

На правах рукописи

РАБЧЕВСКИЙ Андрей Николаевич

**МЕТОДЫ И АЛГОРИТМЫ ПОДДЕРЖКИ ПРИНЯТИЯ
РЕШЕНИЙ ПО ПРОТИВОДЕЙСТВИЮ ДЕСТРУКТИВНЫМ
ВОЗДЕЙСТВИЯМ В СОЦИАЛЬНЫХ СЕТЯХ НА ОСНОВЕ
МНОГОФАКТОРНОГО АНАЛИЗА РОЛЕЙ ПОЛЬЗОВАТЕЛЕЙ**

Специальность: 05.13.10 – Управление в социальных
и экономических системах

Автореферат диссертации
на соискание ученой степени
кандидата технических наук

Пермь – 2022

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Пермский государственный национальный исследовательский университет».

Научный руководитель: **Ясницкий Леонид Нахимович**
доктор технических наук, профессор

Официальные оппоненты: **Минаев Владимир Александрович**,
доктор технических наук, профессор,
Федеральное государственное казенное
образовательное учреждение высшего образования
«Московский университет Министерства внутренних
дел Российской Федерации имени В.Я. Кикотя»,
профессор кафедры «Специальные информационные
технологии»

Торопов Борис Андреевич,
кандидат технических наук, доцент
Федеральное государственное казенное
образовательное учреждение высшего образования
«Академия управления Министерства внутренних
дел Российской Федерации», профессор кафедры
«Информационные технологии»

Ведущая организация: Федеральное государственное бюджетное
учреждение науки «Институт проблем управления
им. В.А. Трапезникова Российской академии наук»,
г. Москва

Защита диссертации состоится «16» июня 2022 г. в 14:00 на заседании диссертационного совета Д ПНИПУ.05.01 на базе ФГАОУ ВО «Пермский национальный исследовательский политехнический университет» по адресу: 614990, г. Пермь, Комсомольский проспект, дом 29, аудитория 345.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГАОУ ВО «Пермский национальный исследовательский политехнический университет» (<http://www.pstu.ru>).

Автореферат разослан «__» _____ 20__ г.

Ученый секретарь
диссертационного совета,
кандидат экономических наук,
Доцент

Алексеев Александр Олегович

I. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Современный уровень развития цифровых технологий привел к тому, что социальные сети прочно вошли в повседневную жизнь и стали оказывать влияние на поведение людей. Большинство молодых людей, попавших в преступную среду, суицидальные сообщества, различные протестные, экстремистские и террористические движения, вовлекаются в них посредством социальных сетей. Воздействие на социальные сети осуществляется в виде информационных волн, вызванных определенными социальными явлениями или поводами, информация о которых должна быть вброшена в социальную сеть, распространена среди максимального количества пользователей, усилена многочисленными обсуждениями и поддержана одобрением большого количества пользователей. За всеми этими действиями стоят конкретные пользователи, выполняющие определенные роли. Наибольшее деструктивное воздействие оказывают пользователи, обладающие максимальным уровнем информационного влияния. Существующие методики, используемые для их идентификации, не учитывают некоторые существенные характеристики, в том числе роли наиболее влиятельных пользователей (НВП), из-за чего экспертам, приходится обрабатывать большое количество профилей и тратить много времени на идентификацию пользователей, подлежащих мониторингу или воздействию. Идентификация ролей пользователей позволит выявить структуру воздействия на сеть. Анализ паттернов поведения таких структур по различным информационным поводам позволит выявлять закономерности, проявляющиеся при проведении информационных атак, выявлять признаки целенаправленного воздействия, прогнозировать поведение структур при возникновении информационных волн и блокировать ключевые узлы структур для противодействия распространению их влияния. Таким образом, для поддержки принятия решений при противодействии целенаправленным деструктивным воздействиям на социальные сети **актуальной научной задачей** является разработка эффективных методов идентификации ролей пользователей и уровня их влияния на социальную сеть, а также программного обеспечения для их автоматизации.

Степень разработанности темы. В изучение проблемы идентификации ролей пользователей существенный вклад внесли Е. Глив, Дж. Фюллер, О. Аразий, Х. Велсер, которые формализовали общее понятие ролей пользователей в различных сообществах. Однако для определения социальной роли пользователя каждый автор вводил свое определение, соответствующее задачам его исследования. Для идентификации ролей пользователей П. Брандцаег, М. Чичек, А. Аруларазан, П. Руас и другие ученые и эксперты применяли различные методы кластеризации, в то время как М. Яблоньска, Х. Лин, Ц. Сегалин, П. Вийдженайке, К. Матсумото, М. Сунгхан и многие другие использовали нейросетевую классификацию. Использование кластеризации не гарантирует разбиение множества примеров на заданное количество классов пользователей и сопряжено с трудностями интерпретации результатов кластеризации. Основной проблемой нейросетевой классификации является доступность необходимого количества примеров для обучения нейросетевой модели. В работах Г. Дилмегани, С. И. Николенко, Л. Н. Ясницкого и других экспертов в случае недоступности

примеров данных для обучения нейронных сетей предлагается использование синтетических данных, однако отсутствуют работы, связанные с применением синтетических данных для анализа социальных сетей. Таким образом, общепринятого подхода для определения ролей пользователей социальных сетей не существует.

Значительный вклад в изучение проблемы идентификации мостов в кластерных сетях внесли Х. Вэй, К. Джанг, З. Галман, П. Йенсен, А. Альварес-Сокорро, в работах которых приводятся определения мостов и предлагаются методы их идентификации, однако предложенные определения мостов не соответствуют форме моста, исследуемой в диссертационной работе, и не позволяют использовать существующие методы для идентификации мостов.

Изучению проблемы идентификации ключевых пользователей социальных сетей посвятили исследования Л. Фриман, П. Кэррингтон, Дж. Скотт, С. Вассерман, М. Кастельс, Д. Фэрис, П. Гербаудо, Л. Лю, Д. Чен, Д.А. Губанов, А.Г. Чхартишвили, Б.А. Торопов и многие другие ученые и эксперты, однако их работы были посвящены рассмотрению отдельных аспектов влияния в рамках узких специализированных подходов. Зарубежные авторы, как правило, рассматривают влияние пользователей с точки зрения их положения в структуре социальной сети и используют для этого теоретико-графовый подход. В то же время, Д.А. Губанов и А.Г. Чхартишвили оценивают влияние пользователей с точки зрения их публикационной активности в сети, а Б.А. Торопов применяет теоретико-игровой подход, рассматривая пользователей социальной сети как конкурентных участников игры. Однако комплексные подходы, учитывающие влияние пользователей в нескольких аспектах, а также их роль в единой структуре сети, на данный момент отсутствуют.

Объектом исследования являются цифровые социальные сети.

Предметом исследования – методы идентификации ролей пользователей и уровня их влияния в социальных сетях.

Цель исследования – повышение эффективности принятия решений при противодействии целенаправленным деструктивным воздействиям за счет совершенствования методов идентификации ролей пользователей и уровня их информационного влияния на социальные сети.

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

- 1) разработать метод идентификации ролей пользователей в социальных сетях;
- 2) разработать метод идентификации пользователей, выполняющих роль связующих мостов между ядром социальной сети и изолированными кластерами пользователей;
- 3) разработать метод идентификации НВП на основе многофакторного анализа;
- 4) разработать прототип информационной системы «Сервис анализа распространения контента»;
- 5) исследовать эффективность разработанных методов и алгоритмов в рамках аналитических исследований.

Положения, выносимые на защиту, обладающие научной новизной:

1. Предложен метод подготовки множества данных для обучения и тестирования нейросетевого классификатора, отличающийся тем, что для создания множества данных используется генератор случайных чисел, работающий в пределах диапазонов, заданных экспертом на основании знаний о предметной области, что позволяет классифицировать роли пользователей, вовлеченных в протестное движение в социальных сетях, решить проблему доступности обучающих множеств данных и исключить риск утечки конфиденциальных данных реальных пользователей. (п.6 паспорта специальности 05.13.10 – *Разработка и совершенствование методов получения и обработки информации для задач управления социальными и экономическими системами*).

2. Предложен метод идентификации пользователей социальных сетей, отличающийся использованием метрики центральности по взвешенному вкладу, что позволяет точно вычислять мосты в двухколенных кластерных сетях и идентифицировать узлы сети, соединяющие сообщества пользователей с ядром социальной сети, а также определять уровень их информационного влияния на социальную сеть (п.7 паспорта специальности 05.13.10 – *Разработка методов идентификации в организационных системах на основе ретроспективной, текущей и экспертной информации*).

3. Предложен метод идентификации пользователей социальных сетей, отличающийся комплексным учетом количества социальных связей пользователей и уровня их публикационной активности, что позволяет повысить степень релевантности идентификации НВП и идентифицировать пользователей, обладающих наибольшим потенциалом донесения информации до участников социальной сети (п.7 паспорта специальности 05.13.10 – *Разработка методов идентификации в организационных системах на основе ретроспективной, текущей и экспертной информации*).

Теоретическая значимость работы состоит в разработке новых представлений экспертных знаний, новых алгоритмических моделей интеллектуальной обработки текущей и ретроспективной информации и поддержки принятия решений и развитии теории и методов идентификации ролей пользователей и уровня их влияния в социальных сетях, интегрирующих акциональный, структурный и интерпретационный подходы анализа социальных сетей с использованием нейросетевых технологий.

Практическая значимость работы. Использование синтетических данных для обучения и тестирования нейросетевого классификатора ролей пользователей позволило существенно сократить затраты и сроки подготовки датасета, а дополнительное использование экспертных знаний в алгоритме синтеза позволило повысить точность нейросетевой модели классификатора и качество классификации.

Алгоритм идентификации мостов в кластерных сетях позволил выявлять пользователей, способствующих максимальному росту охвата аудитории, которых существовавшие методики не выявляли или выявляли неточно.

Математическое обоснование модели идентификации НВП позволило повысить производительность труда специалистов при проведении подробных

аналитических исследований за счет повышения релевантности поиска и снижения количества исследуемых пользователей.

Разработана и прошла государственную регистрацию «Программа вычисления мостов в кластерных сетях».

Разработана и прошла государственную регистрацию «База данных для классификации ролей пользователей социальных сетей».

Разработанные методы и алгоритмы внедрены в программный комплекс «поисковая система «SEUS» производства ООО «СЕУСЛАБ» в виде ПО «Сервис анализа распространения контента», что в свою очередь позволило:

- приблизительно в 1000 раз сократить затраты на создание датасета и исключить риск утечки конфиденциальной информации;
- выявить пользователей, играющих важную роль в распространении информации и увеличении охвата аудитории социальной сети, вовлеченной в социальное явление, которые ранее поисковой системой «SEUS» технически не могли быть обнаружены;
- снизить трудозатраты аналитиков при проведении исследований в 10,7 раза, а при проведении более детальных исследований, учитывающих структуру воздействия на сеть, в 42,8 раз.

Методология и методы исследования. Теоретическую и методологическую основу исследования составили труды, основные положения и выводы отечественных и зарубежных учёных, исследователей, экспертов и разработчиков компьютерных технологий анализа социальных сетей. Для исследований в рамках подготовки диссертации были использованы методы системного анализа, статистического и сравнительного анализа, методы эмпирического исследования и логического анализа. Применение данных методов позволило дать обоснование полученным в ходе диссертационного исследования результатам. В диссертации использованы Интернет-ресурсы и информационные данные из докладов, представленных на национальных и международных научно-практических конференциях.

Степень достоверности и апробация результатов. Достоверность теоретических результатов обуславливается их соответствием методологии управления и подтверждается результатами аналитических исследований профилей выявленных ролей пользователей социальных сетей и результатами внедрения разработанного программного в технологический процесс ООО «СЕУСЛАБ».

Апробация полученных результатов была проведена на следующих международных и всероссийских конференциях: XVIII всероссийской научной конференции «Нейрокомпьютеры и их применение», Международной конференции и Шестой всероссийской научно-практической конференции «Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века», XIX всероссийской научной конференции «Нейрокомпьютеры и их применение», 2nd International Conference on Cyber-Physical Systems & Control, The 2021 International Conference on Digital Science, Международном форуме «Наука и глобальные вызовы XXI века», The 10th International Conference — Analysis of Images, Social Networks and Texts.

Публикации. Основные результаты диссертации изложены в 10 работах, в том числе 3 работы в рецензируемых научных изданиях, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, 2 работы в изданиях, индексируемых в Scopus, а также получено 1 свидетельство о государственной регистрации программы для ЭВМ и 1 свидетельство о государственной регистрации базы данных.

II. СОДЕРЖАНИЕ РАБОТЫ

Введение содержит обоснование актуальности исследуемого вопроса, приведены объект, предмет, поставлена цель и необходимые задачи для достижения данной цели, положения о научной новизне, сведения о результатах внедрения разработанного метода, апробация и публикации.

В первой главе представлен анализ существующих подходов и методов выявления НВП в социальных сетях, проводится их сравнительный анализ и делается вывод о том, что существующие методы выявляют влияние только в отдельных аспектах и не дают адекватной информации о реальной влиятельности пользователей в социальных сетях. Показано, что наиболее активные пользователи часто не имеют большого количества связей, а пользователи, у которых много связей, мало активны в сети. Предлагается комплексный подход, учитывающий одновременно количество публикаций по целевой тематике и количество социальных связей пользователей с другими участниками социальной сети. Представлен оригинальный метод выявления НВП, основанный на вычислении значения потенциального уровня влияния (ПУВ), отражающего максимально возможное количество актов донесения информации от исследуемого узла до других узлов сети.

Пусть в сети из n узлов имеется узел p_k , тогда количество связей с другими узлами графа для этого узла можно выразить как

$$C(p_k) = \sum_i^n a(p_i, p_k), \quad (1)$$

где $a(p_i, p_k) = 1$ тогда и только тогда, когда узлы p_i и p_k связаны между собой и 0 в противном случае. Если x_k – количество постов, y_k – количество репостов, а z_k – количество комментариев, опубликованных пользователем k , тогда количество материалов m , опубликованных пользователем k , будет равен

$$m_k = x_k + y_k + z_k, \quad (2)$$

а потенциальный уровень влияния l пользователя k можно выразить как

$$l_k = C(p_k) m_k. \quad (3)$$

Для исключения влияния разницы масштабов множителей в произведении предложено нормализовать множители по максимальным значениям их величин. Нормализованный ПУВ можно выразить как

$$l_k = \frac{C(p_k) m_k}{p_{\max} m_{\max}}. \quad (4)$$

Далее был предложен алгоритм вычисления группы НВП, чей суммарный уровень влияния равняется половине суммы уровней влияния всех пользователей.

На практике это может означать, что при блокировании этих пользователей уровень влияния, а значит и объем передаваемой в сети информации, уменьшится в 2 раза. Для этого был рассчитан рейтинг ПУВ путем ранжирования по убыванию значения для каждого пользователя и, последовательно двигаясь по рейтингу вниз, подсчитывалась сумма ПУВ для каждого пользователя, начиная с лидера рейтинга, и сравнивалась с общей суммой значений ПУВ для всех пользователей. Когда сумма ПУВ достигла значения 50% от общей суммы ПУВ, подсчет был остановлен. Математически это можно выразить так: пусть общий потенциал влияния L всех пользователей равняется

$$L = \sum_{n=1}^N l_n, \quad (5)$$

где N – общее количество пользователей. Тогда половина общего уровня влияния L_n будет выражена как

$$L_n = \sum_{n=1}^N l_n = 0,5L, \quad (6)$$

где n – номер пользователя в рейтинге ПУВ, для которого выполняется это равенство. Таким образом, наиболее влиятельными являются пользователи рейтинга ПУВ с номерами $\{1, 2, \dots, n\}$.

Использование данного метода в реальных аналитических исследованиях показало его практическую значимость за счет повышения релевантности поиска и существенного снижения затрат на анализ профилей пользователей. Вычисление ПУВ для массива пользователей выполняется автоматически. Метод не требует высокой квалификации экспертов, больших вычислительных затрат, прост в реализации как в прототипах, так и в промышленном варианте.

Во второй главе обосновывается актуальность проблемы выявления ролей пользователей социальных сетей, как более точного определения характера влиятельности пользователей, приводится определение ролей пользователей в контексте противодействия протестным движениям в социальных сетях и обзор существующих методов. Указано, что основными методами являются кластеризация множеств пользователей и нейросетевая классификация, однако готовых методов для выявления ролей пользователей в контексте противодействия не найдено.

Представлено описание *метода нейросетевой классификации* ролей пользователей. Указано, что основной проблемой, с которой приходится сталкиваться специалистам при проектировании нейросетевых моделей, это доступность и адекватность обучающих множеств. Приведены примеры создания датасетов для классификации ролей, а также примеры использования готовых датасетов. Сделан вывод о том, что создание датасета с помощью разметки профилей пользователей связано с очень высокими трудозатратами, готовые датасеты для решения задачи, поставленной в данном исследовании, недоступны.

Предложено создание датасета с использованием синтетических данных, алгоритм генерации которых учитывает диапазоны значений и соотношения параметров, полученные на основании экспертных знаний. При проектировании нейросети были приняты следующие входные и выходные параметры:

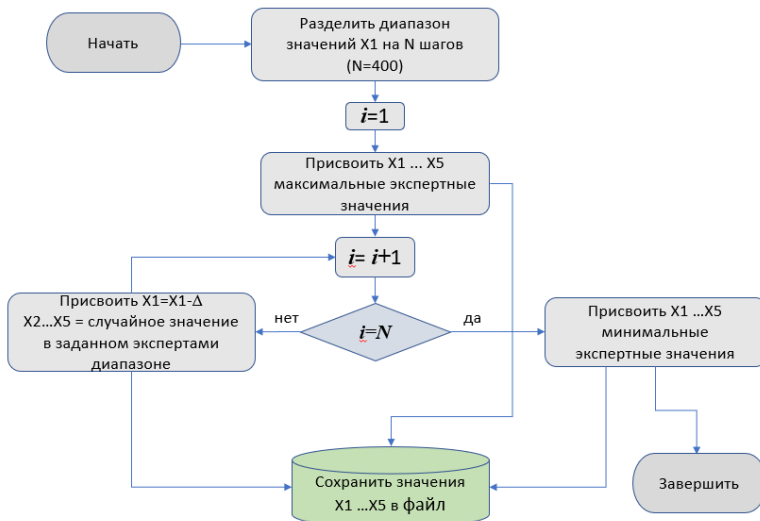
- X1 – Возраст аккаунта¹
 X2 – Количество друзей
 X3 – Количество опубликованных постов
 X4 – Количество опубликованных репостов
 X5 – Количество опубликованных комментариев.

Выходными данными нейросетевой модели являлись:

- D1 – равно 1, если пользователь является Постером и 0, если нет.
 D2 – равно 1, если пользователь является Репостером и 0, если нет.
 D3 – равно 1, если пользователь является Комментатором и 0, если нет.
 D4 – равно 1, если пользователь является Универсалом и 0, если нет.
 D5 – равно 1, если является Пассивным участником и 0, если нет.

Для генерации множества использовалась функция случайного выбора (рандомизации) значения из диапазона значений. Обозначим эту функцию как $R(X_{\min}; X_{\max})$.

В общем виде алгоритм генерации множества представлен на рисунке 1.



Пассивными считаются участники, которые имеют не высокий рейтинг активности, при этом значения не должны выходить за пределы максимальных и минимальных значений. В таблице 1 представлены формулы, использованные при генерации множества для роли Пассивный участник. Через Δ обозначен шаг уменьшения возраста аккаунта

Рисунок 1 - Общий вид алгоритма генерации множества. Для каждой роли экспертами задаются диапазоны значений и зависимости между X2 - X5

Таблица 1 – Формулы генерации множества для роли Пассивный участник

n	X1	X2	X3	X4	X5
1	$X1_{\max}$	$X2_{\max}$	$X3_{\max}$	$X4_{\max}$	$X5_{\max}$
2	$X1_{\max}-\Delta$	$R(X2_{\min}; X2_{\max})$	$R(X3_{\min}; X3_{\max})$	$R(X4_{\min}; X4_{\max})$	$R(X5_{\min}; X5_{\max})$
3	$X1_{\max}-2\Delta$	$R(X2_{\min}; X2_{\max})$	$R(X3_{\min}; X3_{\max})$	$R(X4_{\min}; X4_{\max})$	$R(X5_{\min}; X5_{\max})$
..
N	$X1_{\min}$	$X2_{\min}$	$X3_{\min}$	$X4_{\min}$	$X5_{\min}$

Постером является активный участник, основным видом активности которого является создание постов. Пусть p_i – количество постов, r_i – количество репостов, а k_i – количество комментариев, опубликованных пользователем i , тогда общее количество материалов этого пользователя m_i можно выразить как

¹ Возраст аккаунта исчисляется в днях с момента регистрации аккаунта в социальной сети.

$$m_i = p_i + r_i + k_i. \quad (7)$$

На основании представлений экспертов, полученных при анализе поведения пользователей, Постером является пользователь, количество постов которого, составляет не менее 60% от всех опубликованных им материалов, количество репостов – не более 37% и количество комментариев – не более 3%. То есть Постер должен удовлетворять следующим условиям:

$$p_i \geq \alpha m_i, r_i \leq \beta m_i, k_i \leq \gamma m_i, \text{ где } \alpha = 0,6, \beta = 0,37, \gamma = 0,03. \quad (8)$$

В этом случае значения r_i и k_i можно выразить как

$$r_i \leq \delta p_i, k_i \leq \varepsilon p_i, \text{ где } \delta = \frac{\beta}{\alpha}, \varepsilon = \frac{\gamma}{\alpha}. \quad (9)$$

Набор формул для роли Постер представлен в таблице 2.

Таблица 2 – Формулы генерации множества для роли Постер

n	X1	X2	X3	X4	X5
1	X1 _{max}	X2 _{max}	X3 _{max}	R(0; δX3)	R(0; εX3)
2	X1 _{max} -Δ	R(X2 _{min} ; X2 _{max})	R(X3 _{min} ; X3 _{max})	R(0; δX3)	R(0; εX3)
3	X1 _{max} -2Δ	R(X2 _{min} ; X2 _{max})	R(X3 _{min} ; X3 _{max})	R(0; δX3)	R(0; εX3)
..
N	X1 _{min}	X2 _{min}	X3 _{min}	X4 _{min}	X5 _{min}

Аналогичные вычисления были выполнены для оставшихся ролей. Всего для каждой роли было сгенерировано по 400 примеров, все они были объединены в общее множество из 2000 примеров, перемешаны и разделены на 2 множества: обучающее и тестовое, 1700 и 300 примеров соответственно.

Подготовленный датасет был использован для обучения и тестирования нейросетевой модели на платформе Нейросимулятор 5.0 Nsim5sc² (доступ www.LbAi.ru). В результате многочисленных итераций наилучший результат показала нейросеть персептронного типа с пятью входными нейронами, одним скрытым слоем с семью нейронами и пятью выходными нейронами. В качестве активационных функций всех нейронов использовался тангенс гиперболический. Для оценки погрешности в нейросимуляторе использовалась формула:

$$E = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - y_n)^2}{N}}}{\max(d_n) - \min(d_n)} 100\%, \quad (10)$$

где N – количество элементов выборки, d_n – заявленная роль n -го пользователя, а y_n – его роль, оцененная с помощью нейронной сети. Ошибка тестирования нейросетевой модели классификации ролей пользователей представлена в таблице 3.

² Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 5.0: Свидетельство о государственной регистрации программы для ЭВМ № 2014618208 от 12.07.2014 г.

Таблица 3 – Результат тестирования нейросетевой модели классификатора ролей пользователей на базе Нейросимулятора Nsim5-10

№	Наименование роли	Ошибка %
Y1	Постер	10,3%
Y2	Репостер	10,2%
Y3	Комментатор	2,9%
Y4	Универсал	16,5%
Y5	Пассивный участник	6,3%

Нейросетевая модель на основе синтетического датасета была обучена и протестирована в других нейросетевых пакетах (TensorFlow, Apple Create ML, Orange Data Mining) и во всех случаях наилучший результат показала нейронная сеть с теми же гиперпараметрами.

Валидация нейросетевой модели, созданной на основе искусственно синтезированного датасета, была проведена в виде классификации реальных пользователей и последующего аналитического исследования выявленных ролей пользователей. Валидация была выполнена на нескольких выборках пользователей, активно публикующих различные виды материалов (посты, репосты, комментарии) в социальных сетях во время наиболее активной фазы (10 – 15 дней) политических акций протестного характера, в том числе протестных акций вокруг выборов президента республики Беларусь и так называемого «Дворца Путина». В процессе валидации было получено полное совпадение результатов экспертной нейросетевой классификации и результатов аналитических исследований данных реальных пользователей социальной сети, выполненных экспертами-аналитиками.

Применяя метод вычисления значения ПУВ, в каждом классе были выявлены НВП. Было установлено, что для инфоповода «Жыве Беларусь» из 28,4 тысяч пользователей оперативный интерес представляют не более 100 пользователей, а для инфоповода «Дворец Путина» из 35,5 тысяч человек – не более 200. Такое существенное снижение количества пользователей, подлежащих контролю, позволяет резко повысить эффективность работы правоохранительных органов.

Показано, что нейросетевой метод классификации применим в случае, когда эксперты могут предсказать адекватные значения диапазонов и соотношений параметров предметной области. Важная особенность нейросетевого метода классификации состоит в его эффективности при использовании в потоковых онлайн приложениях оценки параметров пользователей, которые применяются при подготовке информации для принятия решения о блокировке, подавлении активности или постановке на мониторинг ключевых пользователей в социальной сети. Использование синтетических датасетов является универсальным методом и может рекомендоваться для использования в тех случаях, когда получение реальных данных для датасета невозможно, затруднено или в случае необходимости соблюдения конфиденциальности данных. Кроме того, можно предложить использование данного метода для увеличения количества примеров в датасетах, когда реальных примеров недостаточно для качественного обучения нейросети. Синтетические данные для обучения и тестирования нейросетевой

модели были зарегистрированы в качестве базы данных³ и доступны для использования по запросу.

В третьей главе описывается феномен наличия в графах пользователей, соединяющих кластеры пользователей и ядро социальной сети, названными в данной работе «мостами». Даны результаты аналитических исследований таких

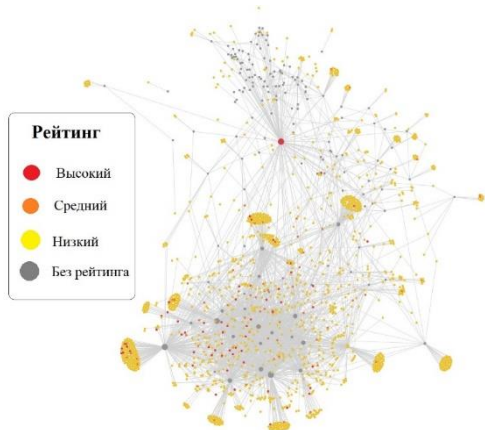


Рисунок 2 - Пример графа кластерной сети, где некоторые мосты не имеют рейтинга (окрашены серым цветом)

пользователей, их роли в социальном явлении и высоком вкладе мостов в общий уровень активности в социальной сети. Представлен обзор существующих методов выявления мостов и сделан вывод, что такие пользователи не могут быть выявлены на основе параметров активности, так как они могут не иметь активности по теме социального явления, а применение существующих теоретико-графовых методов выявления мостов не может гарантировать точное выявление мостов.

Пример графа, содержащего мосты, соединяющие кластеры с ядром сети, представлен на рисунке 2. Далее дано определение моста, особой методики построения графа для его выявления и предложена новая метрика центральности по взвешенному вкладу, точно выявляющая мосты согласно заданному определению. Автором исследуется частный случай, когда граф социальных связей включает только два колена, а кластеры подключены через одного единственного пользователя, при этом узлы кластера не имеют связи между собой. Для данного частного случая мостом считается узел социальной сети, удовлетворяющий следующим требованиям:

- узел, который подключает кластер к ядру сети;
- узлы кластера связаны только с мостом и не имеют связи между собой;
- мост связан с узлами кластера и узлами ядра.

Граф социальных связей узла, отвечающего определению моста представлен на рисунке 3.

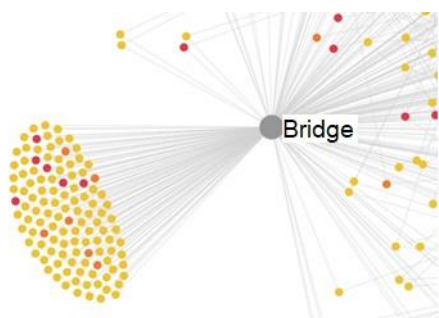


Рисунок 3 - Пример графа для узла, являющегося по определению мостом

Уровень влияния моста определяется количеством узлов в кластере и суммарным уровнем активности каждого узла в кластере, то есть суммарным уровнем активности кластера, подключаемого мостом к ядру сети. Далее предлагается описание метрики «Центральность по взвешенному вкладу». Пусть степень влияния моста на общий уровень публикационной активности временной

³ Рабчевский А.Н., Заякин В.С. База данных для классификации ролей пользователей социальных сетей. Свидетельство государственной регистрации базы данных №2021621533 от 15.07.2021. 2021.

социальной сети определяется как суммарный уровень активности кластера, который соединен с ядром через мост.

Рассмотрим граф, представленный на рисунке 4. Применение определения моста для данного графа дает следующие результаты:

- для красного узла связи с зеленым и серыми узлами учитываться не будут, так как они имеют связи с другими узлами в сети, а значит значение суммарного рейтинга красного узла, также как и для серых, будет равно 0,

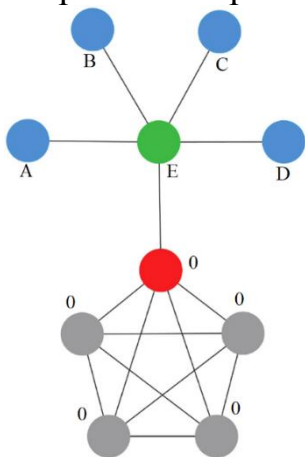


Рисунок 4 - Схема сети узлов с учетом весов

- для синих узлов, имеющих связь только с зеленым узлом, значение так же будет равно 0, так как у зеленого узла также есть другие связи,
- для зеленого узла связь с красным узлом также дает 0, а связи с синими узлами дают значения весов этих узлов,
- вес зеленого узла будет равен сумме весов синих узлов $E=A+B+C+D$.

Таким образом, получено одно ненулевое значение для зеленого узла во всей сети, величину которого можно обозначить как «Центральность по взвешенному вкладу» (ЦВВ). Значение ЦВВ = 0, если узел не является мостом, ≥ 1 – если узел является мостом, а самый влиятельный узел в роли моста имеет максимальное значение ЦВВ.

Далее приведена теоретико-множественная модель вычисления ЦВВ. Если рейтинг публикационной активности пользователя u обозначить как $r(u)$, то суммарный рейтинг некоторого подмножества пользователей V будет вычисляться по формуле

$$R(V) = \sum_{u \in V} r(u). \quad (11)$$

Тогда вес кластера, подключаемого мостом, произвольного пользователя графа равняется

$$W(u) = R(S(u)) = \sum_{w \in S(u)} r(w), \quad (12)$$

где $S(u)$ это множество пользователей, связанных только с пользователем u . ЦВВ определяется как отношение веса кластера, к общему весу всех кластеров в сети, который может быть рассчитан как

$$R = \sum_{u \in V} W(u). \quad (13)$$

То есть ЦВВ может быть выражена как

$$C_{wc}(u) = \frac{W(u)}{R}. \quad (14)$$

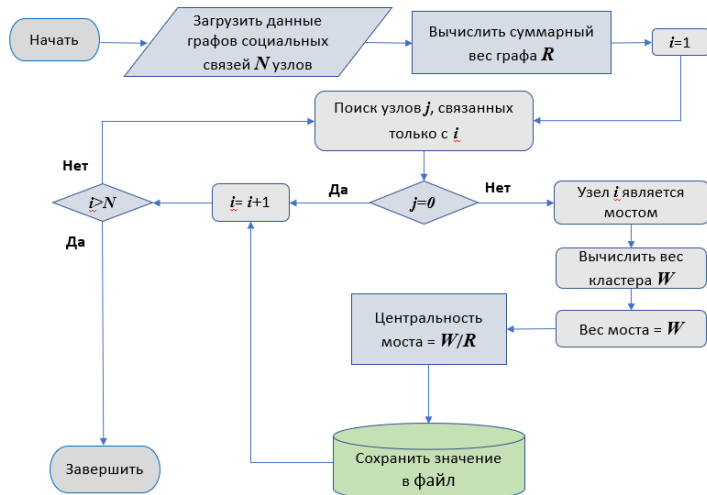


Рисунок 5 - Алгоритм идентификации мостов и вычисления ЦВВ в кластерной сети

Алгоритм идентификации мостов и вычисления значения ЦВВ, представленный на рисунке 5, реализован в виде программы вычисления мостов в кластерных сетях⁴. Для оценки эффективности ЦВВ было выполнено ее сравнение с другими методами, наиболее пригодными для выявления мостов – центральностью по промежуточности и центральностью по вкладу. Для этого из 10 случайных графов были удалены 10 самых влиятельных мостов и связанные с ними вершины, а также те вершины, которые оказались изолированными после

удаления мостов. Далее оценивалось уменьшение веса графа и веса всех кластеров в результате такого удаления, рассчитанное с помощью различных метрик центральностей (см. таблицу 4).

Таблица 4 – Изменение метрик центральности в результате удаления из графа 10 мостов с их вершинами

Метрика	Изменение веса графа	Изменение веса кластеров
Центральность по промежуточности	-54,4%	-76,0%
Центральность по вкладу	-52,7%	-66,4%
Центральность по взвешенному вкладу	-57,2%	-80,9%

Представленные в таблице 4 данные показывают, что ЦВВ наиболее эффективна с точки зрения изменения веса графа и веса кластеров.

Главная особенность ЦВВ состоит в том, что она однозначно определяет является ли узел мостом в описанных ранее сетевых конфигурациях. Также было выполнено исследование структуры, которую составляют наиболее влиятельные Постеры, Репостеры, Комментаторы, Универсалы и Мосты, на основании которого сделан вывод о важном значении мостов в структуре воздействия на социальные сети и не менее важном значении разработки эффективной методики для их идентификации.

Мосты способствуют расширению размеров сети, увеличению количества пользователей, вовлеченных в социальное явление, и увеличению общего уровня активности социальной сети. Блокирующее воздействие на самые влиятельные мосты может существенно изменить характеристики всей сети и снизить общий уровень активности социальной сети в данном социальном явлении. Таким образом, воздействие на наиболее влиятельные мосты является эффективным

⁴ Рабчевский А.Н., Заякин В.С. “Программа вычисления мостов в кластерных сетях” Свидетельство государственной регистрации программ для ЭВМ № 2021616086 от 16.04.2021. Пермь: ООО “СЕУСЛАБ,” 2021.

способом снижения уровня охвата аудитории и публикационной активности социальной сети.

В четвертой главе представлены результаты внедрения многофакторного анализа ролей пользователей в технологию построения аналитических отчетов поисковой системы SEUS. Дается описание технологии создания прототипа информационной системы с использованием нейросетевого пакета Нейросимулятор 5.0, пакета визуального программирования Orange Data Mining, табличного процессора MS Office 2017 и графического пакета Gephi.

На основании положительных результатов использования прототипа принято решение разработки программного приложения, предназначенного для промышленной эксплуатации.

Представлены описание функций и выходных данных приложения «Сервис анализа распространения контента». UML схема имплементации приложения в общую структуру поисковой системы SEUS представлена на рисунке 6. Сервис поддерживает экспорт 10 различных файлов формата *.csv:

- Коллекция публикаций и обсуждений;

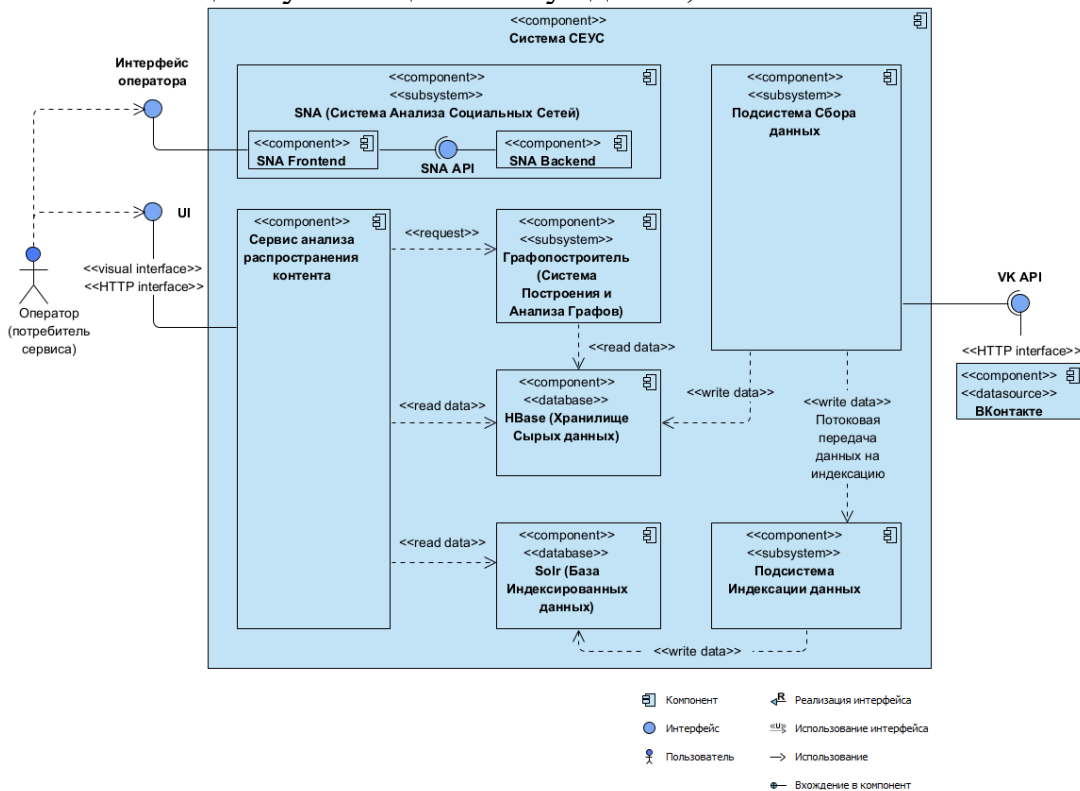


Рисунок 6 - Архитектура сервиса анализа распространения контента

- Выборка пользователей;
- Массив выявленных связей распространения между публикациями и обсуждениями;
- Массив выявленных социальных связей между пользователями;
- Массив ключевых пользователей по ролям постера, репостера, комментатора и универсала;
- Массив ключевых постеров;
- Массив ключевых репостеров;
- Массив ключевых комментаторов;

- Массив ключевых универсалов;
- Массив ключевых мостов.

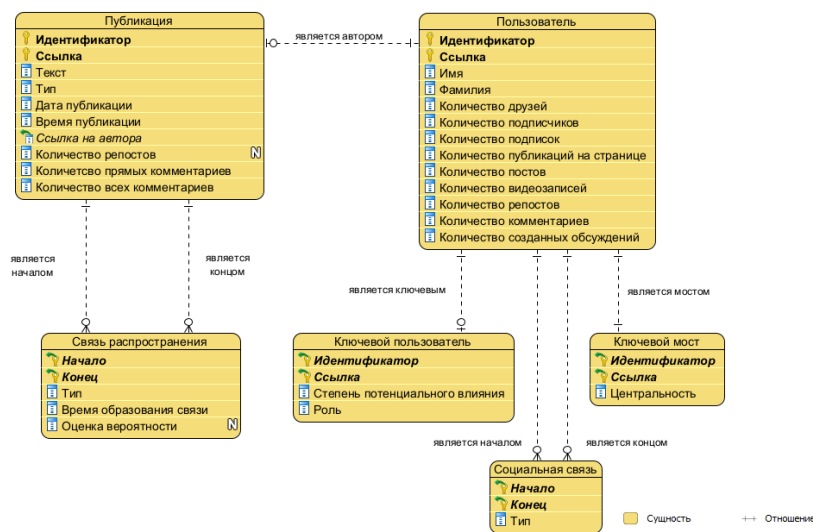


Рисунок 7 - Логическая модель выходных данных

Модель отношений между экспортируемыми из Сервиса сущностями приведена на рисунке 7 (ER-диаграмма в нотации Crow's foot). Далее представлены результаты внедрения многофакторного анализа ролей пользователей и уровня их информационного влияния. Основные полученные технико-экономические показатели отражены в акте внедрения и в разделе

«практическая значимость работы».

III. ЗАКЛЮЧЕНИЕ

В рамках проведенного исследования были получены следующие результаты, обладающие научной и практической значимостью:

1. Предложенный экспертный способ подготовки синтетических множеств данных для обучения и тестирования нейросетевого классификатора ролей пользователей социальных сетей позволил приблизительно в 1000 раз сократить затраты на создание датасета и исключить риск утечки конфиденциальной информации.

2. Предложенный метод идентификации пользователей, выполняющих роль мостов между ядром социальной сети и изолированными кластерами пользователей на основе расчета метрики ЦВВ, позволил выявить пользователей, играющих важную роль в распространении информации и увеличении охвата аудитории социальной сети, вовлеченной в социальное явление, которые ранее технически не могли быть обнаружены в поисковой системе «SEUS».

3. Предложенный метод идентификации НВП на основе учета публикационной активности пользователей и количества их социальных связей позволил повысить релевантность выявления НВП и снизить трудозатраты аналитиков при проведении исследований в 10,7 раза, а при проведении более детальных исследований, учитывающих структуру воздействия на сеть, в 42,8 раз.

4. Внедрение программного продукта «Сервис анализа распространения контента» позволило увеличить выручку ООО «СЕУСЛАБ» за 2021 год на 12% за счет поступлений оплаты лицензий.

Комплексное применение всех предложенных методов позволило выявить важную информацию об элементах структуры воздействия на сеть, которая может быть использована для поддержки принятия решения о воздействии на

пользователей социальной сети в контексте противодействия протестным движениям в социальных сетях.

IV. ОСНОВНЫЕ ПУБЛИКАЦИИ

Публикации в журналах, включенных в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени кандидата наук

1. **Rabchevsky, A. N.** Comparison of methods for identifying user roles in online social networks = Сравнение методов идентификации ролей пользователей социальных сетей / A. N. Rabchevsky, L. N. Yasnitsky, V. S. Zayakin. – текст: непосредственный. – DOI: 10.15593/2499-9873/2021.2.06 // Прикладная математика и вопросы управления / Applied Mathematics and Control Sciences. – 2021. – № 2. – С. 93–111. – Ст. на англ. языке.

2. **Рабчевский А. Н.** Выявление мостов в кластерных сетях и оценка уровня их информационного влияния / **А. Н. Рабчевский**, В. С. Заякин, Е. А. Рабчевский. – текст: непосредственный // Информационные системы и технологии. – 2021. – № 5 (127). – С. 21–30.

3. **Рабчевский А. Н.** Оценка потенциального уровня информационного влияния пользователей в социальных сетях / **А. Н. Рабчевский**, Е. А. Рабчевский. – текст: непосредственный // Информационные системы и технологии. – 2022. – № 1 (129). – С. 114–122.

Публикации в изданиях, индексируемых в Scopus

4. **Rabchevsky A. N.** Modelling the structure of protest movement advocacy in social media using graph and neural network analysis = Моделирование структуры пропаганды протестного движения в социальных сетях с помощью анализа графов и нейронных сетей / **A. N. Rabchevsky**, E. G. Ashikhmin, E. A. Rabchevsky. – текст: электронный. – DOI: 10.1007/978-3-030-89477-1_1 // Lecture Notes in Networks and Systems. – 2022. – Vol. 342. – P. 3–15. – Ст. на англ. языке.

5. **Rabchevskiy A. N.** Creating and using synthetic data for neural network training, using the creation of a neural network classifier of online social network user roles as an example = Создание и использование синтетических данных для обучения нейронной сети, на примере создания нейросетевого классификатора ролей пользователей онлайн-социальных сетей / **A. N. Rabchevskiy**, L. N. Yasnitskiy. – текст: электронный. – DOI: 10.1007/978-3-030-93677-8_36 // Lecture Notes in Networks and Systems. – 2022. – Vol. 381. – P. 412–421. – Ст. на англ. языке.

Свидетельства о регистрации программ для ЭВМ и баз данных

6. Свидетельство о государственной регистрации программы для ЭВМ № 2021616086 Программа вычисления мостов в кластерных сетях / **А. Н. Рабчевский**, В. С. Заякин. Заявка 2021615157, поступ. 13.04.2021. опубл. 16.04.2021; Бюл. №4. – 1 с.

7. Свидетельство о государственной регистрации базы данных № 2021621533 База данных для классификации ролей пользователей социальных сетей / **А. Н. Рабчевский**, В. С. Заякин. Заявка 2021621400, поступ. 07.07.2021. опубли. 15.07.2021; Бюл. №7. – 1 с.

Публикации в прочих изданиях

8. **Рабчевский, А.Н.** Нейросетевая система классификации пользователей социальных сетей и экспертный способ ее создания / А. Н. Рабчевский. – текст: непосредственный // Нейрокомпьютеры и их применение: тез. XVIII Всерос. науч. конф., Москва, 17 марта 2020 г. – 2020. – С. 298–299.

9. Рабчевский, Е.А. Экспертный способ формирования обучающих выборок на примере создания нейросетевой системы классификации пользователей социальных сетей / Е. А. Рабчевский, **А. Н. Рабчевский**, Л. Н. Ясницкий. – текст: непосредственный. –DOI: 10.18127/j19998554-202005-05 // Нейрокомпьютеры: разработка, применение. – 2020. – Т. 22. – № 5. – С. 54–63.

10. Экспертная система для автоматического выявления ролей пользователей в социальных сетях на основе нейросетевых технологий / **А. Н. Рабчевский**, Е. А. Рабчевский, В. С. Заякин, Л. Н. Ясницкий. – текст: непосредственный // Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сб. статей по мат. Междунар. конф. и VI всерос. науч.-практ. конф.; под редакцией Л.Н. Ясницкого. (Пермь, 12–18 октября 2020 г.). – 2020. – С.118 – 127.

11. **Рабчевский, А.Н.** Применение нейро-сетевой фильтрации для оптимизации алгоритмов выявления наиболее влиятельных узлов в социальных сетях / **А. Н. Рабчевский**. – текст: непосредственный // Нейрокомпьютеры и их применение: тез. XIX Всерос. науч. конф., Москва, 30 марта 2021 г. – 2021. – С.158–159.

12. Рабчевский, Е.А. О некоторых аспектах структур пропаганды политического протеста в социальных сетях / Е. А. Рабчевский, **А. Н. Рабчевский**. – текст: непосредственный // Деятельность террористических и экстремистских организаций, направленная на насильственное изменение конституционного строя, как угроза государственной и общественной безопасности стран СНГ. – М.: АТЦ СНГ. – 2021. – С. 193–211.