/jlae.v15i2.643>
- Dao, V. L., Bothorel, C., & Lenca, P. (2020). Community structure: A comparative evaluation of community detection methods. *Network Science*, 8(1), 1–41. <https://doi.org/10.1017/nws.2019.59>
- Dashtipour, P., & Vidaillet, B. (2016). *Christophe Dejours' psychodynamic theory of work and its implications for leadership*. <https://www.semanticscholar.org/paper/Christophe-Dejours%20%80%99-psychodynamic-theory-of-work-Dashtipour-Vidaillet/a2ead5244af431cefaad966d033d06e853cf40bb>
- Daud, N. N., Ab Hamid, S. H., Saadoon, M., Sahran, F., & Anuar, N. B. (2020). *Applications of Link Prediction in Social Networks: A Review*. 166, 102716. <https://doi.org/10.1016/j.jnca.2020.102716>
- Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *Journal of Interactive Advertising*, 8(2), 16–25.
- Davis, A., Gardner, B. B., & Gardner, M. R. (2009). *Deep South: A social anthropological study of caste and class*. Univ of South Carolina Press.
- De La Haye, K., Robins, G., Mohr, P., & Wilson, C. (2011). How physical activity shapes, and is shaped by, adolescent friendships. *Social Science & Medicine*, 73(5), 719–728. <https://doi.org/10.1016/j.socscimed.2011.06.023>
- De La Haye, K., Shin, H., Vega Yon, G. G., & Valente, T. W. (2019). Smoking Diffusion through Networks of Diverse, Urban American Adolescents over the High School Period. *Journal of Health and Social Behavior*, 60(3), 362–376. <https://doi.org/10.1177/0022146519870521>
- Digital 2021: the latest insights into the “state of digital” - We Are Social UK.* (2021, January 27). We Are Social. <https://wearesocial.com/uk/blog/2021/01/digital-2021-the-latest-insights-into-the-state-of-digital/>
- Dijkstra, J. K., Gest, S. D., Lindenberg, S., Veenstra, R., & Cillessen, A. H. N. (2012). Testing Three Explanations of the Emergence of Weapon Carrying in Peer Context: The Roles of Aggression, Victimization, and the Social Network. *Journal of Adolescent Health*, 50(4), 371–376. <https://doi.org/10.1016/j.jadohealth.2011.08.010>
- Driver, M. (2013). The Lack of Power or the Power of Lack in Leadership as a Discursively Constructed Identity. *Organization Studies*, 34(3), 407–422. <https://doi.org/10.1177/0170840612463003>
- Dzansi, J. (2011). Foundations and Investment Performance: The role of non-financial motives. *Unpublished Working Paper*. Jönköping International Business School.
- Einarsen, S., Aasland, M. S., & Skogstad, A. (2007). Destructive leadership behaviour: A definition and conceptual model. *The Leadership Quarterly*, 18(3), 207–216. <https://doi.org/10.1016/j.lequa.2007.03.002>
- eLibrary.ru. (n.d.). Сравнение уровня публикаций российских ученых в базах данных Web of Science, Scopus и RSCI. eLibrary.ru. Retrieved August 30, 2023, from https://elibrary.ru/wos_scopus_rsci.asp?
- Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, 68(4), 435–441. https://www.sciencedirect.com/science/article/pii/S0895435614004995?casa_token=ZiUN0lD32NgAAAAAAtARNoTgR9QBjZx56mXGHsVV-msOW_ZBbExgqwH1tDj9b2DLmAVZpOxMW5xl2yLJiNaWgzC6-7UE
- Esquivel, A. V., & Rosvall, M. (2011). Compression of flow can reveal overlapping-module organization in networks. *Physical Review X*, 1(2), 021025.

- Farasat, A., Nikolaev, A., Srihari, S. N., & Blair, R. H. (2015). *Probabilistic graphical models in modern social network analysis*. 5(1), 1–18. <https://doi.org/10.1007/s13278-015-0289-6>
- Ferligoj, A., Kronegger, L., Mali, F., Snijders, T. A. B., & Doreian, P. (2015). Scientific collaboration dynamics in a national scientific system. *Scientometrics*, 104(3), 985–1012. <https://doi.org/10.1007/s11192-015-1585-7>
- Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1), 36–41.
- Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>
- Freeman, L. C. (1992). The Sociological Concept of “Group”: An Empirical Test of Two Models. *American Journal of Sociology*, 98(1), 152–166. <https://doi.org/10.1086/229972>
- Friedkin, N. E. (1984). Structural cohesion and equivalence explanations of social homogeneity. *Sociological Methods & Research*, 12(3), 235–261.
- Gabriel, Y. (2011). Psychoanalytic approaches to leadership. In *The SAGE Handbook of Leadership* (pp. 393–405). SAGE.
- Gerber, E. R., & Wellens, T. R. (1997). Perspectives on Pretesting : “Cognition” in the Cognitive Interview? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 55(1), 18–39. <https://doi.org/10.1177/075910639705500104>
- Getoor, L., Friedman, N., Koller, D., & Taskar, B. (2002). *Learning Probabilistic Models of Link Structure*. 3, 679–707. <https://doi.org/10.1162/jmlr.2003.3.4-5.679>
- Ghoorjian, K. (2019). *Graph Algorithms for Large-Scale and Dynamic Natural Language Processing*. <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-263914>
- Ghose, A., & Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512.
- Gignoux, J., Chérel, G., Davies, I. D., Flint, S. R., & Lateltin, E. (2017). Emergence and complex systems: The contribution of dynamic graph theory. *Ecological Complexity*, 31, 34–49. <https://doi.org/10.1016/j.ecocom.2017.02.006>
- Glynn, C., Tokdar, S. T., Banks, D. L., & Howard, B. (2015). *Bayesian Analysis of Dynamic Linear Topic Models*. <https://doi.org/10.48550/ARXIV.1511.03947>
- Gou, F., & Wu, J. (2021). *Triad Link Prediction Method Based on the Evolutionary Analysis with IoT in Opportunistic Social Networks*. 181, 143–155. <https://doi.org/10.1016/j.comcom.2021.10.009>
- Hambrick, M. E., & Pegoraro, A. (2014). Social Sochi: Using social network analysis to investigate electronic word-of-mouth transmitted through social media communities. *International Journal of Sport Management and Marketing*, 15(3-4), 120–140.
- Handcock, M. S. (n.d.). *Assessing Degeneracy in Statistical Models of Social Networks*.
- Hansmann, H., & Thomsen, S. (2021). The governance of foundation-owned firms. *Journal of Legal Analysis*, 13(1), 172–230.
- Hasan, M. A., & Zaki, M. J. (2011). A Survey of Link Prediction in Social Networks. In C. C. Aggarwal (Ed.), *Social Network Data Analytics* (pp. 243–275). Springer US. https://doi.org/10.1007/978-1-4419-8462-3_9
- Haupt, M. R., Jinich-Diamant, A., Li, J., Nali, M., & Mackey, T. K. (2021). Characterizing Twitter User Topics

- and Communication Network Dynamics of the “Liberate” Movement During COVID-19 Using Unsupervised Machine Learning and Social Network Analysis. *Online Social Networks and Media*, 21, 100114. <https://doi.org/10.1016/j.osnem.2020.100114>
- He, Z., Zheng, L., & He, S. (2023). A novel approach for product competitive analysis based on online reviews. *Electronic Commerce Research*, 23(4), 2259–2290.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Higgs, M. (2009). The Good, the Bad and the Ugly: Leadership and Narcissism. *Journal of Change Management*, 9(2), 165–178. <https://doi.org/10.1080/14697010902879111>
- Hilt, E. (2008). When did ownership separate from control? Corporate governance in the early nineteenth century. *The Journal of Economic History*, 68(3), 645–685.
- Hoff, P. D. (2009). Multiplicative Latent Factor Models for Description and Prediction of Social Networks. *Computational and Mathematical Organization Theory*, 15(4), 261–272. <https://doi.org/10.1007/s10588-008-9040-4>
- Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460), 1090–1098. <https://doi.org/10.1198/016214502388618906>
- Hoffmann, T., Peel, L., Lambiotte, R., & Jones, N. S. (2020). Community detection in networks without observing edges. *Science Advances*, 6(4), eaav1478. <https://doi.org/10.1126/sciadv.aav1478>
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, 88(1), 100–112. <https://doi.org/10.1037/0021-9010.88.1.100>
- Hogan, R., & Hogan, J. (2007). *Hogan Personality Inventory Manual* (3rd ed.). Hogan Assessment Systems.
- Hogan, R., & Hogan, J. (2009). *Hogan Development Survey Manual* (Second Edition). Hogan Assessment Systems.
- Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109–137.
- Hollands, R. G. (2015). Critical interventions into the corporate smart city. *Cambridge Journal of Regions, Economy and Society*, 8(1), 61–77. <https://doi.org/10.1093/cjres/rsu011>
- Homans, G. C. (2017). *The human group*. Routledge.
- Hopkins, S. B., & Steurer, D. (2017). Efficient Bayesian Estimation from Few Samples: Community Detection and Related Problems. *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, 379–390. <https://doi.org/10.1109/FOCS.2017.42>
- Huang, G. C., Soto, D., Fujimoto, K., & Valente, T. W. (2014). The Interplay of Friendship Networks and Social Networking Sites: Longitudinal Analysis of Selection and Influence Effects on Adolescent Smoking and Alcohol Use. *American Journal of Public Health*, 104(8), e51–e59. <https://doi.org/10.2105/AJPH.2014.302038>
- Huang, Z., Li, X., & Chen, H. (2005). *Link Prediction Approach to Collaborative Filtering*. 141–142. <https://doi.org/10.1145/1065385.1065415>
- Ingold, K., & Fischer, M. (2014). Drivers of collaboration to mitigate climate change: An illustration of Swiss

- climate policy over 15 years. *Global Environmental Change*, 24, 88–98. <https://doi.org/10.1016/j.gloenvch.2013.11.021>
- Ingwersen, P., & Serrano-López, A. E. (2018). Smart city research 1990–2016. *Scientometrics*, 117(2), 1205–1236. <https://doi.org/10.1007/s11192-018-2901-9>
- International Conference on Questionnaire Design, Development, Evaluation and Testing (QDET2)*. (n.d.). Retrieved November 27, 2023, from <https://ww2.amstat.org/meetings/qdet2/>
- Isler, Z., Kiygi-Calli, M., & El Oraiby, M. (2023). Babbling through social media: A cross-country study mapping out social networks using eWOM intentions. *Electronic Markets*, 33(1), 54.
- Javed, M. A., Younis, M. S., Latif, S., Qadir, J., & Baig, A. (2018). Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108, 87–111. <https://doi.org/10.1016/j.jnca.2018.02.011>
- Jobe, J. B., Tourangeau, R., & Smith, A. F. (1993). Contributions of survey research to the understanding of memory. *Applied Cognitive Psychology*, 7(7), 567–584. <https://doi.org/10.1002/acp.2350070703>
- Jugert, P., Leszczensky, L., & Pink, S. (2018). The Effects of Ethnic Minority Adolescents' Ethnic Self-Identification on Friendship Selection. *Journal of Research on Adolescence*, 28(2), 379–395. <https://doi.org/10.1111/jora.12337>
- Kannan, R., Vempala, S., & Vetta, A. (2004). On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3), 497–515.
- Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1), 016107.
- Kaya, & Thomsen, S. (2022). *The Governance of Foundation-Owned Firms*. <https://www.enterprisefoundations.dk/wp-content/uploads/2022/06/The-Governance-of-Foundation-owned-Firms.pdf>
- Kets de Vries, M. F. R. (1989). *Prisoners of leadership*. Wiley.
- Kets de Vries, M. F. R. (1991). *Organizations on the couch: clinical perspectives on organizational behavior and change* (1. ed). Jossey-Bass.
- Kets de Vries, M. F. R., & Miller, D. (1984). *The neurotic organization* (1st ed). Jossey-Bass.
- Kim, A., & Maltseva, D. (2021). Stability evaluation of the Russian sociologists online community: 2011–2018 years. <https://doi.org/10.48550/ARXIV.2110.08756>
- Kim, K., Lee, W.-R., & Altmann, J. (2015). SNA-based innovation trend analysis in software service networks. *Electronic Markets*, 25, 61–72.
- Kim, S.-M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. 423–430.
- Kinne, B. J. (2013). Network Dynamics and the Evolution of International Cooperation. *American Political Science Review*, 107(4), 766–785. <https://doi.org/10.1017/S0003055413000440>
- Kinne, B. J. (2014). Dependent Diplomacy: Signaling, Strategy, and Prestige in the Diplomatic Network. *International Studies Quarterly*, 58(2), 247–259. <https://doi.org/10.1111/isqu.12047>
- Kirimtat, A., Krejcar, O., Kertesz, A., & Tasgetiren, M. F. (2020). Future Trends and Current State of Smart City Concepts: A Survey. *IEEE Access*, 8, 86448–86467. <https://doi.org/10.1109/ACCESS.2020.2992441>
- Kiuru, N., Burk, W. J., Laursen, B., Salmela-Aro, K., & Nurmi, J. (2010). Pressure to drink but not to smoke: Disentangling selection and socialization in adolescent peer networks and peer groups. *Journal of Adolescence*, 33(6), 801–812. <https://doi.org/10.1016/j.adolescence.2010.07.006>

- Kleinberg, J. M. (1999). Hubs, authorities, and communities. *ACM Computing Surveys*, 31(4es), 5. <https://doi.org/10.1145/345966.345982>
- Kochkarov, A. A., Kochkarov, R. A., & Malinetskii, G. G. (2015). Issues of dynamic graph theory. *Computational Mathematics and Mathematical Physics*, 55(9), 1590–1596. <https://doi.org/10.1134/S0965542515090080>
- Koller, D., & Friedman, N. (2010). *Probabilistic Graphical Models: Principles and Techniques* (Nachdr.). MIT Press.
- Kretschmer, D., & Leszczensky, L. (2022). In-Group Bias or Out-Group Reluctance? The Interplay of Gender and Religion in Creating Religious Friendship Segregation among Muslim Youth. *Social Forces*, 100(3), 1307–1332. <https://doi.org/10.1093/sf/soab029>
- Kronegger, L., Mali, F., Ferligoj, A., & Doreian, P. (2012). Collaboration structures in Slovenian scientific communities. *Scientometrics*, 90(2), 631–647. <https://doi.org/10.1007/s11192-011-0493-8>
- Kruschke, J. K. (2010). Bayesian data analysis. *WIREs Cognitive Science*, 1(5), 658–676. <https://doi.org/10.1002/wcs.72>
- Kuo, T.-T., Yan, R., Huang, Y.-Y., Kung, P.-H., & Lin, S.-D. (2013). Unsupervised Link Prediction using Aggregative Statistics on Heterogeneous Social Networks. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 775–783. <https://doi.org/10.1145/2487575.2487614>
- Leguia, M. G., Levnajić, Z., Todorovski, L., & Ženko, B. (2019). Reconstructing Dynamical Networks via Feature Ranking. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(9), 093107. <https://doi.org/10.1063/1.5092170>
- Leifeld, P., & Cranmer, S. J. (2019). A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Network Science*, 7(1), 20–51. <https://doi.org/10.1017/nws.2018.26>
- Leifeld, P., & Cranmer, S. J. (2022). The stochastic actor-oriented model is a theory as much as it is a method and must be subject to theory tests. *Network Science*, 10(1), 15–19. <https://doi.org/10.1017/nws.2022.7>
- Li, W.-J., Yeung, D. Y., & Zhang, Z. (2011). Generalized Latent Factor Models for Social Network Analysis. *International Joint Conference on Artificial Intelligence*, 1705. <https://api.semanticscholar.org/CorpusID:6389369>
- Liang, X., & Liu, A. M. M. (2018). The evolution of government sponsored collaboration network and its impact on innovation: A bibliometric analysis in the Chinese solar PV sector. *Research Policy*, 47(7), 1295–1308. <https://doi.org/10.1016/j.respol.2018.04.012>
- Liben-Nowell, D., & Kleinberg, J. (2007). The link-Prediction Problem for Social Networks. *Journal of the American Society for Information Science and Technology*, 58(7), 1019–1031. <https://doi.org/10.1002/asi.20591>
- Lipman-Blumen, J. (2010). Toxic Leadership: A Conceptual Framework. In F. Bournois, J. Duval-Hamel, S. Roussillon, & J.-L. Scaringella (Eds.), *Handbook of Top Management Teams* (pp. 214–220). Palgrave Macmillan UK. https://doi.org/10.1057/9780230305335_23
- Liu, X., Kertkeidkachorn, N., Murata, T., Kim, K.-S., Leblay, J., & Lynden, S. (2018). Network Embedding Based on a Quasi-Local Similarity Measure. In X. Geng & B.-H. Kang (Eds.), *PRICAI 2018: Trends in Artificial Intelligence* (pp. 429–440). Springer International Publishing. https://doi.org/10.1007/978-3-319-97304-3_33

- Liu, Y., Zhao, C., Wang, X., Huang, Q., Zhang, X., & Yi, D. (2016). The Degree-Related Clustering Coefficient and its Application to Link Prediction. *Physica A: Statistical Mechanics and Its Applications*, 454, 24–33. <https://doi.org/10.1016/j.physa.2016.02.014>
- Lorrain, F., & White, H. C. (1971). Structural equivalence of individuals in social networks. *The Journal of Mathematical Sociology*, 1(1), 49–80.
- Lospinoso, J., & Snijders, T. A. (2019). Goodness of fit for stochastic actor-oriented models. *Methodological Innovations*, 12(3), 205979911988428. <https://doi.org/10.1177/2059799119884282>
- Lotker, Z. (2021). Introduction to Evolving Social Networks. In Z. Lotker, *Analyzing Narratives in Social Networks* (pp. 167–185). Springer International Publishing. https://doi.org/10.1007/978-3-030-68299-6_11
- Lü, L., & Zhou, T. (2011). Link Prediction in Complex Networks: A Survey. *Physica A: Statistical Mechanics and Its Applications*, 390(6), 1150–1170. <https://doi.org/10.1016/j.physa.2010.11.027>
- Luce, R. D., & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika*, 14(2), 95–116.
- Lundberg, J., Tomson, G., Lundkvist, I., Skar, J., & Brommels, M. (2006). Collaboration Uncovered: Exploring the Adequacy of Measuring University-Industry Collaboration through Co-authorship and Funding. *Scientometrics*, 69(3), 575–589. <https://doi.org/10.1007/s11192-006-0170-5>
- Lungeanu, A., & Contractor, N. S. (2015). The Effects of Diversity and Network Ties on Innovations: The Emergence of a New Scientific Field. *American Behavioral Scientist*, 59(5), 548–564. <https://doi.org/10.1177/0002764214556804>
- Maccoby, M. (2000). Narcissistic leaders: The incredible pros, the inevitable cons. *Harvard Business Review*, 78(1).
- Maltseva, D., & Batagelj, V. (2020). Towards a systematic description of the field using keywords analysis: main topics in social networks. *Scientometrics*, 123(1), 357–382. <https://doi.org/10.1007/s11192-020-03365-0>
- Marcot, B. G., & Penman, T. D. (2019). Advances in Bayesian network modelling: Integration of modelling technologies. *Environmental Modelling & Software*, 111, 386–393. <https://doi.org/10.1016/j.envsoft.2018.09.016>
- Matveeva, N., & Ferligoj, A. (2020). Scientific collaboration in Russian universities before and after the excellence initiative Project 5-100. *Scientometrics*, 124(3), 2383–2407. <https://doi.org/10.1007/s11192-020-03602-6>
- Mehra, A., Marineau, J., Lopes, A., & Dass, T. (2009). The coevolution of friendship and leadership networks in small groups. *Predator's Game-Changing Designs: Research-Based Tools*, 145–162.
- Mercken, L., Sleddens, E. F. C., De Vries, H., & Steglich, C. E. G. (2013). Choosing adolescent smokers as friends: The role of parenting and parental smoking. *Journal of Adolescence*, 36(2), 383–392. <https://doi.org/10.1016/j.adolescence.2012.12.004>
- Meyer-Bäse, A., Roberts, R. G., Illan, I. A., Meyer-Bäse, U., Lobbes, M., Stadlbauer, A., & Pinker-Domenig, K. (2017). Dynamical Graph Theory Networks Methods for the Analysis of Sparse Functional Connectivity Networks and for Determining Pinning Observability in Brain Networks. *Frontiers in Computational Neuroscience*, 11, 87. <https://doi.org/10.3389/fncom.2017.00087>
- Mirshahvalad, A., Lindholm, J., Derlén, M., & Rosvall, M. (2012). Significant Communities in Large Sparse Networks. *PLoS ONE*, 7(3), e33721. <https://doi.org/10.1371/journal.pone.0033721>
- Moed, H. F., Markusova, V., & Akoev, M. (2018). Trends in Russian research output indexed in Scopus and

- Web of Science. *Scientometrics*, 116(2), 1153–1180. <https://doi.org/10.1007/s11192-018-2769-8>
- Mohan, A., Venkatesan, R., & Pramod, K. V. (2017). *A Scalable Method for Link Prediction in Large Real World Networks*. 109, 89–101. <https://www.sciencedirect.com/science/article/pii/S0743731517301600>
- Mohorko, A., & Hlebec, V. (2016). Degree of cognitive interviewer involvement in questionnaire pretesting on trending survey modes. *Computers in Human Behavior*, 62, 79–89. https://www.sciencedirect.com/science/article/pii/S0747563216301820?casa_token=JkvjyJdpsU4AAAAA:KdJJiMyz_RPIFh_GyyO-jFW-Ul5lxwq8JGsTfLGHPi4U7CCaEPGalooldZai40ZaVRXoRjjQkt4
- Mohrenberg, S. (2017). Studying Policy Diffusion with Stochastic Actor-Oriented Models. In B. Hollstein, W. Matiaske, & K.-U. Schnapp (Eds.), *Networked Governance* (pp. 163–188). Springer International Publishing. https://doi.org/10.1007/978-3-319-50386-8_10
- Mokken, R. J. (1979). Cliques, clubs and clans. *Quality & Quantity*, 13(2), 161–173.
- Molokwu, B. C. (2021). *Social Network Analysis: A Machine Learning Approach*. University of Windsor, Canada.
- Moradabadi, B., & Meybodi, M. R. (2018). *Link Prediction in Stochastic Social Networks: Learning Automata Approach*. 24, 313–328. <https://doi.org/10.1016/j.jocs.2017.08.007>
- Mørup, M., & Schmidt, M. N. (2012). Bayesian Community Detection. *Neural Computation*, 24(9), 2434–2456. https://doi.org/10.1162/NECO_a_00314
- Muniz, C. P., Goldschmidt, R., & Choren, R. (2018). Combining Contextual, Temporal and Topological Information for Unsupervised Link Prediction in Social Networks. *Knowledge-Based Systems*, 156, 129–137. <https://doi.org/10.1016/j.knosys.2018.05.027>
- Mydosh, J. A. (1993). *Spin glasses: an experimental introduction*. CRC Press.
- Nam, T., & Pardo, T. A. (2011). Conceptualizing smart city with dimensions of technology, people, and institutions. *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, 282–291. <https://doi.org/10.1145/2037556.2037602>
- Nasiri, E., Berahmand, K., Samei, Z., & Li, Y. (2022). Impact of Centrality Measures on the Common Neighbors in Link Prediction for Multiplex Networks. *Big Data*, 10, 138–150. <https://doi.org/10.1089/big.2021.0254>
- Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), 066133.
- Newman, M. E. J. (2013). *Spectral community detection in sparse networks*. <https://doi.org/10.48550/ARXIV.1308.6494>
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.
- Nguyen, G. H., Lee, J. B., Rossi, R. A., Ahmed, N., Koh, E., & Kim, S. (2018). *Continuous-Time Dynamic Network Embeddings*. 969–976. <https://doi.org/10.1145/3184558.3191526>
- Nowicki, K., & Snijders, T. A. B. (2001). Estimation and Prediction for Stochastic Blockstructures. *Journal of the American Statistical Association*, 96(455), 1077–1087. <https://doi.org/10.1198/016214501753208735>
- Obholzer, A. (1996). Psychoanalytic contributions to authority and leadership issues. *Leadership & Organization Development Journal*, 17(6), 53–56. <https://doi.org/10.1108/01437739610152966>
- Özcan, A., & Öğüdücü, Ş. G. (2016). *Temporal Link Prediction Using Time Series of Quasi-Local Node Similarity Measures*. 381–386. <https://doi.org/10.1109/ICMLA.2016.0068>
- Pan, Y., & Fond, M. (2014). Evaluating multilingual questionnaires: A sociolinguistic perspective. *Survey*

- Research Methods*, 8(3), 181–194. <https://ojs.ub.uni-konstanz.de/srm/article/view/5483>
- Paranjape, A., Benson, A. R., & Leskovec, J. (2017). *Motifs in temporal networks*. 601–610.
- Park, D.-H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.
- Pattison, P., & Robins, G. (2002). 9. Neighborhood-Based Models for Social Networks. *Sociological Methodology*, 32(1), 301–337. <https://doi.org/10.1111/1467-9531.00119>
- Peixoto, T. P. (2014). Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Physical Review E*, 89(1), 012804.
- Perianes-Rodriguez, A., Waltman, L., & Van Eck, N. J. (2016). Constructing Bibliometric Networks: A Comparison Between Full and Fractional Counting. *Journal of Informetrics*, 10(4), 1178–1195. <https://doi.org/10.1016/j.joi.2016.10.006>
- Persson, C., Bohlin, L., Edler, D., & Rosvall, M. (2016). Maps of sparse Markov chains efficiently reveal community structure in network flows with memory. *arXiv Preprint arXiv:1606.08328*.
- Pike, T. W. (2010). *Collaboration Networks and Scientific Impact among Behavioral Ecologists*. 21(2), 431–435. <https://doi.org/10.1093/beheco/arp194>
- Pittel, B., Spencer, J., & Wormald, N. (1996). Sudden Emergence of a Giantk-Core in a Random Graph. *Journal of Combinatorial Theory, Series B*, 67(1), 111–151. <https://doi.org/10.1006/jctb.1996.0036>
- Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks. 284–293.
- Pranckutè, R. (2021). *Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World*. 1–59. <https://doi.org/10.3390/publications9010012>
- Psorakis, I., Roberts, S., & Sheldon, B. (2010). Efficient Bayesian Community Detection using Non-negative Matrix Factorisation. <https://doi.org/10.48550/ARXIV.1009.2646>
- Redhead, D., & Von Rueden, C. R. (2021). Coalitions and conflict: A longitudinal analysis of men's politics. *Evolutionary Human Sciences*, 3, e31. <https://doi.org/10.1017/ehs.2021.26>
- Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, 74(1), 016110. <https://doi.org/10.1103/PhysRevE.74.016110>
- Rettinger, A., Wermser, H., Huang, Y., & Tresp, V. (2012). *Context-Aware Tensor Decomposition for Relation Prediction in Social Networks*. 2(4), 373–385. <https://doi.org/10.1007/s13278-012-0069-5>
- Rhue, L., & Sundararajan, A. (2014). Digital access, political networks and the diffusion of democracy. *Social Networks*, 36, 40–53. <https://doi.org/10.1016/j.socnet.2012.06.007>
- Ricciardi, F., & Za, S. (2015). Smart City Research as an Interdisciplinary Crossroads: A Challenge for Management and Organization Studies. In L. Mola, F. Pennarola, & S. Za (Eds.), *From Information to Smart Society* (Vol. 5, pp. 163–171). Springer International Publishing. https://doi.org/10.1007/978-3-319-09450-2_14
- Richardson, M., & Domingos, P. (2006). *Markov Logic Networks*. 62(1-2), 107–136. <https://doi.org/10.1007/s10994-006-5833-1>
- Roesler, C., & Broekel, T. (2017). The role of universities in a network of subsidized R&D collaboration: The case of the biotechnology-industry in Germany. *Review of Regional Research*, 37(2), 135–160. <https://doi.org/10.1007/s10037-017-0118-7>
- Romero, D. M., Meeder, B., & Kleinberg, J. (2011). *Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter*. 695–704.

- Rosenthal, S. A., & Pittinsky, T. L. (2006). Narcissistic leadership. *The Leadership Quarterly*, 17(6), 617–633. <https://doi.org/10.1016/j.leaqua.2006.10.005>
- Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4), 1118–1123.
- Rosvall, M., & Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PloS One*, 6(4), e18209.
- Rosvall, M., Delvenne, J., Schaub, M. T., & Lambiotte, R. (2019). Different Approaches to Community Detection. In P. Doreian, V. Batagelj, & A. Ferligoj (Eds.), *Advances in Network Clustering and Blockmodeling* (1st ed., pp. 105–119). Wiley. <https://onlinelibrary.wiley.com/doi/10.1002/9781119483298.ch4>
- Ruan, J., & Zhang, W. (2007). An Efficient Spectral Algorithm for Network Community Discovery and Its Applications to Biological and Social Networks. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 643–648. <https://doi.org/10.1109/ICDM.2007.72>
- Schaefer, D. R., Haas, S. A., & Bishop, N. J. (2012). A dynamic model of US adolescents' smoking and friendship networks. *American Journal of Public Health*, 102(6), e12–e18.
- Scheffler, M., & Brunzel, J. (2020). Destructive leadership in organizational research: a bibliometric approach. *Scientometrics*, 125(1), 755–775. <https://doi.org/10.1007/s11192-020-03621-3>
- Scheidlinger, S. (1980). The psychology of leadership revisited: An overview. *Group*, 4(1), 5–17. <https://doi.org/10.1007/BF01456483>
- Schmidt, B. (2015). Costs and benefits of friendly boards during mergers and acquisitions. *Journal of Financial Economics*, 117(2), 424–447.
- Schröder, D., & Thomsen, S. (2021, October 10). *Foundation ownership and financial performance—a global analysis*. International Corporate Governance Society Annual Conference, Groningen.
- Schwartz, H. S. (1990). *Narcissistic process and corporate decay: the theory of the organization ideal*. New York University Press.
- Schweinberger, M., & Snijders, T. A. B. (2003). Settings in Social Networks: A Measurement Model. *Sociological Methodology*, 33(1), 307–341. <https://doi.org/10.1111/j.0081-1750.2003.00134.x>
- Seidman, S. B. (1983a). Internal cohesion of LS sets in graphs. *Social Networks*, 5(2), 97–107.
- Seidman, S. B. (1983b). Network structure and minimum degree. *Social Networks*, 5(3), 269–287.
- Seidman, S. B., & Foster, B. L. (1978). A graph-theoretic generalization of the clique concept. *Journal of Mathematical Sociology*, 6(1), 139–154.
- Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of Scientific Collaboration*. The MIT Press.
- Simmel, G. (2010). *Conflict and the web of group affiliations*. Simon and Schuster.
- Singh, A., & Humphries, M. D. (2015). Finding communities in sparse networks. *Scientific Reports*, 5(1), 8828. <https://doi.org/10.1038/srep08828>
- SKOLKOVO. (n.d.). *Портрет владельца капитала в России 2020. Что думают россияне о благосостоянии и о богатых людях*. Бизнес-школа СКОЛКОВО - бизнес-образование, бизнес-обучение в Москве. Retrieved November 25, 2023, from <https://www.skolkovo.ru/expert-opinions/portret-vladelca-kapitala-v-rossii-2020-chto-dumayut-rossiyane-o-blagosostoyanii-i-o-bogatyh-lyudyah/>
- Slaninová, K., Dolák, R., Miškus, M., Martinovič, J., & Snášel, V. (2010). User segmentation based on finding communities with similar behavior on the web site. 3, 75–78.
- Snijders, T. A. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology*,

21(1-2), 149–172.

- Snijders, T. A. (2017). Stochastic actor-oriented models for network dynamics. *Annual Review of Statistics and Its Application*, 4, 343–363.
- Snijders, T. A. B. (n.d.). *Markov Chain Monte Carlo Estimation of Exponential Random Graph Models*.
- Snijders, T. A. B. (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology*, 31(1), 361–395. <https://doi.org/10.1111/0081-1750.00099>
- Souder, D., Reilly, G., Bromiley, P., & Mitchell, S. (2016). A behavioral understanding of investment horizon and firm performance. *Organization Science*, 27(5), 1202–1218.
- Srilatha, P., & Manjula, R. (2016). *Paper Similarity Index based Link Prediction Algorithms in Social Networks: A Survey*. 2, 87–94.
- Steglich, C., Snijders, T. A., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1), 329–393.
- Stokman, F. N., & Sprenger, C. (1989). *GRADAP: Graph definition and analysis package*.
- Strauss, D. (1986). On a General Class of Models for Interaction. *SIAM Review*, 28(4), 513–527. <https://www.jstor.org/stable/2031102>
- Su, L., Lyu, P., Yang, Z., Ding, S., & Zhou, K. (2015). Scientometric cognitive and evaluation on smart city related construction and building journals data. *Scientometrics*, 105(1), 449–470. <https://doi.org/10.1007/s11192-015-1697-0>
- Tan, Q., Liu, N., & Hu, X. (2019). Deep Representation Learning for Social Network Analysis. *Frontiers in Big Data*, 2, 1–10. <https://doi.org/10.3389/fdata.2019.00002>
- Tepper, B. J. (2007). Abusive Supervision in Work Organizations: Review, Synthesis, and Research Agenda. *Journal of Management*, 33(3), 261–289. <https://doi.org/10.1177/0149206307300812>
- Theodoridis, S. (2015). *Machine learning: a Bayesian and optimization perspective*. Academic press. https://www.google.com/books?hl=ru&lr=&id=NHODBAAAQBAJ&oi=fnd&pg=PP1&dq=Theodoridis+S.+Machine+learning:+a+Bayesian+and+optimization+perspective.+Academic+press,+2015.&ots=2TKnMLDj-1&sig=7Zl4HEgG6xsYonXSwAAujJ4_azQ
- Thomsen, S. (1999). Corporate ownership by industrial foundations. *European Journal of Law and Economics*, 7(2), 117–137.
- Thomsen, S. (2016a). Industrial foundations: The Danish model. In *Nordic Ways* (pp. 3–33). Brookings Institution Press.
- Thomsen, S. (2016b). The Nordic corporate governance model. *Management and Organization Review*, 12(1), 189–204.
- Thomsen, S. (2018). Foundation Ownership and Firm Performance: A Review of the International Evidence. In *Corporate Governance in Contention* (pp. 66–85). Oxford University Press. <https://doi.org/10.1093/oso/9780198805274.003.0004>
- Thomsen, S., & Degn, S. M. (2014). *The Charters of Industrial Foundations*. <https://www.tifp.dk/wp-content/uploads/2015/04/Charters06.pdf>
- Thomsen, S., & Kavadis, N. (2022). Enterprise foundations: law, taxation, governance, and performance. *Annals of Corporate Governance*, 6(4), 227–333.
- Thomsen, S., Poulsen, T., Børsting, C., & Kuhn, J. (2018). Industrial foundations as long-term owners. *Corporate Governance: An International Review*, 26(3), 180–196.

- Thurner, P. W., & Binder, M. (2009). European Union transgovernmental networks: The emergence of a new political space beyond the nation-state? *European Journal of Political Research*, 48(1), 80–106. <https://doi.org/10.1111/j.1475-6765.2008.00825.x>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Townsend, A. M. (2013). *SMART CITIES: Big Data, Civic Hackers, and the Quest for a New Utopia*. <https://doi.org/10.48558/BXTE-BQ78>
- Traag, V. A., Waltman, L., & Van Eck, N. J. (2019a). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1), 5233.
- Traag, V. A., Waltman, L., & Van Eck, N. J. (2019b). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1), 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- Van De Bunt, G. G., Van Duijn, M. A. J., & Snijders, T. A. B. (1999). Friendship Networks Through Time: An Actor-Oriented Dynamic Statistical Network Model. *Computational & Mathematical Organization Theory*, 5(2), 167–192. <https://doi.org/10.1023/A:1009683123448>
- van der Pas, S. L., & van der Vaart, A. (2018). *Bayesian community detection*. <https://projecteuclid.org/journals/bayesian-analysis/volume-13/issue-3/Bayesian-Community-Detection/10.1214/17-BA1078.short>
- Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- Vermond, D., De Groot, E., Sills, V. A., Lyratzopoulos, G., Walter, F. M., De Wit, N. J., & Rubin, G. (2022). The evolution and co-evolution of a primary care cancer research network: From academic social connection to research collaboration. *PLOS ONE*, 17(7), e0272255. <https://doi.org/10.1371/journal.pone.0272255>
- Vishnivetskaya, A., & Alexandrova, E. (2019). “Smart city” concept. Implementation practice. *IOP Conference Series: Materials Science and Engineering*, 497, 012019. <https://doi.org/10.1088/1757-899X/497/1/012019>
- Wang, J., Ma, Y., & Yuan, Y. (2021). *Towards Fast Evaluation of Unsupervised Link Prediction by Random Sampling Unobserved Links* (No. arXiv:2002.09165). arXiv. <https://doi.org/10.48550/arXiv.2002.09165>
- Wang, P., Xu, B., Wu, Y., & Zhou, X. (2014). *Link Prediction in Social Networks: the State-of-the-Art* (No. arXiv:1411.5118). arXiv. <https://doi.org/0.1007/s11432-014-5237-y>
- Wang, S., Tang, J., Aggarwal, C. C., Chang, Y., & Liu, H. (2017). *Signed Network Embedding in Social Media*. 327–335. <https://pubs.siam.org/doi/10.1137/1.9781611974973.37>
- Wang, Y., Jin, J., & Goncalves, J. (2022). *Bayesian Inference of Stochastic Dynamical Networks*. <https://doi.org/10.48550/ARXIV.2206.00858>
- Wasserman, S., & Faust, K. (1994a). *Social network analysis: methods and applications*. Cambridge University Press.
- Wasserman, S., & Faust, K. (1994b). *Social Network Analysis: Methods and Applications* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511815478>
- Wassouf, W. N., Alkhatab, R., Salloum, K., & Balloul, S. (2020). Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study. *Journal of Big Data*, 7, 1–24.
- Watts, S. A., & Zhang, W. (2008). Capitalizing on content: Information adoption in two online communities. *Journal of the Association for Information Systems*, 9(2), 3.
- Wu, Z., Lin, Y., Zhao, Y., & Yan, H. (2018). *Improving Local Clustering based Top-L Link Prediction Methods via Asymmetric Link Clustering Information*. 492, 1859–1874. <https://doi.org/10.1016/j.physa.2017.11.103>

- Xu, R., & Wunsch, D. C. (2008). Recent advances in cluster analysis. *International Journal of Intelligent Computing and Cybernetics*, 1(4), 484–508.
- Yang, T., Chi, Y., Zhu, S., Gong, Y., & Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Machine Learning*, 82(2), 157–189. <https://doi.org/10.1007/s10994-010-5214-7>
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- Yuloskov, A., Bahrami, M. R., Mazzara, M., & Kotorov, I. (2021). Smart Cities in Russia: Current Situation and Insights for Future Development. *Future Internet*, 13(10), 252. <https://doi.org/10.3390/fi13100252>
- Zalk, M. H. W. V., Kerr, M., Branje, S. J. T., Stattin, H., & Meeus, W. H. J. (2010). Peer Contagion and Adolescent Depression: The Role of Failure Anticipation. *Journal of Clinical Child & Adolescent Psychology*, 39(6), 837–848. <https://doi.org/10.1080/15374416.2010.517164>
- Zhang, W., Liu, F., Xu, D., & Jiang, L. (2019). Recommendation System in Social Networks with Topical Attention and Probabilistic Matrix Factorization. 14(10), e0223967. <https://doi.org/10.1371/journal.pone.0223967>
- Zhang, Y., Leslie, J. B., & Hannum, K. M. (2013). Trouble Ahead: Derailment Is Alive and Well. *Thunderbird International Business Review*, 55(1), 95–102. <https://doi.org/10.1002/tie.21525>
- Zhou, X., Liu, X., Wang, C., Zhai, D., Jiang, J., & Ji, X. (2021). Learning with Noisy Labels via Sparse Regularization (No. arXiv:2108.00192). arXiv. <http://arxiv.org/abs/2108.00192>
- Zhu, J., & Liu, W. (2020). A Tale of Two Databases: the Use of Web of Science and Scopus in Academic Papers. *Scientometrics*, 123(1), 321–335. <https://doi.org/10.1007/s11192-020-03387-8>
- Барковец, А. А. (2013). Функционирование российской модели эндаумент-фондов. 55–67.
- Горшков, М. К., & Шереги, Ф. Э. (2011). Прикладная социология: методология. М.: Институт Социологии РАН. <http://library.lgaki.info:404/2017/%D0%93%D0%BE%D1%80%D1%88%D0%BA%D0%BE%D0%B2%D0%A1%D0%BE%D1%86%D0%B8%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F.pdf>
- Докука, С. В. (2021). Использование стохастических акторно-ориентированных моделей для анализа коэволюции сетей и поведения. *Мониторинг Общественного Мнения: Экономические и Социальные Перемены*, 2, Article 2. <https://cyberleninka.ru/article/n/ispolzovanie-stohasticheskikh-aktorno-orientirovannyh-modeley-dlya-analiza-koevolyutsii-setey-i-povedeniya>
- Докука, С. В., & Валеева, Д. Р. (2015). Статистические модели для анализа динамики социальных сетей в исследованиях образования. *Вопросы Образования*, 1, Article 1. <https://cyberleninka.ru/article/n/statisticheskie-modeli-dlya-analiza-dinamiki-sotsialnyh-setey-v-issledovaniyah-obrazovaniya>
- Климов, И. А., Мисютина, В. В., Агеев, Д. А., Шибкова, С. А., & Цивилок, Н. В. (2023). Модели управления благотворительными фондами – бенефициарными собственниками бизнес-компаний. МШУ Сколково. <https://fondpotanin.ru/upload/iblock/222/f4yz1ovq7o155bkgvxaki2lwqrnkjh13.pdf>
- Кохут, Х. (2003). Анализ самости: Системный подход к лечению нарциссических нарушений личности / Пер. с англ. А.М. Боковикова. Когито-Центр.
- Мальцева, Д. В., & Fiala, D. (2023). Russian Publications in Web of Science: A Bibliometric Study. (Препринт) *COLLNET Journal of Scientometrics and Information Management*.

- Мисютина, В., Юсуфов, Р., Климов, И., & Абрамов, Р. (2015). *Исследование владельцев капиталов России 2015*. WTC Сколково. [https://wtc.skolkovo.ru/media/documents/SKOLKOVO_WTC-WPS_2015_Report\(RUS\).pdf](https://wtc.skolkovo.ru/media/documents/SKOLKOVO_WTC-WPS_2015_Report(RUS).pdf)
- Павлова, И. А. (2023). ПОСТРОЕНИЕ КАРТЫ СОПРИСУТСТВИЯ КЛЮЧЕВЫХ СЛОВ ПО ТЕМЕ «КАПИТАЛ ЗДОРОВЬЯ» В ПРОГРАММЕ VOSVIEWER. *Векторы Благополучия*, 49(2), 38–54.
- Подольская, А. П., & Харламова, Е. Е. (2016). Целевой капитал как источник финансирования некоммерческой организации. *Финансовая Аналитика: Проблемы и Решения*, 2(284), 31–42. <https://cyberleninka.ru/article/n/tselevoy-kapital-kak-istochnik-finansirovaniya-nekommercheskoy-organizatsii>
- Ребязина, В. А., Шарко, Е. Р., Березка, С. М., & Старков, А. Г. (2022). *Развитие электронной коммерции в России: влияние пандемии COVID-19*. ИД НИУ ВШЭ. <https://doi.org/10.17323/978-5-7598-2611-8>
- Рогозин, Д. М. (2000). Когнитивный анализ опросного инструмента. *Социологический Журнал*, 3-4, 018–070. <https://cyberleninka.ru/article/n/kognitivnyy-analiz-oprosnogo-instrumenta>
- Рогозин, Д. М. (2002). *Когнитивный анализ опросного инструмента*.
- Савченко, П. В., Федорова, М. Н., & Шлихтер, А. А. (2015). Эндаумент как институт социальных инвестиций. *Вестник Института Экономики Российской Академии Наук*, 2, 52–63.
- Соколова, С. Ю. (2011). Фонды целевого капитала в системе обеспечения конкурентоспособности некоммерческих организаций. *Экономический Журнал*, 24, 73–78.
- Таворкин Е. П. (2021). *Основы методики социологического исследования*. НИЦ ИНФРА-М. <https://znanium.com/catalog/document?id=398636>
- Трофимова, И. Н. (2023). *Международное сотрудничество российских исследователей: текущие позиции и тенденции: по данным Web of Science за 2018–2022 гг.* (No. 4). 32(4), Article 4. <https://doi.org/10.17323/1811-038X-2023-32-4-178-198>
- Хорни, К. (2019). *Невроз и личностный рост. Борьба за самореализацию / Пер. с англ. Е.И. Замфир*. Питер.
- Цель под ключ - PBWM.RU. (n.d.). Retrieved November 25, 2023, from <https://pbwm.ru/articles/tsel-pod-klyuch>