

ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ АКТИВНОСТИ В СОЦИАЛЬНОЙ СЕТИ И АКАДЕМИЧЕСКОЙ УСПЕВАЕМОСТИ СТУДЕНТОВ ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

В. В. Сыч

МГТУ им. Н.Э. Баумана;
Институт проблем управления им. В.А. Трапезникова РАН, г. Москва

✉ sycheslavovich@gmail.com

Аннотация. Проанализирована академическая успеваемость студентов МГТУ им. Н. Э. Баумана и её связь с их активностью в социальной сети ВКонтакте. С помощью методов машинного обучения выделены различные траектории успеваемости, отражающие динамику образовательных достижений. На основе анализа подписок студентов идентифицированы маркерные сообщества, характеризующие преобладание студентов, относящихся к определённым категориям успеваемости. С применением теории графов выполнена кластеризация сообществ, выявившая структурные группы интересов студентов. Для каждой траектории построены стохастические векторы долей интересов к кластерным сообществам, позволившие определить кластеры сообществ, статистически значимо связанные с конкретными траекториями успеваемости. Полученные результаты подтвердили наличие связи между цифровым поведением студентов и их академическими результатами, что способствует разработке моделей прогнозирования успеваемости с учётом интересов студентов в социальных сетях.

Ключевые слова: академическая успеваемость, ВКонтакте, машинное обучение, теория графов, прогнозирование успеваемости.

ВВЕДЕНИЕ

В условиях цифровизации образовательного процесса данные, генерируемые студентами в социальных сетях, становятся важным источником информации для анализа их академической и социальной активности. Ранее в ряде исследований была показана возможность дифференцировать студентов по уровню академической успеваемости на основе их активности в социальной сети ВКонтакте [1]. Также изучалась зависимость карьерных достижений от участия в профессиональных социальных сетях [2], а академические успехи рассматривались как распространяющиеся через дружеские и учебные связи [3]. Выявлена инвертированная U-образная зависимость между интенсивностью использования социальных сетей и академической результативностью [4], а также установлено, что психическое здоровье может выступать посредником в этой взаимосвязи [5].

Подавляющее большинство студентов имеет страницы в социальной сети ВКонтакте, на которых накоплен значительный объём информации, отражающий их интересы и социальные связи. Несмотря на то, что соцсети редко используют для поиска серьёзной информации и чаще открывают «для отдыха», долгосрочная структура подписок отражает устойчивые предпочтения и личные ориентиры пользователей, что позволяет связывать их цифровой след с образовательными показателями и использовать его в качестве надёжного предиктора академической успеваемости. Данное исследование направлено на выявление взаимосвязи между академической успеваемостью студентов и их поведением в цифровом пространстве и выполнено с применением методов машинного обучения и теории графов.

Актуальность исследования обусловлена необходимостью разработки эффективных инструментов мониторинга академических достижений и

раннего выявления студентов, находящихся в группе риска. Традиционные методы оценки успеваемости, основанные на ограниченных выборках и опросах, не всегда учитывают многогранные аспекты поведения студентов вне учебных заведений. Анализ цифрового следа в социальных сетях позволяет получить более полное представление о студенте, учитывая его интересы, социальные связи и даже интересы друзей, что может дополнить существующие методы оценки.

Цель исследования состоит в анализе взаимосвязи между траекториями академической успеваемости и цифровыми интересами студентов. Уникальность работы заключается в использовании обширной базы данных: из 76 000 студентов, обучавшихся в МГТУ им. Н.Э. Баумана в период с 2014 по 2023 гг., 48 000 удалось сопоставить страницы в социальной сети ВКонтакте. Данные включают как академические показатели (результаты сессий), так и информацию о подписках и дружеских связях, что позволяет формировать комплексный цифровой портрет каждого студента. Пример графа дружеского коллектива, построенного на основе данных сети ВКонтакте, приведён на рис. 1.

Методологическая база исследования включает алгоритмы кластеризации, стандартизации и анализа главных компонент для выделения типичных траекторий успеваемости, а также методы теории графов для анализа структуры сообществ. Иссле-

дование выполнено на основе данных социальной сети ВКонтакте и электронных записей университета, анонимизированных после сопоставления.

1. ОПИСАНИЕ ДАННЫХ

Анализ проводится на основе двух основных источников информации. Первый источник – академическая база данных, полученная из системы «Электронный университет», охватывающая результаты сессий студентов за период 2014–2023 гг. Для каждого студента зафиксированы оценки по каждой сессии, что позволяет вычислить средний балл по каждому из четырёх курсов обучения. Второй источник – данные из социальной сети ВКонтакте, собранные с помощью VK API. Для каждого пользователя извлекаются списки подписок на сообщества и информация о дружеских связях, что позволяет анализировать интересы студентов и изучать организацию их сетевого окружения. Были выгружены профили, подписанные на сообщества, связанные с университетом, и все их друзья. Подобный подход выгрузки данных использовался в работе [6].

Сопоставление академических данных с профилями ВКонтакте проводилось итеративно. На первом этапе формировалось «ядро» студентов с высокой степенью совпадения данных, где сопоставление осуществлялось на основе строгого со-

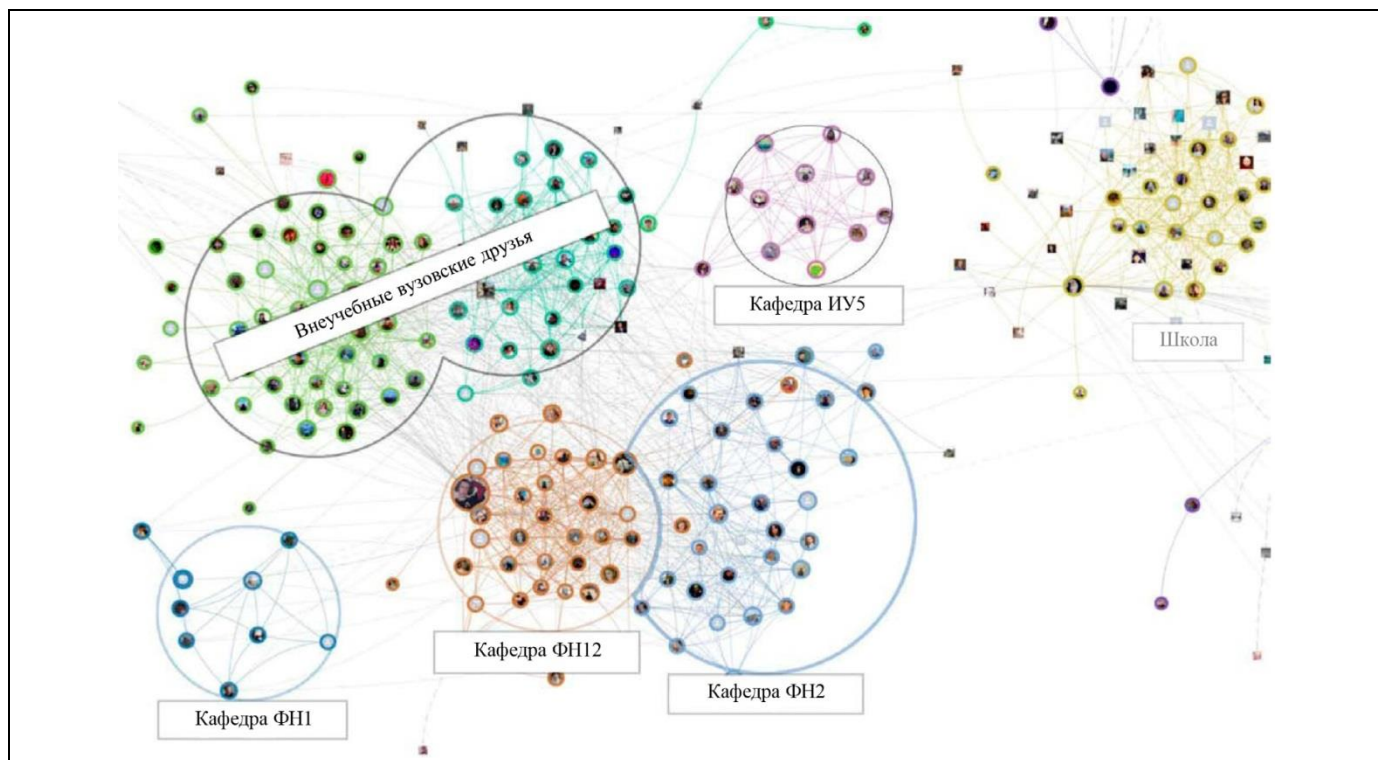


Рис. 1. Граф дружеских коллективов студента по связям из сети ВКонтакте на примере страницы автора (подписаны крупные кластеры)

ответствия Ф. И. О. и даты рождения, дополнительно расчётом расстояния Левенштейна [7] для несовпадающих форм имён. Затем с использованием информации о дружеских связях на втором этапе идентифицировались дополнительные профили: если у кандидата совпадали лишь базовые параметры (например, имя и год рождения), но при этом он был связан с уже идентифицированными студентами из конкретной учебной группы, то его с большей вероятностью относили к этой же группе. Этот процесс основывался на принципе «скажи мне, кто твой друг, и я скажу, кто ты», так как дружеские связи служили дополнительным маркером достоверности. Такой подход позволил включить в исследование тех студентов, у кого исходных совпадений было недостаточно для однозначной идентификации. Итеративное повторение процесса (на первой итерации было получено ядро в 5 000 студентов, на второй около 22 000 студентов, на третьей – 46 000) базируется на принципе: чем больше профилей друзей студента уже подтверждены, тем выше вероятность корректного сопоставления его профиля с академическими данными. Например, в случае, когда множество пользователей имеют одинаковое имя и совпадающую дату рождения, дополнительный признак в виде дружеских связей со студентами, профили которых подтверждены, существенно повышает достоверность идентификации. Такой подход позволяет значительно расширить выборку и обеспечивает её высокую репрезентативность. После проведения процедуры сопоставления данные анонимизировались и только в таком виде использовались для анализа. Подобные методики расширенного сопоставления цифрового следа с образовательными данными ранее демонстрировали высокую диагностическую ценность при исследовании успеваемости старшеклассников и студентов [8, 9].

2. АНАЛИЗ АКАДЕМИЧЕСКОЙ УСПЕВАЕМОСТИ

Для анализа траекторий успеваемости были отобраны студенты, успешно завершившие четыре курса бакалавриата или специалитета без академических задолженностей, в выборку попало 15 495 мужчин и 6 285 женщин. Кластерный подход к сегментации академических результатов ранее применялся при оценке связи между временем, проводимым в социальных сетях, и оценками студентов-биологов [10] и при анализе использования соцмедиа на мобильных устройствах в международной выборке [11], что подтверждает его пригодность для выявления латентных учебных траек-

торий. Для каждого студента i вычислялся вектор средних баллов по курсам

$$x_i = (x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}) \in \mathbb{R}^4,$$

где $x_{i,c}$ – средний балл на курсе c ($c = 1, 2, 3, 4$).

Далее матрица $X \in \mathbb{R}^{N \times 4}$ подвергалась стандартизации:

$$z_{i,c} = \frac{x_{i,c} - \mu_c}{\sigma_c},$$

где μ_c и σ_c – соответственно выборочное среднее и стандартное отклонение по курсу c .

Для разбиения студентов по траекториям академической успеваемости был применён алгоритм k -средних (k -means). Оптимальное число кластеров k выбиралось по коэффициенту силуэта (*Silhouette Score*) – показателю внутрикластерной компактности и межкластерной разделимости, предложенному П. Роуссеу [12]. Максимум метрики (около 0,43) достигался при $k = 2$, однако такое разбиение не отражает разнообразия учебных траекторий. При увеличении k от 6 до 16 коэффициент силуэта стабилизировался в диапазоне 0,21–0,27, что свидетельствует об умеренном, но устойчивом качестве кластеризации.

Для дополнительного подтверждения оптимального числа кластеров был применён метод статистики разрыва (*Gap Statistic*), предложенный Р. Тибширани и соавторами [13]. Этот метод позволяет сравнить качество кластеризации исходных данных с кластеризацией случайно сгенерированных данных, распределённых равномерно в том же диапазоне. Чем выше значение статистики разрыва $Gap(k)$, тем отчётливее структура данных (выделяемые кластеры значимее случайных). Значения статистики разрыва при $k = 6$ и $k = 8$ оказались близкими: $0,994 \pm 0,006$ и $0,911 \pm 0,003$ соответственно. Незначительное уменьшение значения статистики разрыва при переходе от шести к восьми кластерам находится в пределах стандартной ошибки, формируя так называемое «плато». Это свидетельствует о том, что детализация кластеров при увеличении их числа с шести до восьми существенно не ухудшается.

Таким образом, несмотря на умеренные значения метрик качества, выбор $k = 8$ является обоснованным компромиссом между статистической устойчивостью результатов и необходимостью достаточной детализации траекторий успеваемости для последующего анализа её взаимосвязи с цифровыми интересами студентов.

После разбиения каждому студенту была присвоена метка кластера, пронумерованного по воз-

растанию общей средней успеваемости – простого арифметического среднего всех оценок за восемь семестров (табл. 1). Дополнительно отмечается тенденция снижения доли мужчин при переходе от кластеров с низкой успеваемостью к кластерам с высокой успеваемостью.

Траектории представлены на рис. 2 и 3.

Итоговое разбиение характеризуется следующими траекториями:

- **Траектория 0.** Стабильно низкая успеваемость.
- **Траектория 1.** Низкие оценки на начальных курсах с последующим улучшением.
- **Траектория 2.** Первоначально хорошие результаты с последующим снижением.
- **Траектория 3.** Низкая успеваемость с заметным ростом.

– **Траектория 4.** Стабильные хорошие оценки с небольшим проседанием.

– **Траектория 5.** Умеренные оценки, переходящие к устойчивым хорошим результатам.

Таблица 1

Статистика по траекторным кластерам

Кластер	Количество студентов	Средний балл	Доля мужчин, %
0	1910	3,48	90,0
1	2510	3,81	85,7
2	2120	3,92	82,8
3	2870	4,19	76,7
4	1950	4,21	71,6
5	2200	4,43	69,7
6	3500	4,57	52,9
7	4720	4,86	56,5

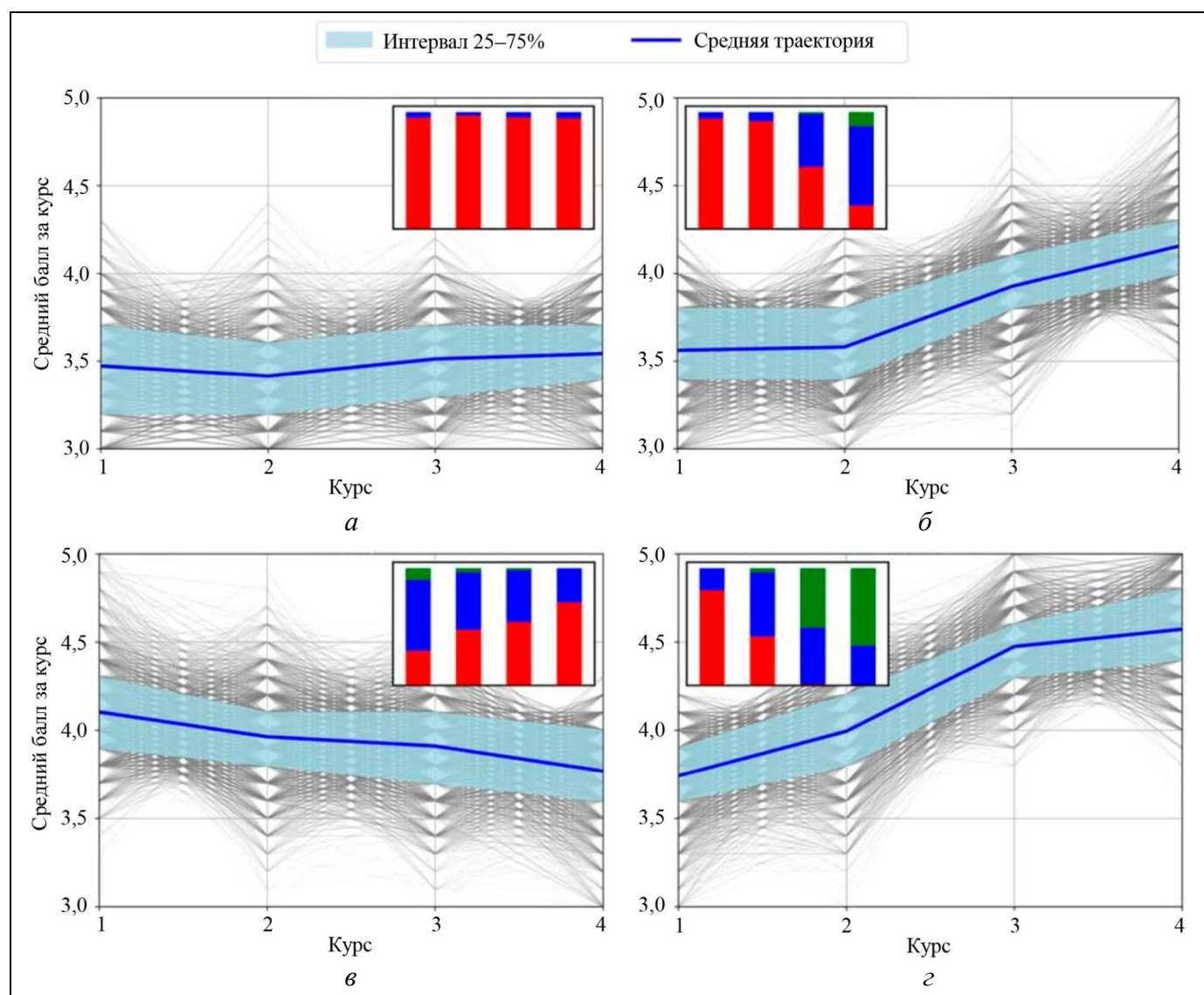


Рис. 2. Кластеры траекторий: а – кластер 0, б – кластер 1, в – кластер 2, г – кластер 3

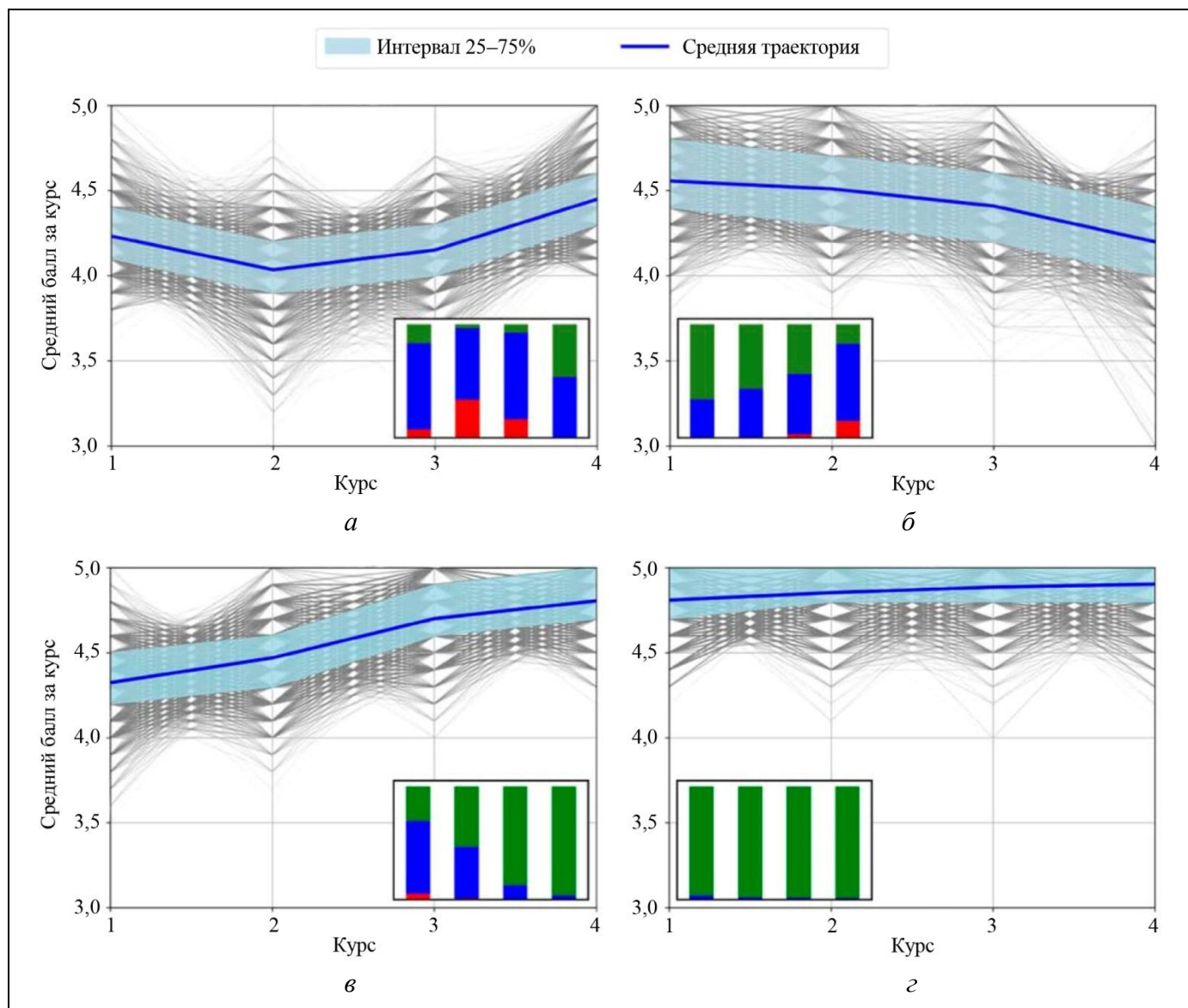


Рис. 3. Кластеры траекторий: а – кластер 4, б – кластер 5, в – кластер 6, г – кластер 7

Для наглядности в каждом кластере представлены диаграммы с пропорциями студентов различных категорий успеваемости: красные столбцы обозначают долю троечников ($\bar{g} < 3,9$), синие – хорошистов ($3,9 \leq \bar{g} < 4,5$), а зелёные – отличников ($\bar{g} \geq 4,5$), где $\bar{g}$ – средний балл студента, вычисленный как арифметическое среднее всех его оценок по сессиям. Эти диаграммы позволяют оценить распределение уровня успеваемости студентов по курсам в каждой траектории.

Для дополнительной визуализации результатов кластеризации применён метод главных компонент (англ. *Principal Component Analysis*, PCA). На рис. 4 точки, соответствующие студентам, отобра-

жены на плоскости, где горизонтальная ось отражает общий уровень успеваемости, а вертикальная – динамику изменений. Центр масс множества отмечен крестиком.

Повторные запуски алгоритма *k-means* подтвердили устойчивость полученного разбиения, а результаты визуализации демонстрируют разделение кластеров, несмотря на частичное наложение их границ в проекции на двумерное пространство. Данная структура обеспечивает надёжную основу для дальнейшего анализа взаимосвязи между цифровыми интересами и академической успеваемостью студентов, так как средние значения траекторий хорошо описывают основные различия между полученными траекторными кластерами.

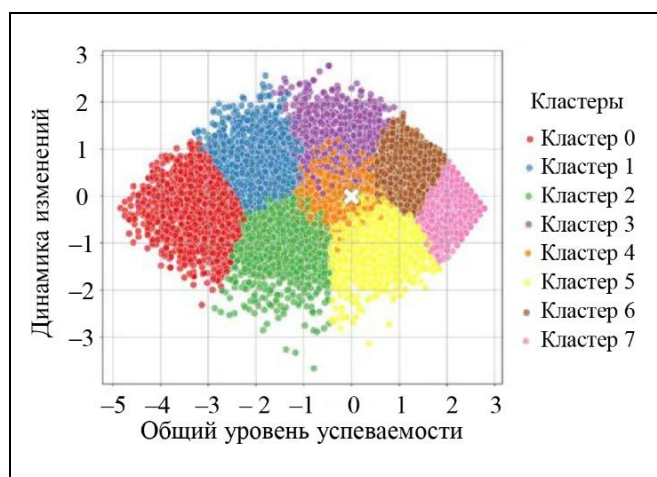


Рис. 4. Кластеры студентов по успеваемости, полученные с помощью метода главных компонент

3. ГРАФОВАЯ МОДЕЛЬ СООБЩЕСТВ В КОНТАКТЕ И ЕЁ РОЛЬ В АНАЛИЗЕ АКАДЕМИЧЕСКИХ ТРАЕКТОРИЙ СТУДЕНТОВ

3.1. Исходные данные

Исходными данными послужили все паблики (сокр. от «публичная страница», также говорят «сообщество» или «группа») ВКонтакте, на которые подписано не менее 50 студентов; таким образом сформирован корпус, содержащий порядка 4500 сообществ. Подход к использованию подписок на сообщества как признаков академической успешности зарекомендовал себя в недавней работе по выявлению лучших студентов средствами продвинутых NLP-алгоритмов [14]. Прежде чем переходить к теоретико-графовому анализу, каждое сообщество проверялось на связь с академическими траекториями посредством однофакторного дисперсионного анализа (англ. *Analysis of Variance*, ANOVA): для 760 пабликов различия между траекториями оказались значимыми при $p < 0,05$, для 350 – при $p < 0,01$.

Пусть $G = (V, E, W)$ – неориентированный граф, где V – множество пабликов, а ребро $(c_1, c_2) \in E$ существует, если найдётся хотя бы один общий подписчик-студент. Сила связи между пабликами оценивалась при помощи взвешенного коэффициента Жаккара (*weighted Jaccard coefficient*), впервые предложенного для оценки сходства флористических составов [15]:

$$w(c_1, c_2) = \ln \left(1 + \frac{|S_{c_1} \cap S_{c_2}|}{|S_{c_1} \cup S_{c_2}|} \right),$$

где S_{c_i} – студенты, подписанные на паблик c_i . Логарифмирование позволяет сгладить крайние значения коэффициента при очень больших или, напротив, почти нулевых пересечениях аудиторий.

Для обнаружения сообществ применялся алгоритм Лувена [16], максимизирующий модулярность графа. При однократном запуске на всём графе при базовом параметре $\text{resolution} = 1$ получаются несколько anomalously крупных и множество маленьких кластеров, поэтому использовалась поэтапная схема. Сначала весь граф кластеризовался при низком значении resolution ; далее слишком большие группы рекурсивно переразбивались тем же методом с $\text{resolution} = 0,7$. Дополнительно были убраны 30 % самых слабых связей, поскольку общее их количество было очень велико. Итеративное объединение результатов обеспечило финальное разбиение на 21 кластер сопоставимого размера.

3.2. Визуализация и интерпретация кластеров

Итоговая структура графа сообществ представлена на рис. 5: слева – полный граф с 21 кластером, справа – кластер 10 в приближении, включающий преимущественно образовательные паблики. Цвета узлов соответствуют преобладающему уровню академической успеваемости подписчиков: красный – троечники, синий – хорошисты, зелёный – отличники. Такое кодирование позволяет сразу оценить однородность составов кластеров по уровню успеваемости.

Отметим, что траекторные кластеры ($n = 8$) отражают динамику изменений успеваемости студентов во времени, тогда как категории «троечники» / «хорошисты» / «отличники» (3/4/5) рассчитываются по индивидуальному среднему баллу за всё время учебы и вводятся отдельно для наглядной интерпретации уровня успеваемости в каждом паблике. Они служат дополнительной «проекцией» данных.

Описание кластеров и распределение студентов по категориям успеваемости приведены в табл. 2 (описания основаны на субъективных обобщениях автора).



Рис. 5. Визуализация графа сообществ ВКонтакте

Таблица 2

Характеристика кластеров пабликов ВКонтакте по результатам кластеризации.

Кластер	Описание пабликов	3/4/5
0	Крупные сообщества с универсальным развлекательным контентом, охватывающие широкую аудиторию	41/32/27
1	Сообщества, специализирующиеся на юмористическом контенте, включая мемы, связанные с киберспортом и видеонграми (CS2, PUBG)	42/33/25
2	Высокопопулярные паблики с мемами, рассчитанные на массовую аудиторию	42/31/27
3	Паблики, посвящённые спортивным трансляциям и киберспорту (футбол, UFC, Dota2)	46/32/22
4	Развлекательные паблики с простыми, широко распространёнными мемами	42/33/25
5	Сообщества с авторскими, уникальными или нишевыми мемами	46/34/20
6	Развлекательные паблики с повторяющимися шаблонными мемами	46/33/21
7	Сообщества, распространяющие постироничные мемы, ориентированные на специфическую аудиторию	40/35/25
8	Юмористические паблики, ориентированные на распространение видеоконтента.	46/34/20
9	Сообщества, представляющие постиронию и абсурдистский юмор	42/34/23
10	Паблики, связанные с образовательными и научными темами: программирование, стажировки, исследования	27/34/39
11	Студенческие паблики, в том числе факультетские сообщества и группы типа «Подслушано»	32/35/33
12	Сообщества, ориентированные на студенческую жизнь, включая мероприятия и активности	27/35/37
13	Паблики с доброжелательными мемами, часто ориентированные на тему домашних животных (особенно котов)	32/34/34
14	Сообщества, посвящённые эстетике, фитнесу, питанию и уходу за собой, преимущественно ориентированные на женскую аудиторию	26/33/41
15	Паблики о моде, индустрии красоты, интернет-магазинах и женских мемах	25/34/41
16	Сообщества, связанные с саморазвитием, литературой, изучением иностранных языков, эстетическими концепциями	30/33/37
17	Зарубежные сообщества, содержащие женские мемы и эстетический контент	29/34/38
18	Сообщества, посвящённые неформальной моде, торговле одеждой и обувью	46/35/19
19	Музыкальные сообщества, ориентированные на рэп-исполнителей и индустрию музыки	45/35/20
20	Игровые паблики (Dota2, киберспорт), а также группы, связанные с торговлей поддержанными автомобилями	46/34/20



4. СВЯЗЬ ЦИФРОВЫХ ИНТЕРЕСОВ СТУДЕНТОВ С ИХ АКАДЕМИЧЕСКИМИ ТРАЕКТОРИЯМИ

4.1. Формализация интересов и вычисление логарифмического отношения шансов

Для каждого студента i построен стохастический вектор интересов

$$p_i = (p_{i,0}, p_{i,1}, \dots, p_{i,20}),$$

обеспечивающий равнозначный учёт всех кластеров – $\left(\sum_k p_{i,k} = 1\right)$, где $p_{i,k}$ – это отношение числа

подписок студента i в кластере k к общему числу его подписок. Этот вектор отражает степень концентрации интересов студента на конкретных кластерах пабликов.

Чтобы выявить различия в цифровых интересах студентов разных академических траекторий, далее использовалась выборка студентов, для которых были известны данные о подписках: после фильтрации осталось 4 500 мужчин и 1 900 женщин. Для оценки применялась метрика логарифмического отношения шансов (*log-odds ratio*) [17], измеряющая отклонение доли подписок на кластер k в академической траектории t от глобального среднего значения:

$$\log_odds_{k,t} = \ln\left(\frac{p_{k,t}}{1 - p_{k,t}}\right) - \ln\left(\frac{p_{k,g}}{1 - p_{k,g}}\right),$$

где

$p_{k,t}$ – усечённая (*trimmed mean*) средняя доля подписок на кластер k среди студентов траектории t . Сначала для каждого студента i из траектории t рассчитывается отношение числа его подписок в кластере k к общему числу его подписок, а затем по полученным отношениям всех студентов траектории t берётся усечённая средняя;

$p_{k,g}$ – глобальная усечённая средняя доля подписок на кластер k по всей выборке студентов. Сначала для каждого студента рассчитывается отношение числа его подписок в кластере k к общему числу его подписок, а затем по всем этим отношениям берётся усечённая средняя.

Чтобы уменьшить искажение из-за нетипично низких или высоких индивидуальных значений, расчёт средних долей для каждого кластера проводился с использованием усечённого среднего: по 10 % самых низких и самых высоких значений долей исключались из анализа. Это позволило полу-

чить более надёжные и информативные оценки, устойчивые к наличию выбросов.

Положительные значения $\log_odds$ говорят о повышенном интересе к кластеру среди студентов конкретной траектории, отрицательные – о пониженной доле подписок по сравнению с глобальной средней $p_{k,g}$, т. е. об «избегании» этого кластера. Для наглядной интерпретации этой метрики можно привести следующие примеры:

- $\log_odds = 0,5$ означает, что шансы встретить подписки на данный кластер среди студентов указанной траектории выше примерно на 65 % по сравнению с общим уровнем, так как $e^{0,5} \approx 1,65$;
- $\log_odds = -1$ означает, что шансы встретить подписки на данный кластер среди студентов указанной траектории примерно в $e^{-1} \approx 0,37$ раза, т. е. в $1/0,37 \approx 2,7$ раза ниже, чем в среднем.

В дальнейшем анализе логарифмические коэффициенты ($\log_odds$) вычислялись отдельно для мужчин и женщин. Причины раздельного расчёта:

• **Неровный гендерный состав траекторий.** В траекторном кластере 0 мужчин около 90 %, в траектории 7 – лишь около 57 %. При агрегации это искажает глобальные средние $p_{k,g}$ и, соответственно, логарифмические коэффициенты.

• **Половая специфика подписок.** Часть пабликов (кластеры 14–17) по определению ориентирована на женскую аудиторию; их «избегание» мужчинами не информативно для академических выводов, но увеличивает дисперсию в общей выборке.

На рис. 6 представлена тепловая карта $\log_odds$, которая показывает выраженные различия интересов студентов-мужчин в зависимости от академической траектории. Под значениями $\log_odds$ указана процентная доля подписок на этот кластер среди студентов данной траектории.

Наблюдения:

• Траектории 0–2 (низкая и падающая успеваемость). Сильное смещение в сторону развлекательных кластеров 0–6 ($\log_odds \approx 0,18$ – $0,51$) и умеренное в сторону уличной моды и рэп-комьюнити. Также «избегание» образовательного кластера 10 и вузовских сообществ 11–12 (до $-0,41$).

• Траектория 7 (стабильные отличники). Картина зеркальная: отрицательные коэффициенты на развлекательный контент и заметный скачок интереса к кластерам 10–12 (образование, вузовские сообщества).

Далее на рис. 7 приведена аналогичная карта для женщин.

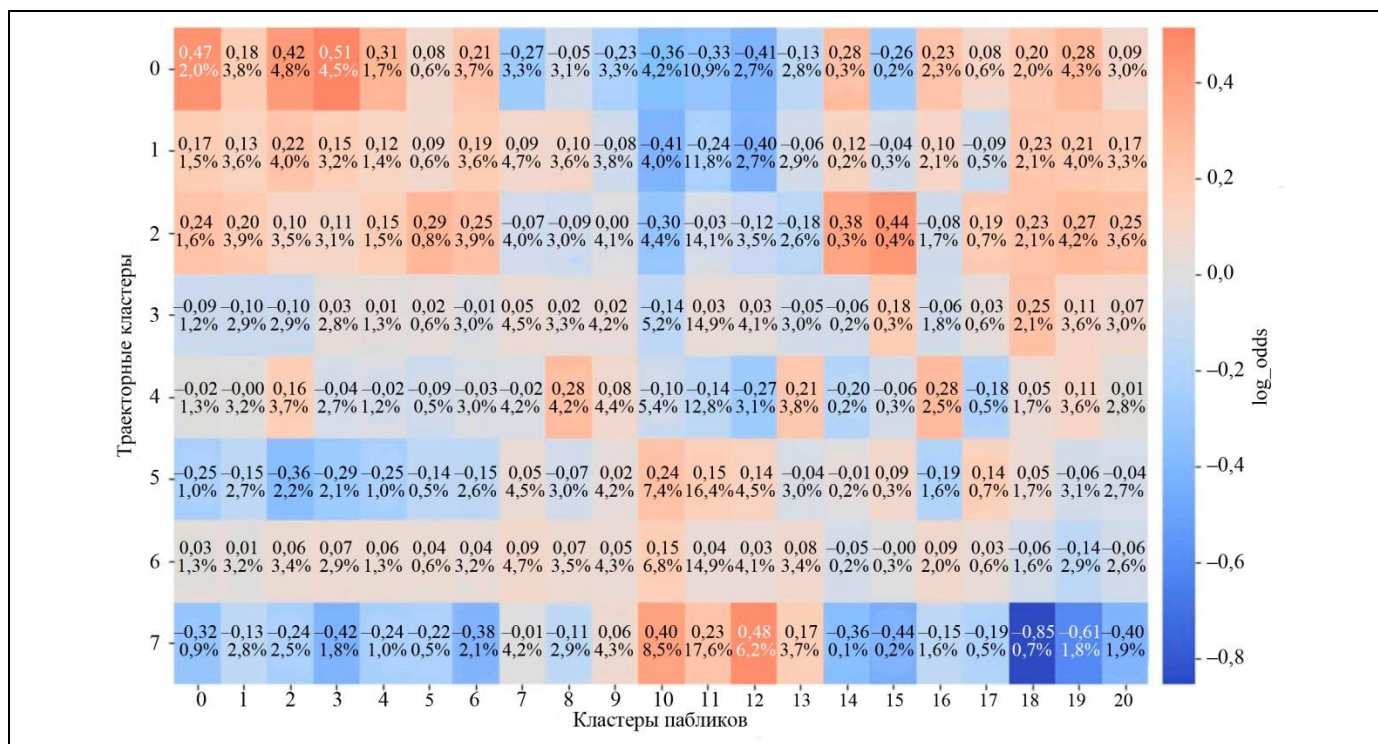


Рис. 6. Тепловая карта по log_odds для мужской выборки студентов

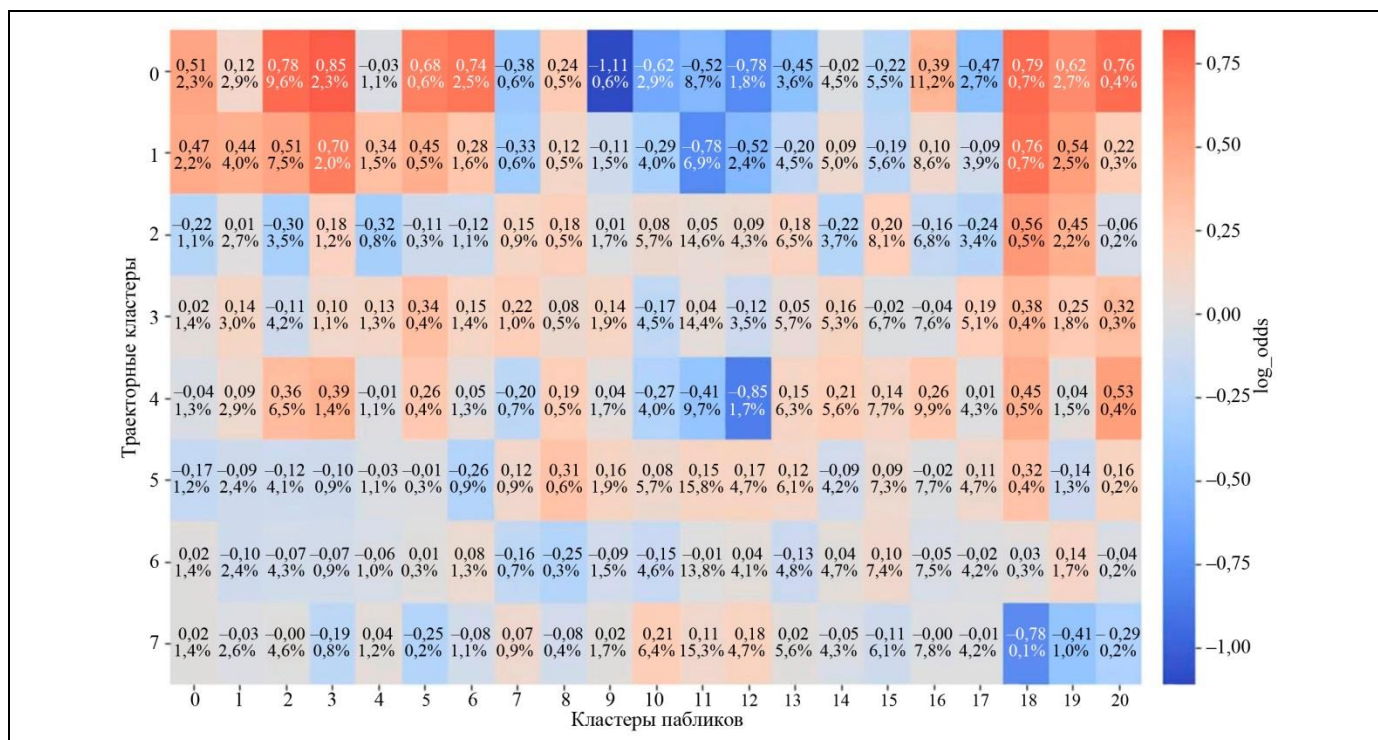


Рис. 7. Тепловая карта по log_odds для женской выборки студентов

По студенткам из слабых траекторий можно сделать аналогичные выводы: высокий интерес к развлекательным кластерам и избегание образовательных. Однако для лучших траекторий (6 и 7) наблюдается более равномерная картина распре-

деления интересов. Несмотря на общую очевидность выводов (развлечения преобладают у слабых студентов, образование – у сильных), внутри каждой траектории наблюдается индивидуальный характер предпочтений. Эти различия по цифровым



интересам могут быть использованы при построении прогностической модели академической успеваемости.

4.2. Проверка гипотезы о независимости академических траекторий и цифровых интересов

Для формальной проверки гипотезы о статистической зависимости между академической успеваемостью студентов и тематикой их цифровых интересов был проведён анализ таблиц сопряжённости с использованием критерия χ^2 Пирсона отдельно для мужской и женской выборок

(табл. 3). Глобальный χ^2 -тест подтвердил наличие статистически значимой связи между принадлежностью к конкретной траектории успеваемости и профилем подписок студентов как среди мужчин, так и среди женщин.

Более детально различия проиллюстрированы на рис. 8. На рис. 8, а, б представлены частные χ^2 -статистики по каждому из 21 кластера интересов: красный цвет обозначает значимые с точки зрения χ^2 различия ($\alpha = 0,05$), серый – незначимые.

Видно, что связи с академическими траекториями стали значимыми в 10 кластерах для мужчин и 15 – для женщин. На рис. 8, в, г изображены оцен-

Таблица 3

Результаты глобального χ^2 -теста независимости академических траекторий и цифровых интересов

Пол	Выборка, чел.	χ^2 -статистика	Число степеней свободы	p-значение	Критическое значение χ^2 ($\alpha = 0,05$)
Мужчины	~ 4500	557,6	140	1e-51	168,613
Женщины	~ 1900	755,8		1e-85	

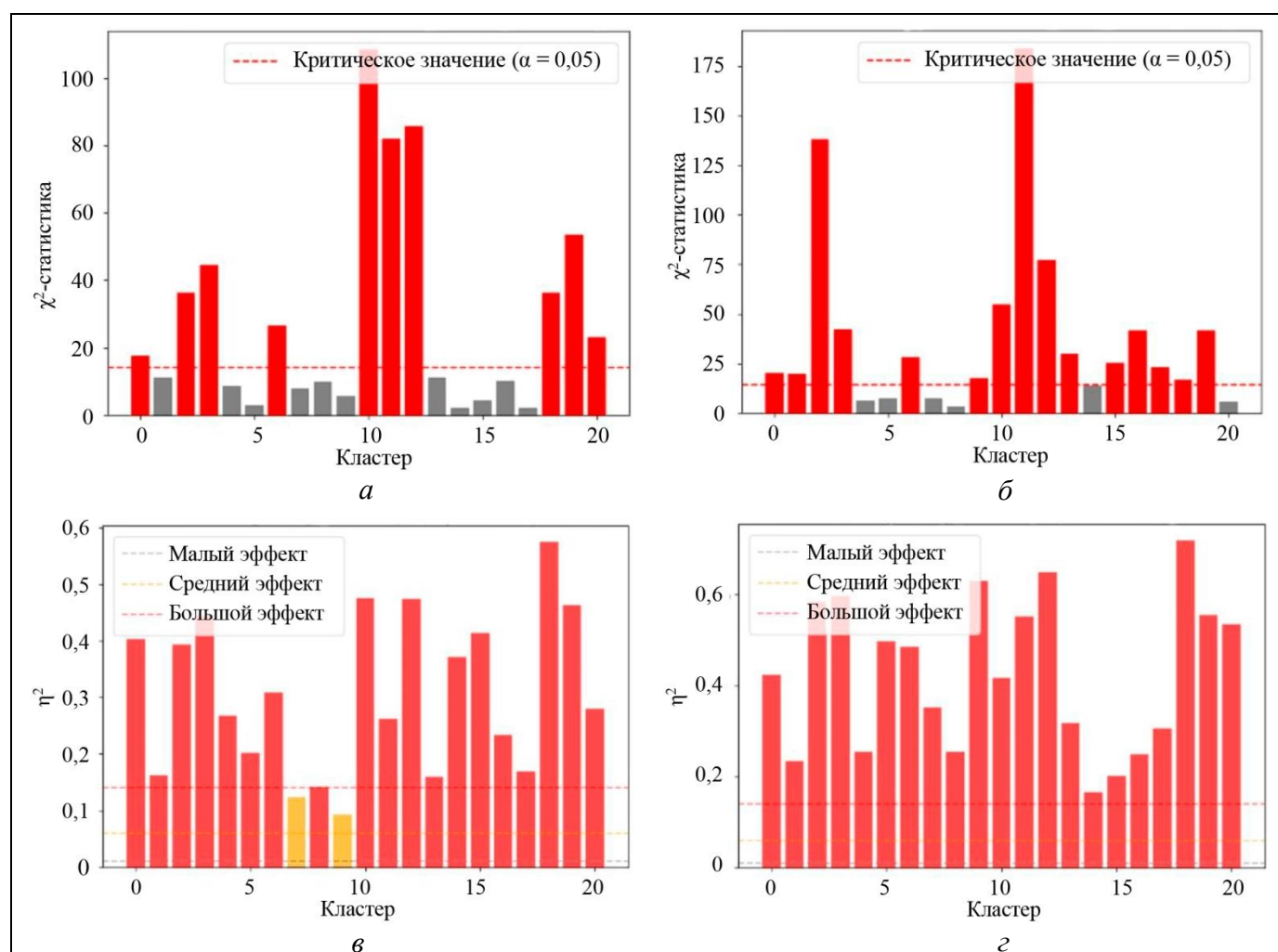


Рис. 8. Частные показатели связи между кластерами цифровых интересов и академическими траекториями: а, б – χ^2 -статистики (красный – $p < 0,05$; пунктир – критическое значение χ^2); в, г – оценки размеров эффекта η^2 (пунктиры – пороги); а, в – мужчины, б, г – женщины

ки размера эффекта η^2 , где пунктирные линии указывают пороги 0,01 (малый эффект), 0,06 (средний) и 0,14 (большой), принятые в работе [18]. Большинство значимых кластеров демонстрирует $\eta^2 > 0,3$, а особенно у женщин около половины кластеров превышают уровень 0,5, что говорит о высокой практической значимости выявленных различий. Наибольшая сила связи с траекторией успеваемости прослеживается в кластерах, связанных с образовательными и студенческими сообществами (номера 10–12), а также с развлекательными и модно-музыкальными пабликами (номера 18–19), что логически коррелирует с ранее наблюдавшимися паттернами подписок. Полученные данные однозначно отвергают нулевую гипотезу о независимости академической траектории и цифровых интересов, подтверждая, что цифровые предпочтения студентов репрезентативны и содержательно связаны с показателями их успеваемости.

ЗАКЛЮЧЕНИЕ

В ходе проведённого исследования зафиксирована статистическая взаимосвязь между тематикой подписок студентов МГТУ им. Н. Э. Баумана в социальной сети ВКонтакте и их академической успеваемостью. Полученные результаты подтверждают, что цифровой след, формируемый через интересы и сетевую активность, может выступать значимым предиктором образовательных траекторий и, возможно, других офлайн-показателей. Формальная проверка гипотезы с использованием критерия хи-квадрат Пирсона показала высокую статистическую значимость ($p < 0,001$) выявленных зависимостей для обеих гендерных выборок, а анализ размера эффекта подтвердил практическую значимость обнаруженных различий.

Исследования других авторов также свидетельствуют о зависимости между активностью в социальных сетях и академической успеваемостью, однако большинство из них ограничивается небольшими выборками и оценкой времени активности в соцсетях.

В перспективе планируется расширение анализа путём включения сетевой структуры дружеских связей студентов, что позволит построить интегрированную модель прогнозирования успеваемости на основе как индивидуальных цифровых интересов, так и их сетевое-структурного положения в студенческой сети (метрики центральности, принадлежность к сообществам, брокерские роли). Такой подход откроет возможности для более точной диагностики академических рисков и форми-

рования персонализированных рекомендаций по академической поддержке – от подбора курсов и консультаций до включения в релевантные учебные сообщества.

ЛИТЕРАТУРА

1. *Смирнов И.Б.* Дифференциация учащихся по академической успеваемости в социальной сети: дис.... канд. пед. наук. – М.: НИУ ВШЭ, 2018. – 210 с. [*Smirnov, I.B.* Differentiation of Students by Academic Performance in a Social Network: Cand. Ped. Sci. Diss. – Moscow: NRU HSE, 2018. – 210 p. (In Russian)]
2. *Nikitkov, A., Sainy, B.* The Role of Social Media in Influencing Career Success // *International Journal of Accounting & Information Management*. – 2014. – Vol. 22, no. 4. – P. 273–294. – DOI 10.1108/IJAIM-02-2014-0009
3. *Dokuka, S., Valeeva, D., Yudkevich, M.* How Academic Achievement Spreads: The Role of Distinct Social Networks in Academic Performance Diffusion // *PLOS ONE*. – 2020. – Vol. 15, no. 7. – Art. no. e0236737. – DOI: 10.1371/journal.pone.0236737
4. *Tafesse, W.* Social Networking Sites Use and College Students' Academic Performance: Testing for an Inverted U-shaped Relationship Using Automated Mobile App Usage Data // *International Journal of Educational Technology in Higher Education*. – 2022. – Vol. 19. – Art. no. 16. – DOI: 10.1186/s41239-022-00322-0
5. *Al Mosharrafa, R., Akther, T., Siddique, F.K.* Impact of Social Media Usage on Academic Performance of University Students: Mediating Role of Mental Health under a Cross-sectional Study in Bangladesh // *Health Science Reports*. – 2024. – Vol. 7, no. 1. – Art. no. e1788. – DOI 10.1002/hsr2.1788
6. *Chkhartishvili, A., Gubanov, D.A., Melnichuk, V.S., Sych, V.V.* Exploratory Data Analysis and Natural Language Processing Model for Analysis and Identification of the Dynamics of COVID-19 Vaccine Opinions on Small Datasets // *Advances in Systems Science and Applications*. – 2023. – Vol. 23, no. 3. – P. 108–126. – DOI: 10.25728/assa.2023.23.3.1381
7. *Levenshtein, V.I.* Binary Codes Capable of Correcting Deletions, Insertions and Reversals // *Soviet Physics Doklady*. – 1966. – Vol. 10, no. 8. – P. 707–710.
8. *Каптур В.В., Губанов А.В., Фещенко А.В. и др.* Взаимосвязь образовательных достижений старшеклассников и их цифрового следа в социальной сети // Педагогика и просвещение. – 2020. – № 4. – С. 37–51. [*Kashpur, V.V., Gubanov, A.V., Feshchenko, A.V., et al.* The Relationship between Educational Achievements of High-School Students and Their Digital Footprint in a Social Network // *Pedagogy and Education*. – 2020. – No. 4. – P. 37–51. (In Russian)]
9. *Smirnov, I.B.* Estimating Educational Outcomes from Students' Short Texts on Social Media // *EPJ Data Science*. – 2020. – Vol. 9. – Art. no. 27. – DOI: 10.1140/epjds/s13688-020-00238-1
10. *Дудина В.А., Анисимовская А.И., Горченко А.Л., Данчышак Т.В.* Оценка влияния социальных сетей и других факторов на успеваемость студентов биологического факультета // Молодой учёный. – 2019. – № 48 (286). – С. 438–440 [*Dudina, V.A., Anisimovskaya, A.I., et al.* Assessment of the Influence of Social Networks and Other Factors on Academic



- Performance of Biology Students // Young Scientist. – 2019. – No. 48 (286). – P. 438–440. (In Russian)]
11. *Giunchiglia, F., Zeni, M., Gobbi, E., et al.* Mobile Social Media Usage and Academic Performance // Computers in Human Behavior. – 2018. – Vol. 82. – P. 177–185. – DOI: 10.1016/j.chb.2017.12.041
 12. *Rousseeuw, P.J.* Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis // Journal of Computational and Applied Mathematics. – 1987. – Vol. 20. – P. 53–65. – DOI: 10.1016/0377-0427(87)90125-7
 13. *Tibshirani, R., Walther, G., Hastie, T.* Estimating the Number of Clusters via the Gap Statistic // Journal of the Royal Statistical Society: Series B (Statistical Methodology). – 2001. – Vol. 63, no. 2. – P. 411–423. – DOI: 10.1111/1467-9868.00293
 14. *Gorshkov, S., Ignatov, D.I., Chernysheva, A.Yu., et al.* Identifying Top-Performing Students via VKontakte Social Media Communities Using Advanced NLP Techniques // IEEE Access. – 2025. – Vol. 13. – P. 962–979. – DOI: 10.1109/ACCESS.2024.3521857
 15. *Jaccard, P.* Étude Comparative de la distribution florale dans une portion des Alpes et du Jura // Bulletin de la Société Vaudoise des Sciences Naturelles. – 1901. – Vol. 37. – P. 547–579. – DOI: 10.5169/seals-266450
 16. *Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.* Fast Unfolding of Communities in Large Networks // Journal of Statistical Mechanics: Theory and Experiment. – 2008. – No. 10. – Art. no. P10008. – DOI: 10.1088/1742-5468/2008/10/P10008
 17. *Barnard, G.A.* Statistical Inference // Journal of the Royal Statistical Society: Series B (Methodological). – 1949. – Vol. 11, no. 2. – P. 115–139. – DOI: 10.1111/j.2517-6161.1949.tb00028.x
 18. *Cohen, J.* Statistical Power Analysis for the Behavioral Sciences: 2nd ed. – Hillsdale, NJ: Lawrence Erlbaum Associates, 1988. – 567 p.
- Статья представлена к публикации членом редколлегии академиком РАН Д. А. Новиковым.*
- Поступила в редакцию 04.05.2025,
после доработки 27.07.2025.
Принята к публикации 28.07.2025.*
- Сыч Владислав Витальевич** – студент, МГТУ им. Н. Э. Баумана; инженер, Институт проблем управления им. В. А. Трапезникова РАН, г. Москва,
✉ sycheslavovich@gmail.com
ORCID iD: <https://orcid.org/0009-0001-6787-8948>
- © 2025 г. Сыч В. В.



Эта статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная.

THE CORRELATION BETWEEN THE SOCIAL NETWORK ACTIVITY AND ACADEMIC PERFORMANCE OF TECHNICAL UNIVERSITY STUDENTS: A CASE STUDY OF *VKontakte*

V. V. Sych

Bauman Moscow State Technical University, Moscow, Russia
Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

✉ sycheslavovich@gmail.com

Abstract. The academic performance of students at Bauman Moscow State Technical University and its correlation to their activity in the *VKontakte* social network are considered. Machine learning methods are used to identify distinct performance paths reflecting the dynamics of educational achievement. The subscription lists of students are analyzed to identify marker communities characterizing the predominance of specific performance categories. Graph-theoretic clustering is applied to reveal structural groups of student interests. For each path, the stochastic vectors of interest shares across the community clusters are constructed and then used to identify the clusters having a statistically significant relation to particular performance paths. The results indicate a correlation between the digital behavior and academic outcomes of students, which contributes to the development of performance prediction models considering student interests in a social network.

Keywords: academic performance, *VKontakte*, machine learning, graph theory, performance prediction.