

# ПОДХОД К КЛАССИФИКАЦИИ TELEGRAM-КАНАЛОВ

Попов В. А.<sup>1</sup>, Чеповский А. А.<sup>2</sup>

DOI: 10.21681/2311-3456-2025-4-73-83

**Цель исследования:** разработка метода определения цифрового профиля Telegram-каналов в сетях информационного взаимодействия и процедуры классификации каналов на основе выделенного цифрового профиля.

**Метод исследования:** метод исследования включает следующие этапы: построение графа взаимодействующих объектов на основании импортированных из сети Telegram данных, определение цифровых профилей вершин на основании их атрибутивных данных и свойств графа, кластеризация вершин на основании выделенных профилей, классификация центров полученных кластеров и исходных Telegram-каналов, вычислительные эксперименты и анализ результатов.

**Полученный результат:** в данной статье вводится определение цифрового профиля Telegram-канала, представленного как одна из вершин графа взаимодействующих объектов. Цифровой профиль задан через нормализованный 5-мерный вектор признаков, полученных на основе атрибутивных данных вершины и свойств графа. Выбранные характеристики отражают свойства Telegram-каналов в построенном графе и метаданные, полученные при импорте из сети. Далее авторы описывают алгоритм кластеризации полученных профилей с использованием настраиваемых параметров. Центры выделенных кластеров классифицируются по 4 предложенным авторами типам, характеризующим роли вершин в графе взаимодействующих объектов. За счет этого производится классификация всех вершин графа – исходных Telegram-каналов анализируемой сети. Предложенный подход дает ценную информацию о ролях Telegram-каналов в сетях информационного взаимодействия.

**Научная новизна:** разработан новый подход к анализу Telegram-каналов: предложен метод создания цифрового профиля Telegram-канала в виде 5-мерного вектора признаков, что позволяет провести анализ и классификацию каналов. Также в рамках подхода предложена основанная на вычислительных методах процедура классификации таких цифровых профилей, которая позволяет выявить основные типы Telegram-каналов скачиваемой подсети по заданной классификации.

**Ключевые слова:** цифровые профили, анализ социальных сетей, безмасштабные сети, модель информационного воздействия, выделение сообществ, задачи классификации.

## Введение

В настоящее время в России мессенджер Telegram является одним из самых популярных приложений для общения и получения информации. Telegram предоставляет широкий набор функций для организации публичных каналов, которые представляют собой информационные ленты. Многие СМИ, информационные сообщества и блогеры имеют собственные Telegram-каналы и регулярно публикуют в них контент, а пользователи подписываются на эти каналы и получают информацию в виде сообщений.

Многие каналы в Telegram связаны между собой: например, каналы упоминают и репостят друг друга. Это позволяет рассматривать множества Telegram-каналов в виде коммуникационных сетей и формировать на основе этого взвешенные графы. В предыдущих работах авторами была предложена (U, M, R)-модель построения таких графов взаимодействующих объектов [1, 2]. Стоит отметить, что полученные при помощи этой модели графы обладают свойствами безмасштабных сетей [3].

Полученные сети информационного взаимодействия можно изучать различными способами. Существует множество исследований, в которых авторы анализируют топологию полученного графа, выделяют неявные сообщества с помощью различных алгоритмов (Louvain, Infomap, BigCLAM и др.) [4].

Также популярны работы, в которых авторы определяют каналы, распространяющие фейковую информацию или занимающиеся противоправной деятельностью. В частности, в работе [5] авторы сосредоточились на изучении поддельных каналов, выдающих себя за официальные каналы знаменитостей или организаций и публикующие сообщения, отличные от сообщений официального канала, и каналов-клонов, которые имитируют официальные каналы, публикуя их точный контент.

Еще одним способом анализа различных объектов является построение их цифровых профилей. Данный метод активно применяется в работе рекомендательных систем. Например, музыкальный

1 Попов Владимир Александрович, аспирант Департамента прикладной математики МИЭМ НИУ ВШЭ, Москва, Россия. E-mail: vapopov@hse.ru

2 Чеповский Александр Андреевич, кандидат физико-математических наук., доцент, Департамент прикладной математики МИЭМ НИУ ВШЭ, Москва, Россия. E-mail: aachepovsky@hse.ru

стриминговый сервис Spotify представляет каждую песню в виде вектора характеристик. Если посмотреть официальный API [6], то можно увидеть, что в данный вектор входит целый ряд параметров, таких как темп, такт, громкость, модальность, инструментальность и др. Сам рекомендательный алгоритм находится в закрытом доступе, но в интернете есть множество упоминаний, что Spotify используют данные профили для нахождения похожих песен. Данный подход также применяют и в работах по анализу видео. Например, в [7] видеозаписи на YouTube сопоставляют вектор из трех компонент, отвечающих за контент видео, социальную и рекомендательную составляющие.

Есть работы, посвященные и Telegram. Так, в [8] авторы анализируют посты каналов, каждому сообщению сопоставляют его повествовательный тип. Выделяют изолированное, продолжающее, возникающее, затухающее, продолжающе-возникающее или продолжающе-затухающее сообщение. После классификации каждого поста в канале авторы подсчитывают относительные частоты категорий сообщений по типу и используют эти значения для представления канала в виде вектора признаков, который называют вектором поведенческого профиля. В дальнейшем, авторы кластеризуют полученные векторы поведенческих профилей, чтобы идентифицировать разнообразные повествовательные профили Telegram-каналов.

Учитывая особенности мессенджера Telegram возможно предложить методы построения цифровых профилей для публичных каналов этой сети, основанные как на метаданных каналов, так и на топологии импортированной подсети. Эти профили могут быть сформированы с использованием построенного по (U, M, R)-модели графа взаимодействующих объектов, а также анализа атрибутивных данных вершин, а именно: текстов постов каналов, реакций подписчиков. Представление каналов в виде цифровых профилей позволяет сравнивать каналы, находить закономерности, проводить классификацию каналов, тем самым решая различные задачи по анализу и поиску каналов в Telegram [9].

Далее в данной работе мы предложим метод построения цифрового профиля Telegram-каналов на основе графа взаимодействующих объектов и процедуру классификации каналов на основе выделяемого этим способом цифрового профиля. Также будет предложен набор типовых профилей для такой классификации и приведены примеры.

### 1. Модель построения графа взаимодействующих объектов

Учитывая особенности мессенджера Telegram, можно выделить следующие ключевые факторы взаимодействия между публичными Telegram-каналами:

репосты между каналами, упоминания одного канала другим и наличие общих внешних URL в постах двух каналов. Учитывая наличие таких связей, Telegram-каналы образуют между собой сеть, которую можно представить в виде взвешенного графа.

Для построения таких моделей сетей каналов изначально необходимо импортировать данные из мессенджера за выбранный временной период. Для этого было разработано программное обеспечение, использующее официальный API Telegram. Данное приложение способно импортировать информацию о каналах в специализированный формат [1, 2], содержащий все необходимые компоненты для дальнейшего построения графов и цифровых профилей Telegram-каналов.

Соответственно, после импорта всех необходимых данных используя алгоритм, описанный в [1], строим граф взаимодействующих объектов сети Telegram  $G(V, E)$ , где  $V$  – список рассматриваемых каналов, а  $E$  – рёбра между ними. Веса полученных рёбер определяются на основе функции, описанной ниже:

$$w(e_{AB}) = 1 \cdot \delta_{e_{AB}}^U + 2 \cdot \delta_{e_{AB}}^M + 3 \cdot \delta_{e_{AB}}^R, \quad (1)$$

где  $\delta_{e_{AB}}^U$  – количество общих уникальных внешних ссылок (URL) в постах каналов  $A$  и  $B$  за выбранный период времени;  $\delta_{e_{AB}}^M$  – количество постов, в которых канал  $A$  упоминал канал  $B$ , плюс количество постов, в которых канал  $B$  упоминал канал  $A$  за выбранный период времени (для каждого поста учитываются уникальные упоминания);  $\delta_{e_{AB}}^R$  – количество репостов каналом  $A$  постов канала  $B$  плюс количество репостов каналом  $B$  постов канала  $A$  за выбранный период времени.

Ранее в [3] было показано, что графы, построенные с использованием описанной выше (U, M, R)-модели, обладают свойствами безмасштабных сетей.

Для дальнейшей иллюстрации применения предлагаемого в данной работе метода определения цифровых профилей и классификации каналов были импортированы несколько сетей:

- Сеть каналов, связанных с образованием, назовем ее **\*\*\*Education\*\*\***. Для формирования этой сети мы импортировали данные из Telegram за двухмесячный период 2025 года (01.01–28.02). Полученный граф состоит из 110 вершин и 1281 ребра.
- Сеть каналов, связанных со спортом, назовем ее **\*\*\*Sport\*\*\***. Для формирования этой сети мы импортировали данные из Telegram за двухмесячный период 2025 года (01.01–28.02). Полученный граф состоит из 514 вершин и 8496 рёбер.
- Сеть каналов, связанных со светской жизнью, назовем ее **\*\*\*Social\_life\*\*\***. Для формирования

этой сети мы импортировали данные из Telegram за двухнедельный период 2025 года (01.02–14.02). Полученный граф состоит из 534 вершин и 3354 рёбер.

- Сеть каналов, связанных с искусством, назовем ее **\*\*\*Art\*\*\***. Для формирования этой сети мы импортировали данные из Telegram за двухмесячный период 2025 года (01.01-28.02). Полученный граф состоит из 823 вершин и 7970 рёбер.

Перейдем теперь непосредственно к определению цифрового профиля для Telegram-каналов.

## 2. Определение цифрового профиля Telegram-канала

Каждая вершина в сети каналов Telegram обладает своими характеристиками, которые зависят как от топологии рассматриваемого графа, так и от атрибутивных данных вершины. Граф построен на некотором временном интервале  $T$  обрабатываемых данных, на нем же рассматриваются и метаданные вершин. Для начала рассмотрим три отдельные группы характеристик.

Первая группа представляет собой показатели, связанные с контентом канала:

- количество подписчиков канала (обозначим показатель за  $A1$ );
- количество постов в указанном временном интервале  $T$  ( $A2$ );
- доля репостов относительного общего количества постов канала ( $A3$ );
- количество упоминаний других Telegram-каналов в среднем за пост ( $A4$ );
- количество внешних ссылок в текстах в среднем за пост ( $A5$ );
- количество эмодзи (графических символов для выражения эмоций, идей или объектов в электронной коммуникации) в текстах в среднем за пост ( $A6$ ).

Вторая группа характеристик представляет собой реакцию пользователей на контент канала, также рассматривается на всем временном интервале  $T$ :

- среднее количество просмотров постов канала (обозначим показатель за  $B1$ );
- среднее количество репостов и пересылок постов канала ( $B2$ );
- среднее количество поставленных реакций пользователей на пост ( $B3$ );
- среднее количество комментариев к постам ( $B4$ );
- рейтинг вовлеченности ( $B5$ ) – среднее значение показателя вовлеченности по всем постам на канале, рассчитываемое как количество пересылок, реакций и комментариев, деленное на количество просмотров поста.

Третья группа относится к свойствам вершины и топологии графа, отражает роль канала в рассматриваемой сети. Для графа взаимодействующих объектов  $G(V,E)$  базовые показатели обозначаются

стандартно:  $N$  – количество вершин,  $w_{ij}$  – вес ребра между вершинами  $i$  и  $j$ . Для каждой вершины можно рассчитать следующие классические меры центральности:

- Центральность по степени (обозначим показатель за  $C1$ ):

$$c_d(i) = \sum_j^N w_{ij}. \quad (2)$$

- Центральность по собственному вектору ( $C2$ ):  
Определяется как решение уравнения:

$$Ac_e = \lambda c_e, \quad (3)$$

где  $A$  – матрица смежности графа  $G$ ;  $\lambda$  – максимальное собственное значение матрицы  $A$ ;  $c_e = [c_{e1}, c_{e2}, \dots, c_{eN}]^T$  (вектор центральностей вершин, элементы которого представляют собой центральности вершин) – собственный вектор, соответствующий собственному значению  $\lambda$ .

- Центральность по близости ( $C3$ ):

$$c_c(i) = \frac{N-1}{\sum_{j \neq i}^N d_{ij}}, \quad (4)$$

где  $d_{ij}$  – кратчайшее расстояние между вершинами  $i$  и  $j$ .

- Нормированная центральность по посредничеству ( $C4$ ):

$$c_b(i) = \frac{2}{N(N-1)} \sum_{(j,k), j \neq k} \frac{\sigma_{j,k}(i)}{\sigma_{j,k}}, \quad (5)$$

где  $\sigma_{j,k}$  – число кратчайших путей от вершины  $j$  до вершины  $k$ .  $\sigma_{j,k}(i)$  – число этих путей, проходящих через вершину  $i$ .

Далее необходимо уменьшить число характеристик вершины, входящих в цифровой профиль канала. Покажем на примерах из приведенных ранее графов процесс этого отбора.

Для каждого Telegram-канала, входящего в четыре сети, которые были описаны в разделе 1, вычислим указанные выше характеристики. Далее, в разрезе каждой из сетей, для всех пар показателей посчитаем коэффициент ранговой корреляции Спирмена. Полученные результаты представлены в виде матриц корреляции (рис. 1).

Корреляция Спирмена для пары характеристик рассчитывалась как коэффициент корреляции Пирсона между рангами двух наборов данных, где ранг – это порядковый номер значений в отсортированных выборках. В случае повторяющихся значений в исходных наборах данных применялся метод среднего ранга: совпадающим значениям присваивается среднее арифметическое их рангов, которые они заняли бы при различии.

В данной работе была выбрана ранговая корреляция Спирмена, поскольку она устойчивее к выбросам и не требует нормального распределения данных.

Как видно из рис. 1, многие показатели коррелируют между собой. Например, центральности по степени (показатель C1), по собственному вектору (C2) и по близости (C3) имеют попарные коэффициенты корреляции близкие к 1, а количество репостов (B2) и просмотров (B1) имеют коэффициент корреляции выше 0,8.

Соответственно, для построения цифрового профиля Telegram-канала из всего набора выберем 5 репрезентативных слабо коррелирующих между собой характеристик, которые представляют различные уникальные особенности канала:

- Количество подписчиков канала (A1);
- Рейтинг вовлеченности (B5);

- Количество эмодзи в текстах в среднем за пост (A6);
- Центральность по степени (C1);
- Центральность по посредничеству (C4).

Выбранные характеристики отражают различные свойства исходных Telegram-каналов. Количество подписчиков показывает насколько популярен канал среди пользователей. Рейтинг вовлеченности измеряет активность и вовлеченность аудитории. Количество эмодзи в текстах относится к стилю написания текстов в канале. Центральности по степени и по посредничеству характеризуют роль, которую играет вершина в графе взаимодействующих объектов.

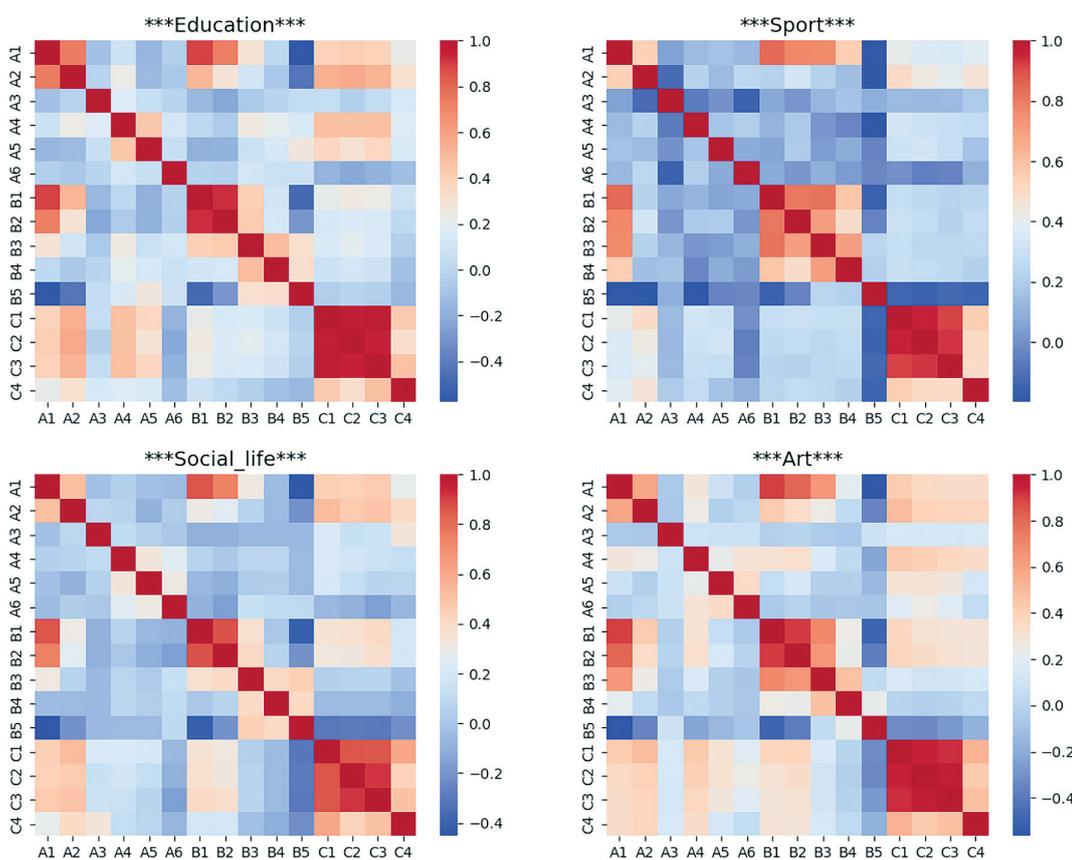


Рис. 1. Матрицы корреляций характеристик каналов

Примеры посчитанных характеристик Telegram-каналов

Таблица 1.

Telegram-канал	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
*university*	14991	0,07	0,16	0,11	0,04
*sport_channel*	306512	0,22	6,02	0,14	0,33
*social_life_channel*	297523	0,01	0,02	0,19	0,60
*art_channel*	34711	0,02	4,23	0,26	0,31

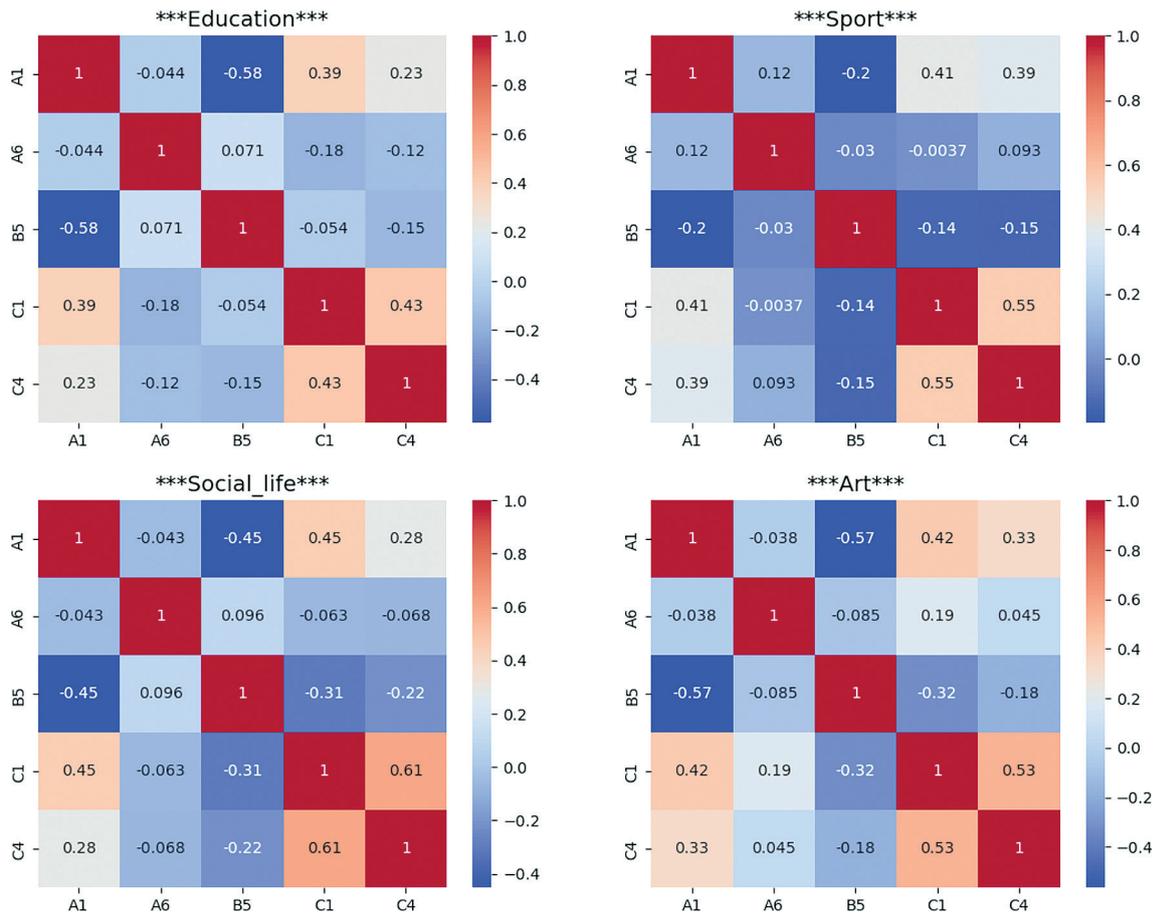


Рис. 2. Матрицы корреляций выбранных 5 характеристик каналов

На рис. 2 приведены матрицы коэффициентов корреляции Спирмена между выбранными пятью характеристиками канала. Стоит отметить, что большинство показателей имеют значение по модулю меньше 0.3. Из исключений количество подписчиков в канале и два показателя центральностей, которые отражают разные характеристики графов, поэтому были выбраны оба.

Таким образом, на основе описанного набора характеристик каждый Telegram-канал в графе взаимодействующих объектов соответствует 5-мерному вектору. В таблице 1 приведены посчитанные характеристики для относительно популярных каналов из 4 представленных выше сетей соответственно: \*university\*, \*sport\_channel\*, \*social\_life\_channel\*, \*art\_channel\*.

Указанные характеристики рассчитываются для каждого Telegram-канала анализируемого графа. Но учитывая наличие в сети каналов-выбросов, для последующего сравнения и анализа полученных 5-мерных векторов необходима их нормализация. Например, каналы со сравнительно большим числом подписчиков, как правило, искажают статистику. Так, в четырех рассматриваемых сетях медианное

значение количества подписчиков равно тысячам, а максимальное – миллионам (табл. 2). Такая диспропорция может привести к искажению набора цифровых профилей.

Таблица 2. Медианное и максимальное значение количества подписчиков в сетях

Сеть Telegram-каналов	Кол-во подписчиков, Максимум	Кол-во подписчиков, Медиана
***Education***	3.377.591	16.644
***Sport***	3.164.182	9.502
***Social_life***	12.002.650	23.289
***Art***	1.304.256	4.115

Поэтому для каждой из выделенных характеристик мы вычисляем среднее значение и стандартное отклонение в разрезе рассматриваемой сети Telegram-каналов и выявляем каналы-выбросы, значения которых отличаются более чем на три стандартных отклонения от среднего. Значения характеристик для этих каналов устанавливаются на уровне трех стандартных отклонений от среднего. Применяя эту

Таблица 3.

Примеры цифровых профилей Telegram-каналов

Telegram-канал	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
*university*	0,01	0,59	0,01	0,13	0,14
*sport_channel*	0,56	0,23	0,61	0,45	1,00
*social_life_channel*	0,19	0,03	0,01	1,00	1,00
*art_channel*	0,07	0,04	0,34	1,00	1,00

процедуру к выбросам, мы избегаем искажения всего набора цифровых профилей, что облегчает анализ характеристик каналов на последующих этапах исследования.

Далее мы нормализуем 5-мерные векторы, путем вычитания из каждой компоненты минимального значения соответствующей характеристики во всем графе и последующего деления каждой компоненты на максимальное значение.

Полученный нормализованный 5-мерный вектор с компонентами на отрезке [0; 1] представляет собой цифровой профиль канала Telegram в заданной сети за указанный период времени. Такое векторное представление характеристик каналов позволяет в дальнейшем проводить сравнительный анализ и их классификацию.

Продолжая предыдущий пример, построим цифровые профили для каналов \*university\*, \*sport\_channel\*, \*social\_life\_channel\*, \*art\_channel\* (Табл. 3). Стоит отметить, что каждая сеть каналов рассматривается отдельно друг от друга, поэтому для четырех Telegram-каналов проводилась своя независимая нормализация.

Как видно из таблицы, профили каналов существенно отличаются и каждый из них имеет свои особенности. Первый канал имеет малое для своего графа количество подписчиков, но высокий рейтинг вовлеченности аудитории. Если изучить сам канал, то так оно и есть: в канале примерно 15 тысяч подписчиков, и аудитория проявляет высокую активность: под постами суммарно ставятся тысячи реакций и оставляются сотни комментариев. Второй канал отличается средним количеством подписчиков, повышенным количеством эмодзи в текстах и наивысшим показателем центральности по посредничеству. А третий и четвертый каналы, учитывая их наивысшие степени центральности, являются важнейшими вершинами в структуре их графов взаимодействующих объектов. Что на практике так и есть: вершина \*social\_life\_channel\* связана с 99 из 534 вершин в графе, а \*art\_channel\* связана с 211 из 823 вершин в сети. Таким образом, цифровые профили Telegram-каналов дают представления об их основных характеристиках.

### 3. Использование цифровых профилей для классификации Telegram-каналов

Предложенный в данной работе метод можно кратко записать следующим образом.

#### Алгоритм смешанной классификации

Шаг 1. Вычисление для каждой вершины  $v_i \in V$  графа взаимодействующих объектов ее показателей:  $A1, A6, B5, C1, C4$ . Далее обозначим их для заданной вершины  $v_i$  как  $v_i^j$ , где  $j = 1, \dots, 5$ .

Шаг 2. Составление для каждой вершины  $v_i \in V$  ее цифрового профиля, вектора  $p_{v_i} = (v_i^1, v_i^2, v_i^3, v_i^4, v_i^5)$ .

$$\alpha_0: V \rightarrow \{p_{v_i}\}. \quad (6)$$

Шаг 3. Кластеризация векторов  $p_{v_i}$ . Получаем  $k$  кластеров  $K_s$ , где  $s = 1, \dots, k$ .

$$\alpha_1: \{p_{v_i}\} \rightarrow K_s. \quad (7)$$

Шаг 4. Поиск в кластерах  $K_s$  их центров. Получаем множество  $\{z_s\}$ , где  $s = 1, \dots, k$ . Так можно считать, что определено отображение:

$$\alpha_2: \{p_{v_i}\} \rightarrow \{z_s\}. \quad (8)$$

Шаг 5. Классификация полученных центров  $\{z_s\}$  по заданному наперед набору из  $g$  классов  $M = \{M_j\}$ , где  $j = 1, \dots, g$ :

$$\alpha_3: \{z_s\} \rightarrow M. \quad (9)$$

Шаг 6. Распространение на все вершины, входящие в кластер результата классификации его центра:

$$\alpha: \{v_i\} \rightarrow M. \quad (10)$$

Таким образом, в итоге действия Алгоритма классификации получаем для каждой вершины исходного графа ее отнесение к одному из заданных классов:

$$\alpha(v_i) = \alpha_3(\alpha_2(\alpha_0(v_i))). \quad (11)$$

Фактически, в представленном алгоритме смешанной классификации заложено последовательное решение задачи кластеризации элементов одного множества и задачи классификации элементов другого множества. Использование кластеризации вместе с классификацией имеет ряд преимуществ. Кластеризация помогает выявить скрытые закономерности и структуры в данных, что дает дополнительную

информацию об объектах. А также центры кластеров более устойчивы к выбросам и случайному шуму, чем отдельные точки, что в целом может повысить точность классификации [10].

Учитывая, что мы рассматриваем графы взаимодействующих объектов, будем при классифицировании учитывать, какую роль играет вершина в графе. Зададим множество  $M = \{M_1, M_2, M_3, M_4\}$ , состоящее из 4-х классов каналов:  $M_1$  – крупные центры;  $M_2$  – мосты и ретрансляторы;  $M_3$  – локальные или узко-направленные центры;  $M_4$  – пассивные каналы.

Реализация шагов 1 и 2 производится за счет описанных в разделе 1 действий. Далее для реализации шага 3 необходимо определиться с используемым алгоритмом кластеризации.

Существует множество алгоритмов кластеризации векторов, но не все они подходят для нашей задачи. В частности, большинство классических алгоритмов, например, k-средних или DBSCAN, требуют заранее определять параметры. Соответственно, использование таких алгоритмов не позволяет сформировать универсальный метод кластеризации цифровых профилей для сетей каналов разного размера и структуры, так как параметры алгоритмов зависят от характеристик сети.

Поэтому для кластеризации цифровых профилей Telegram-каналов мы предлагаем использовать метод агломеративной иерархической кластеризации. Суть данного алгоритма заключается в том, что на изначальном этапе каждый вектор рассматривается как отдельный кластер, далее они итеративно объединяются в кластеры большего размера, пока не будет сформирован один единый кластер. На каждом шаге алгоритма объединяются два кластера, расстояние между которыми минимальное. Таким

образом, иерархическое дерево формируется от листьев к стволу.

Для реализации данного подхода необходимо определить расстояние  $\rho_1$  для пары произвольных векторов  $p_{v_i}$  и  $p_{v_j}$ . Для расчета расстояния между двумя элементами-векторами будем использовать расстояние Чебышева: максимальная абсолютная разность между координатами двух векторов в пространстве, как представлено в формуле (12).

$$\rho_1(p_{v_i}, p_{v_j}) = \max_{t=1, \dots, 5} |v_i^t - v_j^t|. \quad (12)$$

Учитывая проведенную нормализацию при построении цифровых профилей все координаты векторов принимают значения от 0 до 1. Тогда расстояние Чебышева между двумя цифровыми профилями будет принадлежать отрезку [0, 1].

Далее, для определения расстояния  $\rho_2$  между двумя кластерами  $K_1$  и  $K_2$  будем использовать метод полной связи (метод дальнего соседа): расстояние между двумя кластерами определяется как максимум из множества всех попарных расстояний между элементами первого кластера  $K_1$  и элементами второго кластера  $K_2$ . Более формально это записано в виде формулы (13):

$$\rho_2(K_1, K_2) = \max_{p_{v_i} \in K_1, p_{v_j} \in K_2} \rho_1(p_{v_i}, p_{v_j}). \quad (13)$$

Как было указано ранее  $\rho_1(p_{v_i}, p_{v_j}) \in [0, 1]$ . Откуда следует, что и расстояние между двумя кластерами будет также принадлежать отрезку [0, 1].

Помимо того, что метод иерархической кластеризации не требует предварительного указания параметров, алгоритм также раскрывает иерархическую структуру данных. Для этого можно построить дендрограмму – древовидную диаграмму, отражающую каждый шаг процесса последовательного укрупнения

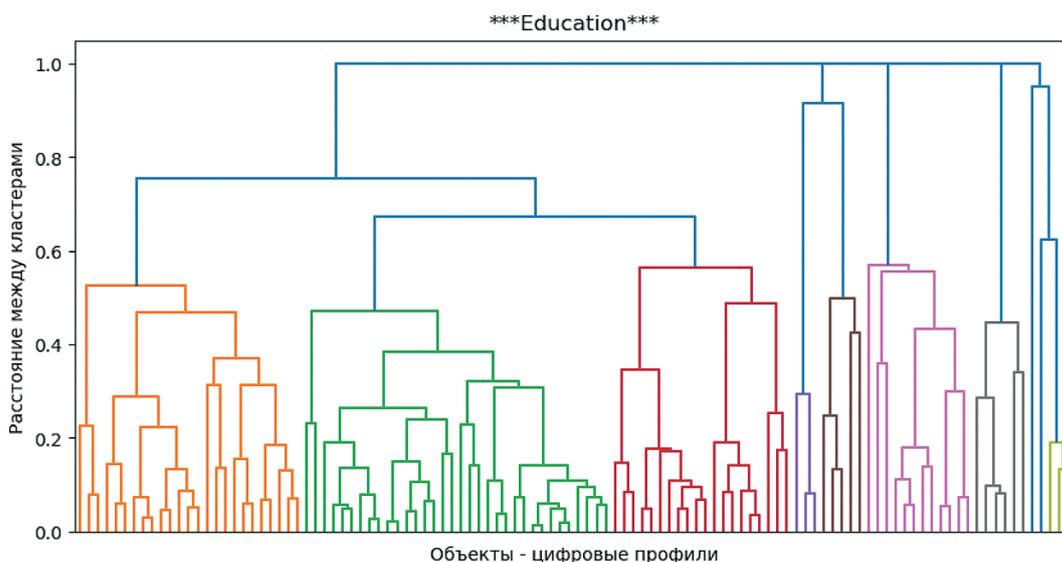


Рис. 3. Дендрограмма кластеризации цифровых профилей Telegram каналов сети \*\*\*Education\*\*\*

кластеров, что позволяет выявлять значимые кластеры сети на основе определенных критериев. Для определения границы между кластерами, а значит и того, когда объекты находятся в разных кластерах, изначально в данном методе задается параметр  $\Delta$ . Экспериментально на графах, полученных из сети Telegram-каналов, наиболее эффективно показало себя значение  $\Delta = 0,6$ .

Для примера применим описанный метод к упомянутой ранее сети каналов, связанных с образованием, к сети *\*\*\*Education\*\*\**. На рисунке 3 представлена полученная дендрограмма кластеризации 110-ти векторов цифровых профилей, которые были разделены на 10 кластеров.

На шаге 4 производим поиск  $\{z_s\}$  – центров кластеров. Пусть  $\{p_{v_i} = (v_i^1, v_i^2, v_i^3, v_i^4, v_i^5)\}$ , где  $i = 1, \dots, |K_s|$  – множество всех векторов, принадлежащих кластеру  $K_s$ . Тогда центром кластера  $K_s$  будет являться следующий вектор:

$$z_s = \frac{1}{|K_s|} \sum_{i=1}^{|K_s|} p_{v_i} = \left( \frac{1}{|K_s|} \sum_{i=1}^{|K_s|} v_i^1, \frac{1}{|K_s|} \sum_{i=1}^{|K_s|} v_i^2, \frac{1}{|K_s|} \sum_{i=1}^{|K_s|} v_i^3, \frac{1}{|K_s|} \sum_{i=1}^{|K_s|} v_i^4, \frac{1}{|K_s|} \sum_{i=1}^{|K_s|} v_i^5 \right). \quad (14)$$

Далее на шаге 5 будем производить классификацию полученных центров кластеров  $\{z_s\}$ , находя, к какому из классов множества  $M = \{M_1, M_2, M_3, M_4\}$  они относятся. Для этого определим эталонный вектор для каждого из этих классов. После чего, с использованием евклидова расстояния  $\rho_z$  найдем расстояние от каждого из центров до каждого из эталонных векторов.

Учитывая, что эти классы соответствуют следующим ролям каналов в сети: крупные центры, мосты и ретрансляторы, локальные или узконаправленные центры, пассивные. Данные типы Telegram-каналов имеют свои особенности, которые можно отразить в цифровых профилях. В таблице 4 приведены значения предложенных авторами цифровых профилей для каждого типа.

Для классификации полученных 10 центров кластеров, найдем евклидово расстояние от них до каждого из четырех предложенных профилей и выберем минимальное (Таблица 5).

Итак, для каждого из центров  $\{z_s\}$  сопоставлен его класс из множества  $M = \{M_1, M_2, M_3, M_4\}$ .

Теперь перейдем к шагу 6 и присвоим всем вершинам исходного графа, входящим в каждый кластер, свой класс, равный классу центра этого кластера.

Таблица 4.

Эталонные цифровые профили для каждого типа Telegram-каналов

Тип Telegram-канала	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
Крупный центр	1	0,25	0	1	0,5
Мост и ретранслятор	0	0,5	0,5	0,5	1
Локальный или узконаправленный	0,25	1	0,5	1	0
Пассивный	0	0	0	0	0

Таблица 5.

Евклидово расстояние от центров кластеров до выбранных векторов цифровых профилей

№ кластера	Крупный центр	Мост и ретранслятор	Локальный или узконаправленный	Пассивный
1	<b>1,40</b>	1,53	1,43	1,41
2	<b>0,85</b>	1,42	1,42	0,94
3	1,10	<b>0,58</b>	1,34	1,31
4	1,33	<b>0,64</b>	1,48	0,81
5	1,51	1,17	<b>1,03</b>	1,20
6	1,16	1,04	1,01	<b>0,61</b>
7	1,41	1,16	1,29	<b>0,25</b>
8	1,39	1,05	1,06	<b>0,55</b>
9	1,51	1,12	1,04	<b>0,71</b>
10	1,67	1,19	1,28	<b>0,93</b>

Кластерные центры сети \*\*\*Education\*\*\*

№	Тип кластера	Количество подписчиков	Рейтинг вовлеченности	Количество эмодзи в текстах	Центральность по степени	Центральность по посредничеству
1	Крупный центр	0,951	0,225	1	0,167	0
2	Крупный центр	0,882	0,035	0,085	0,297	0,104
3	Мост и ретрансл.	0,037	0,271	0,069	0,804	1
4	Мост и ретрансл.	0,034	0,068	0,265	0,182	0,736
5	Локальный/узконапр.	0,053	0,195	1	0,634	0
6	Пассивный	0,027	0,212	0,121	0,552	0,078
7	Пассивный	0,036	0,195	0,115	0,093	0,031
8	Пассивный	0,011	0,304	0,382	0,251	0,011
9	Пассивный	0,001	0,669	0,211	0,089	0,011
10	Пассивный	0,07	0,22	0,902	0,08	0

В нашем рассматриваемом примере мы определим типы полученных 10 центров кластеров, а соответственно и типы всех Telegram-каналов, входящих в эти кластеры (Таблица 6). Как видно из таблицы, в данной сети присутствуют все четыре типа каналов.

После этого экспертным образом была оценена полученная в итоге классификация каналов сети. Разделение каналов по типу их цифрового профиля оказалось актуальным. В крупные центры попали каналы больших СМИ и новостных сообществ, которые в основном сами генерируют свой контент, не делая репосты и не упоминая другие каналы, минимальное количество подписчиков для этих каналов равно 1,2 миллионам. Также интерес представляет наличие двух кластеров крупных центров. При детальном изучении самих каналов выясняется, что в кластер № 2 попали крупные новостные каналы без конкретных направленностей, а в кластер № 1 – каналы, связанные только с военной тематикой. Выделение данных каналов в отдельный кластер также подтверждает релевантность алгоритма кластеризации.

Далее для анализа интересны каналы – мосты и ретрансляторы. Каналы этих кластеров действительно являются распространителями информации: в основном данные каналы распространяют информацию, полученную из других Telegram-каналов или других источников информации. Но интерес заключается в том, что в сети было выявлено два таких кластера. После изучения списков каналов в этих кластерах было установлено, что кластер № 4 включает официальные Telegram-каналы, аффилированные с государством, а кластер № 3 – каналы с альтернативными взглядами. Также стоит отметить, что у каналов в третьем кластере максимальные

значения центральностей, что часто характерно для каналов распространителей информации.

Таким образом, получаем алгоритм смешанной классификации  $\theta(M, \alpha_1, \rho_1, \rho_2, \alpha_2, \alpha_3, \rho_3)$ . Применяя предложенный алгоритм смешанной классификации, можно определить роль и особенности каждого узла в сети каналов Telegram, что полезно для многих задач в контексте анализа социальных сетей.

**Заключение**

В данной статье представлено определение цифрового профиля для Telegram-каналов на основании как топологических свойств вершин графа взаимодействующих объектов, так и на основании характеристик самого канала. Далее описан метод для решения задачи классификации каналов на основании их цифровых профилей. Выделение четырех типов узлов (крупные центры, мосты и ретрансляторы, локальные или узконаправленные центры и пассивные каналы) позволило нам классифицировать и более подробно описать характеристики каналов в сети.

Также были приведены примеры построения цифровых профилей Telegram-каналов и последующая их классификация. Экспертная оценка примеров еще раз подтвердила актуальность данного подхода к классификации, в частности, подчеркнув различие между официальными каналами Telegram, связанными с государственными органами, и теми, которые принадлежат альтернативным новостным агентствам.

Это исследование дает ценные сведения о различных ролях и функциях каналов в Telegram. Таким образом, эти результаты могут быть использованы в будущих исследованиях и прикладных задачах.

## Литература

1. Попов В. А., Чеповский А. А. Модели импорта данных из мессенджера Telegram // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2022. Т. 20. № 2. С. 60–71.
2. Чеповский А. А. Анализ графов взаимодействующих объектов. — М.: Национальный открытый университет «ИНТУИТ». 2022. — 270 с.
3. Попов В. А., Чеповский А. А. О моделях построения графа взаимодействующих объектов в сети Telegram-каналов // Вопросы кибербезопасности. 2024. № 3(61). С. 105–112. DOI:10.21681/2311-3456-2024-3-105-112.
4. Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов // Успехи кибернетики. — 2023. — Т.4. — № 1. — С. 56–64.
5. La Morgia M., Mei A., Mongardini A. M., Wu J.: Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. <https://arxiv.org/abs/2111.13530>. (2021). (Дата обращения: 01.07.2025).
6. Spotify for Developers – <https://developer.spotify.com/documentation/web-api/reference/get-audio-features> (Дата обращения: 05.07.2025).
7. Leopaul Boesinger, Manoel Horta Ribeiro, Veniamin Veselovsky, Robert West: Tube2Vec: Social and Semantic Embeddings of YouTube Channels. <https://arxiv.org/abs/2306.17298> (Дата обращения: 01.07.2025).
8. Willaert T.: A computational analysis of Telegram’s narrative affordances. PLoS ONE 18(11), p. 1–23, (2023). <https://doi.org/10.1371/journal.pone.0293508>
9. Popov, V. A., Chepovskiy, A. A.: Constructing Telegram Channels Digital Profiles. Complex Networks & Their Applications XIII. COMPLEX NETWORKS 2024 2024. Studies in Computational Intelligence, vol. 1189. Springer, Cham. [https://doi.org/10.1007/978-3-031-82435-7\\_7](https://doi.org/10.1007/978-3-031-82435-7_7) (2025).
10. Piernik, M., Morzy, T. A study on using data clustering for feature extraction to improve the quality of classification. Knowledge and Information Systems, 63, 1771–1805 (2021).

## TELEGRAM-CHANNELS CLASSIFICATION APPROACH

Popov V. A.<sup>3</sup>, Chepovskiy A. A.<sup>4</sup>

**Keywords:** digital profiles, social network analysis, scale-free networks, model of information impact, community identification, classification problems.

**The purpose of the study:** development of a method for determining the digital profile of Telegram channels in information interaction networks and a procedure for classifying channels based on the allocated digital profile.

**Method:** includes the following stages: graph of interacting objects construction, based on data imported from the Telegram network, digital profiles determination for vertices based on their attribute data and graph properties, clustering of vertices based on the selected profiles, centers of the obtained clusters and the original Telegram channels classification, computational experiments and analysis of the results.

**Results:** the article introduces the digital profile definition for a Telegram channel, presented as one of the vertices of a graph of interacting objects. The digital profile is defined through a normalized 5-dimensional feature vector based on the attribute data of the vertex and the graph properties. The selected characteristics reflect the properties of Telegram channels in the constructed graph and the metadata obtained during import from the network. The authors then describe an algorithm for clustering the obtained profiles using configurable parameters. The centers of the selected clusters are classified according to 4 types proposed by the authors, characterizing the roles of vertices in the graph of interacting objects. Due to this, all vertices of the graph are classified – the original Telegram channels of the analyzed network. The proposed approach provides valuable information about the roles of Telegram channels in information interaction networks.

**Scientific novelty:** a new approach to the analysis of Telegram channels is developed: a method for creating a digital profile of a Telegram channel in the form of a 5-dimensional feature vector, allowing the analysis and classification of channels. The approach also proposes a procedure for classifying such digital profiles based on computational methods, which allows to identify the main types of Telegram channels of the downloaded subnet according to a given classification.

## References

1. Popov V. A., Chepovskiy A. A. Modeli importa dannyh iz messendzhera Telegram // Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informacionnye tehnologii. 2022. T. 20. № 2. S. 60–71.
2. Chepovskiy A. A. Analiz grafov vzaimodejstvujushhih ob#ektov.: Nacional'nyj otkrytyj universitet «INTUIT». 2022. — 270 s.
3. Popov V. A., Chepovskiy A. A. O modeljah postroenija grafa vzaimodejstvujushhih ob#ektov v seti Telegram-kanalov // Voprosy kibernetiki. 2024. № 3(61). S. 105–112. DOI:10.21681/2311-3456-2024-3-105-112.
4. Chepovskiy A. A. O nejvnyh soobshhestvah na grafe vzaimodejstvujushhih ob#ektov // Uspehi kibernetiki. — 2023. — Т.4. — № 1. — С. 56–64.
5. La Morgia M., Mei A., Mongardini A. M., Wu J.: Uncovering the Dark Side of Telegram: Fakes, Clones, Scams, and Conspiracy Movements. <https://arxiv.org/abs/2111.13530>. (2021). (Data obrashhenija: 01.07.2025).

<sup>3</sup> Vladimir A. Popov, Ph.D. student, School of Applied Mathematics, HSE MIEM, Moscow, Russia. E-mail: vapopov@hse.ru.

<sup>4</sup> Alexander A. Chepovskiy, Ph.D. in Physics and Mathematics, Associate Professor, Department of Applied Mathematics, Moscow Institute of Economics, National Research University Higher School of Economics. Moscow, Russia. E-mail: aachevovskiy@hse.ru

6. Spotify for Developers – <https://developer.spotify.com/documentation/web-api/reference/get-audio-features> (Data obrashhenija: 05.07.2025).
7. Leopaul Boesinger, Manoel Horta Ribeiro, Veniamin Veselovsky, Robert West: Tube2Vec: Social and Semantic Embeddings of YouTube Channels. <https://arxiv.org/abs/2306.17298> (Data obrashhenija: 01.07.2025).
8. Willaert T.: A computational analysis of Telegram’s narrative affordances. PLoS ONE 18(11), p. 1–23, (2023). <https://doi.org/10.1371/journal.pone.0293508>.
9. Popov, V. A., Chepovskiy, A. A.: Constructing Telegram Channels Digital Profiles. Complex Networks & Their Applications XIII. COMPLEX NETWORKS 2024 2024. Studies in Computational Intelligence, vol. 1189. Springer, Cham. [https://doi.org/10.1007/978-3-031-82435-7\\_7](https://doi.org/10.1007/978-3-031-82435-7_7) (2025).
10. Piernik, M., Morzy, T. A study on using data clustering for feature extraction to improve the quality of classification. Knowledge and Information Systems, 63, 1771–1805 (2021).

