

УДК 911.375

ТИПЫ СЕТЕВОГО ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ “ВКОНТАКТЕ” В ГОРОДАХ ВОЛОГОДСКОЙ ОБЛАСТИ

© 2024 г. Н. А. Синицын*

*Московский государственный университет имени М.В. Ломоносова, географический факультет,
Москва, Россия*

**e-mail: nicksinus@yandex.ru*

Поступила в редакцию 05.12.2020 г.

После доработки 23.05.2024 г.

Принята к публикации 18.07.2024 г.

Данные социальных сетей используются специалистами гуманитарных наук для анализа общества уже более десяти лет, однако потенциал этого источника еще не раскрыт до конца. Работа выполнена на основе открытых данных пользователей социальной сети “ВКонтакте”, проживающих в городах Вологодской области. Эти данные — отличный материал для социологических и общественно-географических исследований. Информация собрана с использованием API “ВКонтакте”. Разработана методика очистки данных от фейковых пользователей. Предложена комплексная характеристика пользователя, отражающая его вкусы и интересы по набору сообществ, на которые он подписан. Группа пользователей, склонная подписываться на определенный набор сообществ, названа типом сетевого поведения. Они были выделены с помощью разработанной методики кластеризации графа пользователей и сообществ, в основе которой лежит силовая укладка графа алгоритмом OpenOrd. Одиннадцать полученных типов сетевого поведения можно разделить на две группы: половозрастные и специализированные типы. Сообщества половозрастных типов не имеют основной темы, отличаются большим числом пользователей, важную роль в них играют юмористические ресурсы. Сообщества специализированных типов объединены одной или двумя основными темами, там на порядок меньше пользователей, мало юмористических сообществ. Структура половозрастных кластеров города в Вологодской области зависит от его плотности. На структуру специализированных кластеров кроме плотности влияет состав экономики города. Разнообразие типов сетевого поведения города прямо пропорционально численности населения города, при этом у городов с сопоставимой плотностью, разнообразие связано с ролью сферы услуг в местной экономике.

Ключевые слова: города, Вологодская область, открытые данные, социальная сеть, сетевое поведение, “ВКонтакте”, кластерный анализ, граф, OpenOrd, энтропия Шеннона

DOI: 10.31857/S2587556624040106 EDN: RPEKUO

ПОСТАНОВКА ЗАДАЧИ И ОБЗОР РАНЕЕ ВЫПОЛНЕННЫХ ИССЛЕДОВАНИЙ

Данные социальных сетей используются специалистами гуманитарных наук для анализа общества уже более пятнадцати лет. Нельзя сказать, что это новый источник данных, но потенциал его еще не раскрыт до конца. Данные пользователей сети “ВКонтакте” — отличный материал для социологических и общественно-географических исследований: большой охват респондентов, удобство сбора и анализа данных, но главное — отсутствие промежуточных агентов вроде социологических служб или государственных органов статистического учета (ФСГС, ФНС и др.). Это данные, полученные напрямую.

В статье сделана попытка ответить на следующие вопросы. Распадается ли совокупность пользователей “ВКонтакте” на группы по интересам в зависимости от подписок на сообщества или нет? Другими словами, существует ли тенденция определенной группы пользователей подписываться на ограниченный набор сообществ? Если таковые группы действительно существуют, зависит ли их концентрация в городе от его плотности и структуры экономики?

Вероятно, предлагаемый инструмент может служить измерителем уровня развития человеческого капитала города. Методика рассматривается на примере городов одного региона России, но возможно ее использование для любых территориальных единиц, по которым имеются аналогичные данные.

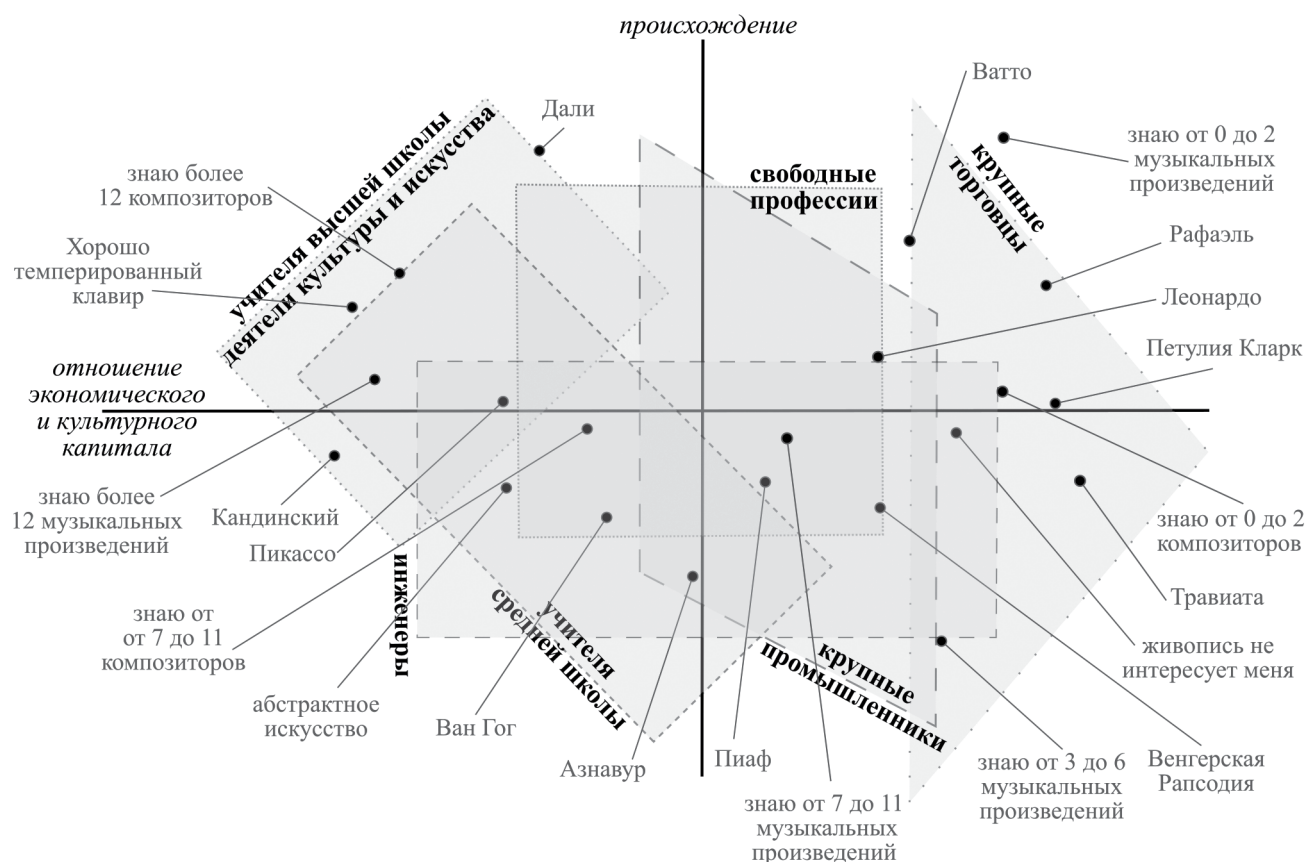


Рис. 1. Пространство П. Бурдьё для высшего класса [по (Bourdieu, 1984), отмечены не все точки].

Тип сетевого поведения — группа пользователей социальной сети, подписанных на сходный (но не одинаковый) набор сообществ. Сообщество (паблик) социальной сети — ресурс в социальной сети, посвященный определенной тематике и предназначенный для информационной коммуникации с пользователями. Спектр сообществ чрезвычайно обширен: официальные страницы крупного бизнеса, стихийные объединения маргинальных субкультур, юмор о школьной жизни, научно-популярные площадки, паблики для молодых мам, порталы про здоровье для пенсионеров — полный список тем, вероятно, сопоставим с общим кругом вопросов, волнующих любого достаточно взрослого современного человека.

Набор сообществ, на которые подписан пользователь в социальной сети, зависит от его вкусов и интересов. В социологии есть несколько работ, где предлагаются разные подходы к типологии людей по этим критериям. Все они, так или иначе, восходят к работе П. Бурдьё «Различение. Социальная критика суждения» (Bourdieu, 1984). Его исследование выполнено по данным социологических опросов на тему эстетических предпочтений французов: среди прочего респондентам нужно было указать авторов музыкальных произведений и фильмов, выбрать те сюжеты в изобразительном искусстве, которые они считают красивыми, пере-

числять любимых актеров, композиторов и художников, описать стиль обстановки своего жилища. По 26 вопросам создано более сотни метрик. Анализировать вместе сотню-другую свойств довольно трудно: П. Бурдьё использует анализ соответствий и переходит к двумерному пространству. Этот инструмент позволяет объединить метрики, высоко коррелирующие друг с другом, и свести их к новым агрегированным латентным переменным (факторам). Это похоже на факторный анализ, когда с помощью метода главных компонент снижают число измерений. Однако за ним стоит несколько иная математика, хотя и довольно похожая: в основе обоих методов лежит сингулярное разложение матрицы. Новые измерения получили адекватную интерпретацию. Горизонтальная ось — соотношение человеческого капитала (тип диплома и число лет обучения) и экономического капитала (доход) человека. Вертикальная ось — происхождение человека (измерялось через занятия отцов респондентов). Люди разных профессий действительно располагаются в разных частях этой координатной плоскости. Рассмотрим для примера высший класс (рис. 1). Похожую схему можно посмотреть в первоисточнике и для среднего класса.

Многоугольниками ограничены области, где расположены респонденты из какой-нибудь

профессиональной группы. Отдельными жирными точками показаны средние координаты респондентов, давших определенный ответ на некоторый вопрос. Например, координаты точки “Пиаф” рассчитаны как средние значения координат респондентов, считающих Э. Пиаф любимым музыкантом. П. Бурдые наложил два пространства друг на друга: пространство профессий и пространство произведений культуры и затем наблюдал за их пересечением.

Крупные торговцы и крупные промышленники знают мало композиторов и музыкальных произведений, среди любимых музыкантов и художников указывают только широко известных классиков и их работы, а также популярных певцов. На другом полюсе — преподаватели высшей школы и деятели культуры и искусства. Знают много композиторов и музыкальных произведений, любят авангард (импрессионизм и абстракционизм) и малоизвестные работы музыкальных классиков. Между двумя полюсами расположились представители свободных профессий и инженеры. У них средние познания в музыке и живописи. Им интересен авангард, но в тоже время они слушают популярный шансон.

Исследование не лишено недостатков. Во-первых, выводы сделаны по диаграммам в двух измерениях, каждое из которых — одна из компонент, получившихся после снижения размерности данных с помощью анализа соответствий. В сумме обе компоненты объясняют около 10% дисперсии исходных данных как для высшего, так и для среднего класса. Это значит, что большая часть различий осталась необъясненной или их можно списать на случайные отклонения в данных (что маловероятно). Во-вторых, показана связь профессии человека и его культурного потребления только для высшего и среднего класса. Низший класс (а это большая часть населения) в метриках П. Бурдые осталась недифференцированной.

Р. Петерсон продолжил работы в направлении, которое обозначил П. Бурдые. Он изучал разницу в потреблении культурных благ между массой и элитой (Peterson, 1992). Судя по его данным, элита всеядна и потребляет блага как высокой, так и низкой культуры. Масса же питается исключительно продуктами низкой культуры (в том числе массовой). Ему, конечно, очень повезло с большим массивом хороших данных — Бюро переписей США провело обширный опрос на тему потребления культурных благ американцами.

Совсем недавно М.М. Соколов с соавторами (2016) показал важность не только профессии (класса), но также пола и возраста для понимания вкусов и интересов человека. Использовались данные электронных читательских карточек библиотек Санкт-Петербурга.

С.И. Сулов (2016, а, б) выделял кластеры среди политических пабликов социальной сети “ВКонтакте” в Санкт-Петербурге. М.В. Хлопотов с соавторами (2019) доказали способность данных “ВКонтакте” адекватно отражать интересы и вкусы человека. Это сделано путем сравнения с независимым источником (порталом “Кинопоиск”). Пользователи из кластеров, выделенных по сходству выставления оценок на “Кинопоиске”, склонны подписываться “ВКонтакте” на паблики, темы которых соответствуют темам их любимых фильмов. Любимые фильмы — те, что получили на “Кинопоиске” высокую оценку пользователя. Студенческий центр анализа данных Дальневосточного федерального университета совместно с компанией SocialDataHub сконструировали графы социальных связей студентов по данным сети “ВКонтакте” в разрезе каждого факультета¹. Сравнивались графы, построенные по дружеским связям и по лайкам. Сделана попытка связать успеваемость студента и его центральность в графе.

Несколько лет назад коллектив под руководством Н.Ю. Замятиной и А.Д. Яшунского собрал данные с помощью парсера через API “ВКонтакте” (Чекмышев, Яшунский, 2014) по всем пользователям этой социальной сети в разрезе возраста, пола, места учебы, работы и списка друзей (Замятина, 2012). В итоге был получен ряд интересных результатов. Показано явление виртуальной урбанизации — завышение отношения числа пользователей к населению в пределах крупного города и занижение в окружающих районах (Ивлиева, Яшунский, 2016). Установлено, что структура имен пользователей позволяет картографировать этнические ареалы с высокой точностью, по крайней мере по сетке муниципальных районов (Замятина, Яшунский, 2018). Предложена методика оценки образовательных миграций: из анкет пользователей-студентов вузов взята информация по специальности, месту рождения и полу, что позволило выделить зоны тяготения крупнейших образовательных центров в разрезе гуманитарных и технических факультетов (Замятина, 2012). Вместе с тем некоторые установки авторов представляются спорными. Например, они не делали никаких попыток фильтровать данные от фейковых пользователей, мотивируя это тем, что “искажения, вносимые за счет анкет, не соответствующих реальным людям, а заводимые программами-роботами для различных целей, не так значительны — анкеты роботов зачастую не содержат практически никаких сведений, тем самым, не влияя на массив содержательных анкет” (Замятина, Яшунский, 2018, с. 122).

¹ Графон на минималках, большие данные в большом университете. <https://theshuttle.ru/big-data-leaders> (дата обращения 25.10.2021 г.).

Еще два интересных географических исследования с использованием открытых данных “ВКонтакте” сделаны для поиска границы зон тяготения Москвы и Санкт-Петербурга на территории Тверской области (Смирнов и др., 2019), а также для анализа аудитории и содержания местных сообществ городов Тверской области (Смирнов, Смирнова, 2019).

Масштабные исследования социальных сетей — довольно трудоемкое дело, этим занимаются большие коллективы. Наиболее существенные результаты по анализу русскоязычных социальных сетей показал гарвардский центр Беркмана. Там выполнили два крупных анализа рунета: один по русскоязычному сегменту Твиттера (Etling et al., 2012), второй — по блогосфере (в основном Живой Журнал) (Etling et al., 2010). Использовали кластерный анализ и анализ графов. Получили кластеры, которые имеют хорошую интерпретацию. Они неплохо отражают основные детали общественной повестки конца нулевых — начала десятых годов.

Эти проекты показали широкие возможности, которые дают открытые данные социальных сетей для научных исследований.

МЕТОДИКА СБОРА И АНАЛИЗА ДАННЫХ

Рассматривались сообщества, на которые подписаны пользователи сети “ВКонтакте”, проживающие в городах Вологодской области с населением более 10 тыс. чел. в 2019 г. Порог численности жителей взят с небольшим запасом относительно официального критерия людности города в Вологодской области. По данным Левада-Центра (Волков, Гончаров, 2014, 2019), самой популярной социальной сетью в России остается “ВКонтакте”, ее аудитория непрерыв-

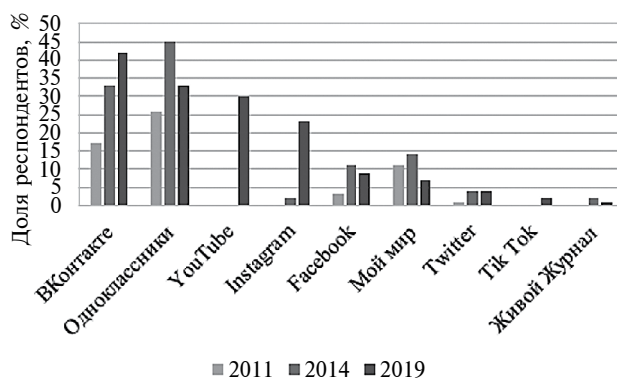


Рис. 2. Доля респондентов, указавших себя среди пользователей разных социальных сетей, % от всей выборки.

но растет с 2011 по 2019 г. (рис. 2). Вологодская область выбрана в качестве рабочего полигона благодаря своей типичности — по многим социально-экономическим показателям она близка к средним значениям по стране (рис. 3).

В мае 2019 г. были собраны данные пользователей сети “ВКонтакте”, которые при регистрации в графе city указали один из следующих городов: Череповец, Вологда, Сокол, Великий Устюг, Шексна, Грязовец, Бабаево, Кадуй, Вытегра, Харовск. Исходя из общего распределения пользователей по возрастам, выгружались данные только для пользователей 1950–2005 гг. рождения. Старше 70 лет довольно мало пользователей, младше 15 лет тоже нет смысла выгружать — формально, во “ВКонтакте” нельзя регистрироваться людям младше 14 лет. “ВКонтакте” предоставляет API для удобной работы с данными пользователей. API (application programming interface — интерфейс программирования приложений) — набор функций в социальной сети для машинного доступа к ее данным. Они нужны для работы приложений социальной сети, но их

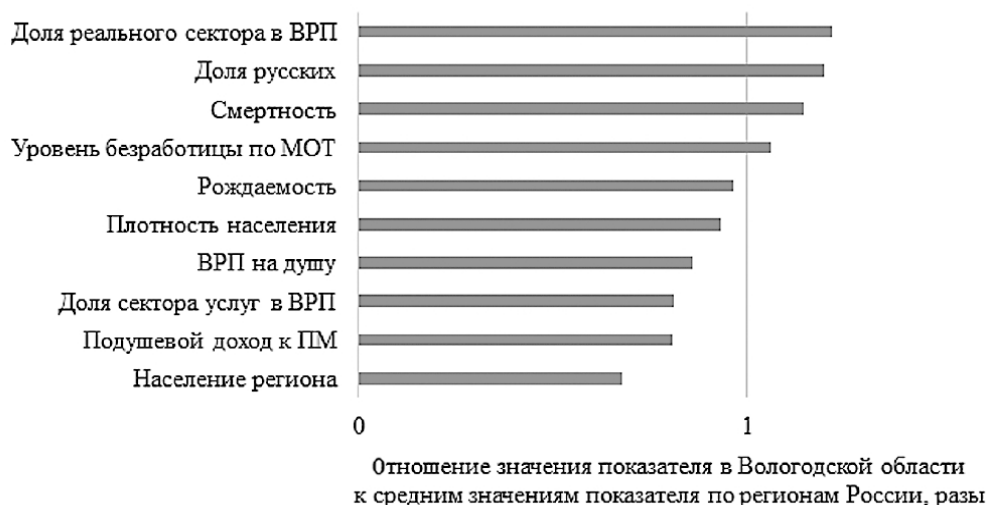


Рис. 3. Некоторые показатели Вологодской области на фоне регионов России, 2020 г.

функционал позволяет выгружать информацию о пользователях². Сбор, анализ и визуализация данных выполнены на языке R.

Первый этап сбора данных — составление списка пользователей по каждому городу с помощью функции `users.search`³. В качестве ответов запрашивались следующие характеристики пользователя: имя, фамилия, `id` (внутренний номер “ВКонтакте”), закрытость страницы, статус, число подписчиков, время последнего посещения.

Второй этап — фильтрация фейковых пользователей. Фейки (*fake* — подделка) — пользователи, использующие персональную страницу во “ВКонтакте” не для личных целей: маркетинг, политические провокации, искусственное повышение внутрисетевого рейтинга страниц известных персон, спам и др. Фильтрация фейков организована в три ступени.

1. Зная дату последнего захода пользователя в сеть и время загрузки, можно посчитать число дней с момента последнего посещения. Если оно больше 14, то пользователь считается фейком. Как правило, фейки создаются под конкретный проект и по его завершении про них забывают. Недостаток фильтра двух недель — теряется часть малоактивных пользователей.

2. Нетрудно выделить тех пользователей, у которых число подписчиков слишком велико. Порог чересчур большого числа подписчиков для каждого возраста свой. Так, для человека 1950-х годов рождения — 200–300 подписчиков, для пользователя 1990-х годов рождения — примерно 1500. Это связано с разным уровнем сетевой активности у разных поколений. Порог нетрудно рассчитать, отсекая хвосты на гистограмме распределения числа пользователей по количеству их подписчиков для каждого возраста. Фейки часто подписываются друг на друга для повышения внутрисетевого рейтинга. Чем выше внутрисетевой рейтинг пользователя, тем выше его место во внутрисетевой поисковой выдаче — это важно для фейков, занимающихся рекламой или продажами.

3. Последняя ступень — фильтр по статусам. Во “ВКонтакте” можно указать свой статус. Планировалось, что люди будут писать там свое настроение, эмоции, переживания, важные вещи, делать заявления, но, как правило, он пустой. Статус размещается на видном месте на странице пользователя — под его именем и фамилией. Этим воспользовались продавцы товаров и услуг — они размещают там объявления о своей деятельности. Составив словарь характерных

Таблица 1. Пример таблицы графа

Пользователь	Сообщество
Вася Сивый	Четкие приколы
Вася Сивый	Новости Нефтеперегоньевска
Алена Солнышко	Цитатная тетрадь
Алена Солнышко	Новости Нефтеперегоньевска
Алена Солнышко	Твой Позитив

слов, сокращений и сочетаний цифр (“куплю”, “тел.”, “руб.”, 921 и др.), можно выделять фейки по содержанию статуса.

Завершающий этап сбора данных — выгрузка списка сообществ, на которые подписан пользователь с помощью функции `users.getSubscriptions`⁴. Не все сообщества в списке сообщают одинаковое количество информации о поведении пользователя. Они располагаются там в порядке популярности. Чем больше пользователь обращал внимание на публикации из паблика (лайкал, репостил, смотрел картинки или видео, слушал аудиозаписи, оставлял комментарии), тем выше располагается паблик в личном рейтинге. Поскольку многие подписаны на десятки и даже сотни сообществ, то было принято решение брать только первые 30 пабликов из списка пользователя.

Изменение половозрастной структуры выборки при выгрузке данных показано на рис. 4. Во “ВКонтакте” есть частотные ограничения при выгрузке информации (3 запроса в секунду), которые можно обойти, используя функцию `execute` (75 запросов в секунду)⁵. Далее переходим к предварительной обработке собранных данных.

Формируем граф в табличной форме (табл. 1). Строка таблицы — ребро графа, им является подписка пользователя на сообщество. Соответственно, пользователи и сообщества — вершины графа. В первом столбце указан пользователь, во втором — паблик, на который он подписан. Ребра графа не имеют весов.

Прежде чем приступить к анализу, нужно подготовить собранную таблицу графа. Удалены сообщества, отражающие локальную повестку населенного пункта (“Сокольские сплетни”, “Подслушано Харовск” и др.). До фильтрации местных сообществ общее число уникальных пабликов было 327 тыс., после фильтрации — 319 тыс. При этом число ребер уменьшилось с 6.1 до 5.3 млн. Число сооб-

² Выполнение запросов к API ВКонтакте. https://vk.com/dev/api_requests (дата обращения 25.10.2020 г.).

³ Список методов: `Users: users.search`. <https://vk.com/dev/users.search> (дата обращения 25.10.2020 г.).

⁴ Список методов: `Users: users.getSubscriptions`. <https://vk.com/dev/users.getSubscriptions> (дата обращения 25.10.2020 г.).

⁵ Список методов: `execute`. <https://vk.com/dev/execute> (дата обращения 25.10.2020 г.).

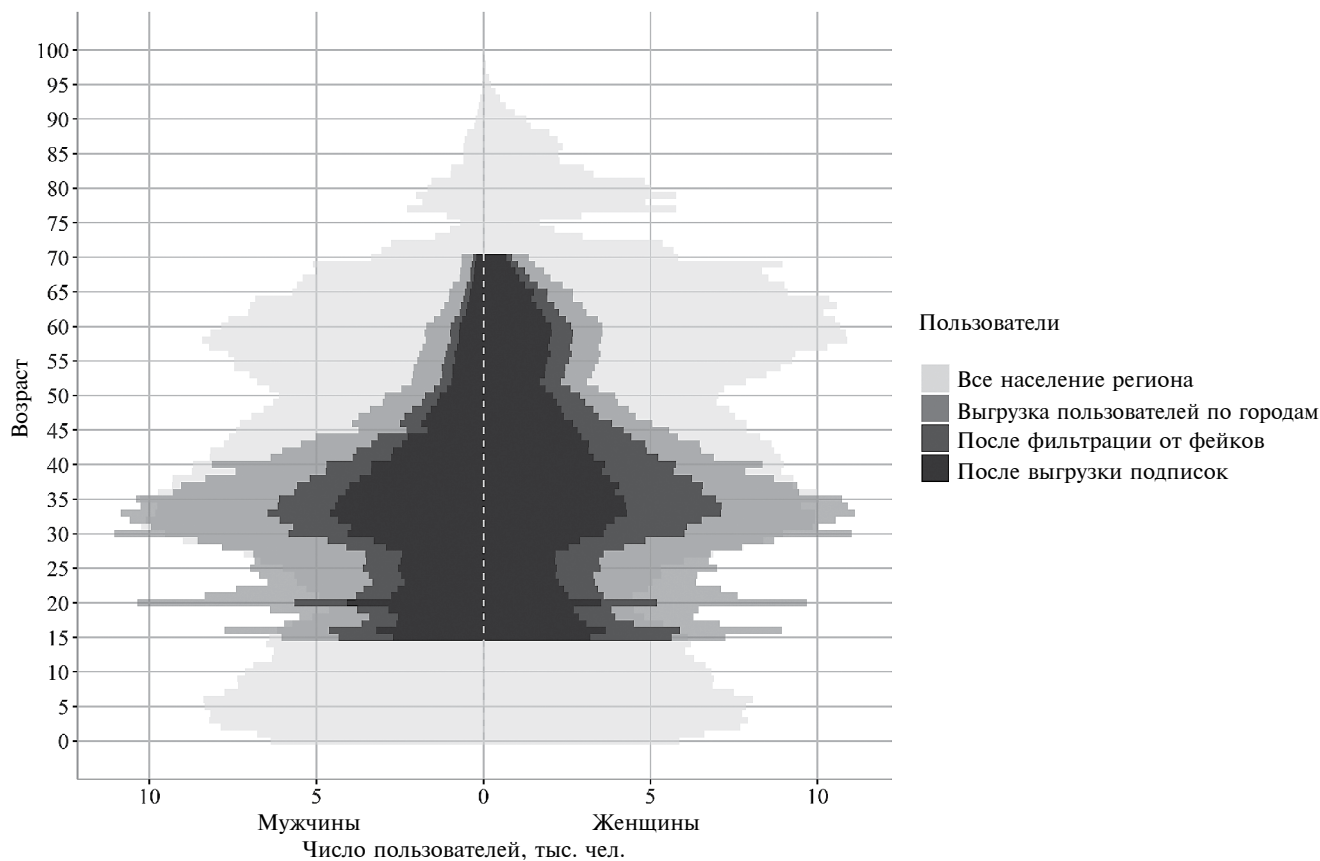


Рис. 4. Трансформация половозрастной пирамиды пользователей выгрузки “ВКонтакте” при загрузке и фильтрации данных.

ществ уменьшилось только на 2.5%, а число ребер — на целых 12.5%. Важно фильтровать данные от локальных пабликов, они обладают большой агрегирующей способностью (на них приходится 1/8 всех ребер графа). Локальные паблики показывают только территориальную принадлежность пользователя (даже не идентичность), а не его пристрастия в потреблении культурных благ. Убираем

сообщества, у которых есть только одна подписка. Они не обладают агрегирующей способностью и бесполезны в укладке графа. Число сообществ существенно сократится — с 319 до 136 тыс.

Наконец, убираем пользователей с одной подпиской или без подписок. Они тоже не обладают агрегирующей способностью при укладке графа (рис. 5).



Рис. 5. Изменение числа пользователей выборки при выгрузке данных.

Затем производится кластеризация графа, методика которой представлена на рис. 6. Важнейшая часть методики — укладка графа. Алгоритм укладки отвечает за концентрацию вершин графа в отдельные кластеры, его задача — расположить на плоскости вершины графа так, чтобы те из них, которые хорошо связаны друг с другом, находились рядом. Для этого используется алгоритм OpenOrd, его старое название — *drl*, под этим именем он до сих пор значится в пакете *igraph* в языке R. Разработан Ш. Мартином с соавторами в 2008 г. (Martin et al., 2011). В настройках алгоритма OpenOrd параметр числа отброшенных ребер задан 0.9 (стандартный — 0.8).

Выделить кластеры на укладке графа напрямую не получится. Во-первых, плотность вершин в разных сгущениях сильно различается, а алгоритмы плотностной кластеризации, например *dbscan*, не могут корректно выделять кластеры с разной плотностью точек. Во-вторых, данные сильно зашумлены (шум — точки, которые не попадают в сгущения вершин на укладке графа), а иерархическая кластеризация и *k-means* плохо работают на зашумленных данных (*k-means* в данном случае вообще не подходит — мы не знаем точно число кластеров). В-третьих, для иерархической кластеризации нельзя брать все точки на укладке (почти 400 тыс. шт.) — матрица расстояний не поместится в оперативную память персонального компьютера. В итоге, берется случайная выборка из 20 тыс. точек; используя *dbscan*, убирается шум (его плотность невелика и везде одинакова); затем с помощью иерархической кластеризации выделяются центры кластеров, к которым приписываются оставшиеся точки по принципу ближайшего соседа.

ТИПЫ СЕТЕВОГО ПОВЕДЕНИЯ

Кластеризация выполнена для графа из 386 тыс. вершин — 250 тыс. личных страниц и 136 тыс. сообществ. Кластеры действительно существуют, они хорошо выражены, сильно отличаются по размеру и плотности (рис. 7). Укладка графа интересна тем, что если провести диагональ из левого верхнего угла в правый нижний, то небольшие кластеры снизу от нее будут состоять из фейков. Хотя фейковых кластеров много, они довольно малочисленные — 35 тыс. пользователей из 250 тыс., или 14%. Пользователь может подписываться на другого пользователя, не только на сообщество — фейки часто подписываются друг на друга для повышения рейтинга. Так появляются кластеры, собранные только из пользователей. Кроме необычных кластеров из пользователей среди фейковых кластеров есть несколько тематических:

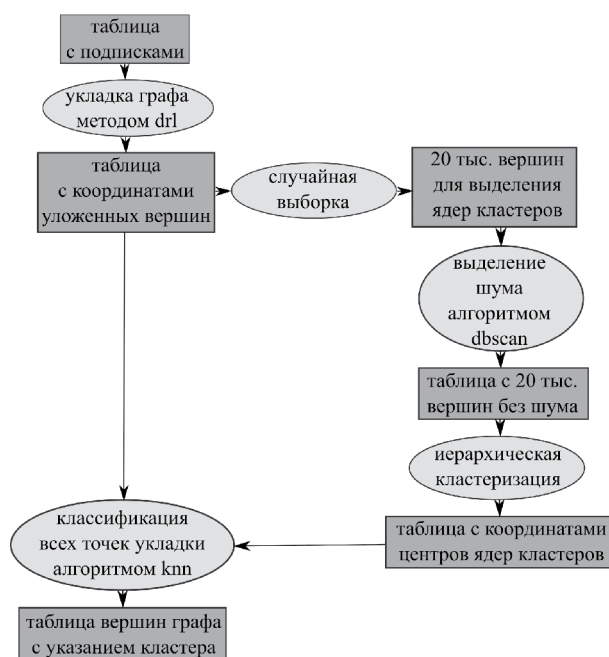


Рис. 6. Методика кластеризации графа.

- кластер, где собраны паблики для поиска работы в других городах России;
- кластер секс-знакомств, эротики и порнографии;
- два кластера из интернет-магазинов и страниц торговых предприятий;
- два религиозных кластера (православный и протестантский).

Сравнительно небольшое число оставшихся фейковых пользователей говорит о том, что методика их фильтрации выполняет свою функцию, хотя и не до конца, в первую очередь потому, что небольшая часть фейков все-таки является рабочими, неспящими. Они концентрируются в отдельные кластеры, которые можно не рассматривать в дальнейшем анализе. Этот прием использовался для исключения маркетинговых аккаунтов при исследовании русскоязычного сегмента Твиттера в центре Беркмана (Etling et al., 2012) — там тоже фейковые аккаунты объединились в один кластер, который потом просто убрали.

Для каждого кластера выполнены следующие действия (табл. 2): построена пирамида пользователей; по каждому городу Вологодской области рассчитана доля пользователей в кластере от всех пользователей в городе; определены темы 30 пабликов, на которые подписано максимальное число пользователей.

Есть два варианта кластеров с реальными (нефейковыми) пользователями: половозрастные и специализированные (рис. 8). В половозрастных кластерах содержание сообществ соответствует полу и возрасту пользователей. Например, в кластере “Пенсионерки” будут

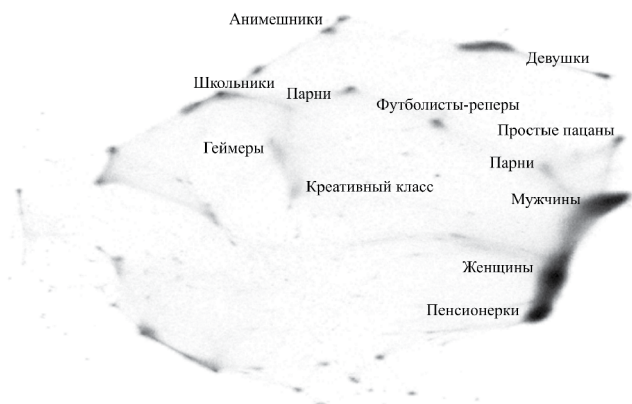


Рис. 7. Кластеры на укладке графа (ребра графа не показаны, фейковые кластеры не подписаны).

паблики о даче, рецептах, здоровье, пенсионерах, новостях, домоводстве. В таких кластерах нет какой-нибудь основной темы или нескольких тем, кроме одного исключения — на первом месте обычно с большим отрывом всегда идут юмористические паблики. Юмор всегда с поправкой на пол и возраст — у “Школьников” про учебный процесс, а у “Женщин” — про быт, отношения и котов.

В специализированных кластерах ситуация иная: там нет или есть буквально единичные примеры юмористических пабликов. Юмор всегда носит узкий специфический характер, например, циничный юмор в кластере “Футболисты-реперы”. Самая главная черта специализированных пабликов — одна или две основные темы, скажем, научпоп у “Креативного класса” или автомобили у “Простых пацанов”.

Большая часть (95%) всех пользователей принадлежит к половозрастным кластерам (“Женщины”, “Мужчины”, “Пенсионерки”, “Девушки”, “Школьники”, “Парни”). Только 5% пользователей попадает в специализированные кластеры (“Футболисты-реперы”, “Геймеры”, “Простые пацаны”, “Креативный класс”, “Анимешники”).

Обычно социологии, как уже упоминалось в обзоре литературы, рассматривают типы культурного потребления через разрез массы и элиты. В данном случае это не совсем верно. Правильнее будет говорить о наличии массы и девиаций: прогрессивных (“Креативный класс”, “Геймеры”, “Анимешники”) и консервативных (“Футболисты-реперы”, “Простые пацаны”). Здесь надо отметить, что разные школы соци-

Таблица 2. Основные характеристики типов сетевого поведения

Тип сетевого поведения/ кластер	Средний возраст, лет	Примерное соотношение мужчин и женщин, %	Основные темы 30 крупнейших сообществ	Категория людности городов, где доля кластера максимальна
“Женщины”	30–40	20/80	Женский юмор, домоводство, отношения, кулинария, ЗОЖ	крупные
“Пенсионерки”	50–60	10/90	Женский юмор, новости, кулинария, пенсионеры, здоровье, домоводство, дача	малые
“Мужчины”	30–40	90/10	Мужской юмор, кино, автомобили, отношения	доля одинакова по городам
“Школьники”	15–18	50/50	Подростковый юмор	малые
“Девушки”	17–21	10/90	Юмор для девушек, поп-культура, красота, эстетика, астрология	крупные
“Парни”	18–25	80/20	Молодежный юмор, футбол, новости, морализаторские паблики*	крупные
“Креативный класс”	30–40	80/20	Научпоп, спорт, новости, музыка, дизайн	крупные
“Геймеры”	20–30	90/10	Компьютерные игры, компьютеры, киберспорт	крупные
“Анимешники”	15–18	10:90	Аниме, рисование, астрология, k-поп	крупные
“Футболисты-реперы”	17–25	90:10	Футбол, циничный юмор, морализаторские паблики, рэп	крупные
“Простые пацаны”	15–25	90:10	Автомобили и музыка	доля не зависит от людности

Примечание. * В морализаторских пабликах публикуют и азартно обсуждают фото неподобающего, на взгляд пользователей, поведения.

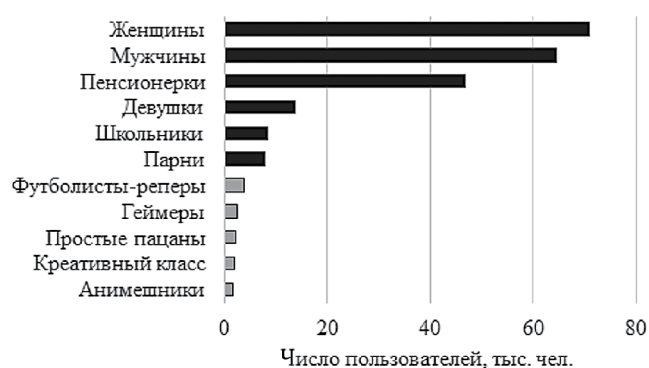


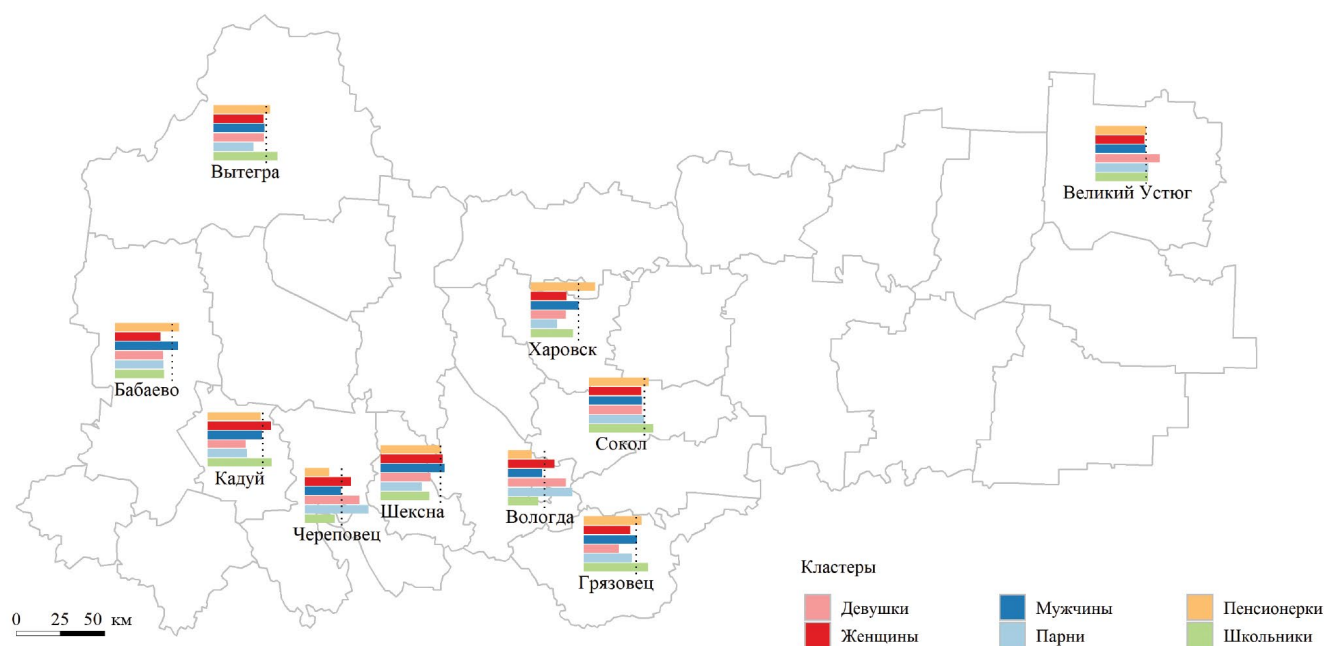
Рис. 8. Распределение кластеров по числу пользователей (черные — половозрастные, серые — специализированные).

ологии трактуют понятие девиации по-своему. Некоторые исследователи (Heckert, Heckert, 2002) говорят о наличии не только отрицательных девиаций (хулиганство, разбой, алкоголизм и т.д.), но и положительных (героизм, добросердечие, талант и т.д.). Термин девиация уместен еще тем, что на соответствующие кластеры действительно приходится сравнительно мало пользователей, на общем фоне это выглядит как отклонения от нормы (см. рис. 8).

ТЕРРИТОРИАЛЬНАЯ СТРУКТУРА ТИПОВ СЕТЕВОГО ПОВЕДЕНИЯ

Для картографирования типов сетевого поведения городов лучше всего подходят значки в виде диаграммы. Длина горизонтальных столбиков зависит от отношения доли кластера в городе к средней доле кластера по городам области. Все столбики пересекает вертикальная пунктирная черта, обозначающая отношение равное 1. Черта задает общий масштаб для всех городов. Это позволяет наглядно показывать соотношение между типами сетевого поведения в городе и сравнивать структуры типов сетевого поведения у городов разной людности между собой.

Структура половозрастных кластеров города зависит от его людности (рис. 9). В крупных городах — Череповце и Вологде (население около 300 тыс. чел.) — повышена доля кластеров “Женщины”, “Девушки” и “Парни”, но сильно понижена доля кластеров “Пенсионеры” и “Школьники”. Доля этих двух кластеров велика в малых городах с населением 10–20 тыс. жителей. Структура половозрастных кластеров средних городов (Великий Устюг и Сокол) строго соответствует средней по региону. Можно говорить об инверсии структуры половозрастных кластеров насе-



Длина горизонтальных столбиков диаграммы отражает отношение доли пользователей кластера в городе к средней доле пользователей кластера по городам области (вертикальная черта — отношение равно 1)

Рис. 9. Структура половозрастных кластеров городов Вологодской области.

ленного пункта при переходе от крупных городов к малым. Возможно, люди в трудоспособном возрасте, из которых состоят кластеры “Женщины”, “Девушки” и “Парни”, уезжают из малых городов в Череповец и Вологду на учебу или на работу, за счет чего их доля в крупных городах больше, чем в малых. Снижение доли пользователей Вологды и Череповца в младших возрастах, вероятно, связано также с тем, что в крупных городах люди этой когорты активно пользуются другими социальными сетями, например “Инстаграммом”. Не понятно, что объясняет повышенную популярность “ВКонтакте” в малых городах у женщин после 50 лет. Мужчины этого возраста там так же безразличны к “ВКонтакте”, как и в больших городах.

Структура специализированных кластеров тоже связана с людностью города (рис. 10). Доля всех специализированных кластеров в Вологде и Череповце повышена по сравнению с малыми и средними городами региона (кроме кластера “Простые пацаны”). Особенно сильно повышена доля “Креативного класса”, “Геймеров” и “Футболистов-реперов”. В промышленном Череповце (черная металлургия и минеральные удобрения) доля “Креативного класса” на 14% меньше, чем Вологде (с ее центральными функциями, информационными технологиями и точным машиностроением) – 0.91 и 1.04% соответственно. Средние города, как и в случае с половозрастными кластерами, являются отражением области. Анализировать географию

специализированных кластеров на примере малых городов не вполне корректно. Характерная доля специализированного кластера – порядка 0.5% населения города. Если на город приходится 3000 пользователей, то среднее число пользователей специализированных кластеров в малом городе – 10–20. При таком масштабе даже случайное добавление 3–4 пользователей способно существенно изменить долю кластера в городе (алгоритм укладки графа не является строго детерминированным).

Тем не менее стоит отметить повышенную долю “Креативного класса” в некоторых малых и средних городах области. Особенно это заметно в Харовске. Дело в сильных лыжных спортивных школах. Пользователи, интересующиеся лыжами (как беговыми, так и горными) часто встречаются среди типичного “креативного класса”, вроде программистов и дизайнеров, но из-за этого в кластер попали также многие спортсмены-лыжники. Важная особенность алгоритма силовой укладки графа – чем больше пользователей с характерным набором подписок на определенные сообщества, тем выше вероятность, что у них хватит агрегирующей способности собраться в отдельное, достаточно большое сгущение. В масштабах Вологодской области лыжников для этого оказалось недостаточно. Однако они выделились в отдельный маленький субкластер, рядом с большой массой “креативных” пабликов (рис. 11). Если проводить кластеризацию в масштабах всей страны или несколь-

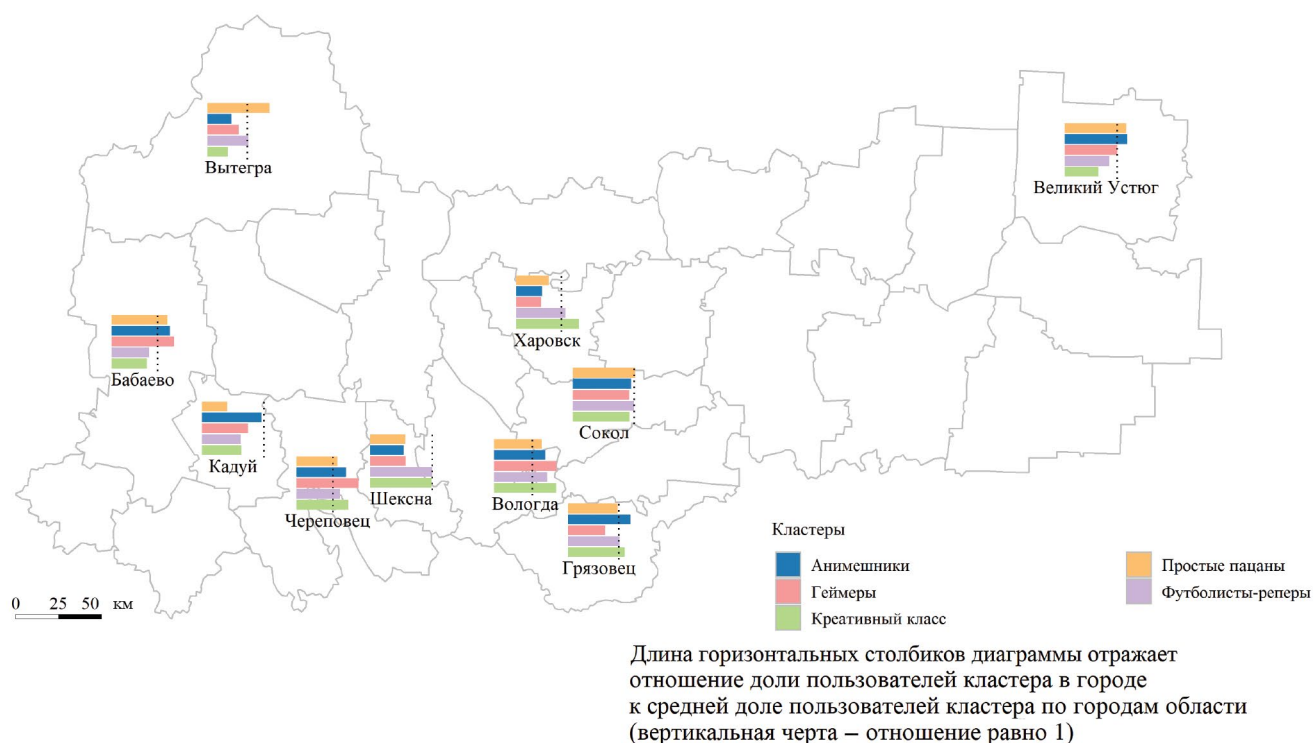


Рис. 10. Структура специализированных кластеров городов Вологодской области.

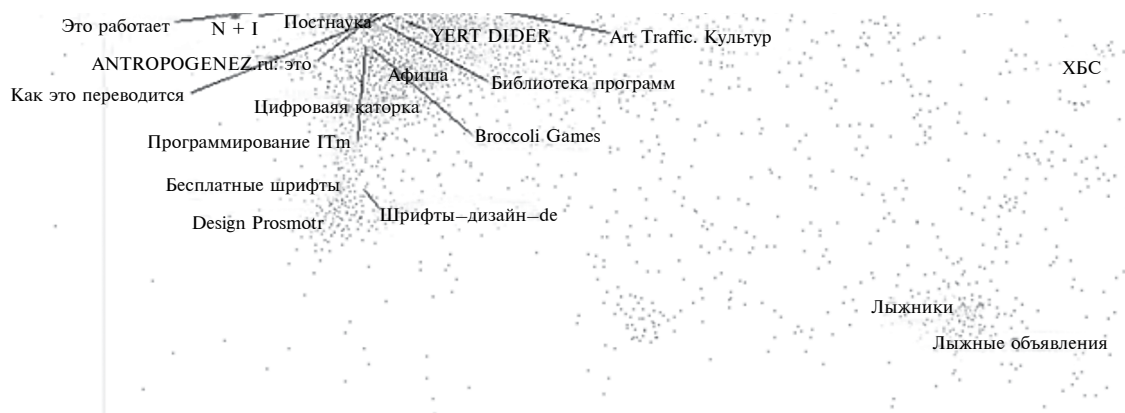


Рис. 11. Сгущение лыжников на периферии кластера “Креативный класс”.

ких регионов, вполне вероятно, что любители лыжного спорта сформируют отдельный большой кластер.

РАЗНООБРАЗИЕ ТИПОВ СЕТЕВОГО ПОВЕДЕНИЯ

Для оценки разнообразия использовалась энтропия Шеннона (Shannon, 1948) как наиболее теоретически обоснованная, хотя допустимо использовать любую меру разнообразия.

$$E = \sum_{i=1}^n w_i \times \log_2 w_i,$$

где E – разнообразие типов сетевого поведения города, $i = 1, 2, \dots, n$ – тип сетевого поведения, n – число типов сетевого поведения, w_i – доля пользователей определенного типа сетевого поведения в численности пользователей города. Соотношение между числом пользователей в городе и разнообразием типов сетевого поведения довольно точно описывается логарифмической зависимостью (рис. 12).

Разнообразие типов сетевого поведения пропорционально логарифму численности населения города. При сопоставимой людности разнообразие выше там, где больше доля сектора услуг в структуре экономики города (рис. 13, 14). Отклонения от линии тренда на графике соотношения разнообразия и населения города (см. рис. 13) связаны именно с ролью третичного сектора. Среди малых городов выше тренда лежат транспортные центры: Вытегра (шлюзы Волго-Балтийского канала), Грязовец (компрессорная станция газопровода Ухта–Торжок, ж/д станция), Бабаево (компрессорная станция газопровода Грязовец–Выборг, ж/д станция). Ниже линии тренда располагаются промышленные центры: Кадуй (электроэнергетика, лесопереработка), Харовск (лесопиление), Шексна (лесопереработка, исправительные колонии). Средние города: Великий Устюг расположен выше линии тренда (туризм, центральные функции),

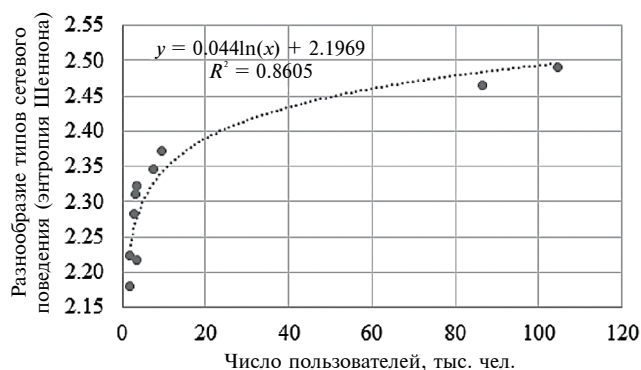


Рис. 12. Соотношение между числом пользователей в городе и разнообразием типов сетевого поведения по городам Вологодской области.

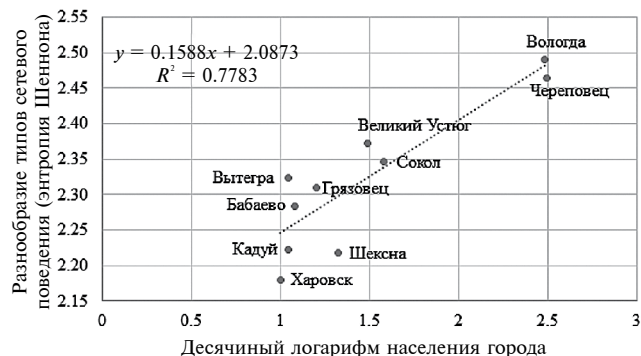


Рис. 13. Соотношение населения города и разнообразия типов сетевого поведения.

а Сокол, где развита лесопереработка, фактически, лежит на ней. Крупные города: Вологда расположена выше линии тренда, а промышленный Череповец – ниже.

ЗАКЛЮЧЕНИЕ

Не существует единой точки зрения о соотношении потребления культурных благ и их социально-экономических факторах. В рамках

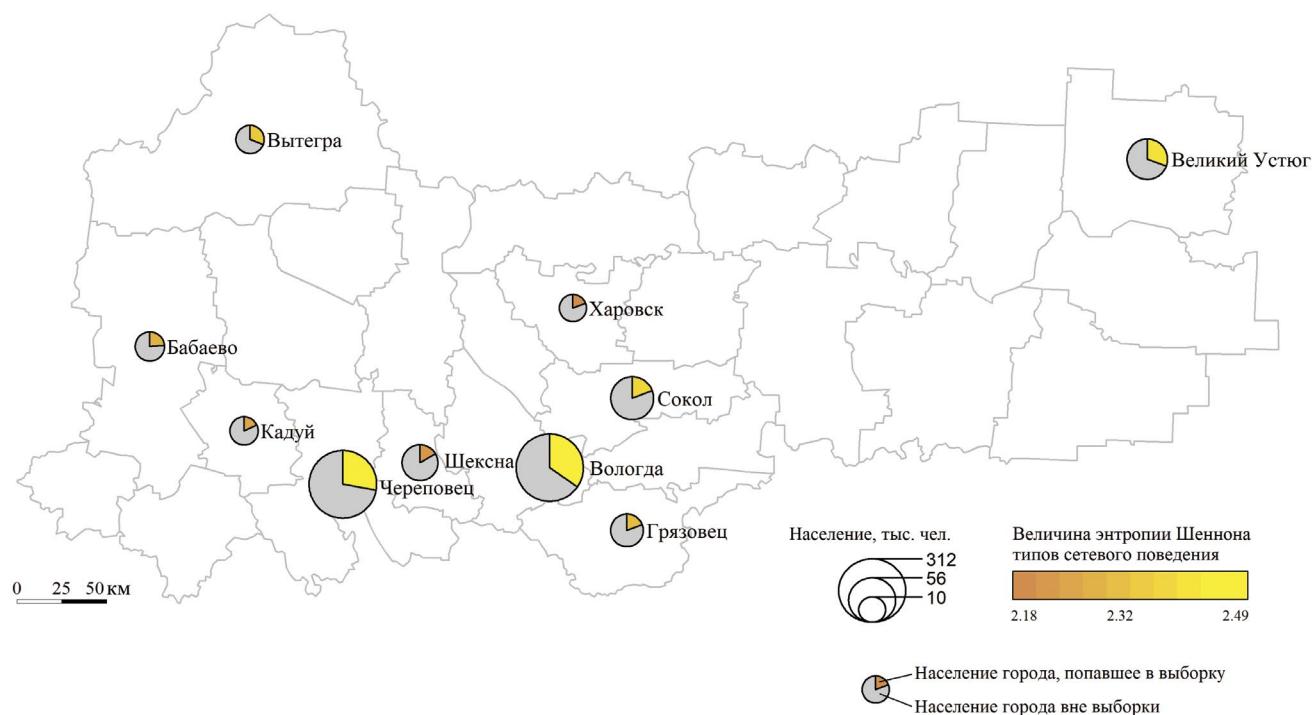


Рис. 14. Разнообразие типов сетевого поведения населения городов Вологодской области.

этого исследования показано, что вместо стандартного разделения на массу и элиту, уместнее говорить о массе и двух типах отклонений от нее — условно-прогрессивных и условно-консервативных.

Трехступенчатая фильтрация позволяет исключать фейковых пользователей социальных сетей по нескольким критериям — времени последнего посещения (две недели и более с даты выгрузки данных), числу подписчиков (фейки имеют завышенное число подписчиков для поднятия своего рейтинга), словарю статусов (фейки часто продают товары и услуги, используя статусы “ВКонтакте” для рекламы). Для корректного агрегирования вершин при кластеризации графа важно исключить сообщества с локальной повесткой и вершины с одним ребром.

В основе методики кластеризации графа пользователей и сообществ лежит его силовая укладка с помощью алгоритма OpenOrd. Выделяются два варианта кластеров (типов сетевого поведения): половозрастные и специализированные. Темы сообществ половозрастных кластеров зависят от пола и возраста типичного пользователя кластера. Единственной главной темой в половозрастном кластере всегда является юмор. Сообщества специализированных кластеров объединены уникальным для каждого кластера небольшим числом основных тем.

Структура типов сетевого поведения пользователей “ВКонтакте” в городах Вологодской

области зависит от людности города. Это справедливо как для половозрастных, так и для специализированных кластеров. Для объяснения структуры специализированных кластеров пользователей в городе важен состав его экономики. Для малых городов оценки по специализированным кластерам могут быть некорректными.

Разнообразие типов сетевого поведения пользователей в городе прямо пропорционально логарифму людности города с поправкой на долю сектора услуг. При сопоставимой людности разнообразие выше там, где доля сектора услуг больше.

СПИСОК ЛИТЕРАТУРЫ

- Волков Д., Гончаров С. Российский медиа-ландшафт: телевидение, пресса, интернет. М., 2014. 15 с.
- Волков Д., Гончаров С. Российский медиа-ландшафт 2019: телевидение, пресса, интернет и социальные сети. М., 2019. 21 с.
- Замятина Н.Ю. Метод изучения миграций молодежи по данным социальных интернет-сетей: Томский государственный университет как “центр производства и распределения” человеческого капитала (по данным социальной интернет-сети “ВКонтакте”) // Региональные исследования. 2012. № 2. С. 15–28.
- Замятина Н.Ю., Яшунский А.Д. Виртуальная география виртуального населения // Мониторинг об-

- ественного мнения: Экономические и социальные перемены. 2018. № 1. С. 117–137.
- Ивлиева О.Д., Яшунский А.Д. Виртуальная урбанизация // Городские исследования и практики. 2016. № 4. С. 26–36.
- Смирнов И.П., Смирнова А.А. Активность населения как ресурс развития городов Тверской области: опыт оценки по данным сети “ВКонтакте” // Вестн. Тверского гос. ун-та. Серия: экономика и управление. 2019. № 4 (48). С. 108–116.
- Смирнов И.П., Виноградов Д.М., Алексеев А.И. К Москве или к Санкт-Петербургу? Тяготение населения Тверской области по данным сети “ВКонтакте” // Изв. РГО. 2019. Т. 151. № 6. С. 69–80.
- Соколов М.М., Соколова Н.А., Сафонова М.А. Статусные культуры, биографические циклы и поколенческие изменения в литературных вкусах. Анализ полной базы данных читательских абонементов петербургских библиотек за 2014 год // Журн. социологии и социальной антропологии. 2016. № 3. С. 116–135.
- Суслов С.И. Сетевой анализ общественно-политических онлайн-сообществ Петербурга в социальной сети “ВКонтакте”: создание карты и кластеризация // Вестн. СПбГУ. Серия 12: Психология. Социология. Педагогика. 2016а. № 4. С. 69–87.
- Суслов С.И. Сетевой анализ смежности аудиторий политических онлайн-сообществ Санкт-Петербурга в социальной сети “ВКонтакте” // Современная наука и инновации. 2016б. № 2 (14). С. 196–207.
- Хлопотов М.В., Старцева Н.В., Макаренко А.А. Исследование кластеров кинолюбителей и их тематических сообществ в социальных сетях // Вестн. Евразийской науки. 2019. № 2. С. 1–11.
- Чекмышев О.А., Яшунский А.Д. Извлечение и использование данных из электронных социальных сетей // Препринты ИПМ им. М.В. Келдыша. 2014. № 62. С. 3–16.
- Bourdieu P. *Distinction: The Social Critique of the Judgment of Taste*. London: Routledge, 1984. 604 p.
- Etling B., Alexanyan K., Kelly J., Farris R., Palfrey J., Gasser U. *Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization* // Berkman Center Research Publication. 2010. Vol. 11. P. 1–46.
- Etling B., Faris R., Palfrey J., Gasser U., Kelly J., Alexanyan K., Barash V. *Mapping Russian Twitter* // Berkman Center Research Publication. 2012. Vol. 3. P. 1–16.
- Heckert A., Heckert D.M. A new typology of deviance: integrating normative and reactivist definitions of deviance // *Deviant Behavior*. 2002. Vol. 23. № 5. P. 449–479.
- Martin S., Brown W., Klavans R., Boyack K. OpenOrd: An Open-Source Toolbox for Large Graph Layout. In: *Visualization and Data Analysis 2011* / P.C. Wong, J. Park, M.C. Hao, C. Chens, K. Börner, D.L. Kao, J.C. Robert (Eds.). 2011. Vol. 7868.
- Peterson R.A. Understanding audience segmentation. From elite to mass and from snob to omnivore // *Poetics*. 1992. Vol. 21. № (4). P. 243–258.
- Shannon C.E. *A Mathematical Theory of Communication* // *Bell System Technical J.* 1948. Vol. 27. № 3. P. 379–423.

Types of Network Behavior of Users of Social Network “VKontakte” in the Cities of Vologda Oblast

N. A. Sinitsyn*

Lomonosov Moscow State University, Faculty of Geography, Moscow, Russia

**e-mail: nicksinus@yandex.ru*

This study is based on open data of the social network “VKontakte.” The personal data of users from the cities of Vologda oblast were collected using the “VKontakte” API. Several filters were developed to exclude fake users. Complex characteristics of users were created. It reflects his tastes and interests according to his subscriptions for communities. A group of users who tend to subscribe to a certain set of communities is called a pattern of social network behavior. The patterns were defined using the developed method of graph clustering based on the force layout (OpenOrd algorithm). Eleven obtained patterns of social network behavior were divided into 2 groups: age-sex and thematic. Communities of age-sex patterns have no common theme, they have many users, they contain a lot of humorous resources. Communities of thematic patterns have one or two common themes, they are much less populated, they contain a few numbers of humorous resources. The structure of age-sex patterns in a city depends on its population. The structure of thematic patterns of a city is also influenced by the composition of its economy. The diversity of the city’s social network behavior patterns is directly proportional to its population. The diversity is related to the role of services in the local economy for cities with comparable population.

Keywords: cities, Vologda oblast, open data, social network, “VKontakte,” cluster analysis, graph, OpenOrd, entropy, Shannon’s entropy

REFERENCES

- Bourdieu P. *Distinction: The Social Critique of the Judgment of Taste*. London: Routledge, 1984.
- Chekmyshev O.A., Yashunskii A.D. Data mining from electronic social networks and its applications. *Preprinty IPM Keldysha*, 2014, no. 62, pp. 3–16. (In Russ.).
- Etling B., Alexanyan K., Kelly J., Farris R., Palfrey J., Gasser U. Public Discourse in the Russian blogosphere: Mapping RuNet politics and mobilization. *Berkman Center Res. Publ.*, 2010, vol. 11, pp. 1–46.
- Etling B., Faris R., Palfrey J., Gasser U., Kelly J., Alexanyan K., Barash V. Mapping Russian Twitter. *Berkman Center Res. Publ.*, 2012, vol. 3, pp. 1–16.
- Heckert A., Heckert D.M. A new typology of deviance: integrating normative and reactivist definitions of deviance. *Deviant Behav.*, 2002, vol. 23, no. 5, pp. 449–479.
- Ivlieva O.D., Yashunskii A.D. Virtual urbanization. *Gorod. Issled. Praktiki*, 2016, no 4, pp. 26–36. (In Russ.).
- Khlopotov M.V., Startseva N.V., Makarenko A.A. Research of cinemaddict's clusters and their thematic communities in social networks. *Vestn. Evraz. Nauki*, 2019, no. 2, pp. 1–11. (In Russ.).
- Martin S., Brown W., Klavans R., Boyack K. OpenOrd: An open-source toolbox for large graph layout. In *Visualization and Data Analysis 2011. Vol. 7868*. Wong P. C., Park J., Hao M. C., Chens C., Börner K., Kao D. L., Robert J. C., Eds. 2011.
- Peterson R.A. Understanding audience segmentation. From elite to mass and from snob to omnivore. *Poetics*, 1992, vol. 21, no. 4, pp. 243–258.
- Shannon C.E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, vol. 27, no. 3, pp. 379–42.
- Smirnov I.P., Smirnova A.A. Population activity as a source of Tver region city's development: case study of "VKontakte" social network. *Vestn. Tver. Gos. Univ. Ser.: Ekon. Upravl.*, 2019, no. 4, pp. 108–116. (In Russ.).
- Smirnov I.P., Vinogradov D.M., Alekseev A.I. Moscow or St. Petersburg? Attraction of the Tver region population according to the VKontakte network. *Izv. RGO*, 2019, vol. 151, no. 6, pp. 69–80. (In Russ.).
- Sokolov M.M., Sokolova N.A., Safonova M.A. Status cultures, biographical cycles, and generational changes in literary tastes. Analysis of the complete database of readers' subscriptions of St. Petersburg libraries for 2014. *Zh. Sotsiol. Sots. Antropol.*, 2016, no. 3, pp. 116–135. (In Russ.).
- Suslov S.I. Network analysis of St. Petersburg political online communities in the social network "VKontakte": mapping and clusterization. *Vestn. SPbGU. Ser. 12: Psikhol. Sotsiol. Pedagog.*, 2016a, no. 4, pp. 69–87. (In Russ.).
- Suslov S.I. Network analysis of the contiguity of audiences of political online communities in St. Petersburg in the social network "VKontakte". *Sovrem. Nauka Innov.*, 2016b, no. 2, pp. 196–207. (In Russ.).
- Volkov D., Goncharov S. *Rossiiskii media-landshaft: televidenie, pressa, internet* [Russian Media Landscape: Television, Press, Internet]. Moscow, 2014.
- Volkov D., Goncharov S. *Rossiiskii media-landshaft 2019: televidenie, pressa, internet i sotsial'nye seti* [Russian Media Landscape 2019: Television, Press, Internet and Social Networks]. Moscow, 2019.
- Zamyatina N.Yu. Method of studying youth migration according to social Internet networks: Tomsk State University as a "production and distribution center" of human capital (according to the social Internet network "VKontakte"). *Reg. Issled.*, 2012, no. 2, pp. 15–28. (In Russ.).
- Zamyatina N.Yu., Yashunskii A.D. Virtual geography of virtual population. *Monitor. Obshchestv. Mneniya: Ekon. Sots. Peremeny*, 2018, no. 1, pp. 117–137. (In Russ.).