

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра социальной информатики

**РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА ДЛЯ АНАЛИЗА
ОНЛАЙН-СООБЩЕСТВ ВУЗОВ С ПОМОЩЬЮ
СОЦИАЛЬНЫХ ГРАФОВ**

(автореферат бакалаврской работы)

студента 4 курса 451 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
Социологического факультета
Чекмарева Александра Дмитриевича

Научный руководитель
старший преподаватель

_____ М.В. Белоконь
подпись, дата

Зав. кафедрой
кандидат социологических наук, доцент

_____ И.Г. Малинский
подпись, дата

Саратов 2026

ВВЕДЕНИЕ

Актуальность темы. В условиях цифровой трансформации высшего образования присутствие университета в социальных сетях является ключевым фактором формирования репутации и привлечения абитуриентов. «ВКонтакте» – наиболее популярная платформа среди российской молодежи: ежедневная аудитория превышает 50 млн пользователей, значительную часть которой составляют абитуриенты и студенты. Каждый вуз имеет разветвленную сеть онлайн-сообществ (официальные группы, студенческие советы, факультетские страницы, творческие и спортивные объединения, неформальные паблики), однако зачастую эти сообщества существуют изолированно, что снижает цифровую медийность университета. Возникает необходимость в объективном инструменте для измерения степени интеграции сообществ. Таким инструментом выступает анализ социального графа, вершинами которого являются сообщества, а ребрами – текстовые ссылки, репосты и совместные публикации. Среди множества метрик особый интерес представляет разреженность: чем ниже ее значение, тем выше связанность экосистемы, эффективнее распространение информации и цифровая медийность вуза.

Степень изученности. Проблематика анализа университетских сообществ в социальных сетях привлекает внимание зарубежных и российских исследователей. В работе А. Петерсена (2025) на основе сетей совместной встречаемости оценивается брендовый капитал университетов. Группа И. Алешковского и др. (2024) разрабатывает методологию сбора и типологию сообществ ВКонтакте. Исследования С. Докука и др., а также В.В. Сыча посвящены анализу студенческих сетей. Непосредственно построению социальных графов университетских сообществ посвящены работы сотрудников СГУ (2022). Однако целенаправленных исследований, где разреженность социального графа рассматривается как индикатор цифровой медийности вуза на основе данных VK API, до настоящего времени не проводилось.

Объект работы – онлайн-сообщества (группы, публичные страницы) трех университетов: Саратовского государственного университета имени Н.Г. Чернышевского (СГУ им Н.Г. Чернышевского), Саратовского государственного технического университета имени Гагарина Ю.А. (СГТУ им Ю.А. Гагарина) и Калининградского государственного технического университета (КГТУ) в социальной сети «ВКонтакте».

Предмет работы – структура социального графа, образованного связями (текстовые ссылки, репосты, совместные посты) между указанными сообществами, а также метрики его разреженности, плотности, центральности, кластеризации и ассортативности.

Цель работы – разработка программного средства для построения социального графа университетских сообществ на основе данных VK API и проведение сравнительного анализа трех университетов по критерию минимальной разреженности (максимальной связанности).

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Проанализировать теоретические основы построения социальных графов и метрики разреженности.
2. Выявить функциональную типологию университетских сообществ в ВКонтакте.
3. Разработать архитектуру парсера для сбора данных с использованием официального VK API.
4. Реализовать алгоритмы извлечения связей из текстов постов, вложений, репостов и совместных публикаций.
5. Построить взвешенные неориентированные графы для каждого университета.
6. Рассчитать базовые и расширенные метрики графов (плотность, разреженность, центральность, кластеризация, ассортативность, долю изолированных узлов).

7. Выполнить визуализацию графов и тепловых карт матриц смежности.
8. Провести сравнительный анализ полученных результатов и сформулировать рекомендации по повышению связанности сообществ.

Методы исследования. В работе используются теоретические методы (анализ литературы по теории графов и анализу социальных сетей), эмпирические методы (сбор данных через VK API), математические методы (расчет метрик графа), визуализационные методы (построение графов и тепловых карт).

Теоретическая значимость исследования заключается в возможности использования основных положений и выводов данной работы для дальнейшего изучения выбранной проблематики.

Структура работы. Бакалаврская работа состоит из введения, трёх глав, заключения, списка источников и приложений.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом разделе «Теоретические основы анализа социальных графов» рассматриваются основные понятия теории графов применительно к анализу университетских сообществ в социальной сети «ВКонтакте». Формально граф определяется как пара $G = (V, E)$, где V – множество вершин (онлайн-сообщества), E – множество ребер (текстовые ссылки, репосты, совместные посты). Обоснован выбор неориентированного взвешенного графа, поскольку для оценки общей связанности цифровой экосистемы направление ссылки не имеет принципиального значения. Веса ребер дифференцированы: текстовая ссылка (вес 1) – самый слабый вид связи, репост (вес 2) – более тесное взаимодействие, совместный пост (коллаборация) (вес 3) – наиболее сильная форма связи, требующая предварительного согласования между администраторами.

Определены ключевые метрики графа. Плотность вычисляется как отношение фактического числа ребер m к максимально возможному. Плотность принимает значения от 0 до 1, где 1 соответствует полностью связному графу.

Разреженность выбрана в качестве основного критерия цифровой медийности университета: чем ниже разреженность, тем более связанными являются сообщества, эффективнее распространяется информация и выше видимость вуза в цифровом пространстве. Установлены качественные границы: менее 0,30 – отличная связанность, 0,30–0,45 – средняя, более 0,45 – низкая. Средняя степень узла показывает, сколько связей в среднем приходится на одно сообщество. Коэффициент кластеризации измеряет стремление соседей узла образовывать треугольники, то есть плотные локальные группы. Центральность по степени выявляет сообщества-центры, на которые чаще всего ссылаются другие. Центральность по посредничеству показывает, как часто узел служит мостом на кратчайших путях между другими узлами, то есть соединяет различные кластеры. Ассортативность (от -1 до 1) характеризует, связаны ли узлы с похожими степенями: отрицательная ассортативность типична для сетей с несколькими крупными хабами, соединенными с множеством мелких узлов. Доля изолированных узлов является прямым показателем дезинтеграции экосистемы.

Разработана типология университетских сообществ, выделены пять функциональных типов. Официальные сообщества (главные страницы вузов, приемные комиссии, деканаты) создаются и администрируются сотрудниками университета, транслируют новостной контент, выступают главными хабами. Студенческие советы и профсоюзы представляют интересы обучающихся, создаются студентами, часто дублируют информацию официальных групп. Научно-образовательные сообщества (факультеты, кафедры, научные кружки) имеют узкую тематическую направленность, публикуют специализированный контент. Спортивные и творческие объединения (спортклубы, КВН, медиacentры, журналы, волонтерские центры) ориентированы на внеучебную деятельность, обладают высокой вовлеченностью и активно обмениваются ссылками. Неформальные сообщества создаются независимо от университета, генерируют контент пользователями, часто остаются на периферии графа. Показано, что учет типологии необходим для корректной интерпретации метрик:

высокая центральность официальной группы желательна, а высокая центральность неформального сообщества может быть необычной и требовать отдельного анализа.

Обоснована целесообразность использования официального VK API для сбора данных в сравнении с альтернативными методами. Преимущества API: скорость, структурированный ответ, наличие служебных полей «история репостов» и «вложения», легальность (не нарушает пользовательское соглашение VK, п. 6.3). Рассмотрены ограничения API: частота запросов не более 3 в секунду, максимальное количество постов за один вызов – 100 (используется пагинация), ошибки с кодами 5 (невалидный токен), 6 (слишком много запросов), 15 (недостаточно прав), 30 (группа удалена или закрыта). Для работы с API выбрана библиотека «vk_api» на языке Python, также применяется прямое использование библиотеки «requests» для полного контроля над запросами.

Во втором разделе «Разработка программного средства для парсинга и формирования графа» описана архитектура и реализация парсера. Спроектирована модульная структура из восьми последовательных этапов: авторизация, загрузка информации о сообществах, сбор постов, извлечение связей, агрегация, построение графа, расчет метрик, визуализация. Обоснован выбор технологического стека: Python 3.10+, библиотеки «requests», networkx (построение и анализ графов), pandas (обработка табличных данных), matplotlib и seaborn (визуализация). Разработана объектно-ориентированная архитектура с классом VKParser, содержащим 11 методов, в том числе api_call (универсальный вызов API с обработкой ошибок), get_group_info (получение названия, числа подписчиков, screen_name), resolve_screen_name (преобразование короткого имени в числовой ID с кэшированием), get_wall_posts (загрузка постов с пагинацией до 200 записей на сообщество), extract_all_links (поиск ссылок в 8 форматах), extract_links_from_attachments (извлечение скрытых-ссылок из вложений), extract_repost_or_collab (обнаружение репостов и совместных постов через поле copy_history), calculate_advanced_metrics (расчет центральности,

кластеризации, ассортативности), `create_heatmap` (построение тепловой карты матрицы смежности), `visualize_graph` (визуализация графа с цветовым кодированием ребер), `generate_analysis_report` (формирование текстового отчета по вузу). Реализованы алгоритмы определения вершин графа. Формирование выборки производилось вручную в три этапа: поиск сообществ по ключевым словам (для СГУ – «СГУ», «Саратовский университет», «SSU»). Для СГТУ – «СГТУ», «Политех Саратов». Для КГТУ (Калининград) – «КГТУ», «Калининградский технический университет»). Фильтрация по числу подписчиков (не менее 300). Оценка актуальности (посты за последние 3 месяца). В результате для СГУ отобрано 11 сообществ, для СГТУ – 12, для КГТУ – 9. Примеры для СГУ: официальная группа «Саратовский университет | СГУ», «студенческий клуб», «Институт физики», «факультет КНиИТ», «совет студентов и аспирантов», «профком», «журнал «СГУщенька»», «лига КВН» и др.

Установлены правила определения ребер. Три типа взаимодействий с весами: текстовая ссылка (вес 1) – извлекается из текста поста регулярными выражениями, учитывающими полные и короткие URL, упоминания через @, скрытые-ссылки, репост (вес 2) – идентифицируется по наличию поля `reply_history`, совместный пост (вес 3) – определяется, когда `reply_id` текущего поста совпадает с `owner_id` исходного сообщества в `reply_history`. Исключаются петли (ссылки на само себя) и связи через комментарии (из-за многократного увеличения объема парсинга). Практическая проверка на выборке из 50 постов показала, что текстовые ссылки встречаются в 30% постов, репосты – в 10%, совместные посты – в 2%.

После агрегации связей строится взвешенный неориентированный граф. Для каждого сообщества добавляется узел с атрибутами «имя», «Число подписчиков» и «ссылка». Ребра добавляются с весом, равным сумме весов всех взаимодействий между парой. Рассчитываются базовые метрики: плотность, разреженность, средняя степень. Для более глубокого анализа используются расширенные метрики: центральность по степени, центральность по

посредничеству, средний коэффициент кластеризации, ассортативность, выявление изолированных узлов.

Выполнена визуализация графов с использованием силового алгоритма размещения. Размер каждого узла пропорционален квадратному корню из числа подписчиков, что позволяет визуально сравнить аудиторию сообществ. Цвет ребра кодирует вес связи: зеленый (малый вес) – текстовые ссылки, желтый (средний вес) – репосты, красный (большой вес) – совместные посты. Построены тепловые карты матриц смежности. Все результаты сохраняются в текстовые отчеты, сводную таблицу, изображения графов и тепловых карт.

В третьем разделе «Сравнительный анализ графов сообществ СГУ, СГТУ и КГТУ» представлены результаты применения разработанного парсера. Проведен сравнительный анализ вершин графа. Наибольшее число сообществ (12) выявлено в СГТУ, наименьшее (9) – в КГТУ, СГУ насчитывает 11 сообществ. Типологический анализ показал, что СГУ и СГТУ имеют сбалансированную структуру с присутствием всех пяти типов, тогда как в КГТУ минимально представлены неформальные сообщества. Медианное число подписчиков составило около 8 тыс. для СГУ, 5 тыс. для СГТУ, 4 тыс. для КГТУ. Наибольшую аудиторию имеет официальная группа СГУ (27 163 подписчика). Активность сообществ в целом высокая: большинство публикуют посты 2–5 раз в неделю, однако выявлены сообщества с низкой активностью (менее 5 постов за три месяца) – в СГУ это паблик «Ночь открытых дверей», в СГТУ и КГТУ некоторые узкоспециализированные факультетские группы.

Анализ связей и метрик показал следующие результаты. Абсолютное число ребер: СГУ – 42, СГТУ – 39, КГТУ – 32. Плотность графа: КГТУ – 0,889 (88,9% от максимально возможных связей), СГУ – 0,764, СГТУ – 0,591. Разреженность (основной критерий): КГТУ – 0,111, СГУ – 0,236, СГТУ – 0,409. Таким образом, наилучший показатель связанности у КГТУ (отличная связанность), за ним следует СГУ (также отличная, но чуть хуже), а СГТУ демонстрирует среднюю связанность с тенденцией к низкой. Средняя степень

узлов: СГУ – 7,64, КГТУ – 7,11, СГТУ – 6,50. Все три вуза имеют высокие значения, что указывает на развитые сети.

Центральность по степени: в КГТУ три сообщества имеют максимальную нормированную степень 1,000 (официальная группа «КГТУ (бывш. КТИРПиХ)», «ВНЕУЧЕБКА КГТУ», «Точка Кипения КГТУ») – все они связаны со всеми остальными узлами. В СГУ три сообщества имеют степень 0,900 (официальная группа, факультет КНиИТ, Институт физики) – они связаны с 90% остальных узлов. В СГТУ явный лидер – официальная группа со степенью 0,909, затем институт УРБАС (0,818) и объединенный совет обучающихся (0,727). Центральность по посредничеству во всех трех вузах невысока (максимум 0,075 в СГТУ, 0,059 в СГУ, 0,029 в КГТУ), что объясняется близостью графов к полностью связным – при почти полной связности необходимость в «мостах» минимальна.

Коэффициент кластеризации: КГТУ – 0,915, СГУ – 0,885, СГТУ – 0,698. Все значения выше 0,3, что свидетельствует о наличии плотных локальных групп. Наиболее высокая кластеризация у КГТУ, что подтверждает его близость к полностью связному графу. Ассортативность отрицательна во всех трех университетах: КГТУ – -0,296, СГТУ – -0,285, СГУ – -0,188. Это характерно для централизованных сетей, где крупные хабы связаны преимущественно с мелкими узлами. Изолированные узлы: в СГУ и КГТУ изолированных сообществ не обнаружено, в СГТУ выявлено одно изолированное сообщество – «Подслушано в СГТУ», что снижает общую связность.

Визуализация графов и тепловых карт подтвердила количественные метрики. Граф КГТУ практически полностью связный, все узлы тесно переплетены, три хаба имеют максимальную степень. Граф СГУ также плотный, но с равномерным распределением связей без доминирующего центра. Граф СГТУ имеет выраженный центральный узел (официальная группа), вокруг которого группируются остальные сообщества, и отдельно расположенное изолированное сообщество «Подслушано в СГТУ». Тепловые карты показывают: для КГТУ – практически сплошной темный квадрат, для СГУ – темный квадрат с

небольшими светлыми зонами, для СГТУ – плотный блок, соответствующий связанным сообществам, и отдельная строка/столбец для изолированного узла.

ЗАКЛЮЧЕНИЕ

В результате выполнения бакалаврской работы достигнута поставленная цель – разработано программное средство для построения социального графа университетских сообществ на основе VK API и проведен сравнительный анализ СГУ, СГТУ и КГТУ по критерию минимальной разреженности. Все задачи решены в полном объеме.

В теоретической части обоснован выбор разреженности как основного индикатора цифровой медийности: чем ниже значение, тем выше связанность сообществ. Установлены качественные границы: менее 0,30 – отличная связанность, 0,30–0,45 – средняя, более 0,45 – низкая. Разработана типология университетских сообществ (официальные, студенческие советы, научно-образовательные, спортивно-творческие, неформальные). Обосновано использование официального VK API.

В практической части спроектирован парсер на Python с библиотеками requests, networkx, pandas, matplotlib, seaborn. Реализован класс VKParser с методами для авторизации, загрузки постов, извлечения ссылок (8 форматов, включая скрытые), обнаружения репостов и совместных постов, кэширования, расчета метрик и визуализации. Внедрена обработка ошибок API.

Сформированы выборки сообществ: СГУ – 11, СГТУ – 12, КГТУ – 9. Определены ребра с весами: текстовая ссылка (1), репост (2), совместный пост (3). Построены взвешенные неориентированные графы. Основные метрики: КГТУ: плотность 0,889, разреженность 0,111 (отличная), средняя степень 7,11, кластеризация 0,915, изолированных узлов нет. СГУ: плотность 0,764, разреженность 0,236 (отличная), средняя степень 7,64, кластеризация 0,885, изолированных нет. СГТУ: плотность 0,591, разреженность 0,409 (средняя), средняя степень 6,50, кластеризация 0,698, одно изолированное сообщество («Подслушано в СГТУ»). Ассортативность отрицательна во всех вузах (от -0,188 до -0,296), что подтверждает централизованную структуру. Сравнительный

анализ показал, что лучшую связанность имеет КГТУ, затем СГУ, худшую – СГТУ. Визуализация графов и тепловых карт подтвердила эти различия.

Разработанное средство может использоваться PR-отделами для мониторинга и повышения связанности сообществ. Для СГТУ рекомендовано стимулировать перекрестные ссылки и интегрировать изолированное сообщество. Для СГУ и КГТУ – поддерживать достигнутый уровень.