

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

*Крылов Сергей Викторович*

Выпускная квалификационная работа

*Коммуникативные практики формирования ценностно-идеологических комплексов: сравнительный анализ социально-ориентированных общественных движений*

Уровень образования:

Направление **39.04.01 «Социология»**

Основная образовательная программа магистратуры

ВМ.5589.2019 «Социология»

Научный руководитель: доктор

философских наук,

профессор кафедры

социологии культуры и коммуникации

Козловский Владимир Вячеславович

Рецензент:

старший научный сотрудник, Социологический институт РАН - филиал Федерального государственного бюджетного учреждения науки Федерального научно-исследовательского социологического центра Российской академии наук, шеф-редактор, Журнал социальных исследований Laboratorium, Автономная некоммерческая организация «Центр независимых социологических исследований»

**Парфенова Оксана Анатольевна**

Санкт-Петербург

2021

**СОДЕРЖАНИЕ**

|  |  |
| --- | --- |
| Введение | 3 |
| **Глава 1. Теоретические основы изучения коммуникативных практик социально-ориентированных общественных движений** | 8 |
| 1.1. Общественные движения в условиях информационного общества как объект социологического исследования | 8 |
| 1.2. Идентичности членов общественных движений как система ценностно-нормативных комплексов: коммуникативный подход | 20 |
| 1.3. Современные методы исследования коммуникативных практик общественных движений в цифровой среде | 27 |
| **Глава 2. Сравнительный анализ коммуникативных практик социально-ориентированных общественных движений** | 37 |
| 2.1. Программа исследования коммуникативных практик социально-ориентированных общественных движений. Сбор цифровых следов | 37 |
| 2.2. Составление коммуникативного портрета коллективной языковой личности: интеллектуальный анализ текстов | 46 |
| 2.3. Выявление специфики коммуникативных практик формирования ценностно-идеологических комплексов: корреляционный анализ виртуальных сообществ | **53** |
| **Заключение** | **62** |
| **Литература** | **65** |
| **Приложение** | **69** |

**ВВЕДЕНИЕ**

**Актуальность исследования.** В результате развития и повсеместного распространения телекоммуникационных технологий, цифровая среда стала новым пространством общественной жизни. Многие социальные явления в той или иной мере нашли своё отражение в этой среде, и одним из них является деятельность общественных движений. Для общественных движений коммуникативная составляющая в их деятельности играет одну из ведущих ролей, что является фактором, побуждающим их к освоению цифрового пространства и переносу в него большей части своей активности. В частности, Интернет становится для общественных движений инструментом распространения своих убеждений и привлечения в свои ряды новых участников. От того, насколько эффективно ведётся эта деятельность, во многом зависит успешность общественного движения в целом – его способность влиять на жизнь общества. Последствия деятельности общественных движений могут быть весьма велики, при этом, реализуемые ими коммуникативные стратегии воздействия на аудиторию остаются малоизученными и, вероятно, не в полной мере осознаются самими участниками общественных движений, чем и обусловлена актуальность данного исследования. С другой стороны, данная работа имеет методологическую ценность. Цифровая среда является относительно новым пространством социологических исследований, характеризующимся с одной стороны обширными новыми возможностями, а с другой - дефицитом разработанных теоретических и методологических подходов.

**Степень научной разработанности проблемы.**  Вопросы, изучаемые в рамках данного исследования, затрагивают три научно-исследовательских направления: общие теоретических подходы к изучению деятельности общественных движений, вопросы формирования идентичности в процессе коммуникационной деятельности и методологию социологических исследования цифрового пространства.

Теоретические подходы к изучению деятельности общественных движений разделяются на три парадигмы. Наибольший вклад в развитие парадигмы коллективного поведения внесли разработчики теорий коллективного поведения (Н. Селмезер, Г. Блумер, Л. Киллиан, Р. Тернер, Дж. Лофланд), массового общества (С. Липсет, У. Корнхаузер) и относительной депривации (Т. Р. Гарр и Д. Давис). Парадигма коллективного действия объединяет теории, отталкивающиеся от работ М. Олсона, и теорию политического процесса (Ч. Тилли, С. Тэрроу, Д. Макадам). Парадигма Новых общественных движений является продуктом европейской социологической школы и связана с именами таких исследователей, как А. Турен (Франция), К. Оффе и Д. Рухт (Германия), Х. Кризи (Швейцария), А. Мелуччи и А. Пиззорно (Италия), М. Кастеллс (Испания). В отечественной социологии исследованиями общественных движений известны Е.А. Здравомыслова и И.А.Халий, а теоретические подходы к описанию функционирования общественных движений в виртуальном пространстве наиболее активно развивают в своих работах Р.Э. Бараш и А.Ю. Антоновский.

Формирования идентичности в процессе коммуникационной деятельности изучается с средины XX века. Большой вклад в изучение этого явления внесли Э. Эриксон, Жак Лакан, Д. Винникот, Дж. Марсия, П. Бергер и Т. Лукман, Х. Кохут и Г. Салливан, Ч. Кули, Дж. Г. Мид, Г. Брейкуэлл, М. Хайдег­гер. В отечественной социологии разработкой данной темы активно занимается В.И. Карасик.

Методология социологических исследований цифрового пространства является относительно новым и очень разносторонним исследовательским направлением. В рамках российской социологии в этой сфере можно выделить работы В. Н. Нечипуренко, В. И. Дудиной и Д.И. Юдиной, О.Ю. Кольцовой и К.А. Маслинского, Ю. Г. Рыкова, П. А. Мейлахса и Я. Е. Синявской.

**Научная новизна** данного исследования состоит в разработке методики оценки уровня вовлеченности индивида в деятельность общественного движения на основании характеристик его коммуникативного поведения, а также в осуществлении сравнительного анализ коммуникативных практик осуществляемых различными общественными движениями.

**Гипотезой исследования**  является предположение о наличии корреляции между тематической ориентацией социально ориентированных общественных движений, и структурными особенностями коммуникативной деятельности, осуществляемой в посвящённых данным движениям сообществах на интернет-платформе «ВКонтакте».

**Объектом исследования** являются тематические сообщества в социальной сети «ВКонтакте», в рамках которых осуществляется коммуникативная деятельность социально ориентированных общественных движений.

**Предметом исследования** являются структурные особенности коммуникативных практик, направленных на формирование ценностно-идеологических комплексов, осуществляемых в рамках данных тематических сообществ.

**Целью исследования** является выявление взаимосвязи между структурными особенностями коммуникативных практик, направленных на формирование ценностно-идеологических комплексов и тематической ориентацией данных общественных движений.

Для достижения обозначенной целей были поставлены следующие **задачи:**

1. Выявить тематические сообщества в социальной сети «ВКонтакте», в рамках которых осуществляется коммуникативная деятельность социально ориентированных общественных движений.
2. Собрать эмпирические данные, являющиеся цифровыми следами коммуникативной деятельности, осуществляемой социально ориентированными общественными движениям в выявленных тематических сообществах.
3. Выявить особенности коммуникативной деятельности для каждого исследуемого сообщества.
4. Составить коммуникативный портрет коллективной языковой личности для каждого исследуемого сообщества.
5. Составить индивидуальный коммуникативный портрет языковой личности для каждого участника каждого сообщества.
6. Определить уровень идентификации с каждым исследуемым сообществом каждого участника данных сообществ.
7. Выявить особенности коммуникативного воздействия, направленного со стороны каждого исследуемого сообщества на каждого участника данных сообществ.
8. Сопоставить особенности коммуникативного воздействия направленного со стороны каждого исследуемого сообщества на каждого участника данных сообществ с уровнем идентификации этих участников с данными сообществами.
9. Выявить характерные для каждого исследуемого сообщества особенности коммуникативного воздействия, направленного на участников данного сообщества обладающих низким уровнем идентификации.
10. Проанализировать взаимосвязь между особенностями коммуникативных практик, направленных на участников данного сообщества, обладающих низким уровнем идентификации, с тематической ориентацией данного сообщества.

**Теоретико-методологические основания исследования:**

Данное исследование опирается на теорию критической фазы британского исследователя Христиана Фухса, описывающую в рамках парадигмы Новых общественных движений процесс образования общественного движения вследствие формирования соответствующей идентичности в определённых коммуникативных ситуациях. Также используются социолингвистические концепции «коммуникативный портрет» и «языковая личность» и методы сбора и анализа данных относящиеся к сфере цифровой социологии.

**Эмпирическую базу исследования** составляют записи, опубликованные на стене исследуемых сообществ на сайте «ВКонтакте», и комментарии к ним, за период с 01.01.2020 до даты сбора данных (14-22 апреля 2021). Для участия в исследовании выбрано 36 сообществ относящихся к общественным движениям следующих идеологических ориентаций:

1. Отрицание медицины (отказ от прививок, СПИД-диссиденты, COVID-диссиденты) — 6 сообществ.
2. Вегетарианство — 6 сообществ.
3. Борьба с живодёрами и «догхантерами» — 6 сообществ.
4. Феминизм — 6 сообществ.
5. Национализм — 6 сообществ.
6. Идеологически-нейтральные сообщества (контрольная группа) — 6 сообществ.

Для реализации поставленных задач были применены следующие **методы исследования**:

1. Частотный анализ употребимости терминов в текстах, написанных на естественном языке.
2. Тематическое моделирование текстов написанных на естественном языке.
3. Сетевой анализ коммуникативной деятельности виртуальных сообществ.
4. Множественный корреляционный анализ.

**Апробация результатов:**

1. Участие во Всероссийской научной конференции XIV Ковалевские чтения «Цифровое общество – новый формат социальной реальности: структуры, процессы и тенденции развития», с докладом «Препроцессинг текстов загруженных из сети Интернет»
2. Участие в Международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов-2021», на секции "Социология коммуникативных систем", с докладом «Цифровые коммуникации в конструировании идентичности социально ориентированных общественных движений»

**Структура работы:** работа состоит из введения, двух глав, заключения, списка использованной литературы и приложения.

**ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ КОММУНИКАТИВНЫХ ПРАКТИК СОЦИАЛЬНО-ОРИЕНТИРОВАННЫХ ОБЩЕСТВЕННЫХ ДВИЖЕНИЙ**

**1.1. Общественные движения в условиях информационного общества как объект социологического исследования**

В данном разделе мы коротко рассмотрим историю возникновения основных теоретических концепций посвящённых изучению общественных движений и новые тенденции в этой сфере, связанные с влиянием на общественную жизнь телекоммуникационной сети Интернет. Обоснуем актуальность изучения осуществляемой общественными движениями коммуникативной деятельности и выделим подходы к её изучению.

В индустриально развитых странах начала XIX века такое явление социальной жизни как общественные движения, и прежде игравшее заметную роль в истории человечества, стало приобретать массовый характер и оказывать всё большее влияние на происходящие в обществе процессы социального развития. Зародившаяся в тот же исторический период социология, не могла обойти вниманием ставшее столь заметным социальное явление, однако, долгое время социологическому осмыслению подвергались лишь отдельные вопросы, касающиеся возникновения, функционирования и роли общественных движений. Только во второй половине XX века, в контексте осмысления прокатившейся по миру волны становления тоталитарных идеологически ориентированных режимов, американскими социологами была создана первая целостная, научно оформленная концепция общественных движений – парадигма коллективного поведения, остававшаяся доминирующей в социологии с конца 1950-х, до начала 1970-х годов. В рамках этой парадигмы нашли своё место теории коллективного поведения (Н. Селмезер, Г. Блумер, Л. Киллиан, Р. Тернер, Дж. Лофланд), массового общества (С. Липсет, У. Корнхаузер) и относительной депривации (Т. Р. Гарр, Д. Давис). Всем этим теориям свойственно акцентировать дестабилизирующую роль общественных движений в демократическом обществе и социальном порядке. [3, с. 270] Общественные движения здесь трактуются как иррациональное и нескоординированное коллективное поведение в ситуации социального кризиса и связанных с ним явлений фрустрации, депривации и агрессивности. С точки зрения структурного функционализма, нарушения в функционировании социальной системы и противоречия между интересами различных групп порождают различные «структурные напряжения», которые, в свою очередь, провоцируют появление определённых форм коллективного поведения, предстающих в виде общественных движений. Общественные движения, в рамках этого подхода, определяются как усилия групп людей, вовлеченных в совместное действие, с целью «способствовать или препятствовать изменениям в обществе или группе». [12, с. 71]

В целом, можно отметить негативное отношение авторов указанных теорий к описываемому ими явлению. Так, например, теория массового общества, получившая широкую известность после публикации в 1959 г. книги У. Корнхаузера «Политика массового общества», рассматривает общественные движения, как своего рода, социальную патологию, возникающую в аномичном и атомизированном обществе масс. Участники подобных движений описываются как легко поддающиеся манипулированию люди, испытывающие дефицит социальных связей и общения, мотивом которых является желание почувствовать свою значимость. [3, с. 270-271]

В начале 1970-х годов, на смену рассмотренному подходу пришла парадигма коллективного действия. Входящие в её состав теории опираются на «модель рационального актора» М. Олсона и предлагают различные решения сформулированного им «парадокса безбилетника» согласно которому, коллективное действие невозможно без стимулирования или принуждения, так как каждому потенциальному участнику всегда выгоднее отказаться от участия в действии, что позволяет не нести издержек, но пользоваться полученными благами в случае успеха. В рамах этих теорий общественное движение рассматривается как рацио­нальный коллективный актор, выстраивающий свою стратегию для достижения определенной политической цели, а к числу необходимых для коллективного действия ресурсов относятся, в частности, нематериальные символические ресурсы (такие как вовлеченность в движение, солидарность).

Одной из ведущих теорий в рамках рассматриваемой парадигмы является теория политического процесса (Ч. Тилли, С. Тэрроу, Д. Макадам). Она положила начало изучению взаимодействия общественных движений с политическими институтами. Согласно данной теории, общественные движения рассматриваются как рациональные акторы, взаимодействующие в рамках существующей структуры политических возможностей с представителями власти, и отстаивающие перед ними требования социальных групп, не имеющих институционального представительства. [12, с. 72]

В отличие от парадигмы коллективного поведения, в рамках парадигмы коллективного действия, общественные движения рассматриваются не как результат дисфункций социальной системы, а как элемент ее нормального функционирования.

Парадигмы коллективного поведения и коллективного действия являются разработкой американских социологических школ, европейская же мысль, опираясь на иной эмпирический материал и иную интеллектуальную традицию, пошла в несколько другом направлении. В 1970-1980 годы европейскими социологами был сформулирован ряд социологических теорий сформировавших так называемую парадигму «Новых общественных движений». Данная парадигма не претендует на универсальность, ограничиваясь описанием общественных движений ориентированных на поиск и обретение новой идентичности и на новые ценности — ценности постиндустриального общества. К таковым, как правило, относили экологическое, женское, антиядерное движение и движение за мир (но, не движение за права чернокожего населения в США, например, или антитоталитарные движения в странах Восточной Европы). Разработка данной парадигмы связана с именами таких социологов, как А. Турен (Франция), К. Оффе и Д. Рухт (Германия), Х. Кризи (Швейцария), А. Мелуччи и А. Пиззорно (Италия), М. Кастеллс (Испания) и ряда других. [6, с.110-111]

Теории, входящие в состав данной парадигмы, могут значительно отличаться друг от друга, однако, все они сходятся на признании отличия Новых общественных движений от старых по ценностям, специфическим формам коллективных действий, организационным структурам и социальному составу. Совокупность этих отличий сводится к понятию «новая идентичность». [6, с.113]

Рассмотрим некоторые из специфических характеристик Новых общественных движений, в изложении Е.А. Здравомысловой:

1. Ценности. Новые общественные движения ориентируются на проблемы экологии, территориальной общности, городской среды, условий жизни человека и выживания человечества, мира человека (его тело, сексуальность, здоровье) и пр. Подход к этим проблемам базируется на постиндустриальных ценностях, суть которых в установлении новых отношений человека с природой, с самим собой и с социальным окружением, отличных от утилитаризма, свойственного рыночному производству. Наиболее значимыми среди постиндустриальных ценностей являются индивидуальный стиль жизни (life-style), автономия (autonomy) и идентичность (identity). <…> Новые ценности преломляются в организационных структурах и коллективных действиях Новых движений.
2. Организационная структура движений представляет собой внутренний способ функционирования новой идентичности. Структурная децентрализация, самоуправление, горизонтальная сеть коммуникаций являются организационными коррелятами ценностей-целей, которые исповедуют Новые движения. <…>
3. Формы коллективных действий представляют внешний способ проявления идентичности. Они носят принципиально внеинституциональный характер. <…> [6, с.113-114]

Отметим важную для данного исследования роль идентичности в деятельности общественных движений этого типа – она выступает и в качестве смыслообразующей цели (участники общественного движения стремятся «привить» её максимальному количеству людей) и в качестве основного средства (деятельность движений и по форме и по содержанию представляет собой воплощение распространяемой идентичности). Для лучшего понимания сущности этого явления и его связи с участием в общественных движениях обратимся к концепции французского социолога А. Турена, в рамках которой сформулирован понятийный аппарат парадигмы.

Идентичность, по определению Турена - это «осознанное самоопределение социального субъекта». [цит. по: 6, с.116] Центральной характеристикой общественных движений является групповая идентичность. Её формирование происходит в результате анализа социального конфликта, по поводу которого образуется общественное движение. Групповая идентичность — это основанное на общих ценностях, мировоззрении и взаимодействии между участниками движения осознание позиции субъекта в социальной структуре общества, самооценка его роли в историческом процессе. Её основанием может быть как нравственное единение участников движения, так и, например, представление о традиционной общности (религиозной, классовой, территориальной, семейной и т.д.). Через идентичность раскрывается смысл и функции движений. [6, с.117]

Помимо идентичности, признаками, характеризующими Новые социальные движения, являются оппозиция (идентификация истинного противника движения) и тотальность (историческая значимость проблематики социального конфликта между движением и его оппозицией). [6, с.117]

Характерной особенностью изучаемых в данной работе социально ориентированных общественных движений является их нацеленность на социальные преобразования, связанные, в первую очередь, с мировоззрением людей и их ценностями. Даже в тех случаях, когда оппонентом движения является не анонимная социальная сила, а вполне конкретная социальная группа, обоснование необходимости борьбы с ней и усилия направленные на привлечение и мобилизацию сторонников лежат не в утилитарной, а в нравственно-идеологической сфере. При этом политические требования, при их наличии, не являются основой деятельности сообщества. В связи с этим, рассмотрение социально ориентированных общественных движений целесообразно производить через призму парадигмы Новых политических движений, а основной характеристикой участников, считать их идентичность.

Одной из современных теорий описывающих механизм формирования подобной идентичности в контексте социальных коммуникаций является концепт критической фазы британского исследователя Христиана Фухса. Согласно этой модели, образование общественного движения происходит в результате постепенного накопления членами общества наблюдений, фиксирующих наличие некоего негативного феномена (например ухудшения экологии). Когда число таких наблюдений (и связанных с ними коммуникативных актов), достигает некоторого критического уровня, происходит резкое возрастание остроты понимания проблемы, что, в свою очередь, служит толчком к возникновению коммуникативной системы протеста и возникновению общественного движения. До достижения критического уровня, проблема остаётся невидимой для общества, а отдельные высказывания по поводу наблюдаемого негативного феномена являются не протестной коммуникацией, а выражением личного недовольства. [10, с.9-10]

«Критическая фаза этой системы социального протеста, – пишет Фухс, – возникает тогда, когда социальные антагонизмы и проблемы начинают восприниматься как невыносимые, т.е. критическая масса людей недовольна структурами общества, а число оппонентов определенных структур возрастает до такой степени, что это недовольство и воля к изменениям теперь в принципе *воспринимаются*. Такая критическая фаза не является необходимым результатом углубления социального антагонизма (скажем, роста бедности, безработицы, ухудшения экологии), но *является результатом перцепции и осознания* углубления некоторого антагонизма». [цит. по 10, с.10]

Именно накопление достаточного числа наблюдений явлений, которые теперь интерпретируется в рамках протестной коммуникации как ненормальные, не желательные и опасные, внушает участникам коммуникации идею об усугублении ситуации и порождает соответствующие социо-психические установки страха и тревоги, являющиеся «собственными значениями» данной коммуникативной системы и способствующие её самовоспроизводству. Таким образом, возникает цикл самовоспроизводства (автопоэзиса) в ходе которого элементы системы порождают другие элементы системы на основе «собственных значений», уже не нуждаясь в каких-то внешних стимулах и триггерах (например, обсуждение проблем экологии порождает страх и тревогу по поводу её состояния, независимо от действительного наличия экологических проблем). [10, с.11]

Таким образом, в рамках коммуникативной системы генерируемой общественным движением создаются условия для поддержания и распространения ценностно-идеологических комплексов формирующих групповую идентичность участников данного общественного движения.

Такой взгляд на функционирование общественного движения наводит на вопрос сформулированный Р.Э. Бараш и А.Ю. Антоновским в монографии «Новые социальные движения в сетевую эпоху: статьи, интервью, экспертные заключения» [10]:

«Являются ли новые социальные движения новой коммуникативной системой, соразмерной традиционным коммуникативным макросистемам (таким как политика, экономика, религия, наука, искусство, семья, правовая система) и выполняют ли они в связи с этим некоторую особенную задачу-функцию, которую способны осуществить только они, и которая имеет значение не только для их собственного воспроизводства (автопоэзиса), но и поставляет свои достижения и продукты в распоряжение других систем?» [10, с.9]

Однако, в случае нашего исследования данный вопрос остаётся за скобками, поскольку формирование ценностно-идеологических комплексов может в равной мере рассматриваться и как внутренняя, и как внешняя цель коммуникативной деятельности общественного движения.

Рассмотрим современные подходы к изучению общественных движений на предмет выявления методов пригодных для осуществления сравнительного анализа их коммуникативной деятельности.

В первой половине 1990-х годов начинает развиваться новое направление исследований общественных движений, в фокусе внимания которого находится влияние на их деятельность телекоммуникационной сети Интернет. В результате того, что коммуникационные сервисы, предоставляемые посредством сети Интернет, получили повсеместное распространение и стали легкодоступны для широких масс населения, в возникшей виртуальной (цифровой) среде образовалось новое социальное пространство, в котором нашли своё отражение большинство явлений общественной жизни, социальные и культурные практики. Не стала исключением и деятельность общественных движений. Степень, в которой то или иное явление общественной жизни может быть перенесено в виртуальное пространство, определяется тем, насколько велика в нём доля информационного обмена. Например, личная переписка, ранее осуществлявшаяся посредством почтовых отправлений, переместилась в цифровую среду практически полностью, в то время как для деловой переписки, где помимо обмена информацией может иметь значение её фиксация на материальном носителе, услуги почты по-прежнему актуальны. Для оценки возможной степени переноса деятельности общественных движений в виртуальную среду, рассмотрим перечень наиболее важных их компонентов, выделенных польским исследователем П. Штомпкой:

1. Коллективность людей, действующих совместно.
2. Единство в отношении цели коллективных действий, а именно — изменения в обществе, причем цель должна восприниматься участниками однозначно.
3. Коллективность относительно диффузна, с низким уровнем формальной организации.
4. Действия имеют относительно высокую степень стихийности и не принимают институциализированные, застывшие формы. [17, с. 339]

Возможности, предоставляемые виртуальной средой, в разной мере способны обеспечить условия для реализации перечисленных компонентов. Так, при рассмотрении первого компонента, можно отметить, что виртуальное пространство снимает территориальные ограничения, сильно увеличивая возможность людей объединяться в коллектив, однако, при этом их совместная деятельность ограничивается коммуникативными актами. Единство и однозначность восприятия цели достигается при выработке групповой идентичности в процессе коммуникативного взаимодействия участников общественного движения, что в полной мере соответствует возможностям виртуальной среды. Диффузность коллектива требует наличия у его участников возможности свободно вступать во взаимодействие и свободно его прекращать, что, опять же, виртуальной средой в полной мере обесточивается. В свою очередь, диффузность, а также некоторая анонимность участников сообщества, способствуют поддержанию низкого уровня формальной организации. Последний компонент - высокая степень стихийности действий - легко реализуем в виртуальной среде, так как требует только возможностей для быстрого распространения информации и быстрой мобилизации участников движения. Из сказанного видно, что виртуальная среда создаёт благоприятные условия для реализации всех компонентов функционирования общественных движений, накладывая ограничения только на диапазон совместно выполняемых действий. Таким образом, можно сделать вывод о наличии значительного потенциала виртуализации деятельности общественных движений, а в случаях, когда их деятельность может быть сведена к осуществлению коммуникативных актов и индивидуальных действий его участников, нет никаких препятствий для полного её переноса в виртуальную среду.

По мере развития деятельности общественных движений в телекоммуникационной сети Интернет рос и интерес исследователей к этой теме. Произведённый П.В. Колозариди и А.А. Ульдановым анализ посвящённых ей публикаций в международных реферируемых журналах, включенных в базы ScienceDirect и Scopus за период с 1993 по 2014 год, позволил выделить три этапа, определяющих характер публикаций по данной тематике:

* 1. До 2001 года. Формирование и развитие исследовательских направлений рассматривающих Интернет как объект футурологических и специализированных технических исследований.
  2. 2001–2009 годы. Период роста междисциплинарных исследований, связанных с тематикой деятельности общественных движений в Интернете косвенным образом.
  3. 2009 – 2014 годы. Становление общей теоретической рамки для понимания феноменов связанных с деятельностью общественных движений в Интернете в разных областях знания. [9, с.110]

Помимо этого стоит отметить ряд важных для нас наблюдений, сделанных авторами в ходе проведённого исследования:

1. Выделено два подхода к пониманию роли Интернета. Часть исследователей рассматривает его в качестве медиа, что продолжает хорошо разработанную традицию изучения медиавоздействия на аудиторию. С другой стороны, он может рассматриваться в контексте социальных сетей, с применением и развитием ранее разработанных в этой сфере исследовательских подходов к анализу взаимодействия субъектов в виртуальном пространстве. [9, с.106]
2. Исследователи признают заметное влияние Интернета на развитие общественных движений, однако не склонны видеть в появлении виртуальной среды веских оснований для кардинальных изменений в этой сфере. В целом Интернет трактуется как положительный фактор влияния на развитие общественных движений и организаций. Кроме того, воздействие Интернета может трансформировать и сами подходы к изучению общественных движений. [9, с.106-107]
3. После 2005 года происходит смена доминант — от «социальных движений, использующих интернет» к «социальным движениям в интернете». [9, с.118]
4. Так же с 2005 года начинают формулироваться теоретические подходы и понятия, принимающие во внимание, например, уход в виртуальное пространство большой части протестной политической активности, а также факт, что именно в Интернете формируется дискурс и стратегии, которые затем реализуются физически. [9, с.118]
5. Одним из факторов роста внимания к проблемам социальных движений в интернете стало применение онлайн-технологий и технологий computer science в исследовательской деятельности (например анализ социальных сетей, анализ «больших данных»). [9, с.117]

Анализ социальных сетей, как метод исследования любого рода сообществ, опирается на идею представления социальной структуры как графа, узлами которого являются акторы, а рёбрам – связи между ними. Этот подход начал применятся в антропологии и прикладной экономике в 1940-х годах, а в 1954 году был сформулирован как исследовательский инструмент социологом Манчестерской школы Джеймсом Брансом. [15, с.51] В своей работе «Классы и собрания в норвежском островном приходе» он впервые сформулировал понятие социальной сети, охарактеризовав её следующим образом:

«Каждый человек имеет определенный круг друзей, а эти друзья — собственных друзей. Некоторые из друзей одного человека знают друг друга, другие — нет. Я нашел удобным говорить о такого рода полях как о сетях. Под этим мне видится система точек, которые соединены между собой. Точками этой системы являются люди, а линии соединения этих точек указывают, как они взаимодействуют друг с другом». [цит. по: 4, с.159 ]

Концепция социальной сети получила развитие в рамках Манчестерской школы, где на её основе был разработан ряд исследовательских инструментов, в совокупности получивших название антропологический сетевой анализ. Однако к началу 1970-х годов интерес к данному подходу в социальных науках утих, по причинам от части сводившимся к тому, что при необходимости сбора большого количества эмпирических данных и использования сложного математического аппарата, получаемые результаты оказывались малоинтересны, в силу своей тривиальности. Новую жизнь сетевой подход обрёл в начале 1990-х годов, в связи с развитием телекоммуникационной сети Интернет, благодаря которой взаимодействие сетевых акторов получило своё наглядное отображение, свободное от территориальных, временных и социокультурных обстоятельств. Начиная с этого времени, данный инструмент занял прочные позиции в сфере социальных и политологических исследований, и позволил изучать и прогнозировать процессы с использованием неучитывавшихся ранее факторов и особенностей взаимодействия. [15, с.52-53] Наработанный опыт применения сетевого подхода позволяет охарактеризовать его, как достаточно надёжный аналитический инструмент для исследования виртуальных сообществ и использовать для выявления значимых характеристик коммуникационных практик, осуществляемых общественными движениями в виртуальном пространстве.

Применительно к цели данного исследования произведённый теоретический обзор позволяет сделать следующие выводы:

1. В условиях современного информационного общества, деятельность общественных движений в значительной мере осуществляется посредством телекоммуникационной сети Интернет, что делает актуальной задачу изучения представляющих общественные движения виртуальных сообществ и их коммуникативных практик.
2. Социально-ориентированные общественные движения целесообразно рассматривать в рамках исследовательской парадигмы Новых общественных движений.
3. Одной из целей коммуникативной деятельности социально ориентированных общественных движений неизбежно является их самовоспроизводство путём поддержания и распространения групповой идентичности. Наличие иных (внешних по отношению к общественному движению) целей не является обязательным и не принципиально для данного исследования.
4. Ключевой характеристикой участников коммуникативной деятельности осуществляемой общественным движением является их идентичность.
5. При анализе коммуникативных практик, осуществляемых общественными движениями, необходимо рассматривать не только свойства участников коммуникативной деятельности, но и связи между ними, что обуславливает применение сетевого подхода.

**1.2. Идентичности членов общественных движений как система ценностно-нормативных комплексов: коммуникативный подход**

Идентичность, обозначенная в предыдущем разделе как ключевая характеристика участников коммуникативной деятельности, является междисциплинарным понятием с богатой традицией теоретического осмысления. В большой вклад в изучение этого явления внесли Э. Эриксон, Жак Лакан, Д. Винникот, Дж. Марсия, П. Бергер и Т. Лукман, Х. Кохут и Г. Салливан, Ч. Кули, Дж. Г. Мид, Г. Брейкуэлл, М. Хайдег­гер и другие.[13]

В силу своей мультипарадигмальности и большого количества разработанных теоретических моделей, понятие «идентичность» не имеет общепринятого толкования. Ранее, описывая отличительные черты участников Новых общественных движений, мы уже упоминали определение А. Турена («осознанное самоопределение социального субъекта» [цит. по: 6, с.116]), здесь же нам полезно будет обратиться к определению, сформулированному В. Римским специально для анализа воздействия использования сети Интернет на социальную активность и идентичность:

«Идентичность – это состояние сознания индивида, в котором он по совокупностям тех или иных своих личностных характеристик узнает себя, признает устойчивость собственной личности, отделяет себя от окружающей его действительности, определяет свою принадлежность к той или иной социальной группе и, наоборот, признает невозможность принадлежности к другим социальным группам. Для индивидов идентичности – это источники значений, ценностей и основания для совершения ими социальных действий. Идентичность определяется многими совокупными характеристиками личности, имеющими биологическую, психологическую, социальную и культурную природу.» [14, c.86]

Римский также указывает на то, что идентичность индивида, с одной стороны позволяет ему сохранять стабильность личности на протяжении длительного времени, но с другой постоянно меняется, обеспечивая адаптацию к меняющимся социальным условиям. [14, c.86]

Для данного исследования в этом определении особенно важны два пункта:

1. Идентичность является критерием принадлежности к той или иной социальной группе (что даёт нам возможность измерять уровень идентификации).
2. Идентичность является источником значений, ценностей и оснований для совершения социальных действий (что даёт нам возможность рассматривать систему ценностно-идеологических комплексов индивидуума, как элемент его идентичности).

Общим местом в понимании феномена идентичности у большинства исследователей, независимо от их парадигмального предпочтения, является указание на то, что необходимым и определяющим условием для формирования и развития идентичности является участие индивида в социальной коммуникации.

Н.В. Писаренко демонстрирует это в своей статье «Идентичность в аспекте коммуникации: теоретические подходы» [13] на примере анализа более чем 20 теоретических концепций разработанных ведущими западными и российскими исследователями.

«... исследуя феномен идентичности, – Пишет он – подчеркнём, что это не свойство личности, а её отношение. «Соотносительность» идентичности с коммуникативными объектами определяется как её важнейшее феноменологическое свойство. Так как идентичность является продуктом коммуникативных процессов, то исследовать рассматриваемый феномен необходимо с точки зрения всего многомерного пространства идентификаций личности. Феномен идентичности не существует без «Другого», имеет коммуникативный характер, а сам термин «идентичность» нуждается в уточнении посредством категории «коммуникативная». [13, c.121]

Рассматривая вопрос о взаимосвязи индивидуальной идентичности и социальной структуры, П. Бергер и Т. Лукман указывают на то, что идентичность, являясь ключевым элементом субъективной реальности, находится в диалектической связи с обществом: она формируется социальными процессами и, в свою очередь, формирует их. В связи с этим они, также, рекомендуют не использовать понятие «коллективная идентичность», по причине его избыточности. [1, с.279-280] Однако данное понятие (или аналогичное ему по смыслу) часто встречается в работах других исследователей, в связи, с чем полностью отказаться от него мы не можем.

Наличие данной взаимозависимости, выражающееся в свойстве идентичности адаптироваться к меняющимся социальным (в данном случае можно также сказать «коммуникативным») условиям позволяет рассчитывать на то, что фиксируя изменения идентичности можно, до некоторой степени, судить о процессах, происходящих в обществе (в коммуницирующей группе). Применительно к задачам данного исследования это означает возможность судить о свойствах социальной структуры сообществ, образуемых участниками общественных движений, на основании особенностей индивидуальных идентичностей их участников. На основании полученной таким образом оценки свойств сообщества вряд ли можно сделать содержательные выводы об этих свойствах, однако, этого, вероятно, должно быть достаточно для проведения сравнительного анализа.

В виду того, что рассматриваемыми в данном исследовании свойствами общественных движений являются особенности коммуникативных практик, осуществляемых ими с целью формирования ценностно-идеологических комплексов, а ценностно-идеологические комплексы, в соответствии с данным выше определением, являются одним из основных компонентов идентичности, задачу данного исследования можно сформулировать как выявление особенностей процесса идентификации участников общественного движения. Идентификация, это процесс формирования у индивида, или социальной группы, определённой идентичности. Она происходит в результате коммуникативного взаимодействия, которое подразумевает не только обмен информацией, но и обмен смыслами. В результате в коммуницирующей группе выстраивается структура социальных отношений, а также общая система знаний и оценок по отношению к обсуждаемым вопросам. Членство в такой социальной группе подразумевает доверие, лояльность и солидарность с другими её участниками. [18, с.196] Идентичность также включает в себя все социальные роли, которые индивидууму приходится выполнять.[14, с.87]

Мы предполагаем, что участники коммуникативной деятельности, осуществляемой общественным движением, находятся на разных стадиях идентификации. Более того, часть из них может участвовать в коммуникации с позиции противников данного общественного движения. Поскольку формирование идентичности происходит именно посредством коммуникативного воздействия, коммуникативная деятельность группы направленная на любого отдельного её участника может рассматриваться как воздействие, формирующее его идентичность. Таким образом, выявление особенностей процесса идентификации участников общественного движения состоит в выявлении характерных для данного сообщества способов коммуникативного воздействия на его участников, в зависимости от степени их идентификации. При этом анализ особенностей коммуникативного воздействия можно осуществить непосредственно на основании доступных нам эмпирических данных. Однако для решения стоящей перед нами задачи необходимо также определить и степень идентификации каждого участника коммуникативной деятельности с исследуемым сообществом, при том, что единственное, что нам известно о участниках этой деятельности, это следы их коммуникативного поведения.

Рассмотрим вопрос о взаимосвязи между идентичностью и коммуникативным поведением. Данный вопрос подробно изучен исследователями в сфере социолингвистики. Так один из ведущих российских исследователей в этой сфере В.И. Карасик описал концепцию взаимосвязи идентичности и коммуникативного поведения в монографии «Языковой круг. Личность, концепты, дискурс» (2002 г.) [8]. Суть предложенной концепции сводится к следующим семи положениям [8, с.5]:

1. Человек осознает свою идентичность в рамках своей принадлежности этносу и исчисляемой совокупности социальных групп и в границах своей уникальной личности.
2. Такое осознание фиксируется в языковом сознании и коммуникативном поведении и может быть объективировано при помощи используемых в лингвистике специальных исследовательских процедур.
3. Языковое сознание членится на релевантные фрагменты осмысления действительности, которые имеют вербальное выражение и допускают этнокультурное, социокультурное и личностно-культурное измерения.
4. Коммуникативное поведение выражается в текстах, возникающих в ситуациях общения и характеризующих участников общения как принадлежащих этнокультурной и социокультурной общности и как индивидуумов.
5. Типизируемые ситуации общения соотносятся с фрагментами осмысливаемой действительности, при этом выделение и типизация коммуникативных ситуаций и фрагментов мира зависят от степени их значимости для индивидуума, социальной группы и этноса, т.е. в основе значимого выделения тех или иных признаков лежат ценностные приоритеты.
6. Можно выделить а) типы языковых личностей, для которых ведущим моментом будет осознание и переживание своей этнокультурной либо социокультурной, либо индивидуально-культурной идентичности; б) типы концептов, определяющих этнокультуру, социокультуру либо индивидуальную личность; в) типы дискурса, возникающего для поддержания и развития этнокультуры, социокультуры и индивидуально-личностной культуры человека.
7. Языковая личность едина в ее различных проявлениях и аспектах изучения: изучая личность, мы должны прийти к специфическим для этой личности концептам и типам дискурса; моделируя концепты, мы выявляем характеристики типизируемых личностей и типов дискурса; выделяя типы дискурса, мы с иных позиций устанавливаем характеристики личностей и определяем организующие тот или иной дискурс концепты.

Также автор описывает возможность типологизации «языковых личностей» по критериям характерным для различных исследовательских парадигм. В частности, он указывает на два подхода к изучению языковой личности в рамках социолингвистики:

1. Выявление речевых признаков (построение речевого портрета) заранее определённого социального типа личности.
2. Определение статуса и роли участника коммуникации на основании известных речевых признаков. [8, с.10]

Для нашего исследования актуальной является вторая задача. При этом под статусом и ролью участника коммуникации подразумевается его представление о собственном статусе и роли в рамках данного сообщества, либо же его идентичность в целом.

Теперь остаётся решить вопрос о выборе речевых признаков, которые могли бы служить критериями для определения степени идентификации конкретного участника коммуникативной деятельности с рассматриваемым сообществом. Подойти к решению этого вопроса можно прибегнув к понятию «коллективная языковая личность», описанному, например, в учебном пособии И. Н. Горелова и К. Ф. Седова «Основы психолингвистики» (2001 г.) [7]

«Претендуя на языковую исключительность, люди в сходных коммуникативных ситуациях часто ведут себя поразительно одинаково. Это связано и с тем, что говорящие в каждый момент своей речевой биографии демонстрируют особенности группового речевого поведения. Носитель языка как бы фокусирует в себе черты «коллективных языковых личностей». Он может, к примеру, одновременно выступить как языковая личность горожанина, языковая личность студента-словесника, языковая личность двадцатилетнего юноши и т. п» [7, с. 119-120]

Таким образом, определение степени идентификации индивида с социальной группой можно осуществить путём сравнения речевого портрета языковой личности данного индивида с речевым портретом коллективной языковой личности данного сообщества. Одним из основоположников метода речевого портретирования в нашей стране была Т.М. Николаева, разрабатывавшая эту методику в конце 1980-х – начале 1990-х годов. В статье «Социолингвистический портрет и методы его описания» (1991 г.) она вводит понятие «диагносцирующие пятна»:

«... многие языковые парадигмы, начиная от фонетической и кончая словообразовательной, оказываются вполне соответствующими общенормативным параметрам и поэтому интереса не представляют. Напротив, важно фиксировать яркие диагносцирующие пятна» [цит. по 2, с.3]

Речь здесь идёт о том, что при выявлении значимых элементов лексики (например, ключевых слов) решающее значение имеют не характеристики данного элемента в данном тексте (например, частота употребления какого либо слова), а отклонение этих характеристик от нормальных значений для языка вообще. Данное требование к выбору ключевых слов было сформулировано Ю.Н. Карауловым следующим образом: «невысокочастотные в языке, но частотные в тексте» [цит. по 16, стр. 67]

В случае с рассматриваемой нами коммуникативной деятельностью, осуществляемой представителями общественных движений в виртуальном пространстве, диагносцирующими пятнами речевого портрета их коллективной языковой личности могут быть не только аномалии в частоте употребления отдельных слов. Доступные нам эмпирические данные позволяют произвести анализ и других особенностей коммуникативного поведения членов исследуемых сообществ, что будет подробно рассмотрено в четвёртом параграфе этой главы. Здесь же нам важно зафиксировать общий принцип, состоящий в том, что стоящая перед нами задача по выявлению особенностей осуществляемых общественными движениями коммуникативных практик, направленных на формирование ценностно-идеологических комплексов, может быть решена путём сопоставления уровня идентификации каждого участника коммуникативной деятельности с направленным на него коммуникативным воздействием. При этом в уровень идентификации члена сообщества может быть оценен на основании речевых признаков, для выявления которых требуется составить коммуникативный портрет коллективной языковой личности участников данного сообщества. Последний же, в свою очередь, составляется путём выявления таких особенностей коммуникативной деятельности членов сообщества, которые характерны только для данного сообщества и не характерны для всех других.

В следующем разделе мы рассмотрим вопросы, связанные с техническими возможностями реализации процедур описанных в этом и предыдущем разделах.

По данному разделу можно сделать следующие выводы:

1. Ценностно-идеологические комплексы распространяемые участниками общественных движений являются одним из важнейших элементов их идентичности, как членов соответствующих сообществ.
2. Формирование и корректировка идентичности происходит в результате участия индивида в коммуникативной деятельности.
3. Идентичность оказывает влияние на коммуникативную деятельность, вследствие чего особенности коммуникативной деятельности могут рассматриваться как индикаторы той или иной идентичности.

**1.3. Современные методы исследования коммуникативных практик общественных движений в цифровой среде**

В предыдущих разделах был сформулирован ряд задач, для решения которых требуется применение методов сбора, обработки и анализа цифровых данных (в частности текстов написанных на естественном языке) размещённых в телекоммуникационной сети Интернет. Рассмотрим применяемые в социологии методологические подходы к решению подобных задач.

Безопросный сбор данных можно определить, как комплекс исследовательских процедур, направленных на выявление общественного мнения (или фактов поведения), по какому либо вопросу путём анализа так называемых «цифровых следов», т.е. размещённых в интернете данных о действиях пользователей, которые могут служить индикаторами их мнения по поводу исследуемого вопроса. Такими следами могут быть комментарии, репосты, лайки, просмотры тех или иных материалов и т.д. Работу с такого рода данными сопровождает ряд сложностей. О.Ю. Кольцова и К.А. Маслинский в статье «Выявление тематической структуры российской блогосферы: автоматические методы анализа текстов» [11], перечисляют следующий ряд проблем:

1. Проблемы с доступом к генеральной совокупности исходных данных (например, из-за политики ресурса).
2. Проблема операционализации социологических понятий в терминах интернет-данных (например, определение границ объекта или выбор единиц анализа).
3. Потребность в аналитических методах, дающих возможность обрабатывать большие выборки данных. [11, с.114]

Этот список можно дополнить проблемами, обозначенными в статьях Д.И. Юдиной и В.И. Дудиной «Семантическая сеть на биграммах как метод валидизации результатов тематического моделирования в социологическом исследовании» [19] и «Извлекая мнения из сети интернет: могут ли методы анализа текстов заменить опросы общественного мнения» [5]:

1. Неструктурированность данных. [19, с.72]
2. Ограниченная возможность использования качественных методов, предполагающих по большей части «ручную» интерпретативную работу исследователя. [19, с.72]
3. Отсутствие методологических работ позволяющих сравнить данные полученные в результате анализа цифровых следов с репрезентативными результатами хорошо разработанных традиционных опросных методов. [5, с.65-66]
4. Диффицит инструментов для анализа текста написанного на других языках, кроме английского. [5, с.66]

Тем не менее, перспективы развития данной исследовательской области привлекают к ней большой интерес социологов и заставляют искать такие способы работы, которые бы позволили получать значимые социологические результаты. [19, с.72] Кроме того, авторы отмечают, что несмотря на малое количество работ посвящённых сравнению результатов анализа мнений из социальных медиа с результатами социологических опросов, сама методология извлечения мнений из неструктурированных текстов довольно разнообразна. [5, с.65]

Рассмотрим методы анализа текстов размещённых пользователями в телекоммуникационной сети Интернет, которые были применены в исследованиях, описанных в трёх упомянутых выше статьях.

Две статьи Д.И. Юдиной и В.И. Дудиной [5, 22] посвящены анализу комментариев оставленных пользователями интернет-портала YouTube к фильму Фонда борьбы с коррупцией (ФБК) «Чайка». Целью описываемых исследований является сопоставление результатов построения тематической модели анализируемых текстов, с результатами опроса [25] проведённого Левада-Центром по поводу данного фильма [5] и с результатами анализа семантической сети построенной на том же эмпирическом материале[19].

Описывая доступные методы анализа текстов, авторы выделяют два основных подхода к решению задачи извлечения мнений: машинное обучение (тематическое моделирование и различные методы обучения с учителем) и лингвистические методы (на основе словарей и лингвистических анализаторов). Отмечено также существование гибридных методов использующих оба подхода. [5, с.65-66] Проанализировав различные методы анализа текстов на предмет их эффективности, доступности и способности работать с русскоязычными текстами, авторы останавливаются на следующем алгоритме работы:

1. Препроцессинг данных (предварительная обработка), включающий в себя: удаление нетекстовых знаков (включая цифры) и слов длинной менее трёх букв; удаление стоп-слов по словарю и лемматизация слов при помощи библиотеки «Pymorphy2» [22]. Кроме того была выполнена дополнительная обработка для тематического моделирования: удаление низкочастотные слов (с частотой менее пяти); создание матрицы «термин-документ». Также была выполнена дополнительная обработка для сетевого анализа: удаление высокочастотных, но малоинформативных слов (метод на указан) построение списка биграмм с окном (расстоянием в предложении между первым и последним словами включительно) равным 4; ранжирование биграмм на основе оценки их информативности полученной с применением метода отношений правдоподобия; отбор первых 3000 наиболее информативных биграмм; построение матрицы «биграмма-документ». [19, с.74-76]
2. Тематическое моделирование.

Тематическое моделирование производилось с использованием библиотеки “topicmodels” языка R [21]. В качестве метода моделирования был выбран алгоритм на основе латентного размещения Дирихле (LDA) [20] с автоматически высчитываемыми начальными значениями параметров α и β и выборкой Гиббса в качестве алгоритма оптимизации параметров модели. Для нахождения оптимального числа тем использовался показатель внутренней валидности — среднее гармоническое правдоподобие. [19, с. 76]

1. Семантическая модель.

Семантическая модель представляет собой взвешенный ненаправленный граф, построенный на основе матрицы «биграмма-документ» при помощи библиотеки «igraph» языка R [27]. Вершинами графа являются биграммы, а рёбрами – частота их совместного употребления в комментариях. [19, с. 78]

Анализируя полученные результаты, авторы делают следующие выводы относительно эффективности работы предложенных алгоритмов:

1. Тематическое моделирование показывает хорошие результаты при работе с большими объёмами неструктурированных текстовых данных, даже при поверхностном препроцессинге и в отсутствие мощных лингвистических анализаторов.
2. Недостатком тематического моделирования является низкая чувствительность к редко встречающимся темам.
3. Использование сетевого анализа биграмм позволяет компенсировать пропуск тем при тематическом моделирование.
4. Использование сетевого анализа позволяет получить информативную визуализацию структуры дискуссии и более быструю и простую интерпретацию самих тем благодаря устойчивым выражениям. [19, с. 81]

Рассмотрим теперь статью О.Ю. Кольцовой и К.А. Маслинского «Выявление тематической структуры российской блогосферы: автоматические методы анализа текстов» [11].

На материале текстов размещённых на блог-платформе «Живой журнал» авторы стремятся выявить методы разделения текстов на тематические группы, дающие наилучшее качество при решении социологических задач.

Первым методологическим вопросом, на который они обращают своё внимание, являются способы оценки качества работы различных алгоритмов. Этот вопрос также важен для выбора оптимального количества групп, на которое следует разделять коллекцию текстов. Авторы выделяют две группы таких методов: внешние (определение доли «правильно» отнесенных единиц через сравнение с образцом) и внутренние (вычисление ряда параметров, например соотношения внутрикластерной и межкластерной дисперсии). [11, с.120-121].

Недостатком внешних методов оценки является сложность (а порой и не реализуемость) составления образцового корпуса (текстов распределённых на группы вручную). Внутренние методы оценки основаны на формальных допущениях. Кроме того, они ориентированы на специфику конкретного алгоритма и могут быть неприменимы для сравнения алгоритмов между собой. В целом, авторы отмечают, что у этого вопроса нет однозначного решения, а надёжность существующих методов оценки до конца не установлена. Для оценки качества кластеризации авторами была выбрана функция зависимости значения внутрикластерной дисперсии от числа кластеров, а для тематического моделирования – перплексивность (perplexity – параметр рассчитываемый некоторым пакетами тематического моделирования) [Там же]

Вторая рассмотренная авторами проблема анализа текстов, это автоматическое именование («лейбелинг») сформированным алгоритмом групп текстов:

«Само по себе получение списка из нескольких сотен групп, в каждой из которых по несколько тысяч текстов, ничего не прибавляет к знанию исследователя о коллекции текстов и о её тематике, даже когда алгоритм работает качественно и быстро. Если для определения тематики каждой группы требуется вручную перечитать все тексты, автоматизированный анализ обесценивается». [11, с.122]

У этой проблемы также есть несколько решений, заключающихся в том, что алгоритм выдаёт различного рода подсказки, позволяющие исследователю судить о содержании выделенных кластеров (например, ключевые слова, или центроид группы), однако, как замечают авторы, качество этих решений остаётся на невысоком уровне. [11, с.122]

Переходя к задаче объединения в группы текстов, сходных по тематике, авторы рассматривают два подхода: классификация (распределение по заранее обозначенным группам путём сравнения с образцовыми текстами) и кластеризация (формальное вычисление сходства). Причём подробно анализировать алгоритмы классификации авторы считают нецелесообразным, в виду того, что, в рамках проводимого ими исследования, большую ценность имеет возможность выявлять скрытые (не распознаваемые человеком) категории. [11, с.122-123] В связи с этим, всё внимание авторов сосредотачивается на алгоритмах кластеризации, которые, в свою очередь, подразделяются на кластерный анализ и факторный анализ (латентно-семантический анализ и тематическое моделирование). Для выбора конкретных алгоритмов авторы формулируют ряд требований: встроенная оценка качества, способность работать с большим объёмом данных на кириллице, встроенные функции закачки и препроцесинга. Однако ни одного программного продукта удовлетворяющего этим требованиям им найти не удалось. Единственная библиотека, реализующая алгоритмы кластерного анализа и способная работать с большими объёмами данных, которую им удалось выявить – gCLUTO [29]. В части факторного анализа ситуация оказалась значительно лучше и сравнив особенности таких подходов к тематическому моделированию, как латентно-семантический анализ (LSA), вероятностный латентно-семантический анализ (pLSA) и латентное размещение Дирихле (LDA), авторы остановились на алгоритме LDA и реализующей его библиотеке Stanford Topic Modeling Toolbox (TMT) [34]. Далее ими была предпринята попытка сравнить эффективность gCLUTO и LDA, однако при проведении экспериментов с кластеризацией большого количества текстов выяснилось, что предусмотренные в библиотеке gCLUTO методы лейблинга тем не позволяют эффективно осуществить эту процедуру. От дальнейшего использования данной библиотеке пришлось отказаться, остановив свой выбор на алгоритме LDA из библиотеки TMT. [11, с.124-128]

По итогам проведённого исследования авторы делают следующие выводы относительно рассмотренных алгоритмов:

1. Тематическое моделирование имеет следующие преимущества над кластерным анализом: в явном виде формируются относящиеся к темам списки слов (что упрощает задачу лейблинга) а для каждого текста определяется соотношение тем по важности.
2. По отношению к кластерному анализу лучше развиты внешние меры качества.
3. Оба подхода могут с успехом применятся в исследовательской деятельности, но тематическое моделирование предпочтительнее для большей части задач. [11, с.136-137]

В нашем исследовании задача интерпретации выявленных тем не стоит, но важно иметь возможность определить предпочтения участников исследуемых сетевых сообществ в их отношении. При наличии относящихся к темам списков слов эта задача решается наиболее простым способом. Таким образом, на основании двух рассмотренных исследований и стоящих перед нами задач, можно заключить, что тематическое моделирование методом латентного размещения Дирихле является наиболее предпочтительным способом выявления тем, обсуждаемых в исследуемых сообществах.

Метод LDA относится к классу генеративных вероятностных моделей. Принцип действия этого алгоритма состоит в том, что в корпусе текстов предполагается существование заданного количества скрытых тем к каждой из которых каждый текст корпуса принадлежит с некоторой вероятностью (сумма вероятностей для каждого текста равна 1). Эквивалентной интерпретацией будет представление каждого текста, как смеси тем, каждая из которых занимает в нём определённую долю. Также, как смесь тем, рассматриваются и слова, каждое из которых принадлежит каждой теме с разной вероятностью. Лежащий в основе алгоритма математический аппарат позволяет распределить вероятности для слов так, чтобы максимизировать принадлежность каждого текста минимальному количеству наиболее выраженных в нём тем. [11, с.127]

Рассмотрим теперь возможности сетевого анализа при исследовании виртуальных сообществ.

В качестве примера применения сетевого анализа к виртуальным сообществам рассмотрим статью Ю. Г. Рыкова, П. А. Мейлахса и Я. Е. Синявской посвящённую исследованию сетевой структуры виртуального сообщества СПИД-диссидентов в социальной сети «ВКонтакте».[23] В ходе исследования авторами были выявлены механизмы распространения убеждений соответствующих теме сообщества от наиболее активных их носителей («ядра сообщества») к представителям группы риска. Анализируемый авторами механизм распространения убеждений соответствует ранее описанному нами механизму формирования идентичности участников общественных движений, что делает данную работу особенно ценной для нашего исследования.

На основании произведённого авторами рассматриваемого исследования литературного обзора статей на тему влияния сетевой структуры сообщества на особенности распространения убеждений среди его участников [23, с. 4-6] можно сделать следующие, важные для нашего исследования, выводы:

1. Сети дружбы и общения в рамках виртуальных сообществ, как правило, соответствуют структуре «ядро-периферия».
2. Индивидуальные показатели интегрированности в социальную сеть (например, мера центральности) положительно коррелируют с показателями приверженности убеждениям распространяемым в сообществе (например, с вероятностью успешного отказа от курения в сообществе бросающих курить).
3. Есть тенденция к образованию кластеров из участников сообщества со схожими характеристиками по отношению к теме сообщества.
4. Лидерами и моральными авторитетами и наставниками для большей части участников сообщества может являться небольшое количество наиболее опытных (и активных) его участников.
5. В группах взаимопомощи частота контактов между участниками положительно коррелирует с количеством высказываний слов поддержки.
6. Поведенческие явления (например, эмоции по отношению к чему либо, или практики потребления) могут передаваться от одного пользователя другому посредством личного или опосредованного общения. Этот эффект называют «социальное заражение».

Выявленные закономерности позволяют сделать вывод о эффективности сетевого анализа виртуальных сообществ для прогнозирования происходящих в данных сообществах социальных процессов. Однако следует заметить, что приведённые закономерности выявлены в отдельных исследованиях по отношению к отдельным сообществам и возможно являются частными случаями, нехарактерными для подобных сообществ в целом.

Непосредственно само описанное в статье исследование сообщесва СПИД-диссидентов осуществлялось по следующей методике [23, с. 7-10]:

1. Сбор данных обо всех записях со стены сообщества и комментариях к ним, анкетных данных всех активных пользователей, отношениях «дружбы» между пользователями, и о коммуникативных взаимодействиях между пользователями на стене сообщества.
2. Построение сетей дружбы и коммуникационных взаимодействий.
3. Качественный контент-анализ записей на стене и комментариев, с целью выявления таких атрибутов пользователей, как ВИЧ-статус и отношение к убеждениям СПИД-диссидентства.
4. Анализ распределения атрибутов пользователей по сетям дружбы и коммуникационных взаимодействий.

Описанная методика позволила авторам исследования описать структуру сообщества, в частности выявить ядро сообщества, состоящее из наиболее убеждённых пользователей и группу риска, состоящую из сомневающихся пользователей активно взаимодействующих с ядром.

В следующем разделе мы опишем пути решения стоящих перед нами исследовательских задач, с применением рассмотренных здесь методов.

По данному разделу можно сделать следующие выводы:

1. Методы социологического исследования «цифровых следов» активно развиваются в рамках цифровой социологии.
2. В сфере цифровой социологии существует ряд методологических проблем не имеющих однозначного решения в силу новизны этого исследовательского направления.
3. Оптимальным решением для выявления тематической структуры из набора неструктурированных текстов является алгоритм тематического моделирования на основе латентного распределения Дирихле.
4. Сетевой анализ позволяет получать значимые социологические результаты при исследовании коммуникативной деятельности виртуальных сообществ.
5. Механизм идентификации с сообществом новых его участников может быть описан в рамках концепции социального заражения.

**ГЛАВА 2. СРАВНИТЕЛЬНЫЙ АНАЛИЗ КОММУНИКАТИВНЫХ ПРАКТИК СОЦИАЛЬНО-ОРИЕНТИРОВАННЫХ ОБЩЕСТВЕННЫХ ДВИЖЕНИЙ**

**2.1. Программа исследования коммуникативных практик социально-ориентированных общественных движений. Сбор цифровых следов**

В данном разделе, на основе сформулированных в предыдущей главе задач и подходов к их решению, мы опишем программу исследования коммуникативных практик социально ориентированных общественных движений и сбор эмпирического материала, выполненный в соответствии с этой программой.

Одним из наиболее массовых и популярных в России сервисов предоставляющих возможности для личного общения, формирования сообществ по интересам и организации совместных действий, является интернет-платформа «ВКонтакте». Всё разнообразие возможностей по организации и регулированию коммуникативной деятельности, предоставляемых данной интернет-платформой сгруппировано вокруг двух функциональных категорий: личные страницы и сообщества.

Личные страницы являются своего рода проекцией конкретных людей в виртуальное пространство образованное интернет-платформой «ВКонтакте». Доступ к их использованию имеет, как правило, только один человек (далее – «пользователь» или «владелец страницы»). Функциональные возможности личных страниц сводятся к размещению информации предназначенной для идентификации и самопрезентации пользователя, публикации от имени, которое приписано личной странице, частных или публичных сообщений, а также использованию функций сайта предназначенных для организации и регулирования коммуникативной деятельности (в частности функций по созданию и управлению сообществами).

Сообщества, это разделы сайта «ВКонтакте», создаваемые и управляемые пользователями от имени их личных страниц, функционал которых позволяет организовывать и регулировать коллективную коммуникативную деятельность.

Одним из распространённых сценариев использования функционала сайта «ВКонтакте», является создание сообществ, предназначенных для информирования общения и координации совместной деятельности участников социально ориентированных общественных движений. В таких сообществах, как правило, размешена информация описывающая идеологию и цели соответствующих общественных движений, представлены коллекции мультимедийных файлов, или текстовых документов, относящихся к их деятельности, а также организовано обсуждение пользователями тем, относящихся к сфере интересов соответствующих общественных движений.

Функционал сайта предоставляет три организационные формы коллективной коммуникации в рамках сообщества:

1. Обсуждение отдельных тем в специально выделенном для этого разделе.
2. Обсуждение новостей опубликованных в виде записей на стене сообщества (записи опубликованные на стене сообществ отображаются в ленте пользователей подписавшихся на данные сообщества).
3. Комментирование фотографий и видеозаписей размещённых в соответствующих коллекциях сообществ.

Руководители сообществ могут регулировать коммуникативную деятельность пользователей путём установления правил ограничивающих возможность просматривать или оставлять сообщения в каждом из описанных разделов для определённых групп, либо же для всех пользователей. Также руководители сообществ могут устанавливать запрет на публикацию или на просмотр материалов сообщества для отдельных пользователей (например, в качестве наказания за нарушение правил). Кроме того, у организаторов есть возможность удалять отдельные сообщения и настроить автоматическое удаление сообщений содержащих определённые слова (например, ненормативную лексику).

В разных сообществах реализуются разные сценарии использования описанного функционала. Кроме того, поскольку состав участников коммуникативной деятельности, осуществляемой в рамках подобных сообществ, и её цели, относительно постоянны во времени, можно ожидать, что в каждом из них сложилась определённая практика общения, характеризующаяся структурно-функциональными и речевыми особенностями.

В данном исследовании мы сосредоточим своё внимание на общении, происходящем в комментариях к записям, публикуемым на стене сообщества. Этот выбор обусловлен тем, что данное коммуникационное пространство, наиболее доступным для пользователей и активно привлекает к себе их внимание. Записи, публикуемые на стене, появляются в информационной ленте пользователей. Таким образом, их могут увидеть даже те участники сообщества, которые редко посещают его страницу. Вместе с тем, стоит ответить, что в связи с высокой динамикой публикации таких записей, обсуждение в комментариях к ним, как правило, не продолжается длительное время. Активная его фаза занимает от нескольких часов, до недели, в то время как дискуссии в обсуждениях иногда могут длиться годами.

Важно заметить, что, поскольку, круг лиц активно или пассивно участвующих в коммуникативной деятельности, осуществляемой в рамках подобных сообществ, может быть очень неоднороден и, помимо непосредственных членов соответствующих общественных движений, часто включает в себя людей, которые недавно заинтересовались идеями этих движений, случайно увидели публикации со стены сообщества, либо же являются противниками данных общественных движений, к функциям осуществляемой в них коммуникативной деятельности добавляется функция распространения ценностно-идеологических комплексов, исповедуемых данными общественными движениями. Исходя из этого, можно предположить, что практика общения, сложившаяся в подобных сообществах, находится в тесной взаимосвязи с содержанием исповедуемой участниками общественного движения идеологией, с одной стороны являясь следствием присущей им коллективной идентичности, а с другой – имея своей целью распространить данную идентичность на других, участвующих в коммуникации, пользователей.

Таким образом, **гипотезой** лежащей в основе данного исследования является предположение о наличии корреляции между тематической ориентацией социально ориентированных общественных движений, и структурными особенностями коммуникативной деятельности, осуществляемой в посвящённых данным движениям сообществах на интернет-платформе «ВКонтакте».

**Объектом** исследования являются тематические сообщества в социальной сети «ВКонтакте», в рамках которых осуществляется коммуникативная деятельность социально ориентированных общественных движений.

**Предметом** исследования являются структурные особенности коммуникативных практик направленных на формирование ценностно-идеологических комплексов, осуществляемых в рамках данных тематических сообществ.

**Целью** исследования является выявление взаимосвязи, между структурными особенностями коммуникативных практик направленных на формирование ценностно-идеологических комплексов и тематической ориентацией данных общественных движений.

Исходя из поставленной цели, можно сформулировать следующие **задачи**:

1. Выявить тематические сообщества в социальной сети «ВКонтакте», в рамках которых осуществляется коммуникативная деятельность социально ориентированных общественных движений.
2. Собрать эмпирические данные, являющиеся цифровыми следами коммуникативной деятельности осуществляемой социально ориентированными общественными движениям в выявленных тематических сообществах.
3. Выявить особенностей коммуникативной деятельности для каждого исследуемого сообщества.
4. Составить коммуникативный портрет коллективной языковой личности для каждого исследуемого сообщества.
5. Составить индивидуальный коммуникативный портрет языковой личности для каждого участника каждого сообщества.
6. Определить уровень идентификации с каждым исследуемым сообществом каждого участника данных сообществ.
7. Выявить особенности коммуникативного воздействия направленного со стороны каждого исследуемого сообщества на каждого участника данных сообществ.
8. Сопоставить особенности коммуникативного воздействия направленного со стороны каждого исследуемого сообщества на каждого участника данных сообществ с уровнем идентификации этих участников с данными сообществами.
9. Выявить характерные для каждого исследуемого сообщества особенности коммуникативного воздействия направленного на участников данного сообщества обладающих низким уровнем идентификации.
10. Проанализировать взаимосвязь между особенностями коммуникативных практик направленных на участников данного сообщества обладающих низким уровнем идентификации, с тематической ориентацией данного сообщества.

Для реализации поставленных задач были разработаны программные решения (скрипты) на языках программирования R и Python, при помощи которых был произведён сбор эмпирического материала и последующий его анализ.

Рассмотрим **первый этап** социологического исследования, заключающийся в выборе исследуемых сообществ и сборе эмпирического материала.

В ходе первого этапа проведённого исследования был произведён разведывательный анализ сообществ на интернет-платформе «ВКонтакте», с целью выявления достаточно широко представленных разными сообществами социально ориентированных общественных движений. Выбор сообществ происходил на основании следующих критериев:

1. Тематическая направленность сообщества, позволяющая отнести происходящую в нём коммуникативную активность к деятельности того или иного социально ориентированного общественного движения.
2. Высокая частота публикаций записей на стене сообщества.
3. Активное комментирование участниками сообщества записей публикуемых на его стене.
4. Наличие достаточного количества других сообществ относящихся к деятельности этого же (или близкого по тематике) общественного движения.

В результате было выбрано 5 общественных движений, для каждого которых было обнаружено не менее 6 посвящённых ему сообществ. Дополнительно, в качестве контрольной группы, было выбрано 6 сообществ идеологически-нейтральной тематики. Таким образом, общее количество выбранных для исследования сообществ составило 36 штук. На основании их тематической принадлежности к общественным движениям, данные сообщества можно сгруппировать в 6 категорий:

1. Отрицание медицины (отказ от прививок, СПИД/ВИЧ-диссидентство, COVID-диссидентство) — 6 сообществ.
2. Вегетарианство — 6 сообществ.
3. Борьба с живодёрами и «догхантерами» — 6 сообществ.
4. Феминизм — 6 сообществ.
5. Национализм — 6 сообществ.
6. Идеологически-нейтральные сообщества (контрольная группа) — 6 сообществ.

Информация о выбранных сообществах представлена в **Приложении 1**.

После выбора исследуемых сообществ был произведён сбор эмпирического материала, включающего в себя отобранные по определённым критериям текстовые сообщения, опубликованные на «стене» в исследуемых сообществах и комментарии пользователей к ним.

Для сбора эмпирических данных, являющихся цифровыми следами коммуникативной деятельности осуществляемой на стене сообществ социальной сети ВКонтакте, на данном сайте было зарегистрировано приложение, предназначенное для осуществления взаимодействия между зарегистрированным на сайте пользователем и API данного сайта. Также был разработан скрипт, осуществляющий автоматические запросы к API сайта данной социальной сети [24]. Исходными данными для работы скрипта являются:

1. Идентификатор зарегистрированного приложения и список запрашиваемых им у пользователя прав доступа.
2. Токен (access\_token), являющийся ключом доступа, идентифицирующим пользователя обращающегося к приложению.
3. Номер версии API для работы с которой написан скрипт.
4. Минимальное количество комментариев оставленных пользователями к записи на стене сообщества, необходимое для того, чтобы данная запись попала в выборку (в данном исследовании было установлено значение 5)
5. Дата в формате Unixtime, начиная с которой записи на стене сообщества могут попасть в выборку (в данном исследовании было установлено значение соответствующее 00:00:00 01.01.2020).
6. Максимальное количество записей со стены, собираемых из одного сообщества (в данном исследовании было установлено значение 1000).
7. Список идентификационных номеров исследуемых сообществ.

После запуска скрипт начинает выполнять запросы к API сайта «ВКонтакте», форматировать полученный ответ и записывать собранные данные в табличной форме.

Размещённые на стене сообществ сайта «ВКонтакте», текстовые записи имеют трёхуровневую структуру:

1. Запись, опубликованная на стене,
2. Комментарии к данной записи (комментарии первого уровня)
3. Комментарии, являющиеся ответами к другим комментариям (комментарии второго уровня).

Для доступа к каждому из этих видов записей API сайта «ВКонтакте» предусматривает разные формы запроса и ответа. Кроме того, на сайте установлены ограничения на частоту запросов (не более трёх за секунду)и на количество данных запрашиваемых за одно обращение (не более 100 записей любого уровня). В связи с этим, скрипт предназначенный для сбора данных, был организован в виде трёх вложенных друг в друга циклов:

1. Внешний цикл осуществляет запросы на скачивание записей со стены сообщества (про 100 записей за запрос), фильтрацию полученных записей по заданным критериям и контроль общего количества записей.
2. Второй цикл перебирает скаченные, и прошедшие через фильтр, записи и отправляет запросы на скачивание комментариев первого уровня к каждой из них (по 100 комментариев за запрос)
3. Третий цикл перебирает скаченные комментарии и проверяет наличие комментариев второго уровня у каждого из них. При наличии комментариев второго уровня отправляются запросы на их скачивание (по 100 комментариев за запрос). В общей сложности, в период с 14 по 22 апреля 2021 года, таим образом была собрана информация о 1 162 050 записях, (29 594 записи на стене и 1 132 465 комментариев) из 36 сообществ. Авторами этих записей являются 809 сообществ (учитываются репосты) и 181 895 пользователей.

В отношении каждой записи на стене и каждого комментария первого и второго уровней скриптом собирается следующая информация:

1. Идентификационный номер сообщения.
2. Идентификационный номер автора сообщения (для записей на стене возвращается идентификационный номер сообщества).
3. Идентификационный номер сообщества, в котором оставлено сообщение.
4. Идентификационный номер записи на стене, к которой оставлен данный комментарий (для записей на стене возвращается их собственный идентификационный номер).
5. Идентификационный номер комментария, на который данный комментарий является ответом (для комментариев, не являющихся ответами, возвращается идентификационный номер записи на стене, а для записей на стене – значение «NA»).
6. Дата публикации сообщения.
7. Логическое значение обозначающее является ли данная запись на стене рекламным постом (для комментариев возвращается значение «NA»).
8. Текст сообщения;
9. Количество и тип прикреплённых к сообщению вложений (медиа-файлов, документов, ссылок);
10. Количество комментариев-ответов;
11. Количество лайков;
12. Количество репостов (для комментариев возвращается значение «NA»);
13. Количество просмотров (для комментариев возвращается значение «NA»).

После завершения сбора данных из текущего сообщества, они сохраняются в файле с форматом .csv и удаляются из памяти скрипта, после чего начинается сбор данных из следующего сообщества.

В работе скрипта используются следующие библиотеки языка программирования R:

«vkR» [39] – для работы с API «ВКонтакте».

«RCurl» [32] ­– для отправки запросов на сервер (в случаях, когда в vkR не предусмотрена нужная функция; например для скачивания комментариев второго уровня).

«jsonlite» [30] – для обработки json-ответов от сервера.

«dplyr» ­­[28] – усовершенствованные функции работы с таблицами.

«stringr» [35] ­– операции с регулярными выражениями (для формирования названий файлов при сохранении).

В данном разделе была изложена программа исследования и описан алгоритм работы скрипта осуществляющего сбор информации. В следующем разделе мы произведём анализ собранных данных, с целью составления коммуникативного портрета коллективной языковой личности изучаемых сообществ.

**2.2. Составление коммуникативного портрета коллективной языковой личности: интеллектуальный анализ текстов**

Данный раздел посвящён анализу собранных данных, направленному на выявление особенностей коммуникативной деятельности каждого исследуемого сообщества и составление коммуникативного портрета коллективной языковой личности этих сообществ.

Для решения обозначенных в прошлом разделе задач, был разработан скрипт осуществляющий анализ собранных со стены сообщества текстов. Обработка текстов данным скриптом осуществляется в 4 этапа:

1. Загрузка данных.
2. Препроцессинг.
3. Частотный анализ текстов.
4. Тематическое моделирование.

Исходными данными для работы скрипта являются:

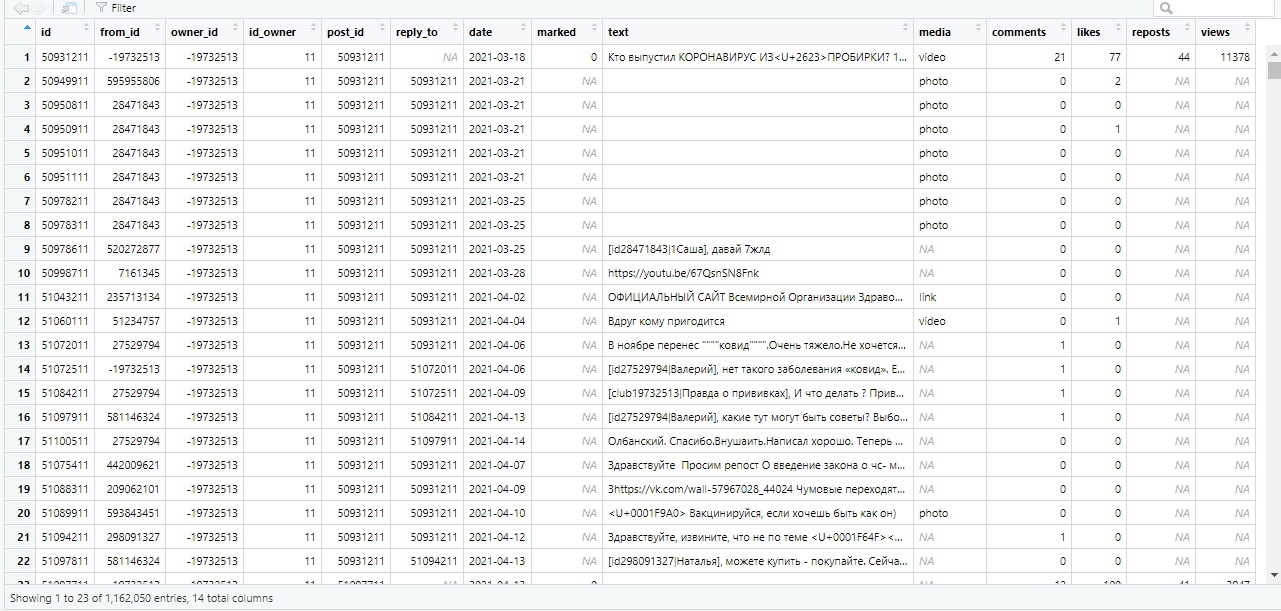
1. Полное имя директории в которой хранятся файлы созданные скриптом осуществляющим сбор данных.
2. Количество используемых для обработки ядер процессора.
3. Значение, задающее нижнюю границу абсолютной частоты употребления слов (сохраняются только слова, употребляемые чаще данного значения, хотя бы в одном сообществе).
4. Значение, задающее нижнюю границу относительной частоты употребления слов (сохраняются только слова с относительной частотой употребления (на 100 000 слов) превышающей данное значение).
5. Значение, задающее критерий типичности слова для одного из сообществ (слово считается типичным для сообщества, если отношение его частоты употребления в данном сообществе к медианной частоте его употребления среди всех сообществ превышает данное значение) .
6. Значение, задающее нижнюю границу для критерия, характеризующего неравномерность употребления слова в разных сообществах (для выявления общей лексики).

Рассмотрим каждый из этапов работы скрипта.

1. **Загрузка данных.**

На этом этапе информация из всех сохранённых файлов загружается в среду языка R и объединяется одну общую базу данных. Для удобства дальнейшей обработки, каждому сообществу дополнительно присваивается двузначный индекс, первая цифра которого обозначает порядковый номер тематической группы, а вторая – порядковый номер сообщества внутри тематической группы. Этот же индекс, в виде двух дополнительных последних цифр, включается в состав идентификационных номеров комментариев, записей опубликованных на стене и комментариев которым адресуется ответ. Это позволяет избежать возможных совпадений номеров из разных сообществ и упрощает определение принадлежности номеров к сообществам, в процессе работы.

В результате получена таблица из 14 столбцов и 1162050 строк. После исключения из таблицы записей относящихся к удалённым комментариям, в ней осталось 1158133 строки. Фрагмент этой таблицы изображён на рисунке 1. Видно, что некоторые комментарии состоят только из медиа вложения и не содержат текстовых записей.



**Рис. 1.** *Фрагмент таблицы содержащей весь собранный для целей исследования эмпирический материал.*

1. **Препроцессинг**

Для эффективной работы алгоритмов анализа текстов, необходимо произвести препроцессинг (предварительную обработку) анализируемых данных. Стандартная процедура обработки текста включает в себя такие этапы как перевод всех букв в нижний регистр, очистка текста от символов, не являющихся буквами и единичными пробелами, разделяющими слова, удаление стоп-слов, стемминг или лемматизацию. В данном случае, исходя из целей исследования, в процедуру был внесён ряд изменений. В результате текст подвергается следующим этапам обработки:

1. Перекодировка символов в формат UTF-8, для совместимости с некоторыми из применённых в дальнейшем алгоритмов.
2. Перевод всех букв в нижний регистр.
3. Очитка текста от выделенных тегами обращений к другим пользователям, ссылок на веб-сайты, эмотиконов и особых символов, удаление меток окончания строки («/n»), и всех оставшихся символов, не являющихся буквами. Данные процедуры осуществляются с использованием регулярных выражений.
4. Лемматизация (приведение всех слов к стандартной (словарной) форме; например слово «кроватью» заменяется на «кровать»). Для осуществления этой процедуры используется библиотека «pymorphy2» [22] работающая в среде языка программирования Phyton, а также библиотека «reticulate» [33], обеспечивающая взаимодействие между R и Phyton. Нами было принято решение производить лемматизацию до удаления стоп-слов, поскольку скорость работы алгоритма лемматизации была достаточно высока, чтобы не требовалось снижать объём текстов, при этом унификация словоформ повышает качество очистки текстов при удалении стоп-слов.
5. Удаление стоп-слов по словарю. За основу списка стоп-слов был взят встроенный словарь библиотеки «tm» [38], словарь из Яндекс.Директ [26]. Метод удаления стоп-слов по словарю является довольно грубым, поэтому нами использовался сильно сокращённый список, из которого были удалены все существительные, отрицательные частицы, личные местоимения и слова с выраженной эмоциональной окраской. В словарь были добавлены единичные буквы, а также некоторые слова и буквосочетания характерные для исследуемых текстов, но не несущие полезной смысловой нагрузки.
6. Объединение отрицательных частиц со следующим за ними словом. Это делается, чтобы в дальнейшем рассматривать соответствующие слова с противоположным смыслом. Для сохранения возможности последующего обратного разделения слов, был использован твёрдый знак, в качестве разделителя (например, словосочетание «не хочу» преобразовывается в слово «неъхочу»).
7. В завершении препроцессинга все тексты преобразуются в так называемый «аккуратный формат» (Tidy text). В этом формате все слова текста, по порядку, располагаются в таблице, причём на каждое слово приходится отдельная строка, а в столбцах указывается информация о них (например, номер комментария, из которого данное слово было взято). Этот формат позволяет использовать очень эффективные (с точки зрения производительности) алгоритмы обработки данных.

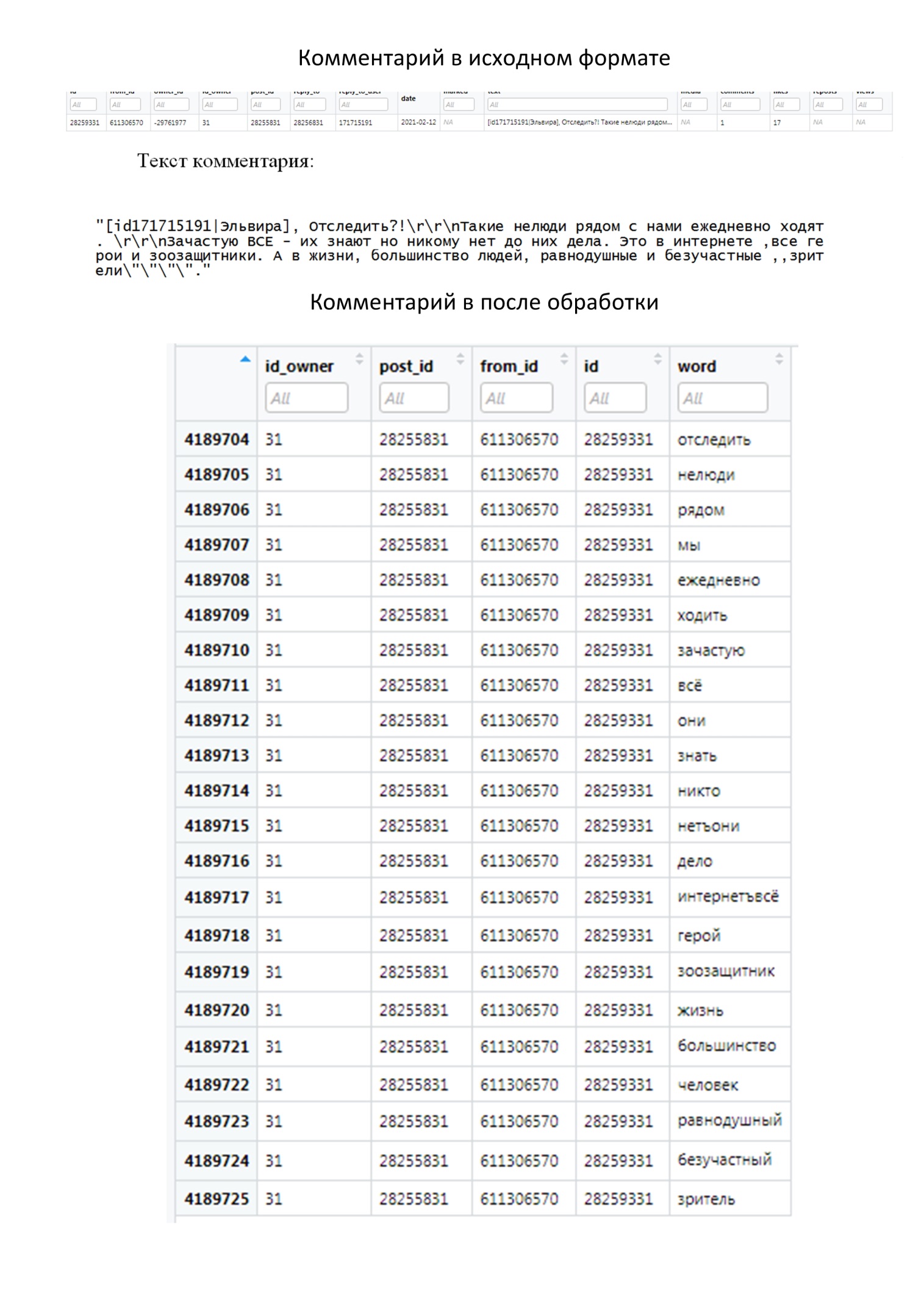
Образец текста до и после описанной обработки представлен на рисунке 2 (стр. 51).

1. **Частотный анализ слов.**

Для выполнения анализа частоты встречаемости терминов (слов) в собранных текстах, была сформирована матрица термин-документ, в которой в которой строки относятся к употребляемым во всех текстах терминам, а столбцы – к исследуемым сообществам. Значение на пересечении строки и столбца означает, сколько раз данный термин был употреблён в данном сообществе. Подобные матрицы, создаваемые, например, с помощью функций библиотеки «tidytext» [37], представлены не в табличной форме, а в специальном формате данных, позволяющем эффективно хранить и обрабатывать большие по размеру, но очень разряженные матицы. В связи с тем, что объём собранных текстов был очень велик, было принято решение исключить из дальнейшего анализа термины, которые не встречаются чаще, чем 10 раз или чаще чем 15 раз на 100 000 слов, хотя бы в одном сообществе. Эта мера позволила сократить список анализируемых терминов более чем на 92%, при менее чем 10% потере объёма текста.

В отношении оставшихся терминов произведено преобразование частоты их употребления из абсолютных значений в относительные (частота встречаемости на 1000 слов данного сообщества) и выделение следующих типов лексики для каждого сообщества:

1. Уникальная лексика – термины, которые встречаются в текстах только в одного из исследуемых сообществ.
2. Специфическая лексика – термины, которые встречаются в текстах более чем одного, но менее чем половины исследуемых сообществ.
3. Характерная лексика – термины, чья максимальная относительная употребимось по текстам исследуемых сообществ не менее чем в 10 раз превышает медианную.
4. Общая лексика – термины, которые примерно с равной частотой употребляются в текстах исследуемых сообществ. В качестве критерия используется коэффициент, определяемый как отношение максимального значения относительной частоты употребления слова к наибольшему значению, которое меньше, чем 7/8 из совокупности значений. Предельное значение для данного коэффициента (ниже которого слова считаются общей лексикой) нами было принято равным 4.
5. Уникальная, специфическая и характерная лексики являются элементами коммуникативного портрета коллективной языковой личности каждого из исследуемых сообществ. Уровень их употребления отбельными участниками сообществ, в рамках данного исследования, считается нами основным показателем степени их идентификации с соответствующими общественными движениями.



**Рисунок 2.** *Образец текста до и после препроцессинга.*

1. **Тематическое моделирование.**

Перед построением тематической модели тексты были дополнительно подготовлены путём удаления из них, выделенных на предыдущем этапе, слов общей лексики. В качестве единицы документа для данного вида анализа была принята запись на стене и все относящиеся к ней комментарии. Сформирована соответствующая матрица термин–дкумент («слово» – «запись на стене») которая и использовалась для построения тематической модели.

Построение тематической модели осуществлялось с помощью библиотеки «topicmodels» [21] реализующей алгоритм на основе латентного распределения Дирихле. Для оценки качества тематического моделирования использовалось 2 показателя:

1. Оценка когерентности тем, вычисляемая функцией «CalcProbCoherence» из библиотеки «textmineR» [36]. С ростом числа тем этому показателю свойственно расти.
2. Косинусное расстояние между темами вычисляемое функцией «cosine» из библиотеки «lsa» [31]. С ростом числа тем, это показатель сначала растёт, а потом снижается.

Кроме того, был введён штраф за количество тем.

(1)

Итоговое качество модели рассчитывалось по формуле 1, где:

S – безразмерный показатель сравнительного качества тематических моделей;

Hmed – медианная когерентность тем;

Lmin – минимальное расстояние между темами;

k – количество тем.

Скрипт рассчитывал данный показатель для моделей разным числом тем (от 5 до 160), после чего выбиралось число тем, при котором значение показателя было максимальным.

В результате произведённого тематического моделирования для каждого сообщества было определено оптимальное число тем, и распределение вероятностей отношения каждого слова к каждой теме. Первый из этих показателей характеризует коммуникативный портрет коллективной языковой личности каждого из исследуемых сообществ, а второй позволяет определить некоторые параметры индивидуального коммуникативного портрета языковой личности, о чём будет сказано ниже.

**2.3. Выявление специфики коммуникативных практик формирования ценностно-идеологических комплексов: корреляционный анализ виртуальных сообществ**

В предыдущих разделах был описан сбор и анализ данных, благодаря котором удалось выявить такие коммуникативные характеристики сообществ и пользователей, как уровень использования лексики характерной для данных сообществ и интерес к определённым темам. В этом разделе мы сопоставим полученные данные с параметрами коммуникативной активности пользователей и коммуникативного воздействия на них, что соответствует задачам 5-10.

Для решения задач 5-10 была сформирована база данных содержащая подробную информацию о личных коммуникативных качествах пользователя, его взаимоотношениях с другими пользователями и об оказываемом на него коммуникативном воздействии. Такая база данных представляет собой индивидуальный коммуникативный портрет языковой личности.

Для реализации этой задачи был разработан скрипт, анализирующий собранную информацию о коммуникативной деятельности пользователя и формирующий описанную базу данных.

Исходными данными для работы скрипта являются:

1. Полное имя директории, в которой хранятся файлы созданные скриптом осуществляющим сбор данных.
2. Количество используемых для обработки ядер процессора.

В процессе работы скрипт выполняет следующие действия:

1. Подготовка данных для расчёта.
2. Построение портрета коммуникативной личности
3. Корреляционный анализ полученных данных.

Рассмотрим подробнее каждый из этапов работы скрипта.

1. **Подготовка данных для расчёта.**

На этом этапе в качестве исходных данных используется та же таблица, что была сформирована при загрузке данных, на предыдущем этапе расчётов. Однако, в данном расчёте нас интересуют не тексты комментариев, а данные для активности пользователей. Для дальнейших расчётов эти данные надо извлечь и переформатировать в удобный для работы формат. Для этого скриптом выполняются следующие действия:

1. Удаление из таблицы строк, соответствующих удалённым с сайта комментариям.
2. Формирование списка уникальных идентификационных номеров всех авторов опубликованных текстов (сообществ и пользователей)
3. Разделение списка на отдельные списки для каждой группы.
4. Строится матрица отображающая количество комментариев оставленных каждым пользователем в каждой группе.
5. Построение матрицы, отображающей количество комментариев, оставленных каждым пользователем в каждом из сообществ.
6. Построение матрицы, отображающей количество общих пользователей у сообществ.
7. Построение списка матриц отображающих количество комментариев, оставленных каждым пользователем под каждой записью на стене, в каждой группе.

На этом подготовительный этап заканчивается и начинается формирование таблицы с данными, характеризующими индивидуальный коммуникативный портрет языковой личности каждого пользователя.

1. **Построение портрета коммуникативной личности.**

Таблица сданными, выполняющая роль портрета коммуникативной личности содержит информацию трёх типов о каждом пользователе:

1. Данные о коммуникативной активности пользователя.
   1. Индикационный номер пользователя.
   2. Количество комментариев оставленных каждым пользователем в каждой группе.
   3. Кол-во комментариев являющихся ответами другим пользователям (не сообществам).
   4. Доля комментариев являющихся ответами другим пользователям (не сообществам).
   5. Количество уникальных пользователей, которым отвечал пользователь (за исключением сообществ).
   6. Относительное количество уникальных пользователей, которым отвечал пользователь (за исключением сообществ).
   7. Среднее количество комментариев-ответов, приходящееся на одного собеседниеа.
   8. Количество комментариев с медиаволжениями.
   9. Доля комментариев с медиавложениями.
   10. Количество записей на стене каждой из групп, прокомментированных пользователем хотя бы один раз.
   11. Среднее количество комментариев оставленных под каждой записью на стене.
2. Данные о коммуникативном воздействии на пользователя:
   1. Количество полученных лайков.
   2. Среднее количество лайков на комментарий.
   3. Количество других уникальных пользователей отвечавших данному.
   4. Среднее количество уникальных пользователей отвечавших данному, приходящееся на 1 комментарий.
   5. Количество полученных ответов,
   6. Среднее количество ответов, получаемых на комментарий.
3. Данные из предыдущего раздела, о речевом поведении пользователя:
   1. Средняя длина комментария.
   2. Уровень употребления специфичной для данного сообщества лексики.

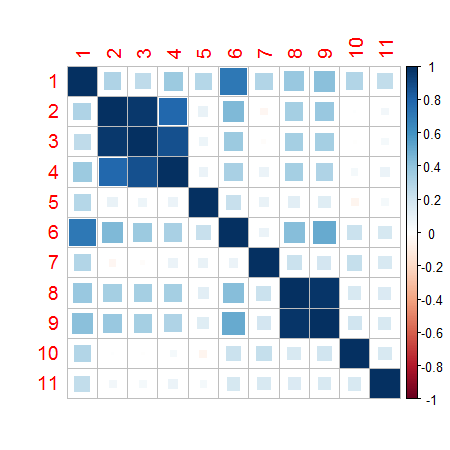
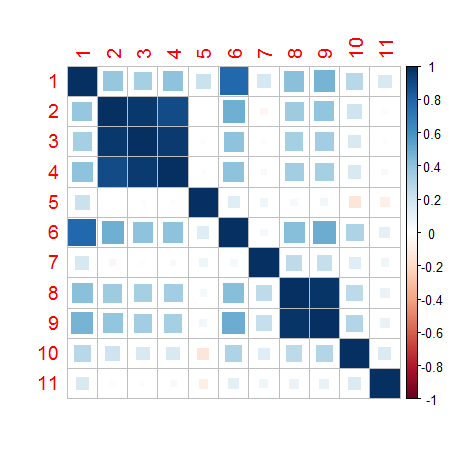
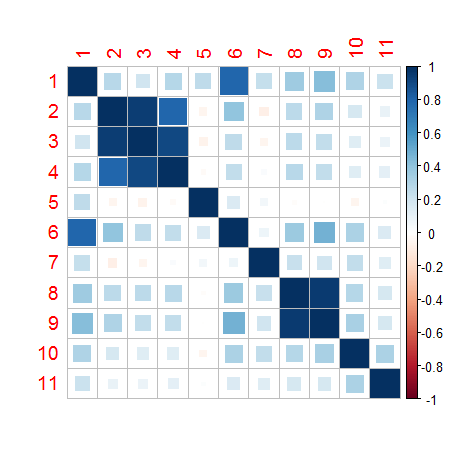
На этом этапе возникли сложности технического характера, из-за которых не удалось добавить в таблицу информацию о трёх сообществах и о предпочитаемых пользователями темах, выявленных при тематическом анализе. Поэтому в дальнейшем эти данные в расчёте участвовать не будут.

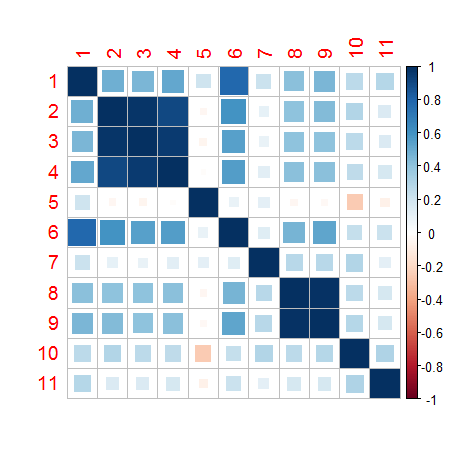
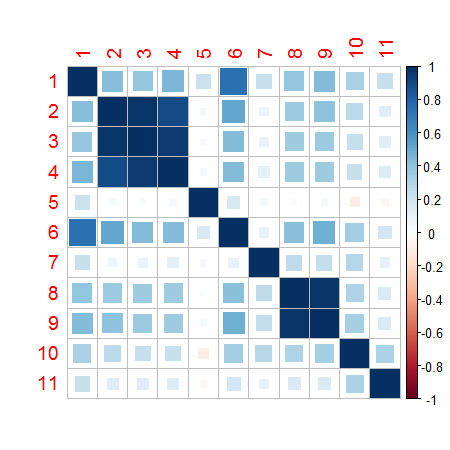
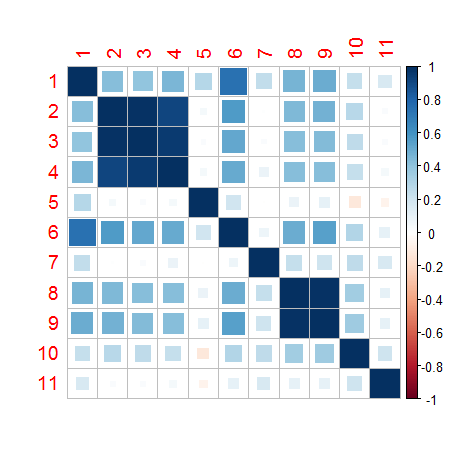
Собранных таким образом данных достаточно для проверки гипотезы, о наличии взаимосвязи между речевым поведением пользователя (что, в рамках данного исследования интерпретируется как мера идентификации пользователя с сообществом), его коммуникативным поведением (что можно понимать, как ролевую функцию в данном сообществе) и оказываемым на него коммуникативным воздействием.

Для проведения такого анализа из данных были исключены значения, представленные в абсолютных величинах (кроме числа комментариев), поскольку они напрямую зависят от числа оставленных пользователем комментариев. Оставшийся список характеристик выглядел следующим образом:

1. Количество комментариев, опубликованных пользователем в данном сообществе.
2. Доля комментариев являющихся ответами другим пользователям (не сообществам)
3. Относительное количество уникальных пользователей (собеседников), которым отвечал данный пользователь (за исключением сообществ).
4. Среднее количество комментариев адресованное каждому собеседнику.
5. Доля комментариев с медиавложениями.
6. Среднее количество комментариев оставленных под каждой записью на стене.
7. Среднее количество лайков на комментарий.
8. Среднее количество собеседников приходящееся на 1 комментарий
9. Среднее количество ответов, получаемых на комментарий.
10. Средняя длина комментария.
11. Уровень употребления специфичной для данного сообщества лексики.

Для выявления взаимосвязей между этими характеристиками по каждой из 33 оставшихся в работе групп был произведён ранговый корреляционный анализ, с применением коэффициента Спирмена. В результате получено 33 матрицы с данными о взаимосвязи между рассмотренными характеристиками. Эти данные удобно визуализировать в виде так называемых «термокарт». На рисунке №3 представлены 6 термокарт относящихся к первому сообществу каждой из тематических групп. Представленные данные демонстрируют картину типичную и для всех остальных сообществ.





**Рис. 3.** *Образцы корреляционных «термакарт». Верхний ряд, сообщества: «Правда о прививках», «Вегетарианство, сыроедение, ЗОЖ», «Живодёрам.НЕТ».*

*Нижний ряд, сообщества: «Подслушано Феминизм», «Правые», «Это Питер, детка! [Типичный Питер]». Номера столбцов и строк соответствуют списку на стр. 56-57.*

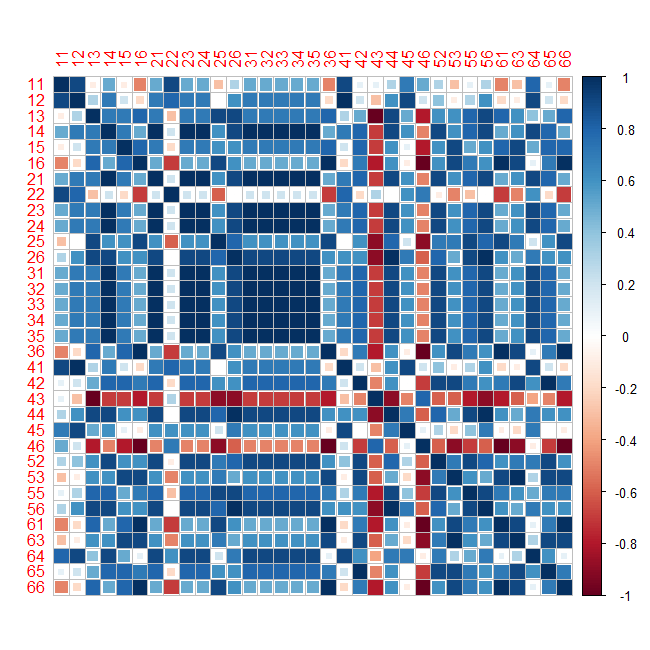
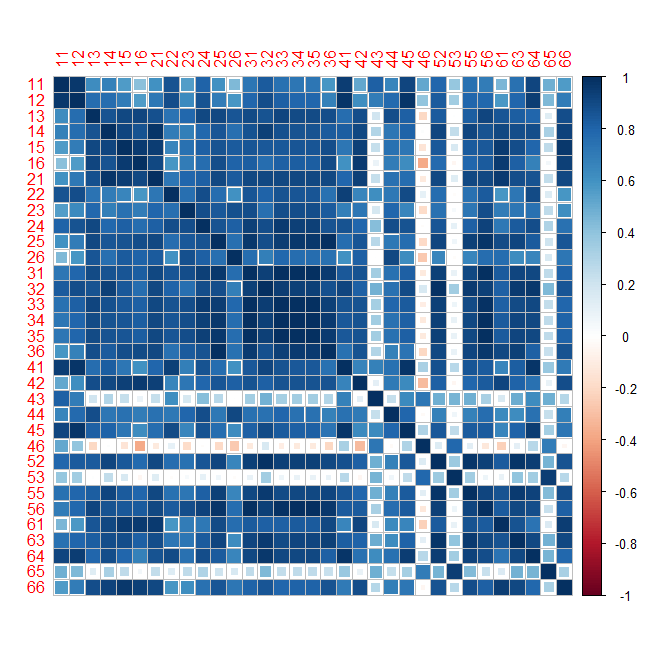
По представленным данным видно, что предполагаемая теоретической моделью данного исследования корреляция между уровнем использования характерной для сообщества лексики и коммуникативным воздействием на пользователя не находит подтверждения. Также весьма слабо влияют на коммуникативное воздействие и другие рассмотренные параметры. В целом, результаты анализа выглядят довольно тривиальными, однако значения коэффициента Спирмена, выявленные для слабо коррелирующих характеристик, всё же не настолько малы, чтобы их проигнорировать. Они часто превышают значение 0,3, что позволяет охарактеризовать данную взаимосвязь, как умеренную. Кроме того, среди коэффициентов имеющих значение меньше чем 0,3 большая часть остаётся статистически значимой (p <<< 0.01).

Основной гипотезой данного исследования является предположение о наличии взаимосвязи между тематической направленностью социально ориентированных общественных движений и характеристиками осуществляемой ими коммуникативной деятельности. Коммуникативная деятельность общественных движений, в той мере, в какой она была изучена в рамках данной работы, характеризуется представленными матрицами с данными о корреляции изученных параметров. Если основная гипотеза верна, то данные в матрицах сообществ, относящихся к одной тематической группе, буду более похожи друг на друга, чем данные из матриц, относящихся к разным тематическим группам. Проверить это также можно с помощью корреляционного анализа.

Для проведения второго этапа корреляционного анализа была сформирована таблица, состоящая из 33 колонок, соответствующих исследуемым сообществам и 110 строк, соответствующих значениям из рассмотренных ранее корреляционных матриц (за исключением 11 строк относящихся к диагоналям). По каждой строке была подчитана величина дисперсии, значения которой менялись в диапазоне от 0.000362 до 0.0135.

Расчёт корреляции по всем значениям таблицы давал почти равномерное заполнение корреляционной матрицы значениями близкими к единице, однако, путём удаления из расчёта строк с низкой дисперсией удалось проявить структуру коммуникативного разнообразия. Дветерм карты соответствующие разным уровням фильтрации строк с низкой дисперсией представлены на **рисунке 4**.

Подтверждением исходной гипотезы была бы картина, стоящая из квадратов 6х6 клеток, расположенных вдоль диагонали термокарты. Некоторые элементы подобной структуры присутствуют и здесь, но несовпадение границ образовавшихся квадратов с границами тематических групп, большое количество исключений и отсутствие отличительных черт у контрольной группы, свидетельствуют в пользу того, что различия и сходства в структуре коммуникативной деятельности сообществ обусловлены не их идеологической ориентацией, а какими-то иным факторами, для выявления которых требуется дополнительное исследование.

**Рис. 4.** *Образцы корреляционных «термакарт».*

*Левое изображение по лучено по параметрам с дисперсией больше 0,01; правое ­– при дисперсии больше 0,008.*

Последним вопросом, который стоит рассмотреть в рамках данного исследования, является выявление наиболее дисперсных характеристик коммуникативной деятельности исследуемых общественных движений. Значения дисперсии этих характеристик были нами получены при подготовке ко второму этапу корреляционного анализа. Выпишем их наименования:

1. Доля комментариев являющихся ответами другим пользователям (не сообществам) и среднее количество собеседников, приходящееся на 1 комментарий.
2. Относительное количество уникальных пользователей (собеседников), которым отвечал данный пользователь (за исключением сообществ) и среднее количество комментариев оставленных под каждой записью на стене.
3. Относительное количество уникальных пользователей (собеседников), которым отвечал данный пользователь (за исключением сообществ) и среднее количество ответов, получаемых на комментарий.
4. Среднее количество комментариев адресованное каждому собеседнику и Среднее количество комментариев оставленных под каждой записью на стене.
5. Среднее количество комментариев адресованное каждому собеседнику и Среднее количество ответов, получаемых на комментарий.
6. Среднее количество комментариев оставленных под каждой записью на стене и Относительное количество уникальных пользователей (собеседников), которым отвечал данный пользователь (за исключением сообществ).
7. Среднее количество комментариев оставленных под каждой записью на стене и среднее количество комментариев, адресованное каждому собеседнику.
8. Среднее количество собеседников приходящееся на 1 комментарий и доля комментариев являющихся ответами другим пользователям (не сообществам)
9. Среднее количество ответов, получаемых на комментарий и относительное количество уникальных пользователей (собеседников), которым отвечал данный пользователь (за исключением сообществ).
10. Среднее количество ответов, получаемых на комментарий и среднее количество комментариев, адресованное каждому собеседнику.

Данные соотношения, как было показано в нашем исследовании, являются наиболее вариабельными (отличающимися от сообщества к сообществу), что даёт возможность рассмотреть их в качестве основы для построения системы классификации сообществ в будущих исследованиях.

По результатам проведённого исследования можно сделать следующие выводы:

1. Исследование коммуникативной деятельности осуществляемой (в частности общественными движениями) в сети Интернет, является перспективным и динамично развивающимся исследовательским направлением, уже сейчас предлагающим исследователю большое разнообразие методологических подходов, однако требующее более тщательной теоретической разработки.
2. Общественные движения широко представлены в виртуальном пространстве и активно осуществляют в нём свою деятельность, что формирует перспективное поле для их исследований.
3. Подтверждений гипотезе о наличии взаимосвязи между идеологической ориентацией общественных движений и особенностями их коммуникативной деятельности выявлено не было. Однако, разработанная методика исследования коммуникативной деятельности в сети Интернет, позволила выявить наличие нескольких разновидностей подобных особенностей, не связанных с идеологической ориентацией, что указывает на перспективы для дальнейших исследований.

**Заключение**

Развитие и повсеместное распространение телекоммуникационных технологий является, в настоящее время, одним из наиболее существенных факторов, обуславливающих изменения в общественной жизни. Сформировав новое социальное пространство, характеризующееся своими, весьма специфическими, возможностями и ограничениями, Интернет оказал глубокое влияние на образ жизни людей, на их коммуникативные и потребительские привычки, на профессиональную и досуговую деятельность. И тенденции эти продолжают развиваться. Испытала на себе его влияние и идеологическая сфера жизни. Конкурируя с традиционными СМИ, размывая границы между межличностной и групповой коммуникацией, давая возможность каждому желающему, независимо от времени и места, получить доступ к массовой аудитории, Интернет стал мощным инструментом для формирования ценностей и убеждений. Отчасти в силу этого, отчасти в силу изначально коммуникативной природы самого это явления, широкое распространение в интернете получила деятельность разнообразных общественных движений. Характерной отличительной чертой такой деятельности является её непосредственная направленность на осуществление социальных изменений, в то время, как в остальных случаях такие изменения оказываются, в основном, побочными эффектами от реализуемой деятельности. Стремление изменять социальную реальность, в купе с новыми телекоммуникационными возможностями, даёт повод рассматривать общественные движения как перспективный источник социальных трансформаций.

Данное обстоятельство обуславливает исследовательский интерес к этому явлению социальной жизни. Изучение общественных движений, начавшееся в середине XX-го века, с появлением Интернета, как значимой части социальной реальности, получило существенный импульс к развитию. В первой главе данной работы были рассмотрены некоторые современные тенденции в этой сфере. Приведён обзор современных теоретических и методических подходов. В связи со спецификой цели данного исследования, особое внимание было уделено рассмотрению социо-лингвистических концепций формирования идентичности, и методическим вопросам изучения цифровых следов. По итогам рассмотрения социо-линвистических концепций были выделены понятия «коммуникативный портрет» и «языковая личность», описаны способы их изучения и показана взаимосвязь с понятием «идентичность». Анализ работ посвящённых социологическим исследованиям Интернета позволил выделить методы, подходящие для решения задач данного исследования: частотный анализ текстов, тематическое моделирование и сетевой анализ виртуальных сообществ. Также, в первой главе, был произведён обзор парадигмальных подходов к исследованию общественных движений, и обосновано применение к объекту данного исследования парадигмы Новых общественных движений.

Объектом данного исследования стали 36 сообществ социальной сети «ВКонтакте», 30 из которых были посвящены деятельности пяти различных общественных движений, а оставшиеся 6 идеологически нейтральных сообществ составляли контрольную группу. Коммуникативная деятельность участников данных сообществ исследовалась на предмет выявления структурных особенностей и их взаимосвязи с идеологической ориентацией сообщества. В качестве эмпирического материала были собраны тексты написанные участниками исследуемых сообществ в комментариях к записям, опубликованным на стене сообществ, за период с 01.01.2020 по 22.04.2021. Для сбора и анализа данного материала были специально разработаны программные решения – скрипты на языках программирования R и Python.

Перед началом анализа тексты были обработаны путём приведения всех букв к нижнему регистру, удаления всех символов, кроме букв, лемматизации и удаления стоп-слов. Анализ текстов происходил по двум направлениям: в ходе частотного анализа для каждого сообщества была выделена характерная лексика. Теоретическая модель данной работы предполагала рассмотрение интенсивности употребления участниками сообществ характерной для этих сообществ лексики, как маркера степени идентификации их с данным сообществом. Вторым направлением анализа текста было построение его тематической модели при помощи алгоритма основанного на методе скрытого распределения Дирихле. Целью этого анализа было распределение участников сообщества по группам, на основании общих интересов к обсуждению определённых тем.

Параллельно с анализом текста производился анализ коммуникативного поведения участников сообществ и коммуникативного воздействия, которое оказывается на них со стороны сообщества. Данный анализ состоял в сборе следующей информации о коммуникативных характеристиках каждого участника сообщества: количество написанных комментариев, количество прокомментированных постов, количество комментариев являющихся ответами другим пользователям, интенсивность использования медиавложений. В качестве оказываемого на участников сообществ коммуникативного воздействия рассматривалось количество обращённых к ним комментариев и обратившихся пользователей, а также количество полученных лайков. После того, как данная информация была дополнена данными о средней длине комментария и уровне употребления характерной для сообщества лексики, был произведён корреляционный анализ всех собранных данных, с целью выявить взаимосвязь между коммуникативным поведением пользователя, уровнем его идентификации с сообществом и оказываемым на него коммуникативным воздействием. Однако выявленные корреляции оказались существенно слабее взаимосвязей между другими характеристиками, что ставит под вопрос возможность их рассмотрения в качестве факторов регулирующих функционирование сообществ. Для проверки гипотезы о наличие взаимосвязи между особенностями коммуникативных практик и тематической ориентацией сообщества, был проведён корреляционный анализ этих практик друг с другом. В результате анализа выявлены тенденции унификации коммуникативных практик в различных группах сообществ, однако принцип разделения на эти группы не соответствовал тематической ориентации сообществ. По результатам анализа выделены наиболее вариабельные соотношения характеристик сообществ и сделан вывод о возможности рассмотрения их в качестве основы для построения классификации.

**ЛИТЕРАТУРА**

1. *Бергер П., Лукман Т.* Социальное конструирование реальности. Трактат по социологии знания. / Пер. Е.Д. Руткевич. М.: Медиум, 1995. 323 с.
2. *Бойко Б. Л.* Принципы моделирования речевого портрета носителя социально-группового диалекта (к проблеме создания речевого портрета человека на войне) // Вестник военного университета. 2008. №2. С.114-121
3. *Бутина М.В.* Дискуссия о сущности общественных движений // Известия АлтГУ, 2013. №4 (80). С. 268—272.
4. *Городнова, А. А.* Развитие информационного общества: учебник и практикум для академического бакалавриата / А. А. Городнова. М.: Юрайт, 2017. 234 с.
5. *Дудина В. И., Юдина Д. И.* Извлекая мнения из сети Интернет: могут ли методы анализа текстов заменить опросы общественного мнения? // Мониторинг общественного мнения: Экономические и  социальные перемены. 2017. № 5. С.  63—78.
6. *Здравомыслова Е.А.* Парадигмы западной социологии общественных движений / Е.А. Здравомыслова. СПб.: Наука, 1993. 172 с.
7. *Горелов И. Н., Седов K. Ф.* Основы психолингвистики. Учебное пособие. / И. Н. Горелов, K. Ф. Седов. Третье, переработанное и дополненное издание. Издательство "Лабиринт", М., 2001. 304с.
8. *Карасик В.И.* Языковой круг, личность, концепты, дискурс – Волгоград: Перемена. 2002. 477 с.
9. *Колозариди П.В., Ульданов А.А.* Интернет и общественные движения: анализ развития тематики в базах ScienceDirect и Scopus // Социологический журнал, 2015. №3. С. 105—125.
10. Новые социальные движения в сетевую эпоху: статьи, интервью, экспертные заключения: монография. / под ред. Р.Э. Бараш М: Русское общество истории и философии наук, 2020. 282 с.
11. *О.Ю. Кольцова, К.А. Маслинский.* Выявление тематической структуры российской блогосферы: автоматические методы анализа текстов. // Социология: методология, методы, математическое моделирование. 2013. № 36. С.  113—139.
12. *Павлова Т.В.* Институциональные подходы к изучению социальных движений // Политическая наука, 2009. №3. С.  71—83.
13. *Писаренко Н. В.* Идентичность в аспекте коммуникации: теоретические подходы // Вестн. Том. гос. ун-та. 2005. №286. 118-122
14. *Римский В.* Воздействие сети Интернет на социальную активность, формирование и развитие идентичностей // Вестник общественного мнения. Данные. Анализ. Дискуссии. 2009. №1. С. 86—96
15. *Степанова Е.А.* Теория и методология сетевого подхода в исследованиях российских онлайн-сообществ // Южно-российский журнал социальных наук, 2018. №2. С. 50—63.
16. *Чурилина Л.Н.* Антропоцентрический принцип в исследовании лексической структуры художественного текста // Известия РГПУ им. А. И. Герцена. 2003. №5. С. 63-74
17. *Штомпка П.* Социология социальных изменений/ П. Штомпка . Пер, с англ, под ред. В.А. Ядова. М.: Аспект Пресс, 1996. 416 с.
18. *Штомпка П.* Социология. Анализ современного общества: Пер. с польск. СМ. Червонной. М.: Логос, 2005. 664 с.
19. *Юдина Д.И., Дудина В.И.* Семантическая сеть на биграммах как метод валидизации результатов тематического моделирования в социологическом исследовании // Журнал социологии и социальной антропологии. 2016. Том XIX. № 4 (87) С. 71–83.
20. *Blei D. M., Ng A. Y., Jordan M. I.* Latent dirichlet allocation. //The Journal of machine Learning research. 2003. №3. pp. 993-1022.
21. *Hornik K., Grun B.* topicmodels: An R package for fitting topic models.//Journal of Statistical Software. 2011. №40(13). pp. 1-30.
22. *Korobov M.* Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts. 2015. pp. 320-332
23. *Rykov Y. G., Meylakhs P. A., Sinyavskaya Y. E.* Network structure of an AIDS-Denialist online community: identifying core members and the risk group //American Behavioral Scientist. 2017. Т. 61. №. 7. С. 688-706.

**Электронные ресурсы:**

1. Знакомство с API ВКонтакте //VK Developers. URL: https://vk.com/dev/first\_guide. (Дата обращения 05.05.2021).
2. Реакция общества на фильм «Фонда борьбы с коррупцией» // Левада-центр. 2015. URL: https://www.levada.ru/2015/12/23/reaktsiya-obshhestva-na-film-fonda-borby-s-korruptsiej/. (Дата обращения: 05.05.2021)
3. Яндекс.Директ // https://direct.yandex.ru/. (Дата обращения: 21.04.2021)
4. *Csardi G., Nepusz T.* The igraph software package for complex network research. //InterJournal, Complex Systems 1695. 2006. URL: http://igraph.sf.net. (Дата обращения: 05.05.2021)
5. dplyr: A Grammar of Data Manipulation. URL: https://cran.r-project.org/package=dplyr. (Дата обращения 25.04.2021).
6. gCLUTO - Graphical Clustering Toolkit. URL: http://glaros.dtc.umn.edu/ gkhome/cluto/gcluto/overview. (Дата обращения 25.04.2021).
7. jsonlite: A Simple and Robust JSON Parser and Generator for R. URL: https://cran.r-project.org/package=jsonlite. (Дата обращения 27.04.2021).
8. lsa: Latent Semantic Analysis URL: https://cran.r-project.org/package=lsa. (Дата обращения 03.05.2021).
9. RCurl: General Network (HTTP/FTP/...) Client Interface for R URL: https://cran.r-project.org/package=RCurl. (Дата обращения 07.04.2021).
10. reticulate: Interface to 'Python' URL: https://cran.r-project.org/package=reticulate. (Дата обращения 13.04.2021).
11. Stanford Topic Modeling Toolbox // The Stanford Natural Language Processing Group. URL: http://nlp.stanford.edu/software/tmt/tmt-0.4/. (Дата обращения 25.04.2021).
12. stringr. URL: https://github.com/cran/stringr. (Дата обращения 25.04.2021).
13. Thomas W. Jones. textmineR: Functions for Text Mining and Topic Modeling. // CRAN. 2019. URL: https://cloud.r-project.org/web/packages/textmineR/index.html. (Дата обращения 04.05.2021).
14. tidytext: Text Mining using 'dplyr', 'ggplot2', and Other Tidy Tools. URL: https://cran.r-project.org/package=tidytext. (Дата обращения 27.04.2021).
15. tm: Text Mining Package. URL: https://cran.r-project.org/package=tm. (Дата обращения 27.04.2021).
16. vkR: Access to VK API via R. URL: https://cran.r-project.org/package=vkR. (Дата обращения 27.04.2021).

**Приложения**

**Приложение 1.** Перечень исследуемых сообществ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1. Отрицание медицины** | **2. Вегетари-анство** | **3. Борьба с живодёрами «догхан-терами»** | **4. Феминизм** | **5. Национа-лизм** | **6. Идеологи-чески-нейтральные сообщества (контрольная группа)** |
| 11. Правда о прививках  N= 103 000 чел.  URL: https://vk.com/club19732513 | 21. Вегетарианство, сыроедение, ЗОЖ  N= 75 000 чел.  URL: https://vk.com/club13164379 | 31. Живодёрам.НЕТ  N= 28 000 чел.  URL: https://vk.com/club29761977 | 41. Подслушано Феминизм  N= 36 000 чел.  URL: https://vk.com/public69629007 | 51. Русский Националист  N= 40 000 чел.  URL: https://vk.com/public183369260 | 61. Это Питер, детка! [Типичный Питер]  N= 1 112 000 чел.  URL: https://vk.com/public31516466 |
| 12. Прививки: «за» и «против»  N= 50 000 чел.  URL: https://vk.com/club38532412 | 22. Вегетарианство - норма жизни!  N= 57 000 чел.  URL: https://vk.com/club138950 | 32. Охотники на живодёров. Антидогхантеры.  N= 29 000 чел.  URL: https://vk.com/public133071341 | 42. le mot de la fin | последние слова  N= 16 100 чел.  URL: https://vk.com/club125021495 | 52. Правые  N= 149 000 чел.  URL: https://vk.com/public23486475 | 62. Интеллектуальный юмор  N= 3 123 000 чел.  URL: https://vk.com/public28950133 |
| 13. ВИЧ СПИД - ВЕЛИЧАЙШАЯ МИСТИФИКАЦИЯ XX ВЕКА  N= 18 000 чел.  URL: https://vk.com/club5940425 | 23. Первый Вегетарианский,  N= 467 000 чел.  URL: https://vk.com/public58656979 | 33. Неравнодушные  N= 36800 чел.  URL: https://vk.com/club23241512 | 43. по ту сторону изгороди. / кейли рей арт.  N= 73 400 чел.  URL: https://vk.com/public161037110 | 53. Русский национализм в схемах и мемах  N= 10 000 чел.  URL: https://vk.com/club165978802 | 63. Я КИНОМАН  N= 937 000 чел.  URL: https://vk.com/public79555295 |
| 14. ИММУННЫЙ ОТВЕТ - Всероссийская акция  N= 10000 чел.  URL: https://vk.com/club182362555 | 24. Я ВЕГЕТАРИАНЕЦ  N= 117 000 чел.  URL: https://vk.com/public23505064 | 34. Защита животных от живодеров. Антидогхантеры Уфы  N= 5600 чел.  URL: https://vk.com/public93692040 | 44. Школа феминизма  N= 20 000  чел.  URL: https://vk.com/club27506840 | 54. Спутник и Погром  N= 113 000 чел.  URL: https://vk.com/public40399920 | 64. My Story  N= 356 000 чел.  URL: https://vk.com/my\_storry |
| 15. МедАльтернатива.инфо  N= 40000 чел.  URL: https://vk.com/club94239677 | 25. Центр защиты прав животных «ВИТА»  N= 25 000 чел.  URL: https://vk.com/club3349701 | 35. О живодерах в России.  N= 5400 чел.  URL: https://vk.com/club185097559 | 45. Феминизм. Кавказ  N= 8900 чел.  URL: https://vk.com/public141676240 | 55. Я – РУССКИЙ  N= 84 000 чел.  URL: https://vk.com/club72048 | 65. «Бумага»  N= 92 000 чел.  URL: https://vk.com/public23303030 |
| 16. Медицинское безумие  N= 3000 чел.  URL: https://vk.com/club117920577 | 26. Вегетарианцы / Веганы  N= 75 000 чел.  URL: https://vk.com/public17228 | 36. Гуманные люди, или Жизнь без убийства животных  N= 7600 чел.  URL: https://vk.com/club39108476 | 46. Про самое важное  N= 366 000 чел.  URL: https://vk.com/public55827701 | 56. РОНС - Россия Освободится Нашими Силами  N= 8 700 чел.  URL: https://vk.com/club1000842 | 66. Лингвовести: языки и лингвистика  53 500 чел.  URL: https://vk.com/public46299096 |