Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Национальный исследовательский университет «Высшая школа экономики»

Факультет социологии

Кафедра методов сбора и анализа социологической информации

**Выпускная квалификационная работа по направлению 040100.62 «Социология» подготовки бакалавра на тему: «Анализ структуры распространения новостной информации в социальных сетях»**

Студентка группы № 432

Мурзова Ирина Олеговна

**Научный руководитель**

Старший преподаватель
кафедры методов сбора и анализа
социологической информации,

Полухина Елизавета Валерьевна

Москва, 2014

Оглавление

[Введение 4](#_Toc389828735)

[ГЛАВА 1. ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ ИЗУЧЕНИЯ НОВОСТНОЙ ИНФОРМАЦИИ В СОЦИАЛЬНЫХ СЕТЯХ 14](#_Toc389828736)

[1.1.Современная Россия в Интернете: стремительное увеличение пользователей Сети 14](#_Toc389828737)

[1.2. Социальные сети: структура коммуникации online vs. offline 17](#_Toc389828738)

[1.3. Типология социальных сетей виртуального пространства 19](#_Toc389828739)

[Глава 2. Исследования Интернет-пространств. Основные подходы в России и за рубежом 22](#_Toc389828740)

[2.1. Концепция «лидеров мнений» Пола Лазарсфельда. Соотнесение с онлайн медиа 23](#_Toc389828741)

[2.2. Неавтоматизированные методы 26](#_Toc389828742)

[2.2.1. Нетнография 26](#_Toc389828743)

[2.2.2. Выборочный метод 28](#_Toc389828744)

[2.2.3. Анализ поисковых запросов 28](#_Toc389828745)

[2.2.4. Ручной мониторинг 29](#_Toc389828746)

[2.3. Автоматизированные методы 29](#_Toc389828747)

[2.3.1. Мониторинг социальных сетей 29](#_Toc389828748)

[2.3.2. Средства Data Mining 30](#_Toc389828749)

[2.4. Смешанные методы: совмещение автоматизированных и неавтоматизированных методов 31](#_Toc389828750)

[2.4. Эпоха Больших данных: начало перехода к новому понимаю данных 33](#_Toc389828751)

[2.5. Онлайн исследования в социологии: новые методы анализа данных 38](#_Toc389828752)

[Глава 3. Сетевой анализ как метод изучения виртуального пространства 45](#_Toc389828753)

[3.1. Основные метрики, используемые в сетевом анализе 47](#_Toc389828754)

[Глава 4. Обоснование понятия «ложная новость» 49](#_Toc389828755)

[Глава 5. Описание работы с системой мониторинга социальных сетей в качестве метода сбора данных 54](#_Toc389828756)

[5.1. Процедура анализа текстов: соотнесение с классическим контент-анализом 56](#_Toc389828757)

[5.2. Построение выборки и сбор данных 59](#_Toc389828758)

[Глава 6. Выявление различий в структурах реальных и ложных новостей в социальных сетях 61](#_Toc389828759)

[Глава 7. Определение характеристики «лидеров мнений» в структуре распространения новостей 63](#_Toc389828760)

[7.1. Анализ двух кейсов: подробное рассмотрение ключевых характеристик сети на примере политических новостей 66](#_Toc389828761)

[Кейс 1. Реальная новость: «Министр обороны: Италия может направить миротворцев на Украину» 66](#_Toc389828762)

[Описательная статистика 66](#_Toc389828763)

[Сетевые показатели 70](#_Toc389828764)

[Кейс 2. Ложная новость: «Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель» 74](#_Toc389828765)

[Описательная статистика 74](#_Toc389828766)

[Сетевые показатели 75](#_Toc389828767)

[Заключение 79](#_Toc389828768)

[Библиографический список 83](#_Toc389828769)

[Приложения 86](#_Toc389828770)

# Введение

На сегодняшний день внимание многих ученых, работающих в различных научных сферах, сосредоточено вокруг Интернет исследований. Особенно актуальными становятся работы, посвященные изучению феномена Больших данных. Высокие возможности современных компьютеров, широкое распространение Интернета, возрастающая доступность средств связи для различных групп населения сопровождается ежедневным приростом информации.

В виртуальном пространстве пользователи оставляют огромное количество информационных следов – публичные сообщения, фотографии, заметки, видео, личные данные. Социальные сети при этом, становятся удобной платформой для пользователя (для интеграции всех этих действий на одном сайте), так и для исследователей (для наблюдения за виртуальными процессами). Особое значение в области изучения социальных сетей придается исследованиям структуры виртуальных взаимодействий, информационному потреблению, а также проблемам доверия к информации в виртуальном пространстве.

Пространство социальных сетей предоставляет широкие возможности не только отдельным пользователям, но и различным организациям. Во-первых, это позволяет установить быстрый контакт с потребителем/читателем, а во-вторых, повысить охват распространяемой информации. В виртуальный диалог вовлечены не только производители каких-либо товаров, но и представители СМИ как ведущих национальных каналов и газет, так и отдельных частных изданий. Между такими акторами СМИ происходит борьба за читателей, которые в свою очередь должны буквально прорываться через непрекращающиеся потоки информации.

Однако сама структура таких информационных потоков мало известна. Она имеет сходные характеристики с оффлайновыми сетевыми моделями взаимодействия, но не может быть сведена к ним. Виртуальное пространство отличается высокой степенью анонимности таких сетей и иному распределению связей между акторами.

Огромное количество источников информации, разнообразное изложение одного и того же материала затрудняет ориентацию в анонимном пространстве виртуальной сети. В данном контексте особенно важным становится рассмотрение новостной коммуникационной сети. В виртуальном пространстве существует огромное множество различных новостных агентств, сайтов и отдельных акторов, распространяющих информацию, используя определенные механизмы действий. При повышении числа источников, транслирующих различную, часто противоположную информацию об одном и том же событии, снижается уровень определенности по отношению достоверности этой информации.

Следовательно, необходимо изучение информационных поводов и их структуры. Именно в таких случаях, когда информационный повод получает широкое распространение и обсуждение со стороны пользователей, важно понять, кто является лидером мнений в данной сети, из каких источников предоставляется информация, какова структура взаимодействий участников сети. Такое комплексное представление о распространении информации особенно важно в отношении ложных (содержащих недостоверную информацию) новостей, целью которых является создание информационной напряженности относительно какого-либо объекта/персоны/события.

В этом контексте становится особенно важным создание и развитие новых методов онлайн исследований. Отдельное место при этом занимают автоматизированные методы сбора и анализа информации, которые до сих пор остаются малоизученными и трудно применимыми в социологических исследованиях. Данная проблема тесно связана с недоступностью существующих программ для исследователей и сложностью самостоятельного создания специализированных программ для каждого необходимого вида анализа. Более того, существующие автоматизированные системы сбора информации не предназначены для социологического исследования, что обусловлено минимальным набором функций для анализа полученной информации.

Теоретической основой данной работы является концепция американского социолога Пола Лазарсфельда о «двухступенчатой коммуникации и лидерах мнений»[[1]](#footnote-1). Согласно данной теории распространение информации от СМИ к обществу происходит через лидера мнений – персоне, обладающей большим уровнем доверия у групп населения. Далее посредством социальных сетей и связей охват распространения новости увеличивается. Более того, Лазарсфельд установил, что при коммуникации с лидерами мнений происходит не только увеличения распространения новости, но и влияние этой информации на аудиторию.

Следовательно, в процессе коммуникации важно не средство передачи информации (и, возможно даже не ее содержание), а позиция лидера мнений. Чем больше доверия эта персона имеет в глазах публики, тем больше будет охват.

В данной работе предполагается, что двухступенчатую модель коммуникации Лазарсфельда можно проследить и в виртуальном пространстве социальных сетей. То есть, информация СМИ размещенная на странице человека, обладающего статусом «лидера мнений» (человек, активно ведущий деятельность на страницах социальных сетей, обладающий большим числом виртуальных друзей, подписчиков), во-первых, сразу же получает большой охват аудитории, а во-вторых, повышает вероятность широкого распространения новости (за счет большого числа подписанных пользователей на страницу «лидера мнений»).

Следует отметить, что в данном контексте понятие «лидер мнений» имеет некоторые дополнительные характеристики (в сравнении с концептом Лазарсфельда):

1. **Количество «друзей» как социальный капитал** – число виртуальных подписчиков является одной из наиболее важных характеристик. Так, если у человека, допустим, 1000 друзей в социальной сети, следовательно, любая размещенная информация на данной странице потенциально может быть просмотрена тысячей людей. Если каждый из них сделает «репост» этой информации (разместит специальную ссылку на своей странице с указанием автора новости), то количество людей, которые увидят эту новость, будет увеличиваться в геометрической прогрессии. Далее по принципу «сарафанного» радио новость будет распространяться по аудитории.
2. **Персона «лидера мнений»**

Следует разделить лидеров мнений на две группы: реальные люди и сообщества.

В первом случае «лидер мнений» - это страница реально существующего человека (возможно, публично известного, например, политик, певец и так далее), у которого есть своя аудитория людей, доверяющих ему, которые подписаны на его аккаунт в социальных сетях. Социальная сеть необходима как площадка для общения этого актора с его аудиторией, мгновенного информирования .

Во втором случае – это популярные сообщества (группы) в социальных сетях, которые объединяют большое количество людей. В зависимости от выкладываемого контента эти группы можно разделить на: развлекательные, познавательные и так далее.

Более того, лидерами мнений могут быть аккаунты людей (или сообществ), которые на самом деле являются «фейковыми» пользователями[[2]](#footnote-2). Это специально созданные учетные записи, которые используются для имитации сетевой активности вокруг каких-либо процессов, происходящих в виртуальном пространстве.

**Исследовательская проблема**

В анонимном виртуальном пространстве ведущую роль в распространении информации играют «лидеры мнений» - это те акторы, которые имеют огромное количество «друзей» (социальный капитал), активно ведущие свою страницу, размещающие контент и так далее. Это могут быть как профили известных людей, так и просто активные люди (в частности блоггеры), а также информационные агентства. Размещенная таким актором новость потенциально может быть прочитана всеми, кто подписан на данный профиль. Лидер мнений, участвуя в распространении новостной информации, может способствовать повышению интереса к ней со стороны подписчиков, которые, в свою очередь, также могут разместить ее на своей странице, что автоматически повышает охват новости.

Проблема заключается в то, что на индивидуальном уровне невозможно составить полное представление о составе сети, ее характеристиках и взаимосвязях. Необходимо детальное рассмотрение структуры распространения новостной информации в социальных сетях.

**Исследовательский вопрос**: отличаются ли структура распространения «реальной» и «ложной» новости?

**Цель исследования**

Анализ структуры распространения новостной информации в социальных сетях с применением автоматизированных средств-сбора данных (система мониторинга социальных сетей) и методов сетевого анализа.

Задачи дипломного исследования заключаются в следующем:

1. Определить и операционализировать понятие «ложной новости»
2. Описать основные принципы работы с инструментом мониторинга социальных сетей, определить существующие ограничения метода сбора данных, соотнести с построением выборки
3. Выявить различия в структуре реальных и ложных новостей
4. Определить характеристики «лидеров мнений» - акторов с большим числом связей с остальными участниками сети

**Эмпирический объект исследования**: совокупность информационных (новостных) сообщений, размещенных в социальной сети Twitter в период с апреля по май 2014 года. Единица анализа – единичное сообщение.

**Теоретический объект:** новостная информация («реальная» и «ложная»), распространяемая в социальных сетях

**Предмет исследования**: структура распространения реальных и ложных новостей

В выборку эмпирического исследования вошли 3 «реальных» и 3 «ложных» новости, опубликованных в апреле-мае 2014 года. Новости анализировались за недельный период (7 дней), начиная с момента публикации.

В качестве основной интернет-площадки для исследования распространения новостей была выбрана социальная сеть Twitter. Данный выбор обусловлен особенностями архитектуры данной социальной сети: Twitter – это микроблог, что означает публичный обмен короткими сообщениями (до 140 символов). Такая организация коммуникационной сети используется, с одной стороны, для оперативного информирования подписчиков о чем-либо, с другой стороны, для быстрого реагирования на публикуемые записи[[3]](#footnote-3). Присутствие различных СМИ, а также высокая интеграция с внешними (по отношению к социальной сети Twitter) ресурсами, публичность коммуникаций позволяет выделить данную социальную сеть как особую информационную среду.

Таблица №1

**Кейсы, вошедшие в выборку**

|  |  |  |  |
| --- | --- | --- | --- |
| **Заголовок** | **первоисточник** | **дата публикации** | **Число упоминаний в сети Twitter** |
| **Ложные новости** |
| В Москве после встречи с Путиным пытался застрелиться директор "Первого канала" | http://ivona.bigmir.net/showbiz/stars/391093-Konstantin-Ernst-pytalsja-zastrelit-sja-posle-razgovora-s-Putinym | 03.05.14 | 108 |
| В Донецке сепаратисты расклеили антисемитские листовки: евреи должны заплатить $50 за регистрацию | http://gordonua.com/news/separatism/V-Donecke-separatisty-raskleili-antisemitskie-listovki-evrei-za-50-dolzhny-proyti-registraciyu-18609.html | 16.04.14 | 101 |
| Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель | http://ruskline.ru/politnews/2014/04/28/minoborony\_rf\_nebolshoj\_sboj\_kursa\_rossijskoj\_rakety\_rs24\_yars\_pentagon\_eto\_provokaciya\_pvo\_ssha\_ne\_smogli\_vzyat\_cel/ | 28.04.14 | 390 |
| **Реальные новости** |
| Министр обороны: Италия может направить миротворцев на Украину | http://ria.ru/world/20140504/1006495411.html | 04.05.14 | 460 |
| Фильм Андрея Звягинцева "Левиафан" получил приз Каннского кинофестиваля за лучший сценарий | http://www.1tv.ru/news/culture/259554 | 24.05.14 | 1359 |
| В иркутской школе №11 появилась "Азбука вежливости" | http://vesti.irk.ru/news/society/164374/ | 05.05.14 | 331 |

**Предполагается**, что структуры «ложных новостей» будут похожи по строению и характеристикам «лидеров мнений» между собой, и при этом, будут основательно отличаться от структур «реальных новостей». То есть, по анализируемым данным будет возможно выделить явные сходства и различия в структуре распространения «реальных» и «ложных» новостей.

**Ограничения исследования**

1. **Доступ к информации в социальных сетях**

Политика социальных сетей предоставляет человеку возможность защитить свои данные и размещенную информацию от публичного просмотра. Следовательно, даже если человек активно распространяет и публикует информацию и новости, но при этом закрывает общий доступ к странице, данная единица уже не будет включена в выборку.

Исследователь может работать только с той информацией, которая доступна для публичного доступа.

Более того, общая картина распространения информации может быть искажена за счет того, что люди могут и увидеть новость, отреагировать на нее, но не примут решения распространять ее дальше, либо писать комментарии к ней. Такие «пассивные пользователи» не учитываются в общих подсчетах, так как невозможно с достаточной уверенностью утверждать видели они новость или нет, передали ее другим людям (сообщением или же устно) или нет. Возможно говорить только о потенциальном охвате такой аудитории.

1. **Принцип работы системы мониторинга социальных сетей.**

**Мониторинг социальных сетей**— это выборка текстов пользователей социальных сетей, основанная на тех или иных критериях отбора данных текстов.

Это определенный способ сбора данных, осуществляемый специальным программным обеспечением, в котором исследователь может задать ключевые слова для поиска, дату мониторинга, список исследуемых ресурсов. Все это автоматически выгружается в базу данных и уже может быть предметом анализа. Сама база представляет собой, помимо заданных критериев, конкретные сообщения пользователей, их «координаты» в Сети (никнейм, индивидуальная ссылка на сообщение, география и возраст, если это было указано респондентом). Более того, в этой же базе мы получаем сведения о потенциальном количестве пользователей, увидевших это сообщение, либо отреагировавших на него. Таким образом, мы получаем агрегированную базу текстов, с уже заданными кодами (тегами, словами поиска), которую можем анализировать как количественно, так и качественно.

В силу того, что данное программное обеспечение может осуществлять поиск только по собственной базе ресурсов (ссылок на различные типы социальных медиа), следовательно, нельзя говорить о генеральной совокупности источников.

В данном случае, этим ограничением можно пренебречь. Значительная доля коммуникаций происходит на широко известных платформах (таких как Twitter, Facebook, Livejournal, Вконтакте, Одноклассники), которые неизменно входят в выборку при любом виде мониторинга социальных сетей.

**Основные результаты**

В результате проведения сетевого анализа было выявлено, что оба типа информационных поводов имеют схожие характеристики по статистикам моделей. Акторы слабо связаны между собой, невозможно выделить кластеры.

Различия заключаются в индивидуальных характеристиках акторов каждой сети. В частности, ключевую роль играют такие пользователи, которые являются «лидерами мнений». Такие вершины имеют не только большую мощность узлов, но и большое количество подписчиков в сети Twitter.

 Отдельно необходимо отметить специфику социальной сети Twitter в контексте изучения распространения новостей в данном виртуальном пространстве. В обоих случаях (реальных и ложных новостей) большое количество упоминаний связано с искусственным наращиванием объема публикуемых сообщений.

Речь идет об астротурфинге (astroturfing) – это искусственное формирование обсуждений, путем распространения информации, публикации комментариев по какой-либо теме на различных ресурсах с помощью специально созданных аккаунтов, имитирующих деятельность реального человека[[4]](#footnote-4). Это может быть набор аккаунтов, которые периодически обновляют информацию, ведут свою страницу и внешне ничем не отличаются от других профилей реальных людей. Также посредством специальных программ возможно создание многочисленных профилей и автоматическое распространение информации через данные аккаунты. Такие искусственные «боты» позволяют за короткое время разместить одну и ту же информацию якобы от разных людей.

Высокая доля «фейковых» аккаунтов, ведущих искусственное наполнение информационного пространства сети, является ключевой особенностью распространения и реальных, и ложных новостей в Twitter.

# ГЛАВА 1. ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ ИЗУЧЕНИЯ НОВОСТНОЙ ИНФОРМАЦИИ В СОЦИАЛЬНЫХ СЕТЯХ

## 1.1.Современная Россия в Интернете: стремительное увеличение пользователей Сети

Охват Интернета в России сравнительно маленький: лишь 49% всей территории страны (в западноевропейских странах он равен 79%). Это означает, что практически половина нашей страны просто не имеет возможности выйти в сеть. Основная проблема заключается в труднодоступности некоторых регионов страны, недостаточном финансировании этого вопроса, непонимании его важности и значения.

При этом по данным Фонда Общественного мнения месячная доля пользователей Интернета в России равна 57,8 млн. человек, что составляет 50% населения России старше 18 лет (2012 год), по сравнению с прошлым годом этот показатель вырос на 15%. В день данный показатель составляет 44,3 млн. человек (38% совершеннолетнего населения страны), по сравнению с 2011 годом, показатель вырос на 22%[[5]](#footnote-5).

Все это указывает, что российская аудитория Интернета активна, и с каждым годом она возрастает, охватывая новые населенные пункты страны. При этом важно отметить, что аудитория пользователей мобильного Интернета растет в два раза быстрее стационарного (по данным TNS).

Пользование Интернетом включает в себя множество всевозможных действий, которые при этом могут совершаться одновременно. Так, человек может общаться с кем-либо в социальных сетях, искать необходимую ему информацию, совершать покупки, оплатить счет и так далее. При этом, именно социальные сети играют наиболее интересную и значительную роль в практиках пользования Интернетом. 74,5% россиян посещают хотя бы 1 социальную сеть в месяц. Ежемесячная аудитория наиболее популярных социальных сетей (Вконтакте, Однокласники) составляет приблизительно 30 млн. человек в месяц, при этом, подавляющее большинство пользователей посещают эти сайты не реже 1 раза в неделю (Вконтаке – 77% посетителей, Одноклассники – 81%).

Социальные сети становятся не просто средством общения, но особыми Интернет-площадками, которые можно по-разному использовать для определенных целей и задач различных проектов. Так, например, в сети создаются сообщества по интересам, где выкладывается определенный контент. При этом, человек может состоять в неограниченном количестве сообществ, может также создавать их самостоятельно.

Аудиторию Интернет-площадок можно разделить на 3 группы:

1. Создатели контента

Именно эти люди задают информационный фон посредством публикации различных фотографий, информации, текстов, видео и так далее. Чем больше влияние у этого человека (например, создатель популярного сообщества, политический деятель, звезда шоу-бизнеса и так далее), тем больше будет значимость данного контента в кругу поддерживающих его людей.

1. Активные участники

Постоянно комментируют, реагируют на выкладываемый контент, ведут диалоги, выстраивают темы обсуждений. Это наиболее ценная группа пользователей.

1. Пассивные пользователи

Просматривают страницы сообществ, могут также в них состоять, просматривают обсуждения и комментарии, но сами ничего не пишут и не выкладывают. Такая аудитория сама многочисленная, и даже не смотря на то, что они пассивны, они также могут иметь значимость в контексте различных исследований. Ведь, тот факт, что они состоят в группе, уже может косвенно свидетельствовать о том, что данная тема им интересна, они видят информацию, размещаемую в сообществе, а, значит, также являются объектом влияния.

Благодаря современным средствам связи, информация в Интернет попадет практически сразу после события (или в некоторых ситуация и во время него). Важным становится, как эта информация преподносится. С одной стороны, это могут быть официальные сообщения новостных агентств, а с другой – собственная интерпретация происходящего. Кому в данном случае верить решает сам пользователь, при этом у него появляется возможность сравнения различных источников, но, что более важно, практически всегда может реагировать на это. Он может прокомментировать саму запись, либо написать что-либо самостоятельно, тем самым он вовлекается в общественную дискуссию, не выходя из дома, реагируя непосредственно во время события, то есть часто выдает свое первое мнение, реакцию.

## 1.2. ***Социальные сети: структура коммуникации online vs. offline***

Одна из простейших форм передачи информации – это коммуникация. В то же время это многогранное понятие, включающее в себя различные особенности и характеристики. Научное исследование коммуникаций доказывает наличие сложных механизмов и различных форм, так, например, коммуникации исследуются в психологии, журналистике, политологии, социологии и так далее. Существуют различные теории об общественных коммуникациях, которые описывают структуру и особенности различных форм взаимодействия людей, групп и организаций. При этом, с развитием и усложнением общества появляются новые виды коммуникаций, что приводит к необходимости постоянного развития существующих теорий.

Взаимодействие людей невозможно без установления каких-либо связей, отношений между ними. Так, в течение своей жизни человек неизменно становится участником какой—либо социальной сети. Однако процесс коммуникации не ограничивается только межличностном взаимодействии.

Социальная сеть имеет множество определений, которые похожи по своему смыслу, но различны по своей форме и области применения. Изначально термин «социальная» сеть относится к изучению взаимосвязей между людьми. Впервые этим начал заниматься ученый Я.Л. Морено, которого считают основателем дисциплины сетевого анализа. Однако, самые ранние работы в этом направлении можно увидеть у Г. Зиммеля и Э. Дюркгейма, а также у известного антрополога Рэдклифф-Браунна. Сетевой анализ связан с построением социограмм и графов, визуально представляющих связи между людьми (точки – отдельные индивиды, линии – связи между ними). Такой анализ эффективен при изучении групп для понимания структурных особенностей взаимодействий между индивидами. Существуют различные виды социальных сетей: профессиональные, личные, досуговые, финансовые, и так далее. При этом, один человек может принадлежать одновременно нескольким сетям, входить в новые и уходить (либо исключатся) из них. Социальные сети можно также разделить на формальные и неформальные, горизонтальные и вертикальные.

При исследовании социальных сетей необходимо также рассматривать и позиции отдельных точек, то есть индивидов в ней. Так, можно выделить ядро социальной сети (точка с наибольшим количеством взаимосвязей), определить плотность сети, произвести кластеризацию индивидов. В основе анализа лежит математическая теория графов.

Другое определение социальной сети непосредственно связано с Интернет-платформами и программами. Так, социальную сеть можно определить как специализированно сконструированный сайт, предоставляющий возможность создания личного профиля (страницы), добавления и публикации какого-либо контента (фотографии, видео, заметки, комментарии), а также общения с другими пользователями, зарегистрировавшихся в данной сети.

В качестве примера социальных сетей можно привести наиболее популярные сегодня платформы: Facebook, Вконтакте, Одноклассники, Twitter, Instagram, Livejournal и так далее. У пользователей есть возможность создать собственный аккаунт, заполнить анкету о себе, «наполнить» свою страницу фотографиями, музыкой, видео, добавить «друзей». В данном контексте слово «друзья» было взято в кавычки, так как в виртуальном пространстве не существует вербального различия между друзьями, знакомыми, коллегами и так далее. Так, например, человек в реальной жизни может назвать друзьями троих людей, все же остальные – более слабые связи (знакомые, коллеги), либо более сильные, родственные (жена, муж, родители). В пространстве социальных сетей все эти люди (число которых может исчисляться сотнями) будет именоваться «друзьями».

Основная особенность виртуальных социальных сетей – это их глобальность. Человек может отправлять мгновенные сообщения на другие континенты, участвовать в групповых дискуссиях, искать информацию, публиковать ее на всеобщее обозрение и так далее. Многие реальные практики успешно интегрируются в виртуальное пространство социальных сетей.

Виртуальные сети могут объединить в себе все виды социальных сетей. Здесь люди могут вступать в группы по интересам, профессиональным или учебным нуждам (например, сообщество одногруппников для обмена актуальной информацией), финансовым и так далее и все это на одной платформе (например, Вконтакте). При этом, благодаря высокому развитию технологий и доступности средств связи все большему числу людей с различным материальным положением, число пользователей неизменно растет. Люди получают возможность практически беспрерывного общения, получения информации, мгновенного реагирования на события.

## 1.3. Типология социальных сетей виртуального пространства

Интернет-платформы можно разделить по особенностям строения и целей сайта и предоставляемых (сайтом) возможностей:

1. **Социальные сети** (Facebook, Вконтакте, одноклассники).

Сервис предоставляет также возможность публикации фотографий, видео, аудио, собственных записей, цитирования чужого контента

Цель – рассказать о своих новостях. «Что у Вас нового?» - такой вопрос демонстрируется человеку при просмотре его личной страницы. Социальная сеть как бы побуждает его написать и опубликовать последние новости, мысли, события, произошедшие с человеком. Также предполагается, что этот пользователь просмотрит и новости своих «друзей», возможно, оставит комментарий или выразит одобрение («лайк», «репост»). Все это нацелено на сравнительно длительное пребывание на данных сайтах.

Такие платформы предоставляют наиболее широкие возможности общения и виртуальной активности. Сервисы включают в себя и личные сообщения, возможности комментирования чужих записей, размещения контента практически любых форматов (тест, видео, изображения и так далее).

1. **Микроблоги** (Twitter) – публикация коротких заметок (до 140 символов), сервис для публичного обмена сообщениями

Цель – рассказать, что происходит

Ограничение на количество символов, а также особенности самой платформы сайта формируют особый тип общения: быстрый (в сравнении с социальными сетями) и четкий (человек должен сформулировать свою мысль буквально в 1-2 коротких предложениях). Это констатация факта, события, эмоции в момент их проявления. Это мгновенная (насколько это позволяет наличие доступа к Интернету) реакция на что-либо, происходящее в жизни человека.

Данная платформа не подразумевает возможности личных диалогов, пользователи общаются в комментариях к записи, что подразумевает публичность беседы. В зависимости от личных настроек, такие записи и комментарии к ним могут быть видны всем, либо только друзьям данного пользователя. Однако даже последнее ограничение публичности предполагает видимость такой беседы всем друзьям пользователя, что уже невозможно назвать личным диалогом.

1. **Сервисы для обмена фотографиями и видео** (Instagram) – возможность публикаций своих фотографий или коротких видео, применения различных фото-фильтров

Цель – показать, что происходит

Данный сервис предоставляет возможность только для визуальной передачи информации. При этом, существует ограничение на продолжительности видео (30 секунд). Также как и в микроблогах здесь отсутствует возможность передавать личные сообщения. Человек может выразить свое мнение в поле для комментариев, «лайкнуть» запись или же разместить у себя на странице (с отсылкой к автору).

Сам процесс передачи информации еще более быстрый и ограниченный. Пользователь может сопроводить свое фото какими-либо подписями (комментариями), однако их длина, как правило, тоже ограничена.

Как правило, большинство современных подобных платформ поддерживает функцию интеграции со всеми другими видами сервисов. То есть, размещенная на одном из сайтов информация автоматически дублируется и на другие (при условии согласия пользователя и его регистрации на всех сервисах).

На сегодняшний день, социальные сети становятся объектом внимания многих исследователей. В маркетинге изучаются и совершенствуются принципы общения с клиентами (SMM – Social Media Marketing). В психологии исследуются личные страницы пользователей, их контента, активности и так далее. Появляется новое направление в социальной антропологии – нетнография (netnography), изучающая виртуальные сообщества их внутренние правила и особенности взаимодействия. Для социологических исследований объектом изучения зачастую становятся блоги, где люди в открытой форме излагают свои мысли, мнения.

Главная особенность изучения социальных связей в виртуальном пространстве – получение большого массива данных автоматизированными способами, регулярное отслеживание изменений в режиме реального времени.

# Глава 2. Исследования Интернет-пространств. Основные подходы в России и за рубежом

На сегодняшний день не существует единой классификации методов Интернет-исследований. Во многом это связано с тем, что данная сфера только начинает развиваться, многие методы еще не опробованы. Более того, до сих пор перед исследователями стоит проблема анализа таких данных, обеспечения качества и репрезентативности выборки и так далее.

Однако многие исследователи задумываются о том, как можно разделить существующие методы. Например, в работе И.Ф. Девятко «Инструментарий онлайн-исследований: попытка каталогизации" предлагается предварительная классификация методов онлайн-исследований по типам и источникам данных[[6]](#footnote-6):

В данной классификации основой для различения методов является степень вмешательства исследователя в анализируемое им пространство. Таким образом, реактивные методы предполагают непосредственное участие и взаимодействие исследователя и респондентов (например, при помощи анкеты). То есть, данные, полученные в ходе такого применения метода, отражают реакцию респондентов на интересующие исследователя вопросы. Такой тип метода в целом похож на традиционный анкетный опрос, за исключением того, что все взаимодействие и весь процесс происходит в Интернет-пространстве.

Второй тип – нереактивные методы (малозаметные) – не предполагает влияние исследователя на респондента. То есть, собираются готовые данные, либо производится наблюдение за интересующей группой респондентов в Интернете. Данные методы можно сопоставить с «традиционными» методами включенного/невключенного наблюдения, контент-анализом готовых текстов (например, газетных статей). Основной принцип в данном случае – получить «чистые» данные, не спровоцированные вопросами исследователя, в естественной среде респондентов.

В то же время, в работе «Социальные сети и социометрические исследования: теоретические основания и практика использования автоматизированного инструментария изучения виртуальных сообществ»[[7]](#footnote-7) исследователями была так же предложена классификация онлайн-методов по принципу сбора данных: автоматизированные и неавтоматизированные методы (см. рис. 1).

|  |
| --- |
| Не указаноРис.1 Классификация онлайн-методов по принципу сбора данных |

## Концепция «лидеров мнений» Пола Лазарсфельда. Соотнесение с онлайн медиа

В основе данной работы лежит концепция американского социолога Пола Лазарсфельда о «двухступенчатой коммуникации и лидерах мнений»[[8]](#footnote-8). Согласно данной теории распространение информации от СМИ к обществу происходит через лидера мнений – персоне, обладающей большим уровнем доверия у групп населения. Далее посредством социальных сетей и связей охват распространения новости увеличивается. Более того, Лазарсфельд установил, что при коммуникации с лидерами мнений происходит не только увеличения распространения новости, но и влияние этой информации на аудиторию.

Следовательно, в процессе коммуникации важно не средство передачи информации (и, возможно, даже не ее содержание), а взаимодействие с лидером мнений. Чем больше доверия эта персона имеет в глазах публики, тем больше будет охват.

В данной работе предполагается, что двухступенчатую модель коммуникации Лазарсфельда можно проследить и в виртуальном пространстве социальных сетей. То есть, информация СМИ размещенная на странице человека, обладающего статусом «лидера мнений», во-первых, сразу же получает большой потенциальный охват, а во-вторых, повышает вероятность широкого распространения новости (за счет доверия людей к данной персоне).

Следует отметить, что в данном контексте понятие «лидер мнений» имеет некоторые дополнительные характеристики (в сравнении с концептом Лазарсфельда):

1. Количество «друзей» - число виртуальных подписчиков - одна из наиболее важных характеристик. Так, если у человека, допустим, 1000 друзей в социальной сети, следовательно, любая размещенная информация на данной странице потенциально может быть просмотрено тысячей людей. Если каждый из них сделает «репост» этой информации (разместит специальную ссылку на своей странице с указанием автора новости), то количество людей, которые увидят эту новость, будет увеличиваться в геометрической прогрессии. Далее по принципу «сарафанного» радио новость будет буквально расходиться по аудитории.
2. Персона «лидера мнений»

Следует разделить лидеров мнений на две группы: реальные люди и сообщества.

В первом случае «лидер мнений» - это страница реально существующего человека (возможно, публично известного, например, политик, певец и так далее), у которого есть своя аудитория людей, доверяющих ему, которые подписаны на его аккаунт в социальных сетях. Социальная сеть необходима как площадка для общения этого актора с его аудиторией, мгновенного информирования .

Во втором случае – это популярные сообщества (группы) в социальных сетях, которые объединяют большое количество людей. В зависимости от выкладываемого контента эти группы можно разделить на: развлекательные, познавательные и так далее.

Более того, лидерами мнений могут быть аккаунты людей (или сообществ), которые на самом деле являются «фейковыми» пользователями[[9]](#footnote-9). Это специально созданные учетные записи, которые используются для имитации сетевой активности вокруг каких-либо процессов, происходящих в виртуальном пространстве

**Выводы**

*Российская интернет-аудитория стремительно расширяется с каждым годом, вместе с этим растет число возможностей, предоставляемых пользователям онлайн средой. Различные виды деятельности человека успешно трансформируются в электронный формат, меняя привычные практики людей и общество в целом.*

*Особое значение приобретает изучение различных видов коммуникаций, которые можно отслеживать с помощью изучения социальных сетей. Практически любая интернет-платформа предоставляет стандартный набор опций для пользователя, желающего отобразить свою социальную жизнь в виртуальном пространстве. Такие наиболее популярные сайты как Facebook, Вконтакте, Одноклассники, Twitter, Instagram вошли в ежедневную практику огромного количества людей по всему миру.*

*Через такие интернет-платформы проходят различные потоки неструктурированной информации, которые, также как и в офлайне, могут быть изучены и тереотеризированы. Ведь по мере того, как происходит повсеместная компьютеризация общества, все больший прирост информации, накопление огромных массивов данных возрастает необходимость изучения таких явлений.*

## Неавтоматизированные методы

### Нетнография

Интернет – это глобальная сеть данных, которые используются, создаются, обмениваются миллионами людьми ежедневно. Люди общаются в социальных сетях, форумах, создают свои группы, страницы. То есть, в Интернете постоянно происходят какие-либо социальные действия. Например, создается группа помощи больным детям, где люди могут помочь переводом денег на счет. При этом в Сети очевидно уже существуют определенные нормы и правила, которые регулируют такие действия. Например, создатель группы должен постоянно поддерживать ее активность, но, что наиболее важно, быть честным и убедить в этом других, ведь Интернет все еще является анонимным пространством.

Все это интересно для изучения, и уже сегодня появилось такое направление как нетнография. Хотя ее определение сформировал еще 1990 году Роберт Козинец, антрополог и этнограф, до сих пор это направление не было развито и не пользовалось популярностью. Ученый говорил о нетнографии как о процессе изучения виртуальных культур и сообществ, основанных на Интернет-коммуникациях[[10]](#footnote-10).

Это этнография Сети, где изучаются нормы, правила, сложившиеся обычаи и традиции в Интернет-пространстве. В контексте данного направления можно было бы поставить такие вопросы как: почему люди доверяют друг другу в Интернете? Есть ли какие-то параметры, по которым можно отличить обман? Как формируются и «живут» группы? Как происходит принятие решения о приеме или отказе вступления в сообщество (если оно закрыто). Ведь условно Интернет сообщества можно поделить на «реально существующие» и «существующие только в Сети». Первые состоят из тех, кто действительно существует как группа в офф-лайне, например, студенческая группа, группа выпускников и так далее. То есть, эти люди взаимодействовали друг с другом помимо виртуального пространства. Вторые же, группируются скорее всего изначально в Сети, то есть до этого они никогда не виделись и в будущем также вряд ли будут взаимодействовать вне этой виртуальной группы. Конечно, существуют и такие сообщества, которые, формируясь изначально в Сети, позже устраивают какие-либо встречи, либо, например, создается сообщество, посвященное какой-либо популярной группе перед их концертом в данном городе. Люди могут вступать в это сообщество, общаться, многие даже ищут компанию для концерта, и все они потенциально встречаются в реальности.

В данном случае, такое различие нужно для того, чтобы понять отношения и правила взаимодействия в виртуальных сообществах. Так, если говорить о «реально существующих» группах, то, скорее всего, отношения между их членами сложились в реальном мире и регулируются там же, и в эту виртуальную группу переносятся уже существующие нормы и правила этой группы, а такое сообщество просто служит инструментом для каких-либо организационных вопросов. Более того, такие группы чаще всего закрытые, то есть, чтобы вступить в сообщество, необходимо подать заявку администраторам группы, которые будут решать, принимать этого человека или нет. В данном случае решение будет приниматься, исходя из того, знают ли этого человека в реальности хотя бы кто-то из группы и как он может быть с ней связан.

«Существующие только в Сети» наиболее интересны для нетнографии. Такие группы в большинстве своем анонимны, то есть, участники сообщества изначально не знакомы друг с другом, никак не взаимодействуют в реальности и вряд ли будут. Вступить в такие группы проще, чаще всего они открыты, они больше по численности. Однако их социальные взаимодействия вполне реальны. При этом, понятно, что существуют определенные механизмы взаимодействия, общения, маркеры, по которым люди могут определить обманщиков и так далее. Более того, анонимность Интернет-пространства, знание того, что человек никогда не встретится лицом к лицу с участниками этого сообщества, раскрепощает его, что позволяет исследователю увидеть настоящее мнение человека.

### Выборочный метод

Производится контент-анализ с помощью количественных методов на случайной выборке. Например, в исследовании С. Хиндуйа и Дж. Патчин рассматривались личные профилей подростков в сети My Space[[11]](#footnote-11), исследователи формировали выборку при помощи генератора случайных чисел. Это теоретически возможно, так как каждая страница имеет уникальный номер. Однако авторы статьи столкнулись с множеством ограничений данного метода: закрытые или удаленные страницы, пользователи неподходящего возраста, политика My Space, которая автоматически закрывает публичный доступ к страницам подростков 14-15 лет (то есть, информацию о пользователе могут видеть только его друзья).

### Анализ поисковых запросов

Такие данные можно получить в свободном доступе от различных поисковых машин (Yandex, Google и так далее). Это частота поисковых запросов по определенным тематикам. Так, можно увидеть, что на данный момент волнует аудиторию Интернет сети, сравнить эти цифры по годам или странам.

Так же это может помочь исследователю найти актуальные темы для исследования, составить список ключевых слов, что может быть полезно для концептуализации и операционализации исследования.

### Ручной мониторинг

Сбор информации посредством самостоятельного поиска. То есть, исследователь, при помощи поисковых машин (таких как Google, Yandex, Mail и другие) составляет запрос на интересующую его информацию. Далее он может просмотреть список выданных страниц и выбрать ту информацию, которая ему необходима, скорректировать свой запрос и так далее. Такой метод подходит скорее для предварительного ознакомления с исследовательским объектом, так как объем найденных страниц с упоминанием ключевых слов может превышать миллионные значения, что делает невозможным более детальный анализ объекта.

## Автоматизированные методы

### Мониторинг социальных сетей

На данный момент используется преимущественно в сфере маркетинга и PR, однако, по прогнозам специалистов, этот метод в скором времени выйдет за эти рамки и начнет широко использоваться в социологических исследованиях.

Люди открыто выражают свое мнение по различным вопросам. При этом это мнение не искажено эффектом интервьюера, ситуацией опроса, либо самим вопросом. Более того, это мнение может распространяться по определенным группам и пользователям, что уже может представлять основу для анализа.

**Что такое мониторинг социальных сетей?**

**Мониторинг социальных сетей** — это выборка текстов пользователей социальных сетей, основанная на тех или иных критериях отбора данных текстов. А так же процесс сбора данной выборки и её интерпретации.[[12]](#footnote-12)

Это определенный метод сбора данных, осуществляемый специальным программным обеспечением, в котором исследователь может задать ключевые слова для поиска, дату мониторинга, список исследуемых ресурсов. Все это автоматически выгружается в базу данных и уже может быть предметом анализа. Сама база представляет собой, помимо заданных критериев, конкретные сообщения пользователей, их «координаты» в Сети (никнейм, индивидуальная ссылка на сообщение, география и возраст, если это было указано респондентом). Более того, в этой же базе мы получаем сведения о потенциальном количестве человек, увидевших это сообщение, либо отреагировавших на него. Таким образом, мы получаем агрегированную базу текстов, с уже заданными кодами (тегами, словами поиска), которую можем анализировать как количественно, так и качественно.

Полученная база позволяет применять различные методы работы с информацией. Так, можно увидеть динамику сообщений, всплески и спады количества сообщений, при более подробном контент-анализе можно выявить информационные поводы того или иного всплеска активности. Мы можем проследить за самыми активными участниками и лидерами мнений.

### Средства Data Mining

Буквально это можно перевести как «раскопка данных», также это называют «интеллектуальным анализом данных». Это связано с тем, что с того времени, как началась повсеместная регистрация пользователей и их действий в Интернете, появилось огромное количество данных (при этом как количественных, так и качественных), которые необходимо было анализировать. Например, человек, регистрирующийся на сайте интересующей его компании, предоставляет им как минимум свои социально-демографические характеристики. Более того, на сайтах идет учет посещаемости различных разделов, активности пользователей, возможность реагирования на их вопросы и так далее. Все эти данные требуют нетривиальных решений.

Математическая статистика оказалась в данном случае неэффективной, так как она в основном предназначена для проверки уже имеющихся гипотез, но, что более важно, использует методы осреднения, что дает фиктивные данные на таких выборках (то есть, получается «средняя температура по больнице»). В это время, задача Data Mining найти нетипичные случаи, образцы. Это анализ фрагментов выборки (подвыборки), чтобы найти там свои существующие закономерности.

## Смешанные методы: совмещение автоматизированных и неавтоматизированных методов

Они во многом объединяют или дополняют два вышеперечисленных метода :автоматический и неавтоматический. Это контент-анализ, Интернет-опросы и социометрия.

Так, например, при автоматизированном мониторинге социальных сетей исследователь получает выгрузку текстовых сообщений пользователей. Для расширения своего исследования он может не просто обозначить процентное распределение упоминаний, но также и провести качественный анализ сообщений, определить информационный повод, найти взаимосвязи. Более того, методами социометрии можно буквально увидеть, как распространялось сообщение, сколько людей потенциально могло его увидеть и так далее.

 Интересно отметить, что неавтоматизированные методы используются в основном для гуманитарных исследований, в то время как автоматизированные – для решения частных задач компаний, фирм, организаций. Возможно, наилучшим решением было бы совмещение двух этих методов, однако, на данный момент существуют серьезные ограничения такого подхода.

1. **Автоматизированные методы сбора недоступны для широкой аудитории.** Не все программное обеспечение данного типа, во-первых, просто в использовании, а, во-вторых, доступно по цене. Так, например, сервисы мониторинга социальных сетей предоставляют различные тарифы, исходя из которых, формируются опции доступные для пользователя.
2. **Ограниченность предоставляемой информации пользователями**. В базу данных попадает лишь то, что человек сам захотел написать. В основном, существует проблема таких неуказанных данных как: возраст и место жительства, в некоторых случаях отсутствует и информация о поле (в некоторых случая о поле респондента практически невозможно узнать, особенно если в качестве имени он использует псевдоним, ник).

Более того, человек может умышленно искажать информацию о себе, более того создавать несуществующих людей (например, «фейковые» страницы, где заполнена информация о человеке, его увлечениях, даже ведется какая-то активность, но на самом деле это может оказаться человек другого пола, возраста и так далее).

Сюда же можно отнести «закрытость» многих страниц пользователей, а также групп и сообществ. В данной ситуации, человек сознательно ограничивает круг лиц, а также свою информацию, размещенную на странице от нежелательных ему людей (это могут быть, как все те, кто не является другом этого пользователя, незарегистрированные люди, а также конкретные пользователи в данной социальной сети).

Так как взмывание страниц является незаконным, получить реальную информацию во многих случаях затруднительно. Однако на сегодняшний день, многие люди зарегистрированы не в одной социальной сети, более того, все эти сайты интегрированы друг с другом (то есть, при размещении информации в одной социальной сети, она автоматически копируется на другую страницу, если, конечно, человек выбрал данную опцию). Люди производят огромное количество фотографий, которые тут же «заливают» в сеть. Все это делает создание «фейковых» страниц неэффективным. Однако это лишь предположение.

## Эпоха Больших данных: начало перехода к новому понимаю данных

Современные технологии позволили перейти к качественно новому пониманию данных, информации и возможностей ее анализа. В виртуальном мире ежесекундно происходит огромный прирост новой информации: делаются записи в социальных сетях, регистрируются новые сайты, более того, привычная бумажная бюрократия также принимает электронный формат, превращаясь в огромные базы данных. Это ценнейшая информация, которая может быть использована практически в любой сфере общественной жизни.

Многие ученые из различных областей науки, таких как, биология, математика, информатика, психология, социология, политология, маркетинг видят огромный потенциал использования Больших данных. Более того, огромная мощность современных компьютеров (по сравнению с вычислительными машинами прошлых десятилетий) делает возможным быструю обработку, хранение и накопления таких больших баз данных. На сегодняшний день важной задачей становится понимание того как анализировать эти данные, как защитить персональную информацию пользователей и какие специалисты необходимы для данной работы.

Особенно актуальна эта дискуссия для гуманитарных наук, где исследователи «разделились» на две стороны[[13]](#footnote-13). Одни считают, что в эпоху Больших данных традиционный подход не эффективен и следует «позволить данным говорить за себя», а не пытаться объяснить данное социальное явление с помощью различных теорий. Другие же исследователи наоборот призывают к еще более ответственному подходу к социологическим теориям и необходимости применять их в контексте Больших данных.

Рассмотрим оба подхода более подробно.

Согласно первому из них анализ Большие данные перевернут традиционный подход к социологическим исследованиям и самому объекту познания. В книге Виктора Майера-Шенберга «Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим» предлагается новый подход к имеющимся данным: не искать причинно-следственную связь, а находить корреляции[[14]](#footnote-14). Это обусловлено характером данных, которые исследователь получает для дальнейшего анализа, а именно, их неструктурированность и неорганизованность. Нужно «позволить данным говорить» - это означает, что не нужно заранее продумывать гипотезы и составлять вопросы, это было необходимо в традиционных исследованиях. В ситуации с Большими данными, где очень сложно упорядочить входящие потоки информации, а зачастую и просто невозможно (из-а огромных денежных, трудовых и временных потерь), необходимо «погрузиться» в эти данные, пытаться найти там какие-либо закономерности, «выбросы», что-то новое, что дает новую информацию и решения.

В традиционных исследованиях, при невозможности охватить все случаи, используются выборки, которые позволяют масштабировать небольшое количество данных (относительно общего объема) на всю совокупность. При этом, полученные данные строго структурированы и формализованы, вопросы анкеты четко продуманы и отражают исходные цели и задачи исследования. Сами же результаты, в большинстве случаев, представляются как средние значения и вероятности.

В эпоху Больших данных мы можем исследовать все интересующее нас пространство: «N = все»[[15]](#footnote-15). Современное технологическое оборудование позволяет обрабатывать огромные базы данных, и более того, собирать их в режиме реального времени. Исследователь в кратчайшие сроки получает массив данных, который он может рассматривать с различных сторон, искать какие-либо связи между объектами. Он получает всю совокупность за интересующий его период, при этом, заранее не зная, что может там обнаружить. С одной стороны, это может показаться проблемой, ведь как анализировать данные, не понимая, какие выводы необходимо получить. Но с другой стороны, это дает исследователю огромную свободу для поиска новых решений и связей, сами данные «расскажут» ему о своей специфике и характере.

Наиболее радикальный аргумент о «конце теории» рассмотрен в статье Криса Андерсона «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete»[[16]](#footnote-16), 2008 год. Автор утверждает, что в эпоху Больших данных, когда исследователи получили возможность с высокой точностью отслеживать потоки информации, нет необходимости строить абстрактные теоретические модели. «This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. With enough data, the numbers speak for themselves»[[17]](#footnote-17).

Альтернативный подход призывает исследователей не думать о Больших данных как о панацее и тем более не отказываться от теоретических рамок. В данном случае, ученые говорят о Больших данных как инструменте для познания реальности, а не как о уже существующем знании. Более того, процесс анализа и интерпретации данных очень субъективен, особенно в ситуации исследования социальных медиа, что приводит к необходимости комплексного анализа: «Big Data is at its most effective when researchers take account of the complex methodological processes that underlie the analysis of social data»[[18]](#footnote-18).

В статье Social Science in the Era of Big Data Sandra González-Bailón доказывает, что «старые» теории помогают понять организацию и функционирование социальных отношений в виртуальном пространстве, более того, на основе новых данных дополнить уже существующую теорию. В частности это, в первую очередь, относится к сетевому анализу, где возможности Больших данных позволяют проследить за формированием социальных сетей, их функционированием, формирует представление о структуре связей в виртуальном пространстве, также можно следить за динамикой изменений этих социальных сетей во времени и пространстве.

Автор также предлагает рассмотреть, как Большие данные могут развить существующие теории о поведении масс и социальном влиянии. В частности, речь идет о взаимодействии массы (публики) и медиа. Известный французский социолог Габриель Тард в своих работах подчеркивал, что СМИ формирует общественное мнение, которое зарождается в процессе коммуникации между людьми. В век информационных технологий становится особенно актуально исследовать данную теорию, особенно при учете «раздробленности» современных СМИ. Так, в работе Twilight of the Gods?: How the Internet Challenged Russian Television News Frames in the Winter Protests of 2011-12[[19]](#footnote-19) рассматривается ситуация современных российских медиа. Так, исследователи выяснили, что официальные телевизионные каналы и интернет-газеты сообщают одну и ту же информацию различными способами, что, как следствие, ведет к различному понимаю одного и того же события. Более того, виртуальные СМИ на сегодняшний день имеют ряд значительных преимуществ: возможность быстрого реагирования, распространение «гражданских» журналистов (аналитические статьи пользователей, не принадлежащие каким-либо газетным изданиям), взаимодействие газеты и публики (пользователи могут сами присылать фотографии с места событий, оставлять комментарии) – все это значительно повышает доверие масс к виртуальному изданию.

С другой стороны, в западных странах актуальна дискуссия о правомерности использования Больших данных. В частности, выдвигается тезис о том, что использование пользовательских данных нарушает их частные права и даже при гарантии анонимности и попытке «обезличить» данные остается возможность вычислить конкретного человека. Все это может стать угрозой для репутации человека, а иногда даже его жизни. Главный вопрос заключается в определении границы между частным и публичным: можно ли оперировать публичными данными и сообщениями пользователей? Ведь существуют различные настройки приватности и, если человек хочет ограничить круг лиц, имеющих право просматривать содержание его контента, то он легко может это сделать. С другой стороны, правомерно ли использование каких-либо публичных данных без оповещения и информирования пользователя?

## Онлайн исследования в социологии: новые методы анализа данных

На сегодняшний день анализ социальных сетей и медиа, Интернет-сообществ, пользователей в целом используется в основном в маркетинге. Компания может отслеживать мнение своих потребителей, вступать с ними в дискуссии, реагировать на проблемы, более того, компания может также следить и за своими конкурентами. Все это позволяет им эффективно управлять производством, строить взаимоотношения с потребителями, следить за своим имиджем и так далее.

Для социологии – это все еще практически неопробованный метод, который только входит в исследовательские практики единичными случаями. Однако многие ученые прогнозируют стремительное развитие новых методов работы с информацией в Интернете. Это связано с тем, что в Интернет-пространствах хранится и ежедневно производится и потребляется огромное количество информации. Эти данные идут огромными потоками, которые трудно охватить и собрать самостоятельно, без каких-либо специализированных инструментов. Все это превращается в новое поле для социологических исследований, которое требует как определенных возможностей для сбора информации, так и для ее анализа.

Главной задачей для социологов на сегодняшний день становится обеспечение валидности данных и возможности построения вероятностной выборки. Во многом данная проблема связана с правовыми и этическими нормами. Владельцы сайта не вправе разглашать информацию о своих пользователях (например, возраст, пол, место жительства, если человек сам не предоставил открытого доступа к этим данным), однако, именно такие социально-демографические характеристики важны практически для любого исследования.

Рассмотрим работы современных исследователей в данном направлении.

 Если раньше СМИ вели скорее монолог, излагая новости, события, происходящие в мире, городе или стране, то с момента их перехода в Интернет-пространства начинается взаимодействие с аудиторией. Пользователи оставляют комментарии, вступают в дискуссии, имеют возможность обратиться к экспертам и журналистам, но, что более важно, все это происходит в реальном времени. Дальше люди могут сами разносить полученную или увиденную ими информацию, вовлекая все больше и больше людей в этот виртуальный диалог. Все это может открыто хранится на различных сайтах, что уже дает информацию для исследователя. Первые, «чистые» реакции людей на различные события, происходящие в настоящее время. Такие потоки информации могут стать хорошей основой для социологических исследований. Но существует множество ограничений, тормозящих развитие новых методов в гуманитарных науках.

В работе Хью Макея «Information and the Transformation of Sociology: Inter-activity and Social Media Monitoring»[[20]](#footnote-20) обсуждается выход новых автоматизированных методов сбора информации на первый план, в то время как, такие традиционные методы как опросы и интервью, постепенно теряют свою значимость: «social media monitoring tools come to the fore – because of the vast scope of their data, its granularity, and its capacity to allow so many dimensions of the data to be cross tabulated».

Автор на примере исследовательских практик BBC демонстрирует два метода работы с социальными медиа с помощью продуктов компании Sysomos:

1. Sysomos MAP (Media Analysis Platform)
* Web-аналитика – предоставление количественной информации об особенностях различных сайтов, исходя из активности их пользователей
* Сбор контента – занесение в базу всех найденных сообщений, написанных пользователями по какой-либо тематике
* Анализ настроений – качественный анализ собранного контента

Данный инструмент позволяет в реальном времени увидеть, что обсуждается в Сети, какие мнения являются лидирующими, на каких Интернет-площадках ведутся активные дискуссии (блоги, социальные сети, форумы). Также можно сравнить, как предоставляется и обсуждается информация на разных сайтах, произвести сегментацию пользователей по демографическим характеристикам и их активности.

* Heartbeat - Social Media Monitoring Dashboard

Позволяет произвести более детальный анализ конкретных дискуссий, возникшей вокруг какого-либо события или новости, определить эмоциональную окраску сообщений и мнений. То есть, можно проследить за реакцией пользователей на данную тематику, увидеть, как распространяется мнение или новость, какие вызывает эмоции у различных групп пользователей.

Однако ключевая проблема состоит в том, что все эти инструменты во многом недоступны исследователям. То есть, они принадлежат частным владельцам и это вызывает необходимость оплаты их услуг (по предоставлению этих инструментов сбора информации), либо платы за готовые базы данных. Более того, эти инструменты разработаны и используются в маркетинговых исследованиях. Однако уже сейчас прогнозируется использование данных инструментов и в других сферах.

Например, Департамент защиты в Америке планирует использовать инструменты мониторинга социальных сетей для отслеживания потоков информации в публичных пространствах Интернета, обнаружения дезинформации, а так же попытки предотвращения реальных беспорядков и общественных волнений[[21]](#footnote-21).

Исследовательские практики в России

В России исследование Интернет-пространств только начинает развиваться. Однако уже сейчас, многие исследователи рассматривают новые инструменты и методики, а также размышляют о новых возможностях анализа такой информации.

Например, проект, разработанный в Санкт-Петербургском национальном исследовательском университете информационных технологий и оптики (НИУ ИТМО), «Социодинамика»[[22]](#footnote-22). Целью данного проекта является создание специализированного инструмента, позволяющего проводить социометрический анализ в социальных сетях. Для демонстрации возможностей работы центра были проанализированы сообщества в Интернете, обсуждающие законы о пропаганде гомосексуализма и педофилии. Данная тема активно обсуждалась в различных регионах: с одной стороны, выступали защитники прав меньшинств, с другой – ярые противники гомосексуализма.

Исследователи ставили перед собой задачу рассмотрения плотности связей внутри сообщества и между отдельными пользователями, количество таких связей, анализ количества записей во времени и измерения настроения обсуждений. На основе автоматически собранных сообщений по заданным ключевым словам, были построены различные социограммы (см. Приложение 1). В ходе анализа было выявлено, что пользователи самоорганизуются в сообщества на основе обсуждений, выявлены лидеры мнений, представлена динамика обсуждений и всплесков (резкое увеличение количества сообщений по данной тематике).

Главным же выводом исследования стало доказательство возможности применения методик центра для социометрических исследований. Следовательно, данные инструменты и техники могут стать основой для различных исследований по другим тематикам как в социодинамике, так и в каких-либо других гуманитарных исследованиях.

Другой интересный проект предоставил Фонд содействия изучению общественного мнения.

Исследователи решили провести эксперимент: произвести прогноз на политических выборах президента в Венесуэле в апреле 2013 года (исследование проводилось с 13 марта по 14 апреля)[[23]](#footnote-23). Было проанализировано 6 378 390 упоминаний и 827 385 авторов. Составленный прогноз полностью совпал с официальными результатами голосования (50,7% голосов за Николаса Мадуро и 48,8% за Энрике Каприлеса). Эксперимент подтвердил эффективность данной методики для политического прогнозирования. В дальнейшем исследователи планируют провести подобные исследования и в России.

В работе Анни Петит «Как обычное исследование»[[24]](#footnote-24) также обсуждается возможность применения метода анализа социальных сетей. Автор предлагает рассмотреть новые методы онлайн-исследований в контексте «традиционных» методов социологии. В своей работе Петит доказывает, что в пространстве социальных сетей можно найти те же возможности и ограничения, что и при проведении онлайн-опросов.

1. **Сбор данных.** Вместо анкетных вопросов – поиск уже написанных мнений. Главное условие достижения полноты выборки – максимально возможное количество сайтов, где возможно найти интересующие нас мнения людей. В онлайн-опросах используется либо готовая база (обычно в маркетинговых исследованиях), либо выборка «снежного шара», распространение через рекламу и так далее. В обоих случаях исследователь ищет людей, которые могут дать какую-либо информацию, мнение по интересующей его (исследователя) теме.
2. **Качество данных.** Вне зависимости от методы сбора данных, исследователь может столкнуться с проблемой искажения данных. Респонденты могут намеренно предоставлять ложные данные, некачественно заполнять анкеты (одинаковые ответы в таблицах, слишком быстрое заполнение, несоответствие ответов в одинаковых вопросах или вопросах-ловушка и так далее). Похожие ситуации можно наблюдать и в социальных сетях, например, существуют сообщения, где в обычном, с первого взгляда, тексте пользователя содержатся различные ссылки, не имеющие отношения к обсуждаемой теме.[[25]](#footnote-25)В обоих случаях существуют методы устранения таких данных из выборки.
3. **Выборка**. Существуют различные «традиционные» методы построения выборки, которые применимы и к онлайн-опросам. При использовании метода анализа социальных сетей также необходимо строить выборку, но по несколько иным основаниям. В данном случае, критерием отбора являются различные сообщения пользователей, содержащие ключевые слова по тематике, интересующей исследователя.
4. **Взвешивание**. Также возможно при анализе социальных сетей. Однако, если в случае опроса происходит взвешивание по социально-демографическим характеристикам, то в исследовании социальных сетей основанием становится тип ресурса, на котором размещается информация. Так, например, учитывается направленность сетей: блоги, микроблоги, форумы и так далее. Учитывая статистику пользования людьми определенных типов сайтов, можно производить взвешивание полученных данных.
5. **Шкалирование**. Онлайн-анкета представляет собой совокупность различных шкал, позволяющих буквально преобразовать слова в цифры, что необходимо для дальнейшего анализа. При исследовании социальных сетей можно также преобразовать текстовые данные в цифровые. По эмоционально окрашенным сообщениям можно определить степень согласия или несогласия «респондентов» к тем или иным темам, проблемам.
6. **Переменные**. Совокупность полученных текстов позволяет разбивать их на различные переменные, которые отражают различные аспекты исследуемой темы.

# Глава 3. Сетевой анализ как метод изучения виртуального пространства

Анализ социальных сетей как отдельное направление появилось в конце 20 века, основоположниками которого считаются такие ученые как Милгрэм («феномен маленького мира»), Грановеттер («Сила слабых связей», 1973), Фишер, Вельман, Фриман и многие другие. В своих работах они доказали существование определенных закономерностей в сетях взаимодействий людей. Например, в работе Милгрэма, в ходе проведения масштабного исследования, делается вывод о теории «шести рукопожатий». Согласно данной теории, любой человек соединен с другим, незнакомым ему человеком, посредством 6 (в среднем) общих знакомых. Милгрэм произвел эксперимент, в котором попросил 300 респондентов отправить выданные им письма через своих родственников или знакомых. Позднее, в 2003 году, его эксперимент был повторен социологами Колумбийского университета уже при помощи электронной почты. Результаты подтверждали теорию Милгрэма. Следовательно, это позволяет сделать вывод о том, что структура связей между людьми в виртуальном и реальном пространстве повторяется.

На сегодняшний день сетевой анализ высоко востребован и продолжает быстро развиваться. В значительной этому мере способствуют технологические изменения. С одной стороны, возрастающая мощность вычислительной техники позволяет обрабатывать большие массивы данных в короткие сроки, визуализировать полученные данные, высчитывать различные показатели. С другой стороны, высокая доступность персональных компьютеров и различных средств связи, возрастающая популярность различных интернет-сервисов, тематических сайтов позволяет исследователям получить разнообразные данные без составления анкет и обращения к респондентам.

Интернет – это глобальная сеть данных, состоящая из множества подсетей. Как и в реальном мире, это сложная структура связей между акторами. Основание связи в виртуальном пространстве может быть разнообразным и зависит от исследовательских целей. Так, если целью работы является рассмотрение структуры дружеских связей студентов факультета социологии, то связующим элементом будет рассматриваться наличие в «друзьях» студентов этого факультета.

В данной работе, связью считается наличие ссылки на источник новости. Это может быть как отдельный пользователь социальной сети, так и внешний сайт, опубликовавший новость. В таком случае, рассматривается двухуровневый подход к образуемой сети обмена информацией:

* **1-ый уровень** – интернет-платформа социальной сети (Twitter, Facebook, Вконтакте)
* **2-ой уровень** – внешние источники (сайт, на который сослался пользователь)

Многие новостные сайты предоставляют пользователям возможность быстрого размещения статьи на своей странице, при этом автоматически прикрепляя ссылку на данный ресурс. Таким образом, внешние источники интегрированы в пространство отдельных социальных сетей через таких акторов.

Более того, например, социальная сеть Twitter накладывает ограничение на число символов одного публикуемого сообщения (140 знаков). Следовательно, пользователь не имеет возможности разместить полный текст статьи на своей странице. Однако существует возможность прикрепления ссылки на источник к своему сообщению. Таким образом, пользователь может аргументировать свою позицию (прикрепляя ссылку как доказательство), продемонстрировать свое отношение к обсуждаемой новости, либо же привлечь внимание своих друзей и подписчиков к данной новости.

### Основные метрики, используемые в сетевом анализе

Сетевой анализ позволяет изучать социальные взаимодействия путем выделения структур отношений между индивидом и группой, а также и взаимодействий групп людей. В основе данного подхода лежит математическая теория графов, с помощью которой можно геометрически изобразить структуру исследуемого объекта. Так, например, вершинами графа могут быть отдельные люди, а ребрами – наличие связи между людьми.

Математическая модель графа выглядит следующим образом[[26]](#footnote-26):

G= (V,E)

При этом, V должно быть непустым множеством вершин/узлов, а E – множество ребер (см. пример графов в Приложении).

Таким образом, можно выделить три составляющих совокупности социальной сети[[27]](#footnote-27):

1. Акторы сети – индивиды или коллективы, «действующее» звено сети.
2. Социальные связи – определение основания для связи является одной из ключевых задач сетевых исследований, так как это является непосредственной единицей анализа. Это могут быть отношения дружбы, профессиональные, деловые, родственные и так далее.

Ресурсы – материальные и нематериальные, которые могут быть доступны акторам через установление социальных связей.

Диаметр графа (размер сети) – это максимально возможное расстояние между любыми двумя вершинами графа.

Плотность графа показывает насколько сеть близка к полной, то есть, насколько сильно связаны акторы между собой. Полный граф обладает максимально возможным количеством ребер и значение плотности стремится к 1.

Модулярность измеряет насколько хорошо сеть разбивается на отдельные объединения. Допустимые значения варьируются от 0,7 до 0,8. Высокая модулярность обозначает наличие сложных внутренних структур (подсети).

Коэффициент кластеризации показывает насколько плотно связаны вершины вокруг узлов. Высокое значение коэффициента означает, что граф плотно связан вокруг нескольких узлов. Низкий коэффициент кластеризации свидетельствует об относительно равномерном распределении связей среди всех узлов.

Центральность используется для определения положения конкретного узла относительно других узлов сети. Значение меры центральности варьируется от 0 до 1: от отсутствия связей у узла (изолированный узел) до полной связи узла со всеми вершинами графа. Узел, обладающий большой исходящей мощностью указывает на «властность» данного актора в конкретной сети, именно через него

В исследовании структуры распространения информации необходимо особое рассмотрение узлов, с большим количеством связи. Именно такие акторы могут являться «лидерами мнений» и играть ключевую роль в каждой конкретной сети. Для таких целей существует показатели входящей и исходящей мощи узла. Эти метрики предназначены для направленных графов и обозначают число связей направленных к узлу (входящая мощность) или исходящих от узла (исходящая мощность).

# Глава 4. Обоснование понятия «ложная новость»

В рамках данного исследования такая задача необходима в силу нескольких причин. Во-первых, для теоретического обоснования использования термина, а, во-вторых, из-за отсутствия научного объяснения такого понятия в целом.

В журналистике используется сочетание «газетная утка» для обозначения обмана, фальсификации данных относительно какого-либо события или персоны. Согласно толковому словарю Даля, это «лживая газетная статья, обман»[[28]](#footnote-28). Такая «утка» противопоставляется достоверной новости, описывающей реальные события.

Таким образом, можно разделить новости на ложные и реальные:

Ложные новости – это такие новости, которые заведомо преподносят искаженную, непроверенную, фальсифицированную информацию.

Реальные новости – это такие новости, которые передают объективную и достоверную информацию о событии.

Ложные новости не новое явление для средств массовой информации. Хотя среди филологов нет единого мнения о точной дате происхождения термина «газетная утка» (некоторые утверждают, что это конец 17 века, другие – начало 19 века), ясно, что обманные статьи в средствах массовой информации появились задолго до возникновения интернета. Следовательно, существовала и проблема доверия новостям и способам проверки транслируемой информации.

Одним из способов проверки достоверности новости является проверка ее источника. В телевизионных СМИ можно выделить государственные и частные каналы. Первые контролируются государством, которое несет ответственность перед обществом за объективность и достоверность информации.

В печатных СМИ существует разделение на «качественную» и «популярную» прессу: «качественная пресса рассчитана на подготовленного высокообразованного читателя со средним и высоким доходами, а популярная пресса – на массовую аудиторию»[[29]](#footnote-29). В первом случае, это официально-деловой стиль публикуемых материалов, достоверность источников информации, это аналитические статьи, без использования эмоционально окрашенных высказываний, ненормативной лексики. Именно качественная задает информационный фон в обществе.

Популярная пресса отличается ориентацией на массовую аудиторию, просто изложенным материалом, менее формальным стилем публикаций. Выпускаемые материалы небольшие по объему, содержат множество иллюстраций (для упрощения восприятия), допускается использование оценочных суждений и так далее. В популярной прессе допускается возможность создания ложных новостей (например, в изданиях «желтой прессы»).

В интернете и, в частности, на сайтах различных социальных сетей присутствует огромное множество информационных агентств, групп, изданий. Некоторые из них являются электронной версией печатных газет или даже телевизионных каналов. Например, печатная газета «Известия» имеет электронный сайт газеты (<http://izvestia.ru>), аккаунт в социальной сети Twitter (<https://twitter.com/izvestia_ru>), Вконтакте (<http://vk.com/izvestia>), Facebook (<https://www.facebook.com/izvestia.ru>), Instagram (<http://instagram.com/izvestia_ru>) и Google+ (<http://instagram.com/izvestia_ru>). Публикуемая информация дублируется на перечисленных сайтах и может быть размещена пользователем на его персональной странице в социальной сети.

В интернете доступ к различным новостным изданиям значительно упрощен, нежели в обычной жизни. Так, человек может просмотреть несколько десятков новостных ресурсов в виртуальном пространстве за сравнительно короткое время и, при этом, не выходя из дома и не затрачивая на это денежных средств. Проблема заключается в том, что разнообразие источников информации не означает увеличения ее качества. Разные источники публикуют разную, иногда даже противоречивую информацию. Новости дублируются медиа-агентствами, укорачиваются, искажаются и так далее. Все это значительно усложняет ориентацию в информационных потоках и способствует искаженному восприятию новости.

**Как отличить ложную новость от реальной в условиях разнообразных информационных потоков?**

1. Опровержение, опубликованное на официальных источниках качественной прессы – это один из наиболее эффективных способов. Официальное опровержение, подтвержденное фактами устраняет неопределенность ложной новости
2. Отсутствие подобной новости на сайтах качественной прессы – данное условие не является достаточным. Новость может иметь региональный характер, либо же произошедшее событие не представляется важным для публикации и освещения на каналах качественной прессы (или телевидения). Более того, в ситуации, когда представители авторитетных источников не могут получить достоверных фактов о какой-либо новости, публикация будет откладываться до выяснения всех обстоятельств.
3. Тематический индекс цитируемости сайта, тИЦ (Яндекс) – это специально разработанная технология поиска компании «Яндекс», которая определяет авторитетность сайта с учетом качественных характеристик (а не только количество ссылок на данный сайт). Из описания на сайте Яндекс: «Наш тематический индекс учитывает количество ссылок на сайт с других сайтов, придавая этим ссылкам разный «вес» (то есть значимость) в зависимости от авторитетности ссылающегося сайта».[[30]](#footnote-30)

Таким образом, чем надежнее источник, тем выше его тИЦ. Например, интернет-версия газеты «Известия» тИЦ равен 22000, это число означает количество внешних ссылок на данный ресурс. Однако данный критерий так же не является достаточным условием. Рассмотрим тематические индексы цитирования на исследуемых новостях.

Таблица №2

**Тематический индекс цитирования для анализируемых новостей**

|  |  |  |  |
| --- | --- | --- | --- |
| **Заголовок** | **первоисточник** | **дата публикации** | **тИЦ** |
| **Ложные новости** |
| В Москве после встречи с Путиным пытался застрелиться директор "Первого канала" | http://ivona.bigmir.net/showbiz/stars/391093-Konstantin-Ernst-pytalsja-zastrelit-sja-posle-razgovora-s-Putinym | 03.05.14 | 1500 |
| В Донецке сепаратисты расклеили антисемитские листовки: евреи должны заплатить $50 за регистрацию | http://gordonua.com/news/separatism/V-Donecke-separatisty-raskleili-antisemitskie-listovki-evrei-za-50-dolzhny-proyti-registraciyu-18609.html | 16.04.14 | 900 |
| Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель | http://ruskline.ru/politnews/2014/04/28/minoborony\_rf\_nebolshoj\_sboj\_kursa\_rossijskoj\_rakety\_rs24\_yars\_pentagon\_eto\_provokaciya\_pvo\_ssha\_ne\_smogli\_vzyat\_cel/ | 28.04.14 | 1900 |
| **Реальные новости** |
| Министр обороны: Италия может направить миротворцев на Украину | http://ria.ru/world/20140504/1006495411.html | 04.05.14 | 36000 |
| Фильм Андрея Звягинцева "Левиафан" получил приз Каннского кинофестиваля за лучший сценарий | http://www.1tv.ru/news/culture/259554 | 24.05.14 | 1500 |
| В иркутской школе №11 появилась "Азбука вежливости" | http://vesti.irk.ru/news/society/164374/ | 05.05.14 | 1100 |

Сравнительно высоким значение тематического индекса цитирования обладает только один источник (ria.ru), все остальные вне зависимости от типа распространяемой ими новостной информации (реальной или ложной) имеют уровень тИЦ не более 2000.

Следовательно, такой критерий не может быть определяющим при разделении реальных и ложных новостей.

1. Субъективное восприятие сайта – некоторые сайты массовой прессы содержат большое количество всплывающих окон, рекламы сомнительных препаратов, ссылок на различные сайты с нецензурными материалами и так далее.

Из всех перечисленных способов, официальное опровержение является единственным объективным критерием для разделения реальных и ложных новостей. Однако остальные критерии также могут быть полезны для наиболее продуктивной ориентации в виртуальном пространстве.

# Глава 5. Описание работы с системой мониторинга социальных сетей в качестве метода сбора данных

Система мониторинга социальных сетей предоставляет исследователю возможность собрать интересующие его упоминания в социальных сетях по какой-либо тематике за кратчайшие сроки. Достаточно правильным образом составить поисковый запрос и программа соберет базу сообщений за указанный период с заданными условиями поиска. Полученная база данных состоит из всех найденных системой сообщений пользователей, которые размещены в публичном доступе.

Прежде, чем приступить к анализу, необходимо «очистить» базу от нерелевантных сообщений:

* Удаленные сообщения – это те упоминания, которые по каким-либо причинам недоступны, удалены пользователем или модератором.
* Не связанные с темой исследования – система ищет любые совпадения с заданным в поиске словом, следовательно, чем «проще» искомое слово, тем больше будет нерелевантных сообщений.

Для создания релевантного поискового запроса, необходимо соблюсти некоторые условия:

 Во-первых, подобрать все возможные и распространенные варианты написания слова, названия. Например, в данном случае, ключевое слово – название бренда. Однако было понятно, что не все пользователи пишут данное название правильно. Более того, даже русифицированный вариант написания этого бренда часто встречается с ошибками, либо сокращениями и так далее. Для того, чтобы собрать наиболее полную базу упоминаний, необходимо учитывать все эти особенности написания искомого слова.

Во-вторых, язык поисковых запросов предполагает наличие определенных правил, с помощью которых можно регулировать поиск сообщений. Так, например, можно задать системе искать только точные совпадения (без склонений), либо же исключить какие-либо слова заранее. Для того, чтобы понять что из вышеперечисленного следует сделать, можно воспользоваться пробным поиском в самой системе, либо же самостоятельно попробовать различные комбинации слов в обычных поисковых системах (Google, Mail, Yandex и так далее).

В-третьих, важно установить дату, с момента которой будут собираться упоминания. Например, в данном случае, были необходимы упоминания с момента начала установки «чемодана». Более ранние сообщения, по понятным причинам, не содержат необходимой информации и в любом случае были бы удалены.

При процедуре «очистки» базы исследователь буквально «знакомится» с данными, примерно понимает соотношение публикаций СМИ и непосредственно пользовательских отзывов. При этом, происходит проставление тональности упоминаний: новостные сообщения СМИ относятся к нейтральным, а пользовательские в зависимости от ситуации. В данном случае, конечно, эмоциональная оценка носит достаточно субъективный характер, особенно в случае с большими и подробными сообщениями, где могут быть все краски.

Среди разработчиков систем мониторинга социальных медиа актуален вопрос совершенствования и внедрения такой функции как автотональность. Это означает, что система «сама» проставляет одну из трех тональностей каждому сообщению, исходя из его содержания. Данная функция значительно облегчила и упростила работу с базой упоминаний, но все еще система не может в большинстве случаев правильно определить эмоциональную окраску сообщения. Особенно в случае с «богатым русским языком», где, казалось бы, «позитивные» слова выражают негативную реакцию человека – ирония и сарказм трудны для машинного распознавания. Более того, раскрепощенные анонимностью и неформализованным характером многих социальных сетей, люди могут выражаться, используя матные слова. Однако даже в этом случае, наличие нецензурного слова еще не означает негативную реакцию человека. В каждом случае исследователю необходимо рассматривать очередное упоминание в контексте данного социального события или процесса.

## 5.1. Процедура анализа текстов: соотнесение с классическим контент-анализом

В зависимости от целей исследования уже на этапе очищения базы от нерелевантных сообщений исследователь может провести процедуру тегирования (кодирования) сообщений. Сами теги, присваиваемые каждому сообщению, могут быть определены заранее, либо же в процессе работы с этими текстами. Таким образом, исследователь «разбивает» всю совокупность упоминаний по различным тематикам.

Контент-анализ

Основной метод исследования при мониторинге социальных сетей – контент-анализ. При этом, важно помнить, что в полученной базе присутствует 2 рода текстовой информации: публикации СМИ и пользовательские сообщения. Это совершенно два разных понятия, следовательно, при анализе выбранной темы, необходимо разделить эти два потока.

Классический контент-анализ используется при исследовании СМИ, различных интервью и даже для некоторых анкетных вопросов. При этом существует два вида работы с текстами: количественный (подсчет слов) и качественный (поиск взаимосвязей между словами, событиями). Так, например, можно определить взаимосвязь между какими-либо событиями и особенностями их отражения в СМИ, можно проанализировать эмоции человека, связанные с исследуемым объектом, определить лидеров (например, при подсчете количества упоминаний о какой-либо компании, товаре, продукте, человеке) и так далее.

Социальные сети предоставляют нам огромные потоки текстовой информации ежедневно. Собрав все воедино, исследователь располагает большим массивом мнений, при этом, как со стороны СМИ, так и со стороны пользователей. Конечно, следует помнить, что если публикации СМИ по структуре текста едины и достаточно формализованы, то пользовательские сообщения значительно короче, не формализованы и, как правило, содержат мгновенную реакцию на происходящее (а не развернутое, аргументированное мнение). Таким образом, на первом этапе анализа видится логичным расщепить базу упоминаний по этим двум потокам: публикации СМИ и пользовательские отзывы.

Еще одна особенность анализа при мониторинге социальных сетей – дублирование сообщений. Это означает, что одна и та же новость (как правило, это касается публикаций СМИ) неоднократно повторяется на различных страницах социальных сетей. Это происходит за счет того, что множество людей ссылается на исходный источник информации, либо же одну и ту же ссылку специально распространяют как можно больше в различных группах и так далее. Более того, многие социальные сети предоставляют своим пользователям возможность «репоста», то есть, каждый может опубликовать у себя на странице сообщение исходное другого пользователя, либо даже разместить ссылку на сайт публикации. Таким образом, одна новость умножается в несколько сотен раз, что с одной стороны затрудняет процедуру анализа, но с другой – говорит о популярности новости и ее широком охвате. В данном случае, исследователю необходимо решить, как поступить с дублируемой информацией, что это может дать последующему анализу всей темы.

В классическом контент-анализе исследователь анализирует выборку единичных текстов. То есть, например, если необходимо проанализировать статьи определенной тематики ежемесячного журнала за год, то база будет состоять из 12 статей, каждая из которых будет отдельно проанализирована. В ситуации мониторинга пространства социальных сетей необходимо учитывать ее специфику. Помимо уже упомянутого дублирования информации, необходимо отметить и «временные» особенности публикации материалов. Благодаря современным технологическим возможностям и необходимости мгновенного реагирования на событие, новостные сайты публикуют множество новостей по одной теме несколько раз в день (в зависимости от характера события). Буквально любое изменение, либо новый открывшийся факт тут же попадает на страницы виртуальных газет. Сами статьи, как правило, короткие и отражают основные события, без какого-либо детального анализа. Конечно, есть и подробные публикации, мнения специалистов, интервью, но для таких статей необходимо больше времени, что значит, их публикация происходит через некоторое время после события. Большая же доля новостных публикаций состоит из таких мгновенных сводок с места событий.

## 5.2. Построение выборки и сбор данных

Выборка состоит из корпуса текстов, размещенных в социальной сети Twitter с апреля по май 2014 года, относительно исследуемых информационных поводов. Совокупный объем собранных упоминаний равен 2749 сообщений.

**Первый этап: отбор кейсов.**

В первую очередь были отобраны ложные новости, опубликованные в апреле-мае 2014 года в социальной сети Twitter.

Критерием для отбора таких информационных поводов стало наличие официального опровержения. Данный критерий был выбран для обеспечения надежности результатов. Реальные новости были отобраны так, чтобы обеспечить максимально возможное единое основание для сравнения. Например, в выборку вошли 2 ложные новости политической направленности, следовательно, были выбраны 2 новости относительно политических событий. При этом, учитывалось также и количество сообщений. Конечно, невозможно было найти такую реальную новость, упоминания по которой были бы равны числу упоминаний по ложной новости. Поэтому реальные новости отбирались с числом упоминаний максимально приближенных к количеству сообщений о ложной новости.

**Второй этап: вычленение единицы анализа, сбор текстов.**

Данный этап непосредственно связан с работой в системе мониторинга социальных сетей. В исследовании использовалось программное обеспечение компании Вобот[[31]](#footnote-31). Необходимо было собрать отдельно 6 массивов данных с текстами по каждой исследуемой новости (реальной и ложной).

Перед непосредственным сбором данных необходимо задать поисковый запрос, являющийся единицей счета. Приведем пример, вычленения единицей анализа и процесса подготовки к сбору данных, основанный на материалах, использованных в данном исследовании.

Реальная новость, заголовок: «Министр обороны: Италия может направить миротворцев на Украину».

Так как разные новостные агентства могут немного видоизменять заголовок,[[32]](#footnote-32) то поиск точного совпадения слов с исходным заголовком неэффективен. Необходимо обеспечить попадание максимально большего количества сообщений, относящихся к данной новости. Следовательно, заголовок разделялся на ключевые слова:

**Министр обороны: Италия может направить миротворцев на Украину**

Таким образом, поисковый запрос выглядел так: Италия & миротворцы & Украина. Символ «&» обозначает поиск совпадений внутри всего текста сообщения в любом порядке, независимо от склонения слов и регистра букв.

Более того, поиск осуществлялся в заданный период времени (апрель-май 2014 года), что исключало попадание неактуальных новостей. Так же из выборки были исключены упоминания, не относящиеся к исследуемой новости.

# Глава 6. Выявление различий в структурах реальных и ложных новостей в социальных сетях

Для сравнения 6 моделей графов использовались основные показатели статистик графа:

Таблица №3

**Основные сетевые метрики по всем кейсам**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Тип новости** | **Заголовок** | **Средняя степень** | **Диаметр графа** | **Плотность графа** | **Связные компоненты** |
| **Ложные новости** | В Москве после встречи с Путиным пытался застрелиться директор "Первого канала" | 0,81 | 3 | 0,004 | 53 |
| В Донецке сепаратисты расклеили антисемитские листовки: евреи должны заплатить $50 за регистрацию | 0,952 | 1 | 0,011 | 4 |
| Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель | 1,002 | 4 | 0,002 | 65 |
| **Реальные новости** | Министр обороны: Италия может направить миротворцев на Украину | 0,903 | 2 | 0,002 | 40 |
| Фильм Андрея Звягинцева "Левиафан" получил приз Каннского кинофестиваля за лучший сценарий | 0,982 | 3 | 0,001 | 154 |
| В иркутской школе №11 появилась "Азбука вежливости" | 0,761 | 4 | 0,002 | 106 |

Выше приведена таблица макропоказателей для каждой сети. В данном случае нельзя говорить об отношении направленности существующих различий (больше/меньше), так как графы не одинаковы по своим размерам. Однако при последовательном рассмотрении каждой сети по указанным параметрам можно составить обобщенное представление о структуре распространения информации в социальной сети Twitter в целом (в рамках данных кейсов).

Полученные результаты не позволяют объединять и выводить единую структуру для каждого вида новости (реальной и ложной). Все кейсы можно описать как совокупность низко связных между собой акторов (плотность во всех случаях не превышает 0,1).

Все разобранные сети являются слабо связанными. В моделях отсутствуют эквивалентность вершин, все направления связи односторонние. Это объясняется особенностями рассматриваемых информационных обменов между акторами. Основой для связи являлось размещение пользователем сообщения, содержащего ссылку на новость или на другого пользователя, разместившего соответствующее сообщение на своей странице. Основанием для установление взаимной связи могло бы быть наличие комментариев к публикации данного актора. Однако, проследить такие взаимосвязи достаточно трудно (в рамках данного исследования).

Во всех случаях коэффициент кластеризации равен нулю, что свидетельствует об отсутствии сильно связанных вершин относительно узлов сети. Акторы, участвующие в распространении новостей в большинстве случаев (в рамках разобранных кейсов) связаны только с источником информации, но не между собой.

Макропоказатели по каждой кейсу указывают на диверсифицированность сетей в целом. Акторы слабо связаны друг с другом, отсутствуют кластеры. То есть, в рассмотренных структурах нельзя выделить отдельные группировки акторов. Распределение информации происходит точечно.

Следовательно, среди разобранных кейсов невозможно выделить относительно повторяющейся, устойчивой структуры, каждая из них является уникальным случаем. Однако различия в данных сетях можно проследить на уровне акторов.

# Глава 7. Определение характеристики «лидеров мнений» в структуре распространения новостей

В каждом из разобранных кейсов одну из главных ролей в распространении информации играли искусственно созданные аккаунты, имитирующие деятельность реального пользователя социальной сети.

Данный феномен виртуального пространства имеет специальный термин – астротурфинг (astroturfing) – это искусственное формирование обсуждений, путем распространения информации, публикации комментариев по какой-либо теме на различных ресурсах с помощью специально созданных аккаунтов, имитирующих деятельность реального человека[[33]](#footnote-33).

Такие искусственные аккаунты можно разделить на два типа: «боты» и «фейки»

1. «Боты» – специально написанная программа генерирует большое количество учетных записей (аккаунтов) без полного заполнения профилей. Деятельность таких акторов сети сводится к постоянной публикации новостей, без добавления текста, эмоциональных суждений, зачастую, без прикрепления ссылки на источник.

Персональная страница «ботов» заполнена новостями различных тематик из разнообразных источников.

Можно выделить основные маркеры бот-аккаунтов (Socialbakers):

* Смещенное соотношение числа читаемых (following) пользователей актором и читающих (followers) его. Закономерность можно описать соотношением 50:1. Это означает, что бот-аккаунт имеет 50 подписок на различных пользователей, в то время как на него подписан лишь 1.
* Более 30% твиттов содержат спам-фразы. Это могут быть фразы рекламы сайтов, отдельные слова «купить», «продать». Также возможны словосочетания «читаювзаимно», «взаимныйфолловинг», направленные на привлечение большего числа возможных подписчиков и, следовательно, увеличения подписок самого бот-аккаунта.

Цель – расширение списка подписчиков для большего охвата публикуемы сообщений

* Одна и та же формулировка сообщения повторяется с разных аккаунтов. То есть, происходит перепечатывание одного и того же сообщения, что отражается в большом количестве одинаковых твоттов, но опубликованных разными аккаунтами.

Цель – создание эффекта популярности новости.

* Более 90% публикуемых на странице сообщений – ретвиты других пользователей
* Более 90% публикуемых сообщений содержат ссылки на внешние сайты. При этом, следует дополнить данный маркер. Отличительная особенность таких спам-сообщений то, что ссылки во многих случаях перенаправляют пользователя на сайты, не связанные с информацией, содержащейся в сообщении. Например, бот-аккаунт публикует сообщение, в котором содержится заголовок новости и прикрепляет к нему ссылку. При этом, переход по ссылке перенаправляет читателя на сайт с интернет-магазином, на котором нет ни единого упоминания о заявленной в сообщении новости. Более того, такие ссылки могут направлять не только на сайты с рекламой и интернет-магазинами, а содержать в себе вирусы или нецензурный контент.

Цель – привлечь внимание к определенным источникам.

* Записи бот-аккаунтов не имеют ретвитов. В большинстве случаев, несмотря на мнимую высокую активность аккаунта, сообщения не передаются другими пользователями
* Отсутствие изображения, фотографии, какой-либо уточняющей личной информации (пол, возраст, город проживания, интересы) в профиле.

«Фейки» – это такие акторы сети, которые ведут активную деятельность в виртуальном пространстве, постоянно ведут свою страницу, имеют большое число подписчиков, участвуют в обсуждениях и так далее. Однако это также лишь имитация пользовательской активности. Ведение таких аккаунтов в социальных сетях – оплачиваемая работа. Следовательно, существуют некоторые персоны/компании/агентства, которые платят за подобную деятельность и контролируют ее выполнение, определяют содержание, частоту публикации сообщений/комментариев. Такие фейковые аккаунты сложно обнаружить, особенно неопытному пользователю.

## 7.1. Анализ двух кейсов: подробное рассмотрение ключевых характеристик сети на примере политических новостей

### Кейс 1. Реальная новость: «Министр обороны: Италия может направить миротворцев на Украину»

### Описательная статистика

Общее количество упоминаний по всем социальным сетям составило 717 сообщений за неделю (с 4 по 11 мая). Первоисточником новости является информационное агентство «РИА Новости». Данное медиа-агентство позиционируется как источник качественных аналитических материалов о ситуациях, происходящих во всех сферах российского и зарубежного общества[[34]](#footnote-34). Статьи публикуются профессиональными журналистами, используется деловой стиль речи, отсутствуют оценочные суждения. Целью работы медиа-агенства является: «Оперативно, взвешенно и объективно освещать события в мире, рассказывать международной аудитории о российском взгляде на ситуацию». Таким образом, данный источник возможно отнести к качественной прессе.

Рассмотрим динамику упоминаний, относящихся к данной новости за неделю:

График №1

**Динамика упоминай за неделю**

Динамика нисходящая, 7 мая произошло увеличение сообщений на 9 упоминаний по сравнению с предыдущим днем, однако по сравнению с общим массивом данных и дальнейшим спадом числа упоминаний такой прирост незначителен.

Диаграмма №1

**Распределение долей упоминаний в различных социальных сетях**

Наибольшее число упоминаний было оставлено пользователями социальной сети Twitter. Доля этого ресурса составила приблизительно 50% от остальных социальных сетей.

Динамика количества упоминаний по 3-м основным ресурсам (Twitter, Вконтакте, Facebook) аналогична распределению по всем социальным сетям за недельный период.

График №2

**Динамика упоминаний на 3-х основных ресурсах**

При этом значение охвата аудитории равно 602973. Этот показатель означает суммарное количество подписчиков всех акторов данной сети. То есть, данное значение показывает число пользователей, которые потенциально могут прочитать эту новость. Однако данный показатель не является абсолютно точным.

Во-первых, такое значение по охвату актуально только на момент сбора данных. Охват может расширяться по мере дальнейшего распространения новости. В данном случае, прирост нового числа сообщений составлял менее 10 единиц уже в середине исследуемого периода (8.05). Следовательно, можно сделать допущение о том, что впоследствии этот показатель останется приблизительно таким же[[35]](#footnote-35).

Во-вторых, показатель суммирует число подписчиков каждого актора, но не учитывает возможные совпадения. Так, например, два пользователя, состоящие в подписчиках друг у друга, независимо друг от друга размещают на своих страницах данную новость. Допустим, что у каждого из них по 10 подписчиков, следовательно, суммарный охват будет равен: 20 – 2 = 18. Система мониторинга социальных сетей не имеет автоматического доступа к учетным записям пользователей, что ограничивает возможность получения более точного значения. Однако суммарный охват можно использовать для составления первого представления о величине масштаба распространяющейся новости.

Новость о миротворцах Италии имеет суммарный охват равный приблизительно 600 тысяч пользователей. Следовательно, около полумиллиона пользователей сети Twitter могли увидеть данную новость.

При этом, несмотря на широкий охват, новость практически не нашла отклика среди пользователей, распространявших эту новость.

### Сетевые показатели

Ориентированный граф. Количество узлов – 484, количество ребер – 450.

Рис. 2. Кейс 1. Реальная новость: «Министр обороны: Италия может направить миротворцев на Украину»

RT\_Russian

Lenta\_super

Rianru

Макропоказатели сети.

Сеть децентрализована, показатель плотности равен 0,2%, что является свидетельством отсутствия связи между акторами сети. Также, отсутствуют сильно связные компоненты, сеть невозможно разделить на кластеры. Более того, в большом количестве присутствуют изолированные вершины и диады.

В данном случае в сети присутствуют четыре мощных узла, которые являются аккаунтами: РИА-Новости, RT\_Russian, Lenta\_super и электронная газета «Ведомости». При этом, максимальной мощностью обладает узел RT\_Russian. Данный аккаунт принадлежит сайту телевизионного канала «Russia Today» ([www.rt.com](http://www.rt.com)). Это международный телевизионный канал, высоко интегрированный в пространство социальных медиа. Канал предоставляет альтернативный взгляд на события, происходящие в мире, освещает то, что большинство популярных меда предпочитает «скрывать».[[36]](#footnote-36) Сайт канала имеет огромную популярность, которая характеризуется более полумиллионом зрителей в более чем ста странах мира[[37]](#footnote-37).

Распределение по степени промежуточности между вершинами выглядит следующим образом:

Таблица №4

**Степени промежуточности главных акторов сети**

|  |  |
| --- | --- |
| **Источник** | **Промежуточность** |
| RT\_russian | 100 |
| Rianru | 54 |
| lenta\_super | 51 |

Рис. 3. Распределение степени промежуточности по всем узлам сети

Большая часть вершин (более 450) не обладает степенью промежуточности. В основном это изолированные вершины или диады, которые являются «бот»-аккаунтами. Доля акторов, не связанных с наиболее мощными акторами сети составила 24%.

Таблица №5

**Исходящая мощность главных узлов сети**

|  |  |
| --- | --- |
| **Источник** | **Исходящая мощность** |
| RT\_russian | 100 |
| rianru | 54 |
| lenta\_super | 51 |
| kvedomosti.com | 32 |
| newsdailycomua | 14 |
| news.rambler.ru | 11 |

Рис. 4 Распределение узлов по исходящей мощности

Эти же акторы обладают и максимальным влиянием в сети за счет большого количества связей с акторами сети, а также зачет их социального капитала.

Диаметр сети равен 2, что означает максимальное возможное расстояние между двумя акторами равно двум «шагам». Если не учитывать изолированные вершины, то можно говорить о компактности сети и достаточно быстрой скорости распространения информации.

Вывод: структура данной сети состоит из 4-х ключевых акторов, обладающих высоким социальным капиталом (по большой, в рамках данного случая, мощности узла и числу подписчиков). Новость быстро распространяется от источников к реципиентам. Между акторами практически не устанавливается взаимная связь. Возможно, это объясняется тем, что пользователи сети Twitter, при желании разместить новость на своей странице, обращаются к «лидерам мнений», а не ссылаются на аккаунты своих «друзей».

### Кейс 2. Ложная новость: «Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель»

### Описательная статистика

Общее количество упоминаний за неделю – 530 сообщений. Первоисточник новости – информационно-аналитическая служба «Русская линия». Информации об агентстве практически нет, однако можно проследить явную религиозную направленность: в перечисленных авторах статей большое количество священнослужителей, при описании других сотрудников подчеркивается их национальность (русские) и религиозные взгляды (православие), некоторые также имеют опыт работы с религиозными изданиями. Сайт предлагает большое количество материала на различные тематики: религия, политика, общество и «разное»[[38]](#footnote-38). Также присутствует реклама религиозного содержания, ссылки на «похожие» по направлению сайты.

За рассматриваемый период (с 28.04 по 04.05) ложная новость распространялась преимущественно социальной сети Twitter (93%). Динамика сообщений за неделю нисходящая.

График №3

Динамика сообщений за неделю

Суммарный охват для социальной сети Твиттер равен приблизительно 23 тысячи пользователей. Данный показатель значительно меньше, чем в случае с реальной новостью (где такой же показатель был равен полумиллиону). Конечно, рассматриваемые кейсы не одинаковы по объему упоминаний и количеству акторов, однако такое различие объясняется не только количественными характеристиками этих сетей, но и качественными различиями в акторах.

### Сетевые показатели

Граф ориентированный. Количество узлов – 528, количество ребер – 529.



opernn

swarog09

pillsa5

salty\_1970

ruskline.ru

Рис. 5. Кейс 2. Ложная новость: «Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель»

Рассмотрим показатели промежуточности (betweenness), которые показывают как часто узел появляется в кратчайшем пути между всеми узлами сети. Высокой степенью промежуточности обладают следующие узлы:

Таблица №6

**Степени промежуточности главных акторов сети**

|  |  |
| --- | --- |
| **Источник** | **Промежуточность** |
| ruskline.ru | 3308.0 |
| salty\_1970 | 371.0 |
| pillsa5 | 272.0 |
| swarog09 | 196.0 |
| opernn | 154.0 |
| lagyasul | 146.5 |
| olshan75 | 126.0 |
| Darifma | 112.0 |
| azbukywedy | 112.0 |

Рис. 6 Распределение степенй промежуточностей всех акторов сети

Данных акторов можно назвать «проводниками» сети. Именно через них происходит трансляция новости остальным акторам. Более того, эти акторы так же являются «лидерами мнений» сети, обладая наибольшей мощностью узлов (по сравнению со всеми узлами сети):

Таблица №6

**Исходящая мощность главных узлов сети**

|  |  |
| --- | --- |
| **Источник** | **Исходящая мощность** |
| ruskline.ru | 117 |
| ruskline | 33 |
| salty\_1970 | 27 |
| vk.com | 26 |
| pillsa5 | 17 |
| swarog09 | 15 |
| lagyasul | 13 |
| shtrobl | 13 |
| opernn | 11 |
| olshan75 | 10 |
|  earth-chronicles.ru | 10 |

Рис.6 Распределение исходящей мощности по каждой вершине

Следовательно, в данном кейсе, «проводники» одновременно являются «лидерами мнений». При этом, сайт-источник ложной новости является ядром сети, из которого и происходит дальнейшее распределение информации. При этом, большая часть связей – «простые» бот-аккаунты, которые разместили ссылку на своей странице. Процент изолированных вершин равен 5%.

Выводы: Ключевую роль в распространении новости сыграли «фейковые аккаунты». Их записи обсуждались, а так же ретвитились реальными пользователями. Более того, именно они способствовали суммарному увеличению охвата и увеличению количества связей между акторами, усложнению структуры. Именно от них, как ключевых «лидеров мнений» данной сети, происходили дальнейшие ссылки других акторов сети.

# Заключение

Проведенное исследование позволило составить представление об особенностях распространения новостной информации в социальной сети Twitter.

Была проведена апробация метода сбора данных в социальных сетях при помощи мониторинга социальных медиа. Опыт работы с данной системой позволил в короткие сроки собрать большой массив данных (суммарно число упоминаний 2484), проанализировать описательную статистику по каждому кейсу (динамика за неделю, общее распределение сообщений по ресурсам). Также был проанализирован показатель охвата аудитории, посчитанный данным программным обеспечением. Числовое значение охвата показывает суммарное число пользователей, которые потенциально могли увидеть/прочитать сообщение. В основе расчета лежит число подписчиков у каждого актора исследуемой сети (графа). Однако данный показатель возможно использовать только для составления общего представления о масштабе распространения новости. Проблема заключается в том, что при расчете охвата не учитываются возможные совпадения подписчиков у акторов (одни и те же «друзья» у разных людей).

Таким образом, мониторинг социальных сетей как исследовательский инструмент предоставляет возможность оперативного сбора данных в социальных сетях. Собранная база данных содержит необходимые показатели для составления описательной статистики (количество упоминаний за период, авторов сообщений, тип ресурса, на котором была найдена новость, дата и время публикации). В сетевом анализе данный инструмент может использоваться как инструмент сбора данных для разведывательного исследования, так как для проведения аналитического исследования необходимо более детальное составление выборки с добавлением личных характеристик аккаунтов пользователей. Данную информацию невозможно получить при использовании системы мониторинга социальных сетей, что является одним из ее ограничений.

Ключевых задачей исследования было определение понятия «ложная новость». Такая задача была необходима, во-первых, из-за отсутствия научного определения данного феномена в целом, а, во-вторых, для операционализации понятия и работы с ним на эмпирическом уровне. Таким образом, было обозначено, что под «ложной новостью» подразумевается информационный повод, содержащий «фальсифицированную» информацию, который в последствии был опровергнут в «качественной» прессе (или в государственных СМИ). Были выработаны критерии, позволяющие на индивидуальном уровне ориентироваться в информационных потоках виртуального пространства:

1. Опровержение, опубликованное на официальных источниках качественной прессы
2. Отсутствие подобной новости на сайтах «качественной» прессы
3. Тематический индекс цитируемости сайта, тИЦ (Яндекс)
4. Субъективное восприятие сайта

Из перечисленных критериев, объективным и наиболее достоверным является опровержение, опубликованное источниками «качественной» прессы. В данном случае, информация предоставляется профессиональными журналистами, которые имеют возможность проверки реальности факта.

Остальные критерии не являются однозначными и содержат множество допущений. Например, отсутствие новости в источниках качественной прессы не является свидетельством того, что новость ложная. Это может относиться к региональным новостям, а также информационным поводам, которые не были освещены более крупными изданиями в силу определенных причин (отсутствие интереса аудитории издания к тематике новости, узкая направленность и так далее). Однако данные критерии могут быть полезны для наиболее продуктивной ориентации в виртуальном пространстве.

При анализе общей структуры распространения новостей было рассмотрено 6 кейсов (3 реальных и 3 ложных новости). Было важно понять: возможно ли составить типичную структуру для каждого вида новостей и выделить основные различия между двумя типами информационных поводов. В ходе исследования было выявлено, что каждый кейс обладал уникальной структурой, но не только в силу объективных причин (разные акторы, количество вершин и ребер, тематика новости, дата события), но и по своим индивидуальным сетевым характеристикам. В каждом случае вершины графов слабо связаны, обладают низкой плотностью, невозможно выделить кластеры сети. При этом, в большинстве рассмотренных кейсов присутствует большое количество изолированных вершин и диадных связей, что обусловлено активным участием бот-аккаунтов[[39]](#footnote-39) в распространении новостей. Однако ключевую роль в распространении новостей и увеличения суммарного охвата играют такие пользователи, которые имеют большое число подписчиков на свою страницу.

Данная проблема была описана в последней задаче, целью которой было определение характеристик «лидеров мнений». Для этого было рассмотрено два кейса[[40]](#footnote-40)реальной и ложной новостью относительно политических событий. В обоих случаях можно было выделить ключевых акторов сети – вершины с наибольшей мощностью узла (пользователи или сайт, на который было зафиксировано большое количество ссылок по сравнению с остальными акторами).

В случае реальной, было обнаружено 4 отдельных узла, к 3 из которых были получены ответные реакции пользователей (комментарии к записям). Одновременно с этим, в случае ложной новости можно выделить большее число акторов (но меньшее по их весу) и, при этом, больший процент отклика со стороны других акторов сети. При этом связи между вершинами более частые и сложные, нежели в графе, отражающего структуру реальной новости. В двух изучаемых случаях действовали разные механизмы распространения информации. В первом кейсе новость распространялась в большей степени непосредственно от аккаунтов представителей качественной прессы. Во втором случае, ложная новость, опубликованная на сайте, получила широкий охват благодаря активности «лидеров мнений», с «фейковыми» аккаунтами и с большим числом подписчиков.

Таким образом, новостное пространство сети Twitter в большей степени является «искусственно созданным». С одной стороны, это увеличение объема упоминаний в социальной сети путем большого числа публикаций со страниц «бот»-аккаунтов, с другой стороны, это увеличение масштаба охвата аудитории сети, путем передачи информации через «фейковые» страницы.

Перспективу дальнейшего исследования можно обозначить по нескольким направлениям. Во-первых, это более подробное изучение активности «фейковых» аккаунтов, являющихся «лидерами мнений» в социальной сети, уровень доверия пользователей к таким акторам. Во-вторых, отдельное рассмотрение большого числа «ложных» новостей с возможностью применения статистических моделей для выявления значимых связей и закономерностей.

# Библиографический список

Chris Anderson The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

 URL: http://www.wired.com/science/discoveries/magazine/16-07/pb\_theory

1. Gross R., Acquisti A. Information revelation and privacy in online social networks //Proceedings of the 2005 ACM workshop on Privacy in the electronic society, 2005.

URL:http://www.inf.ed.ac.uk/publications/thesis/online/IM110932.pdf

Kate Crawford, Danah Boyd Six Provocations for Big Data / “A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society” / September 21, 2011

URL: http://softwarestudies.com/cultural\_analytics/Six\_Provocations\_for\_Big\_Data.pdf

1. Kozinets R.V. (1998). «On netnography. Initial reflections on consumer research investigations of cyberculture». *Advances in Consumer Research*,Vol. 25(1), pp. 366–371.
2. Lerman K., Hogg T. Using a model of social dynamics to predict popularity of news //Proceedings of the 19th international conference on World wide web, 2010

URL: http://www.isi.edu/~lerman/papers/wfp0788-lerman.pdf

1. Lerman K., Ghosh R. Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks

URL:http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1509/1839

1. Mackay H. Information and the Transformation of Sociology: Inter-activity and Social Media Monitoring // Journal for a Global Sustainable Information Society, Vol. 11, №1, 2013
2. Sameer Hinduja, Justin W. Patchin Personal information of adolescents on the Internet:A quantitative content analysis of MySpace

URL:http://sfcs.cals.arizona.edu/azsearch/sites/sfcs.cals.arizona.edu.azsearch/files/Hinduja,%202008.pdf

Sandra González-Bailón Social Science in the Era of Big Data / Social Science Research Network, March 22, 2013

URL: http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2238198

1. Sarah Oates, TetyanaLokot Twilight of the Gods?: How the Internet Challenged Russian Television News Frames in the Winter Protests of 2011-12 / Social Science Research Network, June 28, 2013

URL: http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2286727

1. Базенков Н. и др. Обзор информационных систем анализа социальных сетей //Управление большими системами: сборник трудов. – 2013. – №. 41

URL: http://cyberleninka.ru/article/n/obzor-informatsionnyh-sistem-analiza-sotsialnyh-setey

Бершадская Л., Биккулов А., Болгова Е., Чугунов А., Якушев . Социальные сети и социометрические исследования: теоретические основания и практика использования автоматизированного инструментария изучения виртуальных сообществ // «Информационные Ресурсы России» №4, 2012

URL:http://www.aselibrary.ru/digital\_resources/journal/irr/irr3648/irr36483711/irr364837113734/irr3648371137343743/

1. Девятко И.Ф. Инструментарий онлайн-исследований: попытка каталогизации / Онлайн-исследования в России 3.0 / Под редакцией Шашкина А.В., Девятко И.Ф., Давидова С.Г. – М.: Издательский дом «Кодекс» , 2012. С. 29
2. ЗиновьевА. Мониторингсоциальныхсетей — как, чем, зачем?

URL:<http://www.cossa.ru/articles/155/7943/>

1. Лазарсфельд П., Мертон Р. Массовая коммуникация, массовые вкусы и организованное социальное действие //Массовая коммуникация в современном мире: методология анализа и практика исследований. М. – 1999. – С. 138-149.
2. Майер-Шенберг В. «Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим» / Виктор Майер-Шенбергерб Кеннет Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванов и Фербер, 2014. – 240 с.
3. Петит А. «Как обычное исследование» / Онлайн-исследования в России 3.0 / Под редакцией Шашкина А.В., Девятко И.Ф., Давидова С.Г. – М.: Издательский дом «Кодекс» , 2012. С. 131-137

Фонд VoxPopuli Прогноз выборов в Венесуэле

URL: http://vox-populi.ru/venezuala.phtml

# Приложения

|  |
| --- |
| Не указано |

Рис. 1. Социограмма неформального сообщества пользователей ЖЖ, обсуждающих тематику «Законы о пропаганде гомосексуализма и педофилии» (получено с использованием пакета JUNG)[[41]](#footnote-41)

|  |
| --- |
| Не указано |

Рис. 2. Социограмма неформального сообщества, а также каналов распространения информации среди пользователей ЖЖ, обсуждающих тематику «Законы о пропаганде гомосексуализма и педофилии» (получено с использованием пакета Prefuse) [[42]](#footnote-42)

1. Katz E., Lazarsfeld P. F. Personal Influence, The part played by people in the flow of mass communications. – Transaction Publishers, 1970. [↑](#footnote-ref-1)
2. Fake (от англ.) – поддельный. Используется в терминологии Интернет. Также обозначает феномен «поддельной коммуникации»

Глазкова С. А. Особенности коммуникации в поддельных микроблогах сети Интернет //Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики.

URL:http://scjournal.ru/articles/issn\_1997-292X\_2011\_3-1\_08.pdf [↑](#footnote-ref-2)
3. Lerman K., Ghosh R. Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks

URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1509/1839 [↑](#footnote-ref-3)
4. Воат А.А. Использование технологий сплин-докторинг, астротурфинг и инфотейнмен в политической рекламе как институт воздействия на массовое сознание.// Этносоциум и межнациональная культура. 2010. Т.28.№ 4. С.68-77 [↑](#footnote-ref-4)
5. Интернет в России: динамика проникновения / ФОМ, Зима 2012-2013

URL: http://runet.fom.ru/Proniknovenie-interneta/10853 [↑](#footnote-ref-5)
6. Девятко И.Ф. Инструментарий онлайн-исследований: попытка каталогизации / Онлайн-исследования в России 3.0 / Под редакцией Шашкина А.В., Девятко И.Ф., Давидова С.Г. – М.: Издательский дом «Кодекс» , 2012. С. 29 [↑](#footnote-ref-6)
7. Бершадская Л., Биккулов А., Болгова Е., Чугунов А., Якушев . Социальные сети и социометрические исследования: теоретические основания и практика использования автоматизированного инструментария изучения виртуальных сообществ // «Информационные Ресурсы России» №4, 2012

URL:http://www.aselibrary.ru/digital\_resources/journal/irr/irr3648/irr36483711/irr364837113734/irr3648371137343743/ [↑](#footnote-ref-7)
8. Katz E., Lazarsfeld P. F. Personal Influence, The part played by people in the flow of mass communications. – Transaction Publishers, 1970. [↑](#footnote-ref-8)
9. Fake (от англ.) – поддельный. Используется в терминологии Интернет. Также обозначает феномен «поддельной коммуникации»

Глазкова С. А. Особенности коммуникации в поддельных микроблогах сети Интернет //Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики.

URL:http://scjournal.ru/articles/issn\_1997-292X\_2011\_3-1\_08.pdf [↑](#footnote-ref-9)
10. Kozinets R.V. (1998). «On netnography. Initial reflections on consumer research investigations of cyberculture». *Advances in Consumer Research*,Vol. 25(1), pp. 366–371. [↑](#footnote-ref-10)
11. Sameer Hinduja, Justin W. Patchin Personal information of adolescents on the Internet:A quantitative content analysis of MySpace

URL:http://sfcs.cals.arizona.edu/azsearch/sites/sfcs.cals.arizona.edu.azsearch/files/Hinduja,%202008.pdf [↑](#footnote-ref-11)
12. Зиновьев А. Мониторинг социальных сетей — как, чем, зачем?

URL:<http://www.cossa.ru/articles/155/7943/> [↑](#footnote-ref-12)
13. Sandra González-Bailón Social Science in the Era of Big Data / Social Science Research Network, March 22, 2013

URL: http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2238198 [↑](#footnote-ref-13)
14. Майер-Шенберг В. «Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим» / Виктор Майер-Шенбергерб Кеннет Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванов и Фербер, 2014. – 240 с. [↑](#footnote-ref-14)
15. Майер-Шенберг В. «Большие данные: революция, которая изменит то, как мы живем, работаем и мыслим» / Виктор Майер-Шенбергерб Кеннет Кукьер; пер. с англ. Инны Гайдюк. – М.: Манн, Иванов и Фербер, 2014. – 240 с. [↑](#footnote-ref-15)
16. Chris Anderson The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

URL: http://www.wired.com/science/discoveries/magazine/16-07/pb\_theory [↑](#footnote-ref-16)
17. Там же [↑](#footnote-ref-17)
18. Kate Crawford, Danah Boyd Six Provocations for Big Data / “A Decade in Internet Time: Symposium

on the Dynamics of the Internet and Society” / September 21, 2011

URL: http://softwarestudies.com/cultural\_analytics/Six\_Provocations\_for\_Big\_Data.pdf [↑](#footnote-ref-18)
19. Sarah Oates, Tetyana Lokot Twilight of the Gods?: How the Internet Challenged Russian Television News Frames in the Winter Protests of 2011-12 / Social Science Research Network, June 28, 2013

URL: http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2286727 [↑](#footnote-ref-19)
20. Mackay H. Information and the Transformation of Sociology: Inter-activity and Social Media Monitoring // Journal for a Global Sustainable Information Society, Vol. 11, №1, 2013 [↑](#footnote-ref-20)
21. RAWNSLEY A. Pentagon Wants a Social Media Propaganda Machine

URL:http://www.wired.com/dangerroom/2011/07/darpa-wants-social-media-sensor-for-propaganda-ops/ [↑](#footnote-ref-21)
22. Бершадская Л., Биккулов А., Болгова Е., Чугунов А., Якушев . Социальные сети и социометрические исследования: теоретические основания и практика использования автоматизированного инструментария изучения виртуальных сообществ // «Информационные Ресурсы России» №4, 2012

URL:http://www.aselibrary.ru/digital\_resources/journal/irr/irr3648/irr36483711/irr364837113734/irr3648371137343743/ [↑](#footnote-ref-22)
23. Фонд Vox Populi Прогноз выборов в Венесуэле

URL: http://vox-populi.ru/venezuala.phtml [↑](#footnote-ref-23)
24. Петит А. «Как обычное исследование» / Онлайн-исследования в России 3.0 / Под редакцией Шашкина А.В., Девятко И.Ф., Давидова С.Г. – М.: Издательский дом «Кодекс» , 2012. С. 131-137 [↑](#footnote-ref-24)
25. Такие сообщения относятся к спаму, так как оно нацелено на рекламу, а не выражает мнения человека. Например, в сообщении человек рассказывает о своем отношении к недавнему общественному событию, а в конце или даже середине текста размещается ссылка с переходом на магазин лопат. [↑](#footnote-ref-25)
26. Уилсон Р. Введение в теорию графов. [↑](#footnote-ref-26)
27. Князева Е.И. Методологические характеристики и познавательные возможности исследований социальных сетей в социологии

URL:http://www.hse.ru/data/2012/12/03/1302111395/7\_Современная%20Сб.%20памяти%20А.О.%20Крыштановского%202012.pdf [↑](#footnote-ref-27)
28. Даль В. И. Толковый словарь живого великорусского языка. – в, 2008. – Т. 6. [↑](#footnote-ref-28)
29. Попова А. В. Дискурс-картины мира и кортежного взаимодействия элитарных средств информации/АВ Попова

URL:http://elib.bsu.by/bitstream/123456789/20387/1/%D0%9C%D0%BE%D0%BD%D0%BE%D0%B3%D1%80%D0%B0%D1%84%D0%B8%D1%8F%20\_%D0%9F%D0%BE%D0%BF%D0%BE%D0%B2%D0%B0\_.pdf [↑](#footnote-ref-29)
30. Яндекс-помощь

URL: http://help.yandex.ru/catalogue/citation-index/tic-about.xml [↑](#footnote-ref-30)
31. http://wobot.ru/ [↑](#footnote-ref-31)
32. [↑](#footnote-ref-32)
33. http://usconservatives.about.com/od/glossaryterms/g/Astroturf.htm [↑](#footnote-ref-33)
34. Сайт РИА-Новости

URL: http://ria.ru/ [↑](#footnote-ref-34)
35. Данное допущение справедливо при условии, что новые сообщения с этой новостью не будут опубликованы «лидером мнений». В таком случае, величина охвата может значительно измениться. [↑](#footnote-ref-35)
36. Сайт канала «Russia Today»

URL:http://rt.com/about-us/ [↑](#footnote-ref-36)
37. Сайт канала «Russia Today»

URL:http://rt.com/about-us/distribution/ [↑](#footnote-ref-37)
38. Сайт «Русская линия»

URL: http://ruskline.ru/tema/ [↑](#footnote-ref-38)
39. фейковые профили, искусственно созданные аккаунты, имитирующие деятельность реального пользователя [↑](#footnote-ref-39)
40. Кейс 1. Реальная новость: «Министр обороны: Италия может направить миротворцев на Украину»

Кейс 2. Ложная новость: «Минобороны РФ - Небольшой сбой курса российской ракеты РС-24 ЯРС. Пентагон - Это провокация, ПВО США не смогли взять цель» [↑](#footnote-ref-40)
41. Бершадская Л., Биккулов А., Болгова Е., Чугунов А., Якушев . Социальные сети и социометрические исследования: теоретические основания и практика использования автоматизированного инструментария изучения виртуальных сообществ // «Информационные Ресурсы России» №4, 2012

URL:http://www.aselibrary.ru/digital\_resources/journal/irr/irr3648/irr36483711/irr364837113734/irr3648371137343743/ [↑](#footnote-ref-41)
42. 7 Там же [↑](#footnote-ref-42)