

DOI: 10.51790/2712-9942-2023-4-2-05

**АНАЛИЗ ТЕКСТОВ СООБЩЕСТВ СОЦИАЛЬНЫХ СЕТЕЙ****Н. Л. Аванесян<sup>1,a</sup>, В. В. Зенькова<sup>2,b</sup>, А. А. Чеповский<sup>1,c</sup>, А. М. Чеповский<sup>1,3,d</sup>**<sup>1</sup> *Национальный исследовательский университет «Высшая школа экономики», г. Москва, Российская Федерация*<sup>2</sup> *Московский физико-технический институт — национальный исследовательский университет (МФТИ), г. Долгопрудный, Московская область, Российская Федерация*<sup>3</sup> *Российский университет дружбы народов, г. Москва, Российская Федерация*<sup>a</sup> *nlavanesyan@hse.ru*, <sup>b</sup> *zenkova.vv@phystech.edu*<sup>c</sup> *ORCID: <https://orcid.org/0000-0002-0938-3517>,  [aachepovsky@hse.ru](mailto:aachepovsky@hse.ru)*<sup>d</sup> *ORCID: <https://orcid.org/0000-0001-8959-6119>, [chepovskiy-am@rudn.ru](mailto:chepovskiy-am@rudn.ru)*

*Аннотация:* в настоящей работе авторы приводят описание методики статистического анализа текстов Telegram-каналов, основанной на сравнении методами корреляционного анализа автоматически сформированных частотных словарей. Рассматриваются коэффициенты попарной ранговой корреляции для сравнения частотных характеристик текстов различных неявных сообществ на естественном языке. Метод предлагается для оценки качества выделения неявных сообществ на графе, полученном при импорте данных из сети каналов мессенджера Telegram.

*Ключевые слова:* анализ социальных сетей, неявные сообщества графов, автоматический анализ текстов, частотный словарь, ранговая корреляция.

*Для цитирования:* Аванесян Н. Л., Зенькова В. В., Чеповский А. А., Чеповский А. М. Анализ текстов сообществ социальных сетей. *Успехи кибернетики*. 2023;4(2):33–39. DOI: 10.51790/2712-9942-2023-4-2-05.

*Поступила в редакцию: 12.05.2023.**В окончательном варианте: 31.05.2023.***ANALYSIS OF SOCIAL MEDIA COMMUNITY POSTS****N. L. Avanesyan<sup>1,a</sup>, V. V. Zenkova<sup>2,b</sup>, A. A. Chepovskiy<sup>1,c</sup>, A. M. Chepovskiy<sup>1,3,d</sup>**<sup>1</sup> *Higher School of Economics, Moscow, Russian Federation*<sup>2</sup> *Moscow Institute of Physics and Technology (MIPT), Dolgoprudny, Moscow Region, Russian Federation*<sup>3</sup> *Peoples Friendship University of Russia, Moscow, Russian Federation*<sup>a</sup> *nlavanesyan@hse.ru*, <sup>b</sup> *zenkova.vv@phystech.edu*<sup>c</sup> *ORCID: <https://orcid.org/0000-0002-0938-3517>,  [aachepovsky@hse.ru](mailto:aachepovsky@hse.ru)*<sup>d</sup> *ORCID: <https://orcid.org/0000-0001-8959-6119>, [chepovskiy-am@rudn.ru](mailto:chepovskiy-am@rudn.ru)*

*Abstract:* the study is a statistical analysis of Telegram channels based on the correlation analysis of automatically generated frequency dictionaries. We presented the pairwise correlation factors for comparing the word frequency distribution in informal community posts. The method can identify informal communities by analyzing a graph built on the data exported from the Telegram messenger.

*Keywords:* social media analysis, informal graph communities, computerized analysis of texts, frequency dictionary, rank correlation.

*Cite this article:* Avanesyan N. L., Zenkova V. V., Chepovskiy A. A., Chepovskiy A. M. Analysis of Social Media Community Posts. *Russian Journal of Cybernetics*. 2023;4(2):33–39. DOI: 10.51790/2712-9942-2023-4-2-05.

*Original article submitted: 12.05.2023.**Revision submitted: 31.05.2023.***Введение**

Анализ социальных сетей как графов взаимодействующих объектов является одним из самых актуальных направлений в исследовании сетевых систем [1–3]. В отечественной литературе чаще рассматривается построение моделей распространения информации и информационного влияния в социальных сетях [4–7]. Но важнее представляется решение задач информационного воздействия и смежных задач на основе анализа структуры графа социальных сетей [1, 2, 8, 9]. Актуальной и сложной

проблемой такого анализа является задача выделения неявных сообществ в графе социальной сети [9–12].

Под выделением неявных сообществ на графе подразумевается разбиение графа на подграфы, такое, что плотность связей внутри этих подграфов намного выше плотности связей между ними. При этом актуальна задача выделения на графе пересекающихся сообществ, подразумевающих наличие общих вершин, принадлежащих сразу двум или более сообществам. Для такого типа задач в [13] был разработан метод Галактик, который в процессе своего применения обеспечивает выделение пересекающихся неявных сообществ.

Одной из наиболее сложных задач при выделении сообществ на графе является проблема оценки корректности и эффективности работы соответствующих методов [1, 14]. Часто используемые способы оценки качества работы алгоритмов выделения сообществ основаны на использовании случайных графов [15, 16], что не соответствует структуре реальных сложных сетей [8, 14]. Поэтому в обзорной работе основоположников данной области исследований [2] проблема оценки качества выделения сообществ на графах ставится как актуальный вопрос информационных технологий.

В работе [17] было показано, что для качественного анализа сетей Telegram-каналов удобным инструментом является сочетание алгоритмического подхода по выделению сообществ в совокупности с подсчетом психолингвистических характеристик. Это позволяет выделять группы каналов, ведущих активное информационное воздействие, подтверждать корректность выделения неявных сообществ. В данной работе для сравнительного анализа текстов сообществ, выделяемых на графе Telegram-каналов, мы предлагаем использовать методику сравнения частотных словарей текстов этих сообществ. Такая методика была ранее применена в работе [18] для импортированных из сети Twitter данных.

### **Анализируемые сообщества социальных сетей**

В данной работе мы исследуем атрибуты графов, которые получаются в результате импорта данных из сети Telegram-каналов согласно  $(U, M, R)$ -модели, описанной в [19]. Такого типа модели характеризуют распространение информации, имевшее место за импортируемый промежуток времени  $T$ . Кратко опишем содержание модели, предложенной в [19]. В своих сообщениях каналы имеют возможность ссылаться и цитировать публичные записи друг друга, вставлять в тексты сторонние URL-ссылки и упоминания других каналов с автоматическими тегами переходов для пользователей. Для построения графа взаимодействующих объектов этой сети в качестве вершин берутся каналы, а ребра между ними строятся на основании имевших место взаимодействий между каналами за промежуток времени  $T$ . Вес на ребрах такого графа задается как линейная комбинация весов, соответствующих каждому из зафиксированных типов взаимодействия. Само множество вершин строится поиском в ширину, начиная от заданного исходного множества, состоящего из одного или нескольких Telegram-каналов, путем пополнения этого множества за счет анализа на каждом шаге постов за промежуток времени  $T$ . За глубину скачивания отвечает параметр, который задает ограничение для осуществления поиска в ширину. Новые каналы для пополнения множества вершин находятся из постов каналов, составляющих текущее множество. Подробности модели и метода скачивания описаны в [13, 19].

Многие каналы в мессенджере Telegram не имеют единой тематики постов или отражают различные взгляды и интересы. По этой причине канал мессенджера может относиться к нескольким неявным сообществам сети, и необходимо применять алгоритмы для выделения именно пересекающихся сообществ на графах (т. е. одна вершина может относиться к нескольким сообществам).

В данной работе для анализа структуры графа применяется метод Галактик, предложенный в работе [13] и позволяющий выделить пересекающиеся сообщества на графе взаимодействующих объектов, полученном при импорте данных из мессенджера Telegram по указанной выше модели скачивания. Метод состоит в последовательном выполнении следующих шагов: выделение пересекающихся сообществ на взвешенном графе; формирование мета-графа из полученных вершин-сообществ; обработка полученного мета-графа и последующее выделение непересекающихся мета-сообществ на мета-графе. В итоге получается разбиение исходного графа на пересекающиеся сообщества.

Далее исследуются два графа мессенджера Telegram.

Для первого графа  $G_1$  скачивание сети шло, начиная с канала *@kudago* (этот канал соответствует развлекательному сайту <http://kudago.com/msk/>). Начиная с данного канала, производился импорт данных из Telegram, согласно модели работы [19]. При импорте данных для этого графа за временной

интервал  $T$  был взят период с 03.10.2022 по 17.10.2022. Глубина импорта данных для графа  $G_1$  была взята равной 5. Исходный скаченный граф  $G_1$  состоит из 619 вершин и 2973 ребер.

Второй граф  $G_2$  скачивался, начиная с канала @ob\_obraz (этот канал рассказывает о новостях общего и высшего образования РФ). Начиная с данного канала, производился импорт данных из Telegram, согласно той же модели работы. Параметры модели были идентичны параметрам, использованным для первого графа. За временной интервал был взят период с 01.12.2022 по 31.01.2023. Глубина импорта данных для графа  $G_2$  была взята равной 2. Исходный скаченный граф  $G_2$  состоит из 168 вершин и 697 ребер.

Далее к графам  $G_1$  и  $G_2$  был применен метод Галактик и получены разбиения на неявные пересекающиеся сообщества. При анализе графа  $G_1$  в процессе работы алгоритма некоторые вершины и инцидентные им ребра убираются из графа как не участвующие активно во взаимодействии, поэтому после работы метода у графа  $G_1$  осталось 458 вершин. Эти вершины распределились по 8 выделенным пересекающимся сообществам. Аналогично у графа  $G_2$  остались 89 изначальных вершин, из которых были сформированы 8 сообществ в соответствии с методикой работы [13]. Их так же оказалось 8 в данном случае.

### Тексты и частотные словари текстов сообществ

Для каждого из сообществ  $S_i$  были скачаны текстовые сообщения всех пользователей — членов этих сообществ за исследуемый период. Полученные данные объединялись в единый для каждого сообщества массив текстов на естественном (русском) языке за указанный промежуток времени. В первых строках таблиц 1 и 2 приведены размеры созданных массивов текстов для сообществ первого и второго графов соответственно. При этом из текстов удалялись специальные символы, указатели. Методы анализа текстов сообществ социальных сетей были описаны в нашей работе [18].

Характеристики текстов определялись процедурами автоматизированной обработки текстов на естественных языках, описанными в [20, 21]. Лексический анализ позволяет формировать словари буквосочетаний различной длины.

Выделялись методами морфологического анализа отдельные части речи: существительные, глаголы и прилагательные. Частотные словари составлялись для канонических (начальных) форм слов, определяемых по результатам морфологического анализа. Процедуры синтаксического анализа [20] применялись для выделения именных и глагольных групп. Выделенные именные и глагольные группы несут информацию о различных аспектах тематического содержания текста. Частотные словари для словосочетаний составлялись для канонических (начальных) форм слов словосочетаний.

В качестве одной из лингвистических характеристик текста используется псевдооснова, под которой понимается часть слова без некоторых аффиксов. Используемые методы автоматического выделения псевдооснов разработаны в [20]. Лингвистическая характеристика псевдоосновы позволяет анализировать тексты без использования точных словоформ.

По результатам лингвистического анализа для каждого массива текстов сообществ  $S_i$  составлялись частотные словари различных характеристик, размеры которых в единицах записей приведены в таблице 1 для графа  $G_1$  и таблице 2 для графа  $G_2$ .

Таблица 1

Размеры текстов и частотных словарей текстов сообществ графа  $G_1$

Словари текстов	$S_0$	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$
Объем текстов (Кб)	335	39208	13571	3485	1633	37897	2925	10794
Буквосочетаний длиной 3 символа	5358	14161	12252	10092	8058	13785	8763	11958
Буквосочетаний длиной 4 символа	13224	68159	51774	37820	26166	65348	29962	50583
Буквосочетаний длиной 5 символов	17990	155459	106936	72282	44182	148137	53283	107010
Существительных	1373	12890	9100	6426	3785	12641	4507	9292
Глаголов	478	7581	5151	2947	1927	7255	2134	5109
Прилагательных	583	6182	4599	2838	1710	6128	2202	4336
Именных групп	11103	805899	368445	104458	45755	781353	94448	263547
Глагольных групп	2034	267680	105207	30786	13410	256324	24090	83145
Псевдооснов	4243	60280	38746	22625	12957	58487	16508	38054

Таблица 2

Размеры текстов и частотных словарей текстов сообществ графа  $G_2$ 

Словари текстов	S <sub>0</sub>	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	S <sub>4</sub>	S <sub>5</sub>	S <sub>6</sub>	S <sub>7</sub>
Объем текстов (Кб)	988	985	566	782	1021	8914	11881	1981
Буквосочетаний длиной 3 символа	6429	6956	6174	6966	7749	11272	12475	8103
Буквосочетаний длиной 4 символа	18536	20912	17572	20366	25291	44828	54111	26704
Буквосочетаний длиной 5 символов	29077	33816	26633	31953	42117	90963	116542	45980
Существительных	2324	2990	2253	2771	3675	7926	10075	3959
Глаголов	1124	1270	882	1136	1574	4076	5831	1822
Прилагательных	1107	1385	1039	1208	1539	3904	4752	1858
Именных групп	31562	42606	20450	29896	29266	280770	325056	74379
Глагольных групп	9360	8635	4679	6908	7409	79580	99957	18350
Псевдооснов	7830	9706	7048	8850	11516	32135	42875	13887

**Ранговый анализ словарей текстов сообществ**

Сравнение корпусов текстов неявных сообществ возможно осуществлять попарным сравнением частотных словарей различных лингвистических характеристик (всей лексики в целом, отдельных частей речи, псевдооснов слов и подобных показателей), составленных для каждого из исследуемых наборов текстов. Для частотных словарей устанавливаются ранги записей словаря по результатам сортировки по частоте встречаемости занесенной в словарь характеристики. Записи словарей рассматриваются как случайные величины. Тогда для каждой пары словарей можно подсчитать коэффициенты ранговой корреляции, которые будем называть коэффициентами попарной ранговой корреляции. По существу, данные коэффициенты являются оценками наличия монотонной связи между случайными величинами.

Полагаем, что каждая пара словарей имеет одинаковый размер  $n$  записей, который в реальных расчетах ограничивается некоторым заданным значением  $n$  для каждого из рассматриваемых словарей. Данное «обрезание» словарей обеспечивает отбрасывание характеристик с низкими частотами использования (как правило, единичными).

Рассмотрим словари как выборки для двух случайных величин  $X$  и  $Y$ , которые обозначим  $X^n = \{X_i\}_{i=1}^n$ ,  $Y^n = \{Y_i\}_{i=1}^n$ . Определим меру зависимости случайных величин  $X$ ,  $Y$  через средние значения выборок  $\bar{X}^n$  и  $\bar{Y}^n$ . Для ковариации  $\text{cov}(X^n, Y^n)$  тогда имеем:

$$\text{cov}(X^n, Y^n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}^n) (Y_i - \bar{Y}^n), \quad (1)$$

а дисперсия может быть записана в следующей форме:

$$\sigma_{X^n} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}^n)^2}. \quad (2)$$

Для отсортированных по частотам словарей будем рассматривать ранги элементов выборок  $rgX^n$  и  $rgY^n$ . Тогда коэффициент попарной ранговой корреляции для рассматриваемых выборок определяется следующим образом [22]:

$$r = r(rgX^n, rgY^n) = \frac{\text{cov}(rgX^n, rgY^n)}{\sigma_{rgX^n} \cdot \sigma_{rgY^n}}, \quad (3)$$

где числитель и знаменатель определяются формулами (1) и (2) соответственно.

Для ранжированных словарей размерностью  $n$  можем положить среднее рангов:

$$rg\bar{X}^n = \frac{n+1}{2}. \quad (4)$$

Получаем из (3) и (4) конкретное выражение для коэффициента попарной ранговой корреляции через значения рангов записей словарей  $rgX_i$  и  $rgY_i$ :

$$r = \frac{\sum_{i=1}^n (rgX_i - \frac{n+1}{2})(rgY_i - \frac{n+1}{2})}{\sqrt{\sum_{i=1}^n (rgX_i - \frac{n+1}{2})^2} \sqrt{\sum_{i=1}^n (rgY_i - \frac{n+1}{2})^2}}. \quad (5)$$

Формула (5) не накладывает никаких ограничений на порядок элементов, имеющих одинаковые значения. В случае набора  $k$  элементов  $X_{i_1}, \dots, X_{i_k}$ , имеющих одинаковые значения, они могут быть упорядочены согласно произвольной перестановке, и ранг  $i_j$ -го элемента можно задать:

$$rgX_{i_j} = R + \pi(j), \quad (6)$$

где  $R$  – ранг элемента, предшествующего по порядку группе элементов  $X_{i_1}, \dots, X_{i_k}$ ;  $\pi(j)$ ,  $j = 1, \dots, k$  – произвольная перестановка  $j$ -го элемента в группе элементов, имеющих одинаковые значения.

Коэффициент попарной ранговой корреляции  $r$  из (5) может принимать различные значения в зависимости от выбора перестановки  $\pi(j)$  в формуле (6). Для целей однозначности вычисления коэффициента (5) используется ранг, усредненный по всем перестановкам  $\pi(j)$ ,  $j = 1, \dots, k$ . Равные по значению элементы получают одинаковое значение усредненного ранга, не зависящее от их перестановки. Если все частоты внутри каждого из анализируемых словарей не совпадают (все элементы выборок  $\bar{X}^n$  и  $\bar{Y}^n$  различны), формула (5) преобразуется в классическую формулу ранговой корреляции Спирмена [23].

Однозначность и независимость от перестановок элементов с одинаковым значением при определении коэффициента попарной ранговой корреляции достигается использованием в (5) вместо ранга  $rgX_{i_1}$  элемента  $X_{i_1}$ , совпадающего по значению еще с элементами  $X_{i_2}, \dots, X_{i_k}$  выборки  $X^n$ , усредненный по всем перестановкам  $\pi \in S_k$  ранг.

Посредством коэффициента попарной ранговой корреляции сравниваются два частотных словаря размерностью  $m$  над множеством лексических характеристик  $W = \{\omega_i\}$  с частотными характеристиками  $f_i^k$  (где  $k$  – номер словаря).

$$D_1 = \left\{ d_{1,i}^m = \left( \omega_i, f_i^1 \right) \mid \omega_i \in W \right\},$$

$$D_2 = \left\{ d_{2,i}^m = \left( \omega_i, f_i^2 \right) \mid \omega_i \in W \right\}.$$

Анализируемые словари  $D_1$  и  $D_2$  могут иметь достаточно большие размеры. Поэтому мы рассматриваем только первые (по убыванию частоты)  $m$  элементов каждого из словарей. Таким образом, рассматриваются словари  $D_1^m = \{d_{1,i}^m = (\omega_i, f_i^1) \mid rgf_i^1 \leq m\}$  и  $D_2^m = \{d_{2,i}^m = (\omega_i, f_i^2) \mid rgf_i^2 \leq m\}$ . В случае, если в  $D_1^m$  встречается слово  $\omega$ , не встречающееся в  $D_2^m$ , мы полагаем при вычислении (5) его частоту в словаре  $D_2^m$ , равной 0, и наоборот. В данных исследованиях полагаем значение  $m = 10000$ .

Вычисляемый по (5) с учетом вышесказанного коэффициент попарной ранговой корреляции принимает значения на интервале  $[-1, 1]$ . Близкие к 1 значения говорят о монотонной согласованности словарей: если в одном словаре в паре слов одно имеет частоту выше другого, то оно и в другом имеет частоту выше другого. И так для всех пар. Близкие к  $-1$  говорят об обратном эффекте: если в одном словаре слово с частотой выше второго, то в другом – наоборот, его частота будет ниже. То же для всех пар. Если значение близко к 0, то словари несогласованы: соотношение между частотами слов в одном словаре независимо от соотношения с частотами в другом словаре. Цель такого анализа: определить, насколько тексты разных выделенных пересекающихся сообществ близки по психологической направленности, способности к действиям, что может определить корректность разделения каналов (объектов сети) на сообщества по направленности их деятельности.

### Результаты анализа текстов сообществ

Вычислительные эксперименты заключались в подсчете коэффициента корреляции (5) для всех пар частотных словарей лингвистических характеристик текстов выделенных сообществ, перечисленных в таблицах 1 и 2. Это следующие характеристики: буквосочетания длиной 3, 4 и 5 символов, существительные, глаголы, прилагательные, именные группы, глагольные группы, псевдоосновы. Мы исследуем для обоих графов по 8 сообществ, что означает определение 252 коэффициентов корреляции для наборов текстов сообществ каждого из двух графов. Далее конкретные примеры значений коэффициента корреляции приведены в таблицах 3 и 4, где представлены значения коэффициента корреляции (5) для конкретных словарей указанных в заголовке таблицы лингвистических характеристик объединенных текстов разных сообществ, выделенных в одном из рассматриваемых графов.

Сравнение по коэффициенту попарной ранговой корреляции частотных словарей буквосочетаний показывает сильное совпадение частотных распределений для буквосочетаний длиной до 3 символов, что подтверждает утверждение [18] о том, что данные буквосочетания характеризуют язык (все

наборы текстов на русском языке). Наблюдаемая согласованность частотных словарей буквосочетаний длиной 4 и 5 символов, псеодооснов для наборов текстов разных сообществ указывает на близость текстов разных сообществ по содержательной направленности, поскольку эти характеристики определяют, в первую очередь, тематику текстов.

Сравнение частотных словарей частей речи (существительных, прилагательных и глаголов) для различных наборов текстов сообществ по коэффициенту попарной ранговой корреляции показывает разброс значений, из которых невозможно принять решение о сравнении текстов разных сообществ.

Сравнение частотных словарей словосочетаний (именных и глагольных групп) показывает наличие возможных различий между частотными словарями словосочетаний текстов разных сообществ.

Для текстов сообществ первого графа сравнение словарей глагольных групп (Таблица 3) показывает возможность разделить наборы текстов большинства сообществ. Это указывает на различие текстов по их психологической направленности и направленности действия.

Результаты разделения текстов сообществ по именованным группам показаны в таблице 4 для второго исследуемого графа. Словари именованных групп попарно «обратны» по частотам использования словосочетаний в текстах. Это указывает на возможность выделения наиболее часто используемых именованных групп в наборах текстов разных сообществ.

В целом результаты сравнения, представленные в таблицах 3 и 4, показывают возможность рассматривать словосочетания как дифференцирующие признаки сообществ.

Таблица 3

Коэффициенты корреляции словарей глагольных групп текстов сообществ  $S_i$  графа  $G_1$

	$S_0$	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$
$S_0$	1							
$S_1$	-0.694	1						
$S_2$	-0.69	0.1	1					
$S_3$	0.529	-0.153	-0.258	1				
$S_4$	-0.741	-0.292	-0.409	-0.474	1			
$S_5$	-0.69	0.71	0.065	-0.168	-0.198	1		
$S_6$	-0.561	-0.055	-0.16	-0.213	-0.354	-0.029	1	
$S_7$	-0.655	0.197	-0.064	-0.073	-0.122	0.305	-0.102	1

Таблица 4

Коэффициенты корреляции словарей именованных групп текстов сообществ  $S_i$  графа  $G_2$

	$S_0$	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$
$S_0$	1							
$S_1$	-0.574	1						
$S_2$	-0.488	-0.406	1					
$S_3$	-0.598	-0.446	-0.373	1				
$S_4$	-0.609	-0.591	-0.107	-0.539	1			
$S_5$	-0.53	-0.378	-0.414	-0.495	-0.553	1		
$S_6$	-0.483	-0.399	-0.403	-0.466	-0.462	0.14	1	
$S_7$	-0.572	-0.351	-0.147	-0.35	-0.503	-0.28	-0.32	1

### Заключение

В данной работе на основе исследования реальных данных показана возможность оценки корректности выделения пересекающихся сообществ на графе, импортированном из сети Telegram-каналов. Данный подход основан на анализе методами компьютерной лингвистики объединенных корпусов текстов, выявленных у членов выделенных сообществ. Это может быть распространено на иные

графы, имеющие схожие атрибуты. Например, графы, полученные при импорте данных из социальных сетей.

Данная работа в совокупности с работами [17, 18] формулирует принципиально новый подход к оценке корректности выделения пересекающихся сообществ на графах взаимодействующих объектов, основанный на сочетании алгоритмических методов выделения сообществ с лингвистическим анализом текстов из выделенных сообществ.

## ЛИТЕРАТУРА

1. Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов. *Успехи кибернетики*. 2023;4(1):56–64. DOI: 10.51790/2712-9942-2023-4-1-08.
2. Fortunato S., Newman M. E. J. 20 Years of Network Community Detection. *Nat. Phys.* 2022;18:848–850.
3. Kanavos A., Voutos Y., Grivokostopoulou F., Mylonas P. Evaluating Methods for Efficient Community Detection in Social Networks. *Information*. 2022;13(5):209.
4. Базенков Н. И., Губанов Д. А. Обзор информационных систем анализа социальных сетей. *Управление большими системами: сб. трудов*. 2013:357–394.
5. Батура Т. В. Методы анализа компьютерных социальных сетей. *Вестник НГУ. Серия: Информационные технологии*. 2012;10(4):13–28.
6. Батура Т. В., Копылова Н. С., Мурзин Ф. А., Проскуряков А. В. Методы анализа данных из социальных сетей. *Вестник НГУ. Серия: Информационные технологии*. 2013;11(3):5–21.
7. Губанов Д. А., Новиков Д. А., Чхартишвили А. Г. *Социальные сети: модели информационного влияния, управления и противоборства*. М.: Физматлит: МЦНМО; 2010. 228 с.
8. Евин И. А., Хабибуллин Т. Ф. Социальные сети. *Компьютерные исследования и моделирование*. 2012;4(2):423–430. DOI: 10.20537/2076-7633-2012-4-2-423-430.
9. Newman M. E. J. *Networks: An Introduction*. Oxford University Press; 2010. 784 p.
10. Newman M. E. J. Modularity and Community Structure in Networks. *Proc. Natl. Acad. Sci. USA*. 2006;103(23):8577–8582.
11. 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;10:P10008.
12. Rosvall M., Esquivel A., Lancichinetti A., West J., Lambiotte R. Memory in Network Flows and its Effects on Spreading Dynamics and Community. *Nature Communications*. 2014;5:4630.
13. Попов В. А., Чеповский А. А. Выделение неявных сообществ на графе взаимодействия Telegram-каналов с помощью «метода Галактик». *Труды ИСА РАН*. 2022;72(4):39–50. DOI: 10.14357/20790279220405.
14. Fortunato S. Community Detection in Graphs. *Physics Reports*. 2010;486(3):75–174.
15. Lancichinetti A., Fortunato S., Radicchi F. Benchmark Graphs for Testing Community Detection Algorithms. *Physical Review E*. 2008;78:046110.
16. Danon L., Díaz-Guilera A., Duch J. & Arenas A. Comparing Community Structure Identification. *J. Stat. Mech.* 2005;P09008.
17. Чеповский А. А. Об особенностях построения и анализа графов взаимодействующих объектов в сети Telegram-каналов. *Вопросы кибербезопасности*. 2023;1:75–81. DOI: 10.21681/2311-3456-2023-1-75-81.
18. Аванесян Н. Л., Соловьев Ф. Н., Чеповский А. А. Характеристики текстов сообществ социальных сетей. *Вестник НГУ. Серия: Информационные технологии*. 2021;19(1):5–14. DOI: 10.25205/1818-7900-2021-19-1-5-14.
19. Попов В. А., Чеповский А. А. Модели импорта данных из мессенджера Telegram. *Вестник НГУ. Серия: Информационные технологии*. 2022;20(2):60–71. DOI: 10.25205/1818-7900-2022-20-2-60-71.
20. Чеповский А. М. *Информационные модели в задачах обработки текстов на естественных языках*. М.: Национальный открытый университет «ИНТУИТ»; 2015. 228 с.
21. Соловьев Ф. Н. Автоматическая обработка текстов на основе платформы ТХМ с учетом анализа структурных единиц текста. *Вестник НГУ. Серия: Информационные технологии*. 2020;18(1):74–82. DOI: 10.25205/1818-7900-2020-18-1-74-82.
22. Бендат Дж., Пирсол А. *Прикладной анализ случайных данных*. М.: Мир; 1989. 540 с.
23. Деза Е. И., Деза М. М. *Энциклопедический словарь расстояний*. М.: Наука; 2008. 444 с.