

Федеральное государственное автономное образовательное учреждение  
высшего образования  
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

*На правах рукописи*

УДК 519.7

**РАБЧЕВСКИЙ Андрей Николаевич**

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

05.13.10 - Управление в социальных и экономических системах

**ДИССЕРТАЦИЯ**  
на соискание ученой степени  
кандидата технических наук

Научный руководитель:  
доктор технических наук,  
профессор Ясницкий Л.Н.

Пермь  
2022

## ОГЛАВЛЕНИЕ

<b>ВВЕДЕНИЕ.....</b>	<b>4</b>
<b>ГЛАВА 1. ВЫЯВЛЕНИЕ НАИБОЛЕЕ ВЛИЯТЕЛЬНЫХ ПОЛЬЗОВАТЕЛЕЙ .....</b>	<b>13</b>
1.1. Влияние социальных сетей на поведение людей .....	13
1.2. Исследование методов выявления наиболее влиятельных пользователей .....	16
1.2.1. Теоретико-графовый подход.....	16
1.2.2. Теоретико-игровой подход.....	19
1.2.3. Акциональная модель .....	21
1.2.4. Сочетание различных подходов для выявления НВП .....	22
1.2.5. Выявление пользователей, вовлеченных в социальное явление .....	23
1.3. Расчет уровня активности пользователей.....	25
1.4. Метод выявления наиболее влиятельных пользователей .....	26
1.5. Выявление наиболее влиятельных пользователей методом диаграммы Венна.....	37
1.6. Сравнение методов выявления наиболее влиятельных пользователей ...	39
1.7. Выводы по главе 1 .....	43
<b>ГЛАВА 2. ВЫЯВЛЕНИЕ СОЦИАЛЬНЫХ РОЛЕЙ ПОЛЬЗОВАТЕЛЕЙ ..</b>	<b>45</b>
2.1. Обзор методов выявления ролей пользователей.....	45
2.2. Описание ролей пользователей.....	49
2.3. Выявление ролей пользователей методом численной оценки .....	50
2.4. Экспертная нейросетевая классификация ролей пользователей .....	54
2.4.1. Входные и выходные параметры нейросети .....	57
2.4.2. Синтезирование датасета .....	57
2.4.3. Проектирование нейронной сети .....	63
2.5. Выявление ролей пользователей методом кластеризации.....	66
2.6. Выявление наиболее влиятельных пользователей в каждой роли.....	67
2.7. Выводы по главе 2 .....	69
<b>ГЛАВА 3. ВЫЯВЛЕНИЕ МОСТОВ В КЛАСТЕРНЫХ СЕТЯХ.....</b>	<b>71</b>
3.1. Существующие методы выявления мостов в кластерных сетях .....	74
3.2. Метод выявления мостов в ярко выраженных кластерных сетях.....	76

3.2.1. ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДА ВЫЯВЛЕНИЯ МОСТОВ .....	80
3.3. ИДЕНТИФИКАЦИЯ НАИБОЛЕЕ ВЛИЯТЕЛЬНЫХ МОСТОВ .....	88
3.4. МОДЕЛИРОВАНИЕ СТРУКТУРЫ ВРЕМЕННОЙ ПРОТЕСТНОЙ СОЦИАЛЬНОЙ СЕТИ .....	90
3.5. ВЫВОДЫ ПО ГЛАВЕ 3 .....	94
<b>ГЛАВА 4. ВНЕДРЕНИЕ МНОГОФАКТОРНОГО АНАЛИЗА ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ .....</b>	<b>95</b>
4.1. ПРОТОТИП СИСТЕМЫ ВЫЯВЛЕНИЯ НАИБОЛЕЕ ВЛИЯТЕЛЬНЫХ ПОЛЬЗОВАТЕЛЕЙ, МОСТОВ И ОПРЕДЕЛЕНИЯ УРОВНЯ ИХ ИНФОРМАЦИОННОГО ВЛИЯНИЯ .....	95
4.2. ПРИЛОЖЕНИЕ «СЕРВИС АНАЛИЗА РАСПРОСТРАНЕНИЯ КОНТЕНТА» .....	97
4.2.1. СЕРВИС АНАЛИЗА РАСПРОСТРАНЕНИЯ КОНТЕНТА. ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ .....	98
4.2.2. СЕРВИС АНАЛИЗА РАСПРОСТРАНЕНИЯ КОНТЕНТА. ТРЕБОВАНИЯ К ВЫХОДНЫМ ДАНЫМ .....	101
4.3. ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДА МНОГОФАКТОРНОГО АНАЛИЗА ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ .....	105
4.4. ВЫВОДЫ ПО ГЛАВЕ 4 .....	107
<b>ЗАКЛЮЧЕНИЕ .....</b>	<b>108</b>
<b>СПИСОК ЛИТЕРАТУРЫ .....</b>	<b>114</b>
<b>ПРИЛОЖЕНИЕ 2 .....</b>	<b>127</b>
<b>ПРИЛОЖЕНИЕ 3 .....</b>	<b>128</b>
<b>ПРИЛОЖЕНИЕ 4 .....</b>	<b>129</b>
<b>ПРИЛОЖЕНИЕ 5 .....</b>	<b>130</b>
<b>ПРИЛОЖЕНИЕ 6 .....</b>	<b>131</b>
<b>ПРИЛОЖЕНИЕ 7 .....</b>	<b>132</b>
<b>ПРИЛОЖЕНИЕ 8 .....</b>	<b>133</b>
<b>ПРИЛОЖЕНИЕ 9 .....</b>	<b>134</b>
<b>ПРИЛОЖЕНИЕ 10 .....</b>	<b>135</b>
<b>ПРИЛОЖЕНИЕ 11 .....</b>	<b>136</b>
<b>ПРИЛОЖЕНИЕ 12 .....</b>	<b>137</b>

## ВВЕДЕНИЕ

**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, использующиеся 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 научной разработанности темы.** В изучение проблемы идентификации ролей пользователей существенный вклад внесли Е. Глив [1], Дж. Фюллер [2], О. Аразий [3], Х. Велсер [4], 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 [5], М. Чичек [6], А. Аруларазан [7], П. Руас [8] и другие ученые и эксперты применяли различные методы кластеризации, в то время как М. Яблоньска [9], Х. Лин [10], Ц. Сегалин [11], П. Вийдженайке [12], К. Матсумото [13], М. Сунгхван [14] 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 [15–18], С. И. Николенко [19], Л. Н. Ясницкий [20–23] 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 [24], К. Джанг [25], З. Галман [26], П. Йенсен [27], А. Альварес-Сокорро [28], 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 [29, 30], П. Кэррингтон, Дж. Скотт и С. Вассерман [31], М. Кастельс [32], Д. Фэрис [33], П. Гербаудо [34], Л. Лю [35], Д. Чен [36], Д. А. Губанов и А. Г. Чхартишвили [37–40], Б. А. Торопов [41–43] 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» [44], 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» [45], XIX всероссийской научной конференции «Нейрокомпьютеры и их применение» [46], 2nd International Conference on Cyber-Physical Systems & Control, The 2021 International Conference on Digital Science [47], Международном форуме «Наука и глобальные вызовы XXI века» [48], 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 [49–51], в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, 2 работы в изданиях, индексируемых в Scopus [47, 48], а также получено 1 свидетельство о государственной регистрации программы для ЭВМ [52] и 1 свидетельство о государственной регистрации базы данных [53].

**Структура и объем работы.** Диссертация состоит из следующих элементов:

**Введение** содержит обоснование актуальности исследуемого вопроса, приведены объект, предмет, поставлена цель и сформулированы 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.

**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.

**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 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. 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.

Каждая глава оканчивается выводами, полученными в ходе ее изложения. Общие выводы по исследованию сформулированы в **Заключении**.

# ГЛАВА 1. ВЫЯВЛЕНИЕ НАИБОЛЕЕ ВЛИЯТЕЛЬНЫХ ПОЛЬЗОВАТЕЛЕЙ

## 1.1. Влияние социальных сетей на поведение людей

Современный уровень развития цифровых технологий привел к тому, что социальные сети прочно вошли в повседневную жизнь и стали оказывать влияние на поведение людей. Наиболее деструктивными являются такие виды влияния, которые провоцируют пользователей социальных сетей на совершение противоправных или социально опасных действий в реальной жизни. Примером такого влияния могут являться политические события Арабской весны в 2010-2011 гг., движение #Оссурау в США в 2011 году, протесты в Турции, Бразилии и Гонконге (2013-2014 гг.), а также цветные революции на постсоветском пространстве, в которых социальные сети использовались для координации людей в реальных политических акциях. Самыми яркими из последних событий являются использование социальных сетей для организации протестов в Республике Беларусь в связи с выборами президента в 2020 г. и протесты вокруг ареста Навального и так называемого «Дворца Путина» в 2021 г. Изучение механизмов и степени влияния социальных сетей на поведение людей вызвало большой научный интерес. В работах [32–34] авторы утверждают, что все протестные движения неразрывно связаны с созданием автономных сетей связи, поддерживаемых Интернетом. Существенное влияние социальных сетей на уровень мобилизации людей к активным действиям описано в [54], а в работе [55] исследуется организационная динамика, которая возникает, когда коммуникация становится важной частью организационной структуры протестного движения. Использование распределенных неорганизованных социальных сетей, по мнению [56], способствует распространению информации среди протестных групп через межличностные связи и повышает возможности мобилизации участников на конкретные акции. Авторы [57], анализируя протестные тенденции в России по состоянию на 2012 г., приходят к выводу о

повышении эффективности протестов при построении их по сетевому принципу. Авторами отмечается, что при развитии протеста через сеть Интернет увеличивается скорость его распространения, в т.ч. за границы самого государства, что является угрозой нарушения суверенитета. В тоже время, если вначале социальные сети использовались как стихийная неструктурированная среда, ускоряющая распространение информации, то в последнее время стало очевидно, что на социальные сети специально оказываются деструктивные воздействия, чтобы использовать их как эффективный инструмент мобилизации пользователей для участия в протестных акциях. Яркий пример такого воздействия, это протестные акции на выборах президента республики Беларусь, которые организовывались через социальные сети с территории Польши и Литвы.

Каждое социальное явление находит свое отражение в социальных сетях, а социальные сети влияют на эти социальные явления. Это влияние может восприниматься как информационные волны [58], вызванные различными информационными поводами. В качестве информационных поводов могут быть реальные события в общественной жизни или информационные события внутри социальной сети. Они могут носить характер эха реальных событий или создавать информационную основу для реализации информационных событий в реальной жизни. Очевидно, что в целях обеспечения безопасности общества требуется уметь противостоять информационным волнам, уметь создавать их и управлять ими.

Средой для распространения информационных волн являются социальные сети, в недрах которых создаются временные или виртуальные социальные сети, порожденные конкретным социальным явлением, существующие в течение жизни самого социального явления и состоящие из пользователей социальных сетей, которые активно участвуют, поддерживают, интересуются или были бы готовы участвовать в социальном явлении при определенных условиях.

Протестные настроения являются ярким примером устоявшегося социального явления, в котором постоянно появляются информационные

поводы, сопровождающиеся информационными волнами. При решении задач противодействия информационным волнам в протестном движении необходимы ответы на несколько важных вопросов:

- что или кто является источником информационной волны,
- как она распространяется,
- кто или что способствует ее поддержанию или усилению,
- кто или что способствует расширению охвата аудитории,
- кто или что имеет наибольший уровень информационного влияния.

Сумев ответить на эти вопросы, можно управлять скоростью распространения информации, уровнем внимания аудитории, охватом аудитории и т.д. Результатом такого управления может быть снижение или увеличение общественной значимости самого информационного повода и изменение общественного мнения о нем.

Базовой является проблема определения уровня информационного влияния конкретных пользователей, сообществ или сетевых структур. Применение методик для решения данной проблемы облегчает выявление ключевых пользователей, воздействие на которых может изменять параметры распространения информационных волн. В то же время, существующие методики недостаточно эффективно выявляют ключевых пользователей, что вызывает необходимость анализа большого количества профилей пользователей, усложняет работу правоохранительных органов и снижает эффективность их работы по противодействию деструктивным воздействиям на социальные сети. Таким образом, разработка эффективной методики выявления ключевых пользователей и уровня их информационного влияния является весьма актуальной задачей.

## **1.2. Исследование методов выявления наиболее влиятельных пользователей**

Решению задачи выявления наиболее влиятельных пользователей (НВП) в социальных сетях посвящено множество работ, среди которых наиболее ярко представлены несколько основных подходов:

- Теоретико-графовый,
- Теоретико-игровой,
- Акциональный.

Теоретико-графой подход базируется на представлении сети в виде графа и изучении возможности вершин и ребер графа к передаче информации между собой.

Теоретико-игровой подход так же анализирует структуру сети, представляя вершины графа в виде агентов, наделенных определенным набором свойств, и рассматривает возможность объединения вершин графа в коалиции или варианты их конкурентного поведения.

Акциональный подход основан на анализе активности пользователей в сети, считая тем выше уровень влияния пользователя, чем выше степень его активности в сети.

### **1.2.1. Теоретико-графовый подход**

В теоретико-графовом подходе предполагается, что социальная сеть подобна графу, где каждый пользователь является вершиной графа, а его связи с другими пользователями сети, это ребра графа. Для определения наиболее влиятельных узлов используются методы графового анализа и различные метрики центральности.

В 1948 году А. Bavelas [59] впервые применил и математически описал термин «кратчайшие пути» как характеристики взаимодействия между людьми и группами людей, а чуть позднее он же [60], решая проблему сравнения



различных паттернов связей людей в группах, ввел понятие «Относительной центральности» («relative centrality») как отношение суммы всех внутренних путей между членами группы к сумме путей от конкретных людей.

A. Shimbel [61] предложил свою метрику для определения влиятельных узлов в сети и назвал ее «Метрикой напряженности» (measure of the “stress”). Автор предположил, что основной мерой влиятельности узла в сети должна быть сумма кратчайших путей, проходящих через узел, и чем больше значение этой суммы, тем более влиятельным является узел в сети.

A. Flament [62], обсуждая работу A. Bavelas [60], привел примеры сетей разной структуры, имеющих одинаковые метрики относительной центральности, доказав их несовершенство.

M. A. Beauchamp [63] пошел дальше, предложив улучшенный «Индекс центральности», который, как утверждается, является улучшением по сравнению с индексом Бавеласа.

G. Sabidussi [64] проанализировал предыдущие работы [60, 62, 63] и указал на то, что ни одна из метрик не может считаться лучшей чем другие и предложил свою метрику «the trivial centrality index», так как она вполне адекватна, но не так сложна для вычисления.

Большое количество авторов ссылается на L. S. Freeman, который проанализировав предыдущие работы [29], предложил новое семейство метрик центральности [30], основанных на трех разных концепциях: степени, промежуточности и близости.

- «Центральность по степени» это метрика, учитывающая количество связей узла и показывающая уровень его коммуникативной активности.
- «Центральность по промежуточности» это метрика, пропорциональная количеству пар узлов в сети, которые может соединить конкретный узел. Эта метрика является показателем потенциала для контроля над коммуникацией.

- «Центральность по близости» это метрика, пропорциональная количеству кратчайших путей до других пользователей, поэтому она является показателем независимости или эффективности.

Эти метрики легли в основу метода определения выявления влиятельных узлов и, поскольку центральность по промежуточности и центральность по близости требуют заранее знать всю информацию о топологии сетей и не могут применяться к крупномасштабным сетям, стали предметом дальнейших многочисленных исследований.

В развитие метрик центральности по близости и промежуточности, в работе [65] предлагается учитывать сумму информации, которая содержится во всех возможных путях между парами точек.

Комбинация метрик центральности по степени и промежуточности, предложенная [66], позволяет существенно повысить релевантность определения ключевых узлов в сети. Однако, эта мера очень трудоемкая.

В работе [67] делается основной упор на топологию сети и утверждается, что возможны случаи, когда узлы с высокими метриками центральности могут находиться на периферии сети и их влияние на сеть существенно ниже узлов, расположенных в ядре сети. Для того, чтобы определить ядро и периферию сети, авторы предлагают использовать метод K-shell. Расположение узла определяется с помощью анализа декомпозиции K-shell. Этот процесс присваивает целочисленный индекс или центральность  $K_s$  каждому узлу, представляя его местоположение в соответствии с последовательными уровнями (K-shell) в сети. Низкие значения  $K_s$  определяют периферию сети, а самое внутреннее ядро сети соответствует большому  $K_s$ . Однако этот метод не позволяет учесть все узлы в сети и для решения этой проблемы в работе [68] были предложены меры усовершенствования метода K-shell, позволяющие учитывать все узлы в сети.

С бурным развитием социальных сетей возникла необходимость решать задачи определения влиятельных узлов в реальных сетях за приемлемое время, в связи с чем наряду с повышением релевантности методов возникла проблема

повышения эффективности алгоритмов и снижения их вычислительной сложности.

В работе [36] авторы предлагают полу-локальную меру центральности как компромисс между низко релевантной центральностью по степени и другими трудоемкими методами. Различные варианты решения проблемы предложены [69–72] и др.

Для повышения скорости решения задач так же применяются методы случайного блуждания, которые очень легко могут выполняться параллельно несколькими компьютерами. Методы выявления влиятельных узлов на основе случайного блуждания в основном используются при сортировке веб-страниц. Типичными методами являются алгоритм Клейнберга HITS [73], алгоритм Google PageRank [74] и алгоритм LeaderRank [75]. Подробный обзор методов выявления жизненно важных узлов в сети представлен в работе [35], в которой авторы не только описывают наиболее значимые теоретико-графовые методы, но и приводят примеры применения различных подходов для решения различных задач.

### **1.2.2. Теоретико-игровой подход**

Различные метрики центральности рассматривают узлы сети как объекты, а связи между ними как пути передачи информации. Теоретико-игровой подход предлагает рассматривать узлы сети не только как объекты, но и как субъекты, наделенные определенными свойствами и способные объединяться в коалиции или конкурировать между собой. Использование такого подхода позволяет сделать разницу между более влиятельными и менее влиятельными узлами сети более контрастной.

Алгоритм расчета центральности участников сети на основе значения Шепли предложен [43]. Результаты такого расчета учитывают способность участников сети к кооперации для достижения поставленных целей. При этом сам алгоритм достаточно гибок и позволяет оценивать значимость каждого из

участников на основе всей совокупности его качественных и количественных характеристик.

Авторами [42] предлагается дополнить центральность по близости конкретных узлов центральностью по близости нескольких узлов, объединенных в коалицию. Такие объединения имеют большие метрики центральности по близости чем отдельные узлы, а значит имеют более высокую значимость для распространения информации в сети.

В работе А. Суходолова и др. [76] с помощью математических методов показано, что задача нахождения максимально влиятельной группы из  $k$  участников не решается оптимальным образом при включении в эту группу  $k$  наиболее влиятельных по отдельности участников. Для решения задачи также предлагается использование теоретико-игровой концепции — вектора Шепли, который позволяет оценить индивидуальный вклад каждого в формируемую группу, т. е. выяснить, в какой мере каждый может быть полезен как командный игрок.

Б.А. Торопов [41] показывает, что при конкурентном распространении влияний в социальных сетях можно предложить метрики центральности, позволяющие игрокам (интересантам распространения конкурирующих влияний) отбирать в графе вершины – инициаторы влияний, таким образом, что получаемые решения будут максимально приближены к решениям «игры Вороного». Автор предлагает использовать метрики «Конкурентная центральность по близости» и «Изолирующая центральность», и указывает на актуальность поиска эффективных методов и алгоритмов расчета предложенных метрик, а также поиска на графе подмножеств вершин, обладающих наибольшими групповыми значениями конкурентной центральности по близости и изолирующей центральности.

Одним из методов использования кооперативной теории игр является вычисление значений Майерсона как меры центральности вершин в графе [77]. В этом случае в качестве характеристической функции используется число путей определенной длины в подграфах, соответствующих коалициям. В работе

предложена модификация вектора Майерсона для случая, когда в рассмотрение включаются пути в графе, содержащие циклы.

### **1.2.3. Акциональная модель**

Если рассматривать узлы сети как субъекты, участвующие в распространении информации, тогда влияние узлов должна зависеть от их действий, уровня их активности и реакции на их действия.

В работах [38, 39, 78] предложена акциональная модель как новый подход к конструктивному определению влияния пользователей онлайн-социальных сетей, в соответствии с которым влияние вычисляется на основе действий пользователей с учетом установок управляющего органа (центра).

В контексте развития акциональной модели, в докладе [40] предложено вычислять влияние не только пользователей, но и подмножеств пользователей. Влияние и влияние пользователей и подмножеств пользователей также вычисляются на основе реакции на их действия с учетом приоритетов управляющего органа.

Результатом исследований в этом направлении является программа [79], подтвержденная свидетельством о регистрации.

Несколько иной подход демонстрируется в работе [80], где авторы, основываясь на [81] предлагают определять уровень влияния пользователей исходя из их социального капитала, который учитывает количество подписчиков, принятие тематики контента пользователями (количество лайков) и отзывчивость пользователей на контент (количество комментариев). Таким образом, по мнению авторов, пользователи с наибольшим социальным капиталом имеют наибольшее влияние в сети.

#### 1.2.4. Сочетание различных подходов для выявления НВП

Все представленные выше подходы и методы в той или иной мере решают проблему выявления НВП в социальных сетях, но эффективность их применения зависит от типа решаемых задач.

Классический теоретико-графовый метод пытается спрогнозировать процесс распространения контента в глобальной сети, не учитывая ни характеристик самих пользователей, как субъектов, ни параметров передаваемого контента.

Теоретико-игровой метод также решает задачу прогнозирования распространения контента с учетом конкурентного влияния и учитывает такие характеристики пользователей, как возможность вступлений в коалицию.

В обоих случаях исследуются возможности распространения контента по всей доступной сети.

С другой стороны, акциональная модель, пытается выявить наиболее влиятельных пользователей на основе параметров их сетевой активности, то есть по количеству опубликованных материалов.

Однако, высокий уровень активности пользователя без большого количества связей в сети не приводит к высокому влиянию на пользователей в сети. Образно говоря, пользователь громко кричит, но его никто не слышит. С другой стороны, высокие значения метрик центральности при низкой активности также не приводит к высокому уровню влияния пользователей, на что указывают авторы работ [82, 83], поэтому в работах [84, 85] предлагается использовать комплексный учет структурных характеристик и параметров активности пользователей. По мнению авторов работы [86], наиболее оптимальным является использование комбинации акционального и теоретико-графового подходов, поэтому они представляли активность пользователей в виде ориентированного взвешенного графа и затем рассчитывали метрики центральности для каждого пользователя. Действительно, сочетание параметров активности пользователей и их социальных связей более адекватно определяет уровень информационного

влияния пользователей в социальной сети, но в отличие от [87] и других авторов, использующих сложные алгоритмы учета взаимного влияния пользователей, в данной работе был использован более простой и прозрачный метод.

### **1.2.5. Выявление пользователей, вовлеченных в социальное явление**

Как уже было сказано выше, каждое социальное явление порождает в социальной сети виртуальную или временную социальную сеть, которая состоит из пользователей, вовлеченных в это явление в виде публикационной активности. Выявление вовлеченных пользователей выполняется с помощью поисковой системы SEUS, разработанной компанией ООО «СЕУСЛАБ» и широко применяемой сотрудниками правоохранительных органов [88]. Для этого используется контент-анализ социальной сети за определенный интервал времени на определенной географической территории. Признаком вовлеченности является публикация пользователями в социальной сети материалов (постов, репостов, комментариев) целевой тематики. Для поиска материалов используется набор ключевых слов – термов, которые характерны и часто употребляются в исследуемом социальном явлении. Составление наборов ключевых слов (словарей), это очень ответственная процедура, так как от качества словаря зависит релевантность выборки публикаций. Обычно составлением словарей по каждому социальному явлению занимаются эксперты-аналитики.

После того, как словарь составлен, поисковая система выполняет поиск всех публикаций, совершенных в указанный интервал времени, на указанной географической территории, использующих термы, содержащиеся в словаре. Результат выдается в виде многомерного массива с информацией о публикациях и пользователях. Полученная выборка включает посты, репосты, комментарии, видео и ссылки с текстовым содержанием. Пример содержания файла выборки представлен в приложении 1.

Проверка релевантности выборки исследуемому социальному явлению выполняется также средствами поисковой системы SEUS, которая строит график зависимости количества публикаций и вовлеченных пользователей по времени и затем производится проверка соответствия графика реальным событиям в исследуемом социальном явлении. В качестве примера такой проверки на релевантность поиска на рис. 1 и 2 представлено количество пользователей, создавших публикации по протестной тематике, и количество публикаций с распределением по шкале времени в московском регионе в течение 2019 года.

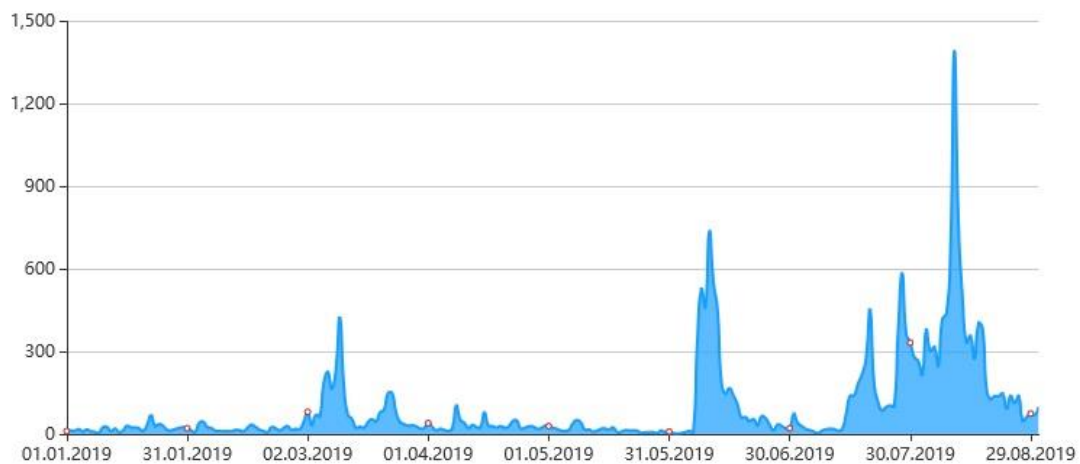


Рисунок 1 - Количество пользователей, вовлеченных в протестное движение

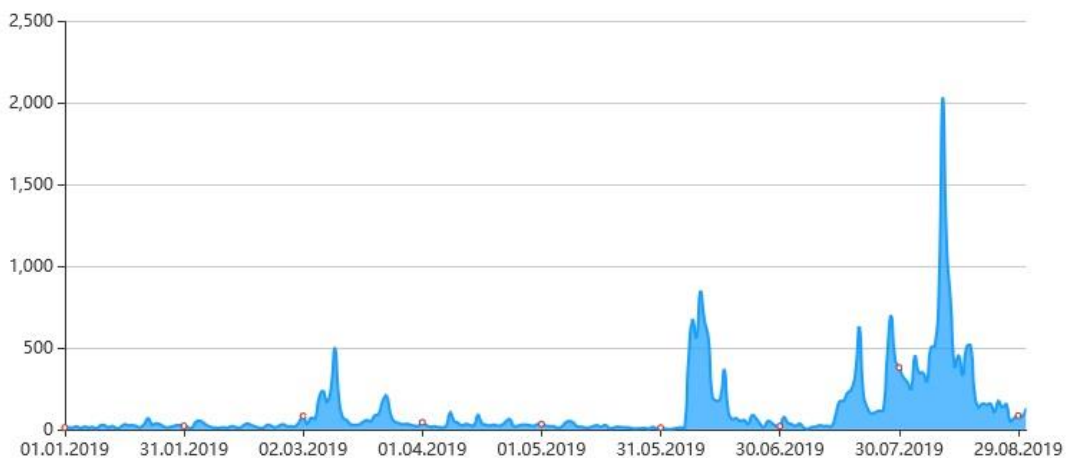


Рисунок 2 - Количество публикаций протестной тематики



Пикам графика соответствуют наиболее крупные протестные акции, которые проходили в Москве в этот период, что говорит об удовлетворительном качестве составленного словаря.

После того, как поисковая система выдала данные обо всех пользователях, совершивших публикации в пределах указанных ограничений по тематике, времени и месту, можно приступать к дальнейшим исследованиям.

### 1.3. Расчет уровня активности пользователей

Далее производится расчет уровня активности пользователей в социальном явлении также с использованием поисковой системы SEUS, которая автоматически считает количество постов, репостов и комментариев, совершенным каждым вовлеченным пользователем и сортирует пользователей по количеству опубликованных материалов и, соответственно, по уровню их активности.

Результаты расчета уровня активности представляются в графическом виде в онлайн приложении, как показано на рисунке 3, и в виде файла рейтинга активности, который может быть экспортирован во внешнее приложение.

The screenshot displays the SEUSLAB application interface. On the left, there is a sidebar with search parameters and filters. The main area shows a list of users ranked by activity level. The top user is 'Алексей Добрый' with a score of 190. Below the list, a detailed profile for 'Алексей Добрый' is shown, including personal information, account details, and a log of published posts with timestamps.

Пользователь	Рейтинг
Алексей Добрый	190
Арви Хэкер	156
Александр Гришаков	106
Юра Шувалов	99
Владимир Новиков	88
Валерий Юрцев	92
Юлия Скурвовская	91

Время	Тип активности
16.06.2019 00:30	Опубликовал пост
12:26	Опубликовал пост
00:46	Опубликовал пост
12:00	Опубликовал пост
07:00	Опубликовал пост
05:33	Опубликовал пост
03:03	Опубликовал пост
01:06	Опубликовал пост
21:30	Опубликовал пост
20:56	Опубликовал пост
20:06	Опубликовал пост
19:50	Опубликовал пост
19:16	Опубликовал пост

Рисунок 3 - Графическое представление рейтинга активности поисковой системы SEUS

Экспортировав рейтинг активности из поисковой системы SEUS, получаем таблицу рейтинга активности для каждого пользователя, где указан идентификационный номер пользователя ID и уровень его активности в социальном явлении. Пример файла рейтинга активности представлен в приложении 2.

Расчет рейтинга активности пользователей в социальном явлении дает информацию о том, насколько активен пользователь в социальной сети и как часто он публикует материалы по целевой тематике. По этим данным уже можно выявлять пользователей, которые играют наиболее заметную роль в социальном явлении, однако, работа с этими данными предполагает ручную обработку и для устоявшихся социальных явлений может потребовать большого количества ресурсов.

#### **1.4. Метод выявления наиболее влиятельных пользователей**

Для целей данного исследования, используя поисковую систему SEUS [89], на основе контент-анализа были собраны данные о пользователях, опубликовавших материалы в социальной сети ВКонтакте по информационным поводам «Жыве Беларусь» (выборы президента республики Беларусь в 2020 году) и «Дворец Путина» в начале 2021 года, а также данные о количестве постов, репостов и комментариев, опубликованных этими пользователями по указанной тематике. Таким образом, были выявлены пользователи, которые вовлечены в целевую тематику и составляют временную протестную социальную сеть.

Для определения социальных связей выявленных пользователей были собраны данные об их друзьях и друзьях их друзей. Такой выбор объясняется тем, что количество связей очень сильно зависит от того, с кем именно связан пользователь. При одинаковом количестве друзей на первом колене, количество друзей на втором колене может различаться на порядки. В тоже время,

дальнейшее увеличение колен в графе не целесообразно, так как приводит к нивелированию разницы в количестве связей. Необходимо также учитывать объем вычислительных ресурсов, необходимых для сбора графовой информации, построения графов и вычисления метрик центральности. С каждым новым коленом вычислительная сложность возрастает на несколько порядков, поэтому 2-х коленный граф является оптимальным с точки зрения выявления различий между пользователями и вычислительной сложностью. Статистические данные по обоим инфоповодам представлены в таблице 1.

Таблица 1 – Общие статистические данные выборки по пользователям, вовлеченным в тематику «Жыве Беларусь» и «Дворец Путина»

№ п/п	Параметр	«Жыве Беларусь»	«Дворец Путина»
1	2	3	4
1	Количество вовлеченных пользователей	28 447	35 502
2	Количество опубликованных материалов	42 148	61 967
3	Количество друзей и друзей их друзей	18 508 073 467	4 161 468 616

По каждой выборке был сформирован рейтинг публикационной активности путем сортировки пользователей по количеству опубликованных ими материалов. Рейтинг публикационной активности по информационному поводу «Дворец Путина» представлен на рисунке 4.



Рисунок 4 - Рейтинг публикационной активности по теме «Дворец Путина»

Как видно из графика, из 35 тыс. пользователей только 500 имеют значимое количество публикаций, что составляет 1,4% от общего количества пользователей в рейтинге. В том числе, из них только первые 100 пользователей имеют уровень активности выше 20 публикаций. Оставшиеся 35,4 тыс. пользователей имеют крайне низкую активность. На рисунке 5 представлен график рейтинга публикационной активности по теме «Жыве Беларусь».

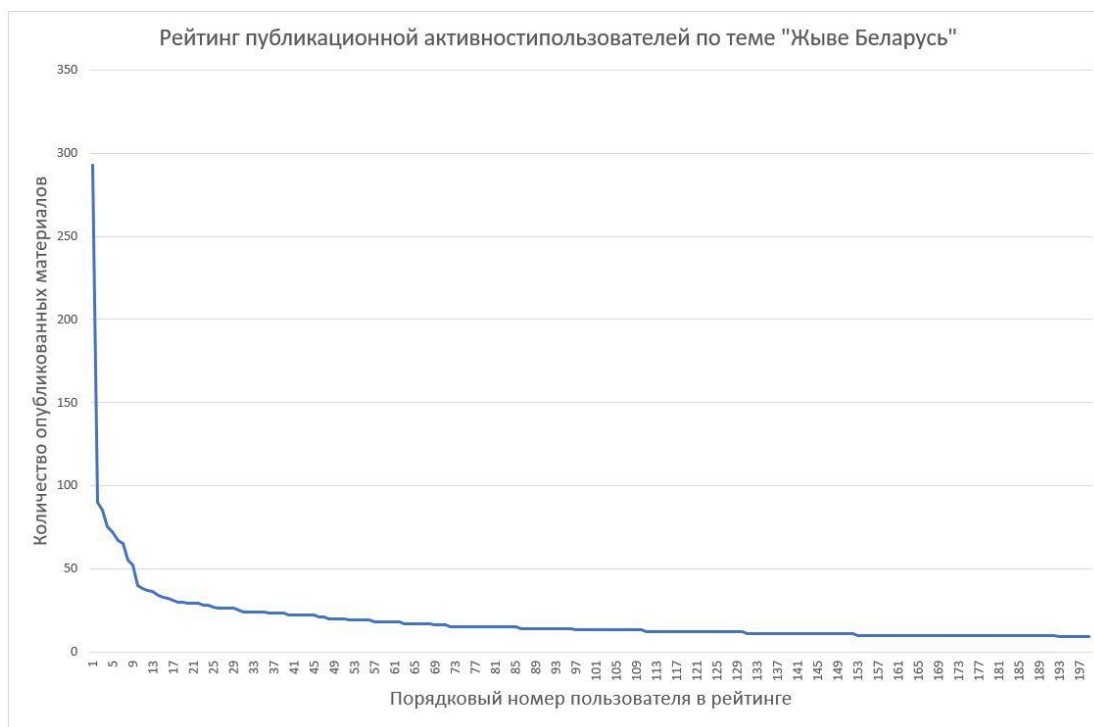


Рисунок 5 - Рейтинг публикационной активности по теме «Жыве Беларусь»

Из данного графика видно, что изменения уровня активности замедляется на уровне первых 150 пользователей, что составляет 0,53% от общего количества пользователей в выборке. Из них половина наиболее активные, то есть примерно 75 пользователей. Сопоставление данных об уровне публикационной активности и количестве социальных связей показало, что наиболее активные пользователи не всегда имеют большое количество связей, в то время как пользователи с большим количеством связей, как правило, имеют невысокий уровень активности, что подтверждается графиками, представленными на рисунках 6 и 7.

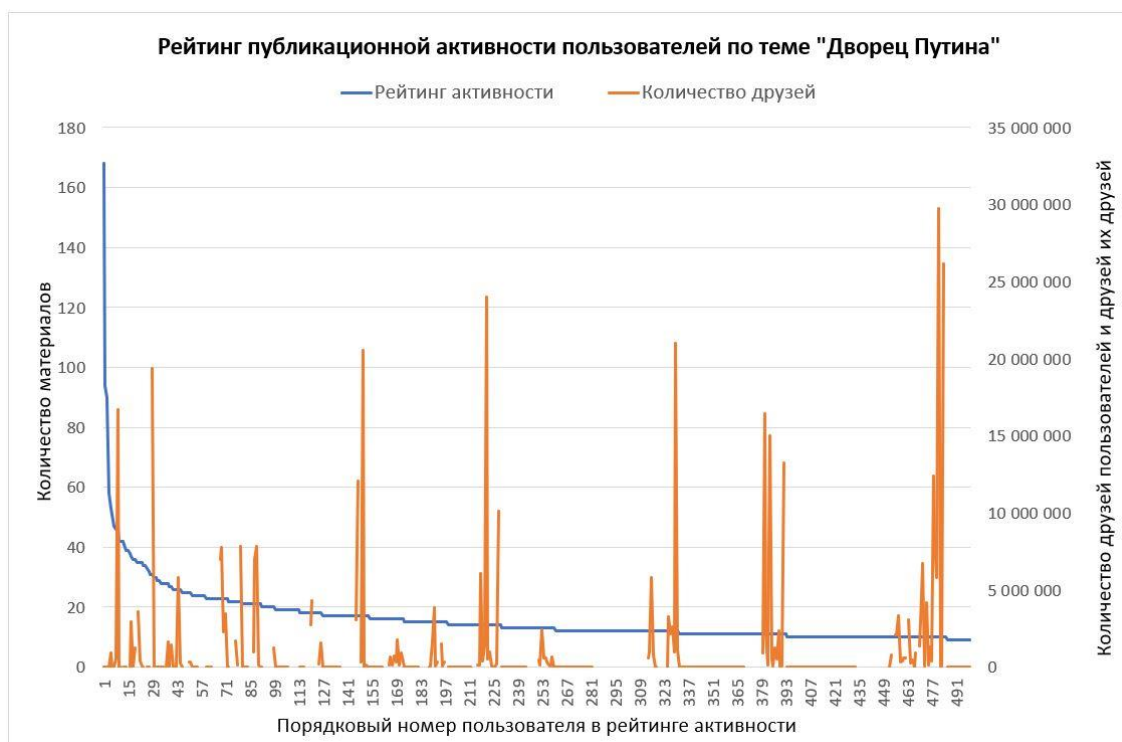


Рисунок 6 - Соотношение уровня публикационной активности и количества социальных связей по инфоповоду «Дворец Путина»

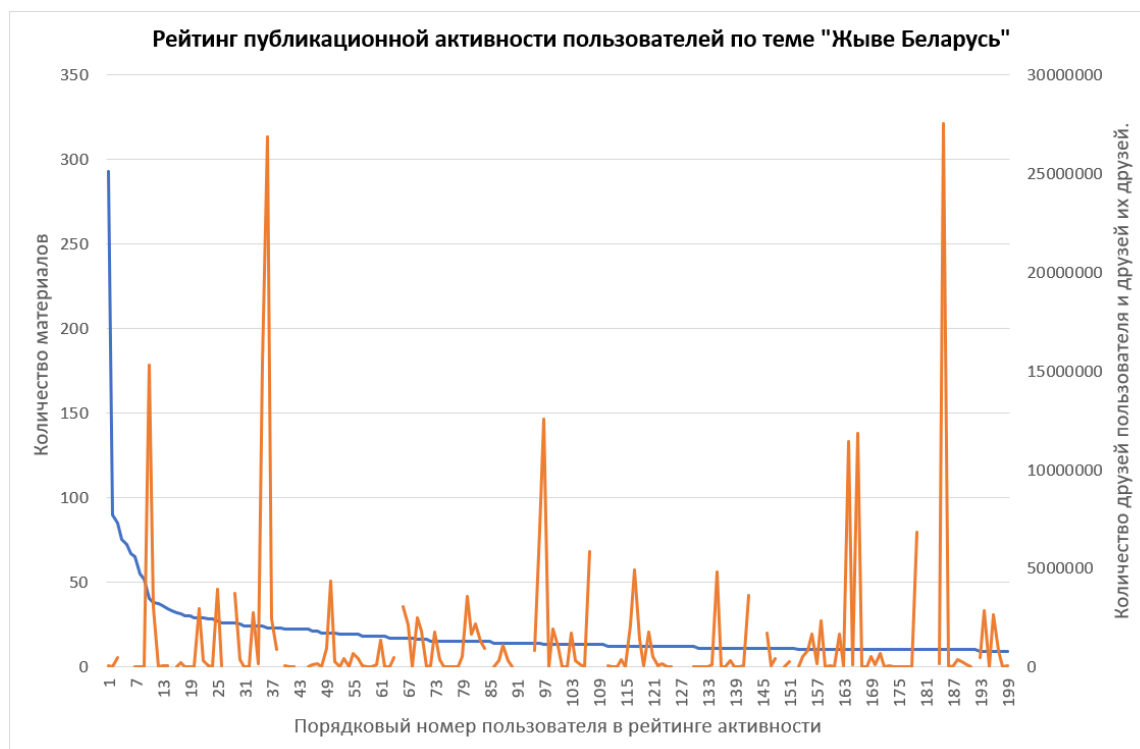


Рисунок 7 - Соотношение уровня публикационной активности и количества социальных связей по инфоповоду «Живе Беларусь»

Следует отметить, что на представленных графиках показаны данные только той части рейтинга публикационной активности, где пользователи имеют

достаточно высокое значение количества связей и высокий уровень активности. У остальных пользователей, при низком уровне активности, количество связей также не велико, поэтому они не представляют интереса для данного исследования. Для вычисления уровня влияния пользователей в социальной сети предлагается применить сочетание параметров публикационной активности и параметров центральности пользователей в графе их социальных связей. Для этого было вычислено произведение уровня публикационной активности пользователя на количество его связей. Таким образом, было получено потенциальное количество актов донесения информации от каждого пользователя до других пользователей социальной сети. Это значение можно интерпретировать как потенциальный уровень влияния (ПУВ). 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, опубликованных пользователем  $k$ ,

$y_k$  – количество репостов, опубликованных пользователем  $k$ ,

$z_k$  – количество комментариев, опубликованных пользователем  $k$ ,

тогда уровень публикационной активности или 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

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

Очевидно, что не каждый пост, репост или комментарий дойдет до потенциального получателя, однако ПУВ это простая, но довольно информативная метрика. Чем выше у пользователя значение ПУВ, тем выше потенциальный уровень его влияния на сеть. Далее были исследованы корреляции между значением ПУВ, уровнем публикационной активности и

количеством социальных связей пользователей. На рисунках 8 и 9 представлены графики соотношения ПУВ и уровня публикационной активности.



Рисунок 8 - Соотношение рейтинга ПУВ и уровня активности по теме «Живе Беларусь»



Рисунок 9 - Соотношение рейтинга ПУВ и уровня активности по теме «Дворец Путина»



Как видно из графиков, представленных на рисунках 8 и 9, несмотря на довольно равномерное распределение уровня активности, в левой части, соответствующей максимальным значениям потенциального уровня влияния, плотность высоких значений активности заметно выше. Разница в уровнях активности на двух графиках не должна вводить в заблуждение. Абсолютные значения уровней активности примерно одинаковые, за исключением выброса на уровне 300 публикаций на рис.8. Этот выброс искажает картину за счет разницы масштабов, но не влияет на результат, так как он находится в зоне низкого потенциального уровня влияния.

На рисунках 10 и 11 представлены графики соотношения ПУВ и количества связей.

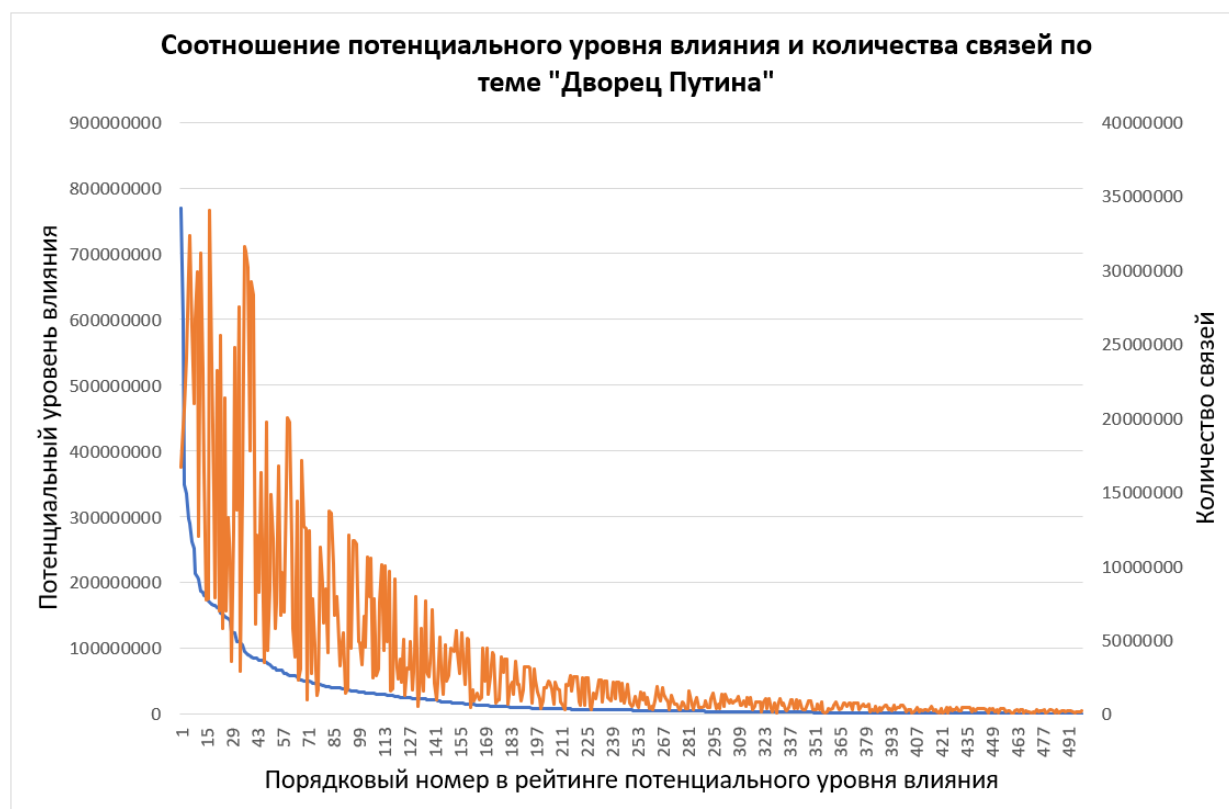


Рисунок 10 - Соотношение рейтинга ПУВ и уровня активности по теме «Дворец Путина»

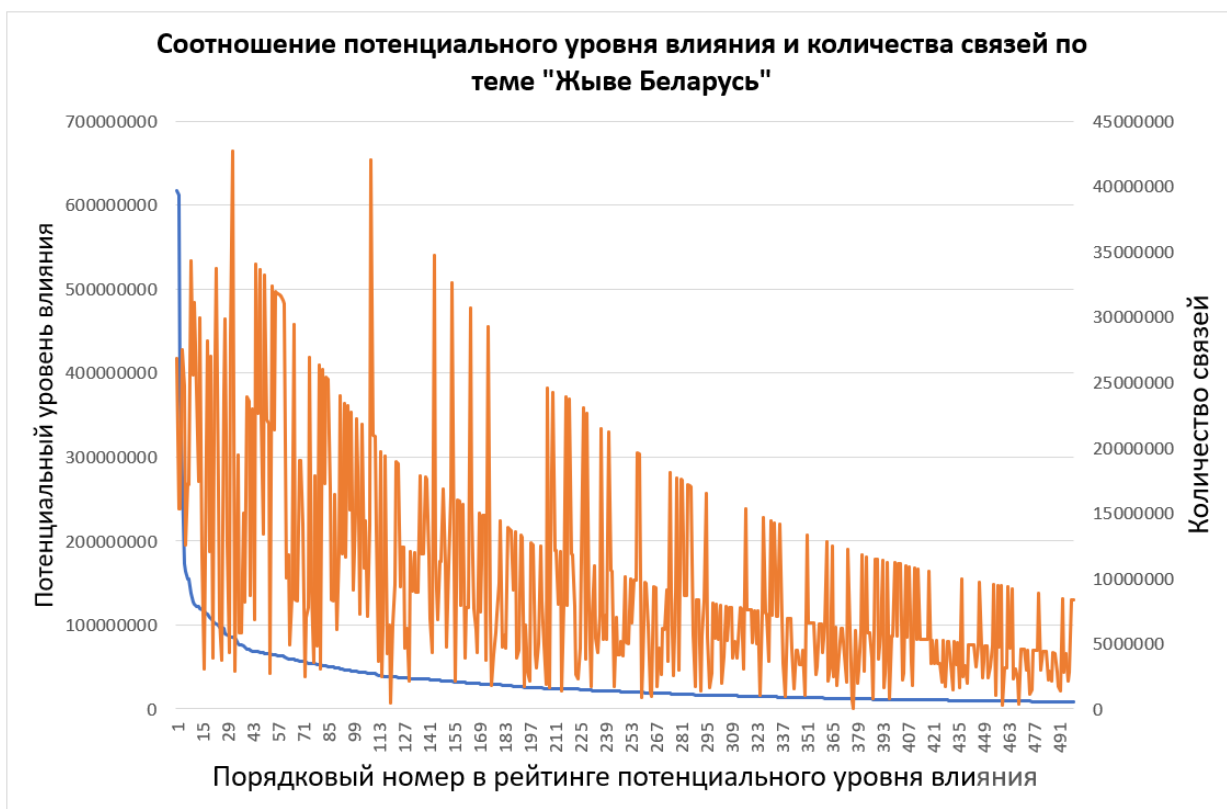


Рисунок 11 - Соотношение рейтинга ПУВ и количества связей по инфоповоду «Жыве Беларусь»

Как следует из графиков, представленных на рисунках 10 и 11, корреляция между потенциальным уровнем влияния и количеством связей более высокая, что вполне объясняется разницей масштабов величин. Количество связей может исчисляться десятками миллионов ( $10^7$ ), в то время как активность исчисляется в сотнях публикаций ( $10^2$ ), то есть разница может составить 5 порядков. Очевидно, что какова бы ни была активность пользователя, количество связей перевешивает эту активность. Для того чтобы устранить этот перекос масштабов была выполнена нормализация количества связей и количества материалов по их максимальным значениям, после чего вычислен нормализованный ПУВ, который математически можно выразить как

$$l'_k = \frac{c(p_k)}{p_{\max}} \frac{m_k}{m_{\max}}. \quad (4)$$

Таким образом, был вычислен нормализованный ПУВ (далее по тексту просто ПУВ), то есть уровень влиятельности каждого пользователя, вовлеченного в целевую тематику. При этом, показатели количества связей

пользователей и их публикационной активности учитывались в относительных величинах, что позволило уравнивать вклад каждого из этих показателей в определение уровня влияния.

Согласно выражению (4), наиболее влиятельным считается узел, у которого значение ПУВ является максимальным. Однако, среди наиболее влиятельных могут быть пользователи, чей контент имеет больший потенциал распространения, чем у других влиятельных пользователей, поэтому для практических целей важно знать не только самого влиятельного, но и группу наиболее влиятельных пользователей. На рисунке 12 представлен график распределения уровня ПУВ для пользователей, вовлеченных в инфоповод «Дворец Путина», отсортированный по убыванию значения.



Рисунок 12 - Распределение значений ПУВ по инфоповоду «Дворец Путина»

Как следует из графика, распределение ПУВ в рейтинге сильно неравномерно. Из 35 тысяч пользователей (см. табл.1) не более 100 пользователей имеют значения ПУВ, не стремящиеся к нулю.

Для выявления группы наиболее влиятельных пользователей был использован метод половинной массы, то есть самыми влиятельными считались те пользователи, 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 потенциального уровня влияния путем ранжирования по убыванию значения ПУВ для каждого пользователя, который представлен на рисунке 12. Далее 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 равенство (6). Таким образом, наиболее влиятельными являются пользователи рейтинга ПУВ с номерами  $\{1, 2, \dots, n\}$ . Результаты вычисления наиболее влиятельных пользователей по обоим инфоповодам представлены в табл.2.

Таблица 2 – Количество наиболее влиятельных пользователей, вовлеченных в тематику «Жыве Беларусь» и «Дворец Путина»

№ п/п	Параметр	«Жыве Беларусь»	«Дворец Путина»
1	2	3	4
1	Кол-во вовлеченных пользователей	28 447	35 502
2	Кол-во наиболее влиятельных пользователей	125	231

В результате применения предложенного комплексного метода из 64 тысяч пользователей нам удалось выявить всего только 356 наиболее влиятельных пользователей. Профили и страницы этих пользователей были переданы специалистам для подробных аналитических исследований с оперативной и политологической точек зрения.

### **1.5. Выявление наиболее влиятельных пользователей методом диаграммы Венна**

Наряду с методом вычисления ПУВ возможно применение метода построения диаграммы Венна. Диаграмма Венна (также используется название диаграмма Эйлера - Венна), это схематичное изображение всех возможных отношений (объединение, пересечение, разность, симметрическая разность) нескольких (часто - трёх) подмножеств универсального множества. На диаграммах Венна универсальное множество  $U$  изображается множеством точек некоторого прямоугольника, в котором располагаются в виде кругов или других простых фигур все остальные рассматриваемые множества.

Диаграммы Венна применяются при решении задач вывода логических следствий из посылок, выразимых на языке формул классического исчисления высказываний и классического исчисления одноместных предикатов, для:

- описания функционирования формальных нейронов Мак-Каллока и сетей из них
- синтеза надежных сетей из не вполне надежных элементов,
- построения управляющих и самоуправляющихся систем и блочного анализа и синтеза сложных устройств,
- получения логических следствий из заданной информации, минимизации формул исчислений.

Диаграммы Венна при помощи  $n$  фигур изображают все  $2n$  комбинаций  $n$  свойств, то есть конечную булеву алгебру. При  $n=3$  диаграмма Эйлера–Венна обычно изображается в виде

трёх кругов с центрами в вершинах равностороннего треугольника и одинаковым радиусом, приблизительно равным длине стороны треугольника.

Применение диаграммы Венна для выявления наиболее влиятельных пользователей позволяет найти пересечение множеств пользователей с наибольшей публикационной активностью и множества пользователей с наибольшим количеством связей. На рисунке 13 представлена диаграмма Венна, использованная для выявления наиболее влиятельных пользователей, активно публикующих посты по тематике инфоповода «Жыве Беларусь».

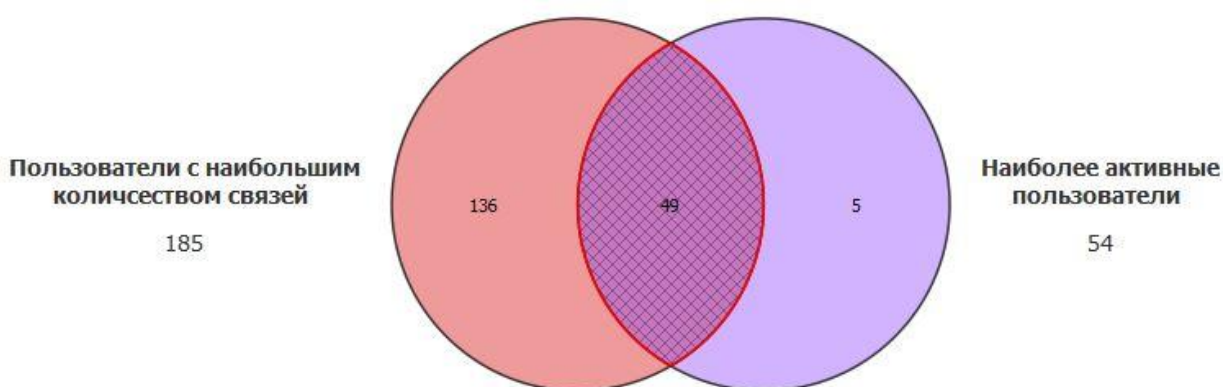


Рисунок 13 - Диаграмма Венна для Постеров по теме «Жыве Беларусь»

На вход диаграммы были поданы множества, которые включали пользователей, у которых не менее 100 друзей и которые опубликовали не менее 10 постов. Указанные ограничения выбраны экспертным путем. Очевидно, что при изменении правил отбора множеств, количество пользователей, входящих в оба множества, также будет изменяться. Это накладывает более высокие требования к уровню квалификации экспертов, использующих данный метод. В тоже время это гибкий, не требующий высоких вычислительных мощностей метод, реализованный во многих пакетах, предназначенных для анализа данных.

## 1.6. Сравнение методов выявления наиболее влиятельных пользователей

В практике компании СЕУСЛАБ для выявления наиболее влиятельных пользователей (НВП) были использованы два способа:

- По уровню потенциального уровня влияния (ПУВ),
- По пересечению множеств самых активных пользователей и пользователей с наибольшим количеством связей методом диаграммы Венна.

ПУВ вычисляется как произведение количества связей пользователя на количество опубликованных им материалов. То есть ПУВ, это потенциальное количество актов доведения материалов пользователя до связанных с ним участников сети. Очевидно, что не все материалы доходят до всех пользователей и это количество зависит от популярности пользователя или популярности темы.

С другой стороны, если среди пользователей выбрать множество наиболее активных, то есть тех, у кого количество опубликованных материалов самое большое, и множество тех пользователей, у которых наибольшее количество связей, то в результате сравнения этих множеств можно найти пользователей, которые входят в оба множества. По сути, это самые активные пользователи с максимальным количеством связей.

В чем разница двух этих подходов и какой подход наиболее адекватный?

Недостаток ПУВ в разнице масштабов величин. Количество связей может исчисляться десятками миллионов ( $10^7$ ), в то время как активность исчисляется в сотнях ( $10^2$ ), то есть разница в 5 порядков. То есть ПУВ в большей степени отображает количество связей, чем количество связей и активность. Для того чтобы устранить этот перекос масштабов необходимо нормализовать количество связей и количество материалов по максимальным значениям и тогда уже вычислять нормализованный ПУВ. С другой стороны, а почему их масштабы должны быть одинаковыми? Что важнее, активность или количество связей? С точки зрения уровня влияния, количество связей важнее, чем активность. Ведь,

пользователь, опубликовавший только один материал, но имеющий миллион подписчиков, имеет большее влияние чем пользователь, опубликовавший 100 постов, у которого 10 000 друзей, читающих не все его посты. С точки зрения ПУВ эти пользователи абсолютно равны, а в реальности уровень влияния абсолютно различен. Из этого следует, что необходимо различать качество связей, друзья и подписчики. Ценность подписчиков гораздо выше, чем ценность друзей. Но если на первом уровне приближения не различать друзей и подписчиков (качество связей), то учитывать уровень влияния двух этих пользователей можно только по количеству связей. Так же точно, невозможно учитывать популярность пользователей и качество контента, так как в данной ситуации необходимая для этого информация отсутствует. Таким образом, остается только числовое значение этих двух параметров, а какое соотношение масштабов между количеством связей и уровнем активности должно быть, является предметом дальнейшего исследования.

Использование пересечения множеств методом построения диаграммы Венна фактически не учитывает масштабы, так как представляет пользователей, являющихся лидерами по количеству связей и количеству опубликованных материалов. Однако, какой из методов наиболее предпочтителен сказать трудно, так как количество связей имеет большее значение на уровень влияния, чем активность пользователя. Для сравнения методов представим результаты пересечения множеств пользователей, выявленных двумя разными методами. В идеале, если эти множества совпадут, то методы идентичны, если полностью не совпадут, это значит, что они показывают разные результаты. На примере выборки по инфоповоду «Живе Беларусь» из 28400 пользователей средствами нейросетевой классификации были отобраны активные пользователи и разделены на роли. Методом пересечения множеств было выбрано 49 наиболее активных Постеров с наибольшим количеством связи. Параллельно для всех Постеров было вычислено значение ПУВ и отобрано несколько разных множеств. Далее результаты отбора сравнивались с множеством из 49 супер-



Постеров на предмет пересечения этих множеств. Результаты пересечения представлены на рис.14 – 17.



Рисунок 14 - Пересечение множеств при ПУВ > 10 тыс.

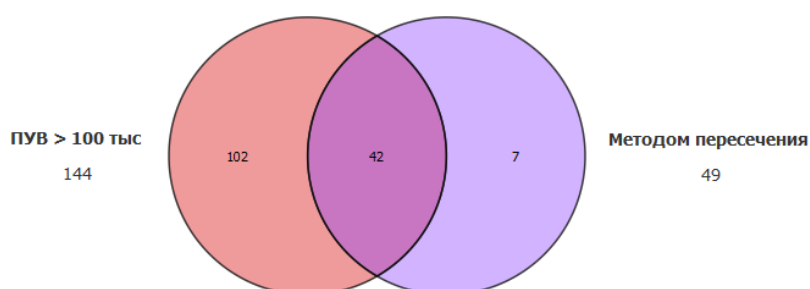


Рисунок 15 - Пересечение множеств при ПУВ > 100 тыс.



Рисунок 16 - Пересечение множеств при ПУВ > 1 млн.



Рисунок 17 - Пересечение множеств при ПУВ > 9 млн.

Из представленных рисунков видно, что при значениях ПУВ от 10 тысяч множество Постеров состоит из 176 пользователей. Это множество полностью включает в себя всех супер-Постеров, выявленных методом пересечения. В тоже время, при значении ПУВ выше 9 млн. множества пользователей, определенных по ПУВ и методом пересечения содержат одинаковое количество, по 49 пользователей, но пересечение составляет только 14 пользователей. То есть, совпадение методов составляет 28,6%. Это означает, что методы измеряют разные величины.

Если исключить влияние масштабов за счет нормализации количества связей и количества опубликованных материалов по максимальному значению и последующим произведением этих величин, то получим более высокий уровень совпадения 44,9% (см. рис.18).



Рисунок 18 - Пересечение множеств при нормализованном ПУВ  $> 0,001$ .

Действительно, нормализация позволяет повысить корреляцию результатов, полученных различными методами, но все равно она ниже 50%.

Получается, что ответ на вопрос какой метод точнее определяет наиболее влиятельных и активных пользователей пока не найден. Фактически, можно констатировать, что если необходимо выявлять пользователей с наибольшим количеством связей с учетом их активности, тогда лучше использовать ПУВ. А если нас интересуют наиболее активные пользователи с наибольшим количеством связей, то нужно применять метод пересечения.

В таблице 3 представлены результаты вычисления пересечения ПУВ с другими суперролями.

Таблица 3 – Уровень совпадения результатов двух методов выявления НВП

№ п/п	Наименование роли	Нормализованный ПУВ	Методом пересечения	Пересечение множеств	Пересечение в %
1	2	3	4	5	6
<b>Живе Беларусь</b>					
1	Постеры	49	49	22	44,9%
2	Репостеры	25	25	15	60,0%
3	Комментаторы	30	30	10	33,3%
4	Универсалы	21	21	10	47,6%
<b>Дворец Путина</b>					
5	Постеры	94	94	74	78,7%
6	Репостеры	51	51	30	58,8%
7	Комментаторы	16	16	7	43,8%
8	Универсалы	31	31	11	30,6%

На основании данных таблицы 3 можно сделать вывод, что для разных ролей и для разных инфоповодов зависимости различаются. Если для Репостеров важны и активность, и количество связей, то для комментаторов количество связей не так важно, как активность. В то время как Постеры и Универсалы в первой выборке показывают совпадение методов около 50%, что говорит о том, что активность для них не так важна, как количество связей, во второй выборке они различаются существенно (80/30).

### 1.7. Выводы по главе 1

Выявление наиболее влиятельных пользователей в социальных сетях является весьма актуальной задачей. Существуют достаточно большое количество методик определения уровня влияния, но каждая из этих методик определяет уровень влияния в каком-то одном аспекте, будь то коммуникативные способности узлов, способность контролировать трафик, публикационная активность и другие.

В данном исследовании было предложено выявлять максимально возможное количество актов донесения информации от каждого пользователя до

других пользователей в сети, поэтому был разработан комплексный метод выявления наиболее влиятельных пользователей на основе вычисления потенциального уровня влияния, который сочетает данные об уровне публикационной активности и количестве социальных связей пользователей в социальной сети. Использование данного метода в реальных аналитических исследованиях показало его практическую значимость за счет повышения релевантности поиска и существенного снижения затрат на анализ профилей пользователей.

Использование метода пересечения множеств с помощью диаграммы Венна также может быть рекомендовано в тех случаях, когда целевым признаком является преобладание связей над активностью или наоборот, так как параметры отбора множеств для построения диаграммы Венна аналитики обычно выбирают вручную на основании своих экспертных знаний о предметной области.

Следует отметить, что вычисление значения ПУВ для массива пользователей не требует от аналитика ввода каких-либо параметров. Все значения вычисляются автоматически. Метод не требует больших вычислительных затрат, прост в реализации как в прототипах, так и в промышленном варианте. В тоже время, в методе пересечения путем построения диаграммы Венна имеется больше степени свободы для аналитика и в некоторых случаях этот метод может дать лучший результат, но при этом требует более высокой квалификации аналитика.

Темой дальнейших исследований может быть изучение корреляции потенциального уровня влияния и уровня реакции на публикации наиболее влиятельных пользователей в виде репостов, комментариев, лайков и просмотров. Значение корреляции могло бы соответствовать степени реализации потенциала информационного влияния пользователей на социальную сеть, а различия в значениях корреляции могли бы дать дополнительную информацию о факторах, которые необходимо учитывать при расчете уровня информационного влияния пользователей в социальных сетях.

## **ГЛАВА 2. ВЫЯВЛЕНИЕ СОЦИАЛЬНЫХ РОЛЕЙ ПОЛЬЗОВАТЕЛЕЙ**

Несмотря на представления авторов работы [56] о том, что социальные сети являются не организованными, очевидно, что одни пользователи в социальных сетях генерируют идеи, другие распространяют эти идеи, третьи комментируют, а четвертые могут накручивать лайки, чтобы поднять популярность идеи. Есть такие пользователи, которые рекрутируют в социальную сеть целые сообщества, способствуя ее стремительному расширению. Анализ показывает, что в каждой информационной волне, связанной с каким-либо информационным поводом, определенные пользователи выполняют определенные роли.

В зависимости от типа социальной сети и исследуемых характеристик пользователей могут использоваться различные определения социальных ролей и методы их классификации. В зависимости от типа социальной сети и исследуемых характеристик пользователей могут использоваться различные определения социальных ролей и методы их классификации.

### **2.1. Обзор методов выявления ролей пользователей**

Прежде чем исследовать методики выявления социальных ролей пользователей в социальных сетях, необходимо дать определение термина «социальная роль». В работе [1] предложено стандартизировать использование термина «социальная роль» в онлайн сообществе как совокупность социальных, психологических, структурных и поведенческих атрибутов, а также предложены стратегии определения социальных ролей пользователей в некоторых онлайн сообществах. Однако, авторы не предлагают строгой классификации социальных ролей пользователей, так как набор и определение социальных ролей зависит и от типа социального сообщества, и от контекста, в котором рассматриваются роли пользователей. Так, [2] исследуют сообщества

инновационных конкурсов, чтобы найти несколько типов ролей пользователей: мастер, социализатор, генератор идей, эффективный участник, пассивный генератор идей, пассивный комментатор. Для определения ролей они используют метрики центральности по степени в сочетании с количеством предложенных идей, а далее несколькими методами кластерного анализа разбивают пользователей на указанные выше группы.

Исследованию Википедии посвящена работа [3], в которой с помощью использования количественных данных об активности пользователей авторы выделяют 12 типов организационных ролей. [4] используют для своего анализа Википедии визуализацию данных, параметры пользователей и контентный анализ и в результате определяют следующий набор ролей пользователей: советники, родители, брокеры, редакторы, менеджеры и вандалы.

Исследуя сайты социальных сетей, [5] использовали данные пользователей и результаты анкетирования, а затем методом кластеризации K-means разбивали пользователей на группы: социализаторы, активные, комментаторы, наблюдатели и случайные пользователи. [6] использовали для исследования сайтов социальных сетей сочетание факторного, кластерного и дисперсионного анализа, чтобы разбить всех пользователей на группы: продвинутые пользователи, спорщики, пользователи для развлечений, случайные и неактивные пользователи. В работе [7] предлагается использовать сочетание метода декомпозиции K-shell для выявления пользователей, наиболее близко расположенных к ядру сети, и метода измерения сходства N-грамм для публикаций, созданных пользователями. Таким образом авторы получают ранжированный список наиболее влиятельных пользователей по каждой теме. [8] выявляли роли пользователей на основе информации профилей пользователей (имя, возраст, пол, национальность, статус отношений и т.д.), данных о публикациях пользователей (посты, комментарии, лайки, тэги и т.д.) и информации о друзьях пользователей и друзьях друзей пользователей. На основе данных о пользователе вычислялось значение переменной Interactions, а на основе анализа графа социальных связей каждого пользователя значение

переменной Metrics. Композиция двух этих переменных составляла вектор, значение которого описывало комплексные характеристики пользователей. А затем выполнялась кластеризация векторов пользователей по ролям тремя различными методами: K-means, SOM и DBSCAN. В результате были выявлены следующие роли: зритель, участник и производитель контента. «Зритель» в основном наблюдает за тем, что происходит в сети, «Участник» чаще взаимодействует с контентом, получая более высокое значение центральности близости, а «Производители контента» имеют большее количество публикаций на своих страницах, что приводит к более высокой степени взаимодействия, чем у двух других профилей.

Оригинальный метод предложил [90]. В нем используется графовый анализ для определения всех триад пользователей в сети, на основе которого строится вектор из 36 элементов, где каждый элемент, это доля условных триад одного из 36 типов. Затем выполняется кластеризация векторов. В результате были определены такие роли: эксклюзивный участник группы, потребитель информации.

Авторы [91] используют для разделения на социальные роли следующие атрибуты пользователей: возраст аккаунта, общее количество открытых тем, общее количество публикаций и общее количество посещений страниц других пользователей. Результаты показывают, что сеть в основном состоит из: социализаторов, генераторов контента, посетителей и пассивных участников, в зависимости от уровня их вклада в сообщество соответственно. Социализаторы открывают новые темы и поддерживают обсуждение в сообществе, а молодые генераторы контента отправляют новые работы по открытым темам. Эти две роли несут ответственность за поток информации в сети. Посетители открывают мало тем и добавляют мало содержания, но все же заходят на сайт в отличие от пассивных пользователей, которые не участвуют в обсуждениях. По утверждению авторов, сочетание информации о характеристиках пользователей на основе метрик центральности и социальной роли пользователей, могло бы существенно обогатить данные о влиятельности пользователей и дать более

полное представление об их реальной роли в структуре социальной сети, а также повысить надежность и достоверность данных.

В работе [9] предлагается использовать нейронные сети для выявления типовых профилей пользователей Instagram. Искомые профили соответствуют психологическому состоянию пользователей: доволен жизнью, чувство собственного достоинства, беспокойство и депрессия. Схожие задачи по выявлению психологических профилей пользователей с использованием нейронных сетей решают [10]. Авторы использовали глубокую нейронную сеть для классификации пользователей по уровню стресса на три группы, а для обучения нейронной сети они создали датасет на основе контент-анализа 350 миллионов сообщений в Twitter. Авторы статьи [11] рассматривают вопрос построения системы распознавания психологического профиля и особенностей характера человека по его фотографиям из социальной сети Facebook и применяют для классификации пользователей сверточные нейронные сети. В результате классификации пользователи были разделены по основным психотипам: экстраверсия, доброжелательность, добросовестность, невротизм, открытость опыту.

Использование глубокой рекуррентной нейронной сети и модели встраивания слов как метод автоматизации обнаружения и выявления социальных ролей в онлайн-сообществах предложили [92].

Метод автоматической классификации пользователей сети Twitter по 8 категориям эмоций предлагают [13]. Для классификации пользователей авторы также используют глубокую нейронную сеть.

Таким образом, в представленных работах для разделения пользователей на классы используются метрики центральности графового анализа и различные методы кластеризации или классификации. В тоже время, очевидно, что представленные авторы исследуют различные аспекты характеристик или моделей поведения пользователей и каждый использует для этого свою классификацию ролей пользователей, что не позволяет создать единый набор ролей пользователей на все случаи жизни. Это же проявляется и в контексте



противодействия деструктивным воздействиям на социальные сети в виде протестного движения и мобилизации пользователей на участие в протестных акциях. В настоящее время поиск в открытых источниках методик выявления ролей пользователей участников протестного движения дает отрицательный результат.

## **2.2. Описание ролей пользователей**

Как было показано в предыдущей главе, существуют различные подходы для определения той роли, которую играют пользователи в каждом социальном явлении и в связанной с ним временной социальной сети. Анализ паттернов поведения пользователей применяется в интерпретационном подходе, в котором набор различных видов и параметров активности интерпретируется как условный профиль пользователя и затем профили пользователей разбиваются на группы по определенным правилам. Далее в рамках этого подхода будут рассмотрены несколько методов выявления социальных ролей пользователей и представлены результаты сравнения этих методов.

Прежде чем приступить к разработке методики выявления ролей пользователей, необходимо дать определение этих ролей и определить параметры, по которым пользователи будут классифицироваться по этим ролям. В контексте противодействия влиянию социальных сетей на протестную активность граждан с точки зрения паттернов поведения пользователей представляет интерес выявление следующих типов ролей:

1. Постер – генератор идей, создатель контента, часто является лидером общественного мнения, при наличии большого количества связей может объединять вокруг себя большое количество пользователей.

2. Репостер – распространитель идей, редко создает контент, в основном репостит уже готовые публикации, нацелен на максимальное распространение чужих публикаций.

3. Комментатор – не создает контент, не репостит, зато оставляет множество комментариев, участвует в осуждениях и спорах. Часто создает лишние комментарии для увеличения популярности тем обсуждения.

4. Пассивный участник – пользователь, который не очень активен в сети в части создания контента, репостов или комментариев, но регулярно посещает различные страницы в социальной сети. Является реципиентом всей информации, созданной Постерами, Репостерами и Комментаторами.

Параметры, которые могут существенно влиять на различие пользователей по ролям и на основании которых может выполняться классификация, представлены ниже:

X1 – Возраст аккаунта

X2 – Количество друзей

X3 – Количество опубликованных постов

X4 – Количество опубликованных репостов

X5 – Количество опубликованных комментариев.

### **2.3. Выявление ролей пользователей методом численной оценки**

Для определения ролей пользователей методом численной оценки использовались параметры уровня и формы их активности X3 – X5, потому что именно эти параметры определяют характер поведения пользователей в сети.

Первая задача состояла в том, чтобы разделить массив пользователей на активных и пассивных. Разделение выполнялось на основе положения о том, что активные пользователи опубликовали существенно больше материалов, чем пассивные. Было принято соотношение материалов, опубликованных активными и пассивными пользователями в пропорции 70/30. Для выявления активных пользователей был рассчитан рейтинг активности путем ранжирования по убыванию суммы количества публикаций (постов, репостов и комментариев) для каждого пользователя. Далее последовательно, двигаясь по рейтингу вниз, подсчитывалась общая сумма материалов, опубликованных пользователями,

начиная с лидера рейтинга, и сравнивалась с общей суммой опубликованных материалов. Когда сумма материалов достигла значения 70% от общей суммы публикаций, подсчет был остановлен. То есть активные пользователи в сумме опубликовали 70% от общего количества публикаций, а пассивные пользователи – оставшиеся 30%. Далее был определен список пользователей, сумма публикаций которых составляет 70% от общей суммы опубликованных материалов. Математически это можно выразить так:

если  $p_i$  – количество постов, опубликованных пользователем  $i$ ,

$r_i$  – количество репостов, опубликованных пользователем  $i$ ,

$k_i$  – количество комментариев, опубликованных пользователем  $i$ ,

то уровень активности  $a_i$  пользователя  $i$  можно выразить как

$$a_i = p_i + r_i + k_i, \quad (7)$$

тогда суммарная активность  $A$  всех пользователей будет равна

$$A = \sum_{i=1}^I a_i, \quad (8)$$

где  $I$  – общее количество пользователей. Общее количество материалов  $A_n$ , опубликованное всеми активными пользователями, можно выразить как

$$A_n = \sum_{i=1}^n a_i = 0.7 \cdot A, \quad (9)$$

где  $n$  – номер пользователя в рейтинге активности, для которого выполняется равенство (9). Таким образом, пользователи рейтинга активности с номерами  $\{1, 2, \dots, n\}$  являются активными, а все остальные – пассивными. Результаты разделения пользователей на активных и пассивных представлены в таблице 4.

Таблица 4 – Количество активных и пассивных пользователей, вовлеченных в тематику «Жыве Беларусь» и «Дворец Путина»

№ п/п	Параметр	«Жыве Беларусь»	«Дворец Путина»
1	2	3	4

1	Кол-во вовлеченных пользователей всего, в том числе:	28 447	35 502
2	Количество активных пользователей	1 257	4 847
3	Количество пассивных пользователей	27 190	30 655

Следующая задача состояла в том, чтобы классифицировать активных пользователей на роли Постер, Репостер и Комментатор. Для такой классификации было принято, что пользователь выполняет ярко выраженную роль, если один вид активности составляет не менее 60% от всех видов его активности. Другими словами, у Постера будет не менее 60% опубликованных им постов из общей суммы опубликованных им материалов, у Репостера – не менее 60% репостов и т.д. То есть:

если  $p_i \geq 0,6 * a_i$  – роль пользователя определяется как Постер,

если  $r_i \geq 0,6 * a_i$  – роль пользователя определяется как Репостер,

если  $k_i \geq 0,6 * a_i$  – роль пользователя определяется как Комментатор.

В результате выполнения вычислений доли  $p_i$ ,  $r_i$  и  $k_i$  в уровне активности  $a_i$  для каждого активного пользователя, было выполнено разделение пользователей по ролям. Большая часть пользователей, которых мы выделили в качестве активных, удовлетворяют правилу отсечки на уровне 60% для какого-либо вида активности. В тоже время, среди активных пользователей выявились и те, кто не удовлетворяет правилу отсечки 60% ни по одному из видов активности. Таких пользователей мы классифицировали как Универсал. Изучение функции Универсалов является темой отдельного исследования. Данные о ролях пользователей, выявленных методом численной оценки, представлены в таблице 5.

Таблица 5 – Общие статистические данные выявленных ролей пользователей, вовлеченных в тематику «Жыве Беларусь» и «Дворец Путина»

№ п/п	Параметр	«Жыве Беларусь»	«Дворец Путина»
-------	----------	-----------------	-----------------

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
1	Кол-во вовлеченных пользователей всего, в том числе:	28 447	35 502
2	Количество пассивных пользователей	27 190	30 655
3	Количество Постеров	81	531
4	Количество Репостеров	74	1050
5	Количество Комментаторов	521	132
6	Количество Универсалов	581	3134

Таким образом, используя метод численной оценки, мы выявили роли пользователей Постер, Репостер, Комментатор, Универсал и Пассивный участник и для каждой из выявленных ролей были определены значения параметров X1 – X5, которые представлены в табл. 6.

Таблица 6 – Значения параметров для каждой роли, полученные методом численной оценки по инфоповоду «Дворец Путина»

<b>№ п/п</b>	<b>Параметр</b>	<b>Название</b>	<b>Постер</b>	<b>Репостер</b>	<b>Комментатор</b>	<b>Универсал</b>	<b>Пассивный</b>
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	
1	<b>X1</b>	Возраст аккаунта (дней)	311-4382	0-4881	0-4183	545-4553	86-5170
2	<b>X2</b>	Кол-во друзей	0-32368259	0-31586803	0-28990781	0-34031059	0-36014728
3	<b>X3</b>	Кол-во постов	2-94	0-94	0-6	0-17	0-1
4	<b>X4</b>	Кол-во репостов	0-13	2-160	1-3	1-19	0-1
5	<b>X5</b>	Кол-во комментариев	0-7	0-48	2-48	0-11	0-1

## 2.4. Экспертная нейросетевая классификация ролей пользователей

Как было показано выше, для разделения пользователей на группы используются в основном методы кластеризации и классификации. В данной главе представлен метод нейросетевой классификации ролей пользователей, основанный на использовании нейросети, обученной на синтетических данных, сгенерированных с учетом экспертных знаний о предметной области.

Для выполнения нейросетевой классификации ролей пользователей необходимо наличие качественных обучающих и валидирующих множеств. Различные эксперты решают эту проблему разными способами. Так, авторы работы [14] разработали специальный датасет из 1000 профилей пользователей, чтобы использовать его для обучения нейронной сети, определяющей социальные роли пользователей Twitter. Создание датасета, построенного на основе 740 тысяч сообщений, представлено [13], а в работе [12] использовался набор данных, состоящий из более чем 1,2 миллиона текстовых сообщений, извлеченных из онлайн-сообщества о высшем образовании в Австралии. Для обучения своей нейронной сети авторы [10] создали датасет на основе контент-анализа 350 миллионов сообщений в Twitter. Во всех этих случаях авторам необходимо было промаркировать реальных пользователей признаками принадлежности к тому или иному классу, что является весьма трудоемкой задачей.

Авторы [9, 11] для анализа онлайн-социальных сетей использовали уже готовые датасеты. Использование готовых датасетов удобно, но может быть связано как с трудностями их получения, так и с неполным соответствием готового датасета той предметной области, для которой планируется его применение. Кроме того, нейронные сети, обученные на одних сетях, могут быть непригодны для других сетей.

Поскольку, как было показано в главе 2.3, распределение ролей между пользователями в социальных сетях крайне неоднородно, для получения качественного датасета необходимо собрать данные огромного количества

пользователей, что весьма дорого и предполагает возможный риск утечки конфиденциальных данных. Готовых датасетов для нейросетевой классификации ролей пользователей в контексте противодействия протестным движениям не существует, поэтому для снижения затрат на подготовку множеств для обучения, тестирования и валидации нейросети было принято решение использовать синтетические данные.

Развитие современных технологий искусственного интеллекта невозможно представить без использования синтетических данных. Как следует из названия, это данные, которые созданы искусственно, а не в результате реальных событий. Они часто создаются с помощью алгоритмов и используются для широкого спектра действий. По утверждению [16, 17], синтетические данные дешевы в производстве и могут быть полезны для разработки моделей искусственного интеллекта, глубокого обучения и тестирования программного обеспечения. Конфиденциальность данных, обеспечиваемая синтетическими данными, одно из наиболее важных преимуществ синтетических данных. Пользовательские данные часто включают личную информацию и личную медицинскую информацию, а синтетические данные позволяют компаниям создавать программное обеспечение, не раскрывая пользовательские данные разработчикам или программным инструментам.

Большинству моделей машинного обучения для большей точности требуются большие объемы данных. Синтетические данные можно использовать для увеличения размера обучающих данных для моделей машинного обучения.

Генерация синтетических данных создает размеченные экземпляры данных, готовые к использованию в обучении. Это снижает потребность в трудоемких усилиях по маркировке данных.

Последние несколько лет генеративные модели, основанные на глубоком обучении, вызывают все больший интерес и предлагают некоторые удивительные улучшения в этой области [93]. Опираясь на огромный объем синтетических данных, хорошо спроектированные сетевые архитектуры и интеллектуальные методы обучения, глубокие генеративные модели

продемонстрировали невероятную способность создавать очень реалистичные фрагменты контента различного типа, такие как изображения, тексты и звуки.

По утверждению автора [18] синтетические данные впервые начали использоваться в 90-х годах, однако избыток вычислительной мощности и места для хранения в 2010-х годах привело к более широкому использованию синтетических данных.

К этому же периоду относится одна из первых работ в этом направлении в России [20], авторы которой задались целью автоматизировать труд менеджера туристической фирмы, в обязанности которого входил подбор туристических маршрутов, максимально удовлетворяющих запросам клиентов. Авторы статьи составили анкету, включающую вопросы клиенту фирмы и рекомендации менеджера по вариантам туристических маршрутов, максимально удовлетворяющих интересам клиента. Множество таких анкет было заполнено менеджером фирмы, который сам придумывал различные ответы клиента на задаваемые вопросы и, используя свой экспертный опыт, давал рекомендации по оптимальным маршрутам для каждого такого виртуального клиента. Таким образом было сформировано синтетическое множество примеров, которое использовалось для обучения нейронной сети, предназначенной для моделирования деятельности менеджера. По-видимому, в России это и был один из первых случаев практического применения нейронной сети, обученной на примерах, синтезированных экспертом.

На сегодняшний день существует целая индустрия по производству синтетических данных. Основные сценарии использования и программные инструменты для создания синтетических данных представлены в публикации [15]. Готовых наборов синтетических данных для решения задачи в рамках данного исследования не существует, но некоторые из подобных сценариев могут быть применены для решения актуальной задачи создания нейросетевого классификатора пользователей онлайн-социальных сетей.



### 2.4.1. Входные и выходные параметры нейросети

В качестве входных данных, на основании которых выполнялась классификация пользователей нами использовались следующие параметры:

X1 – Возраст аккаунта

X2 – Количество друзей

X3 – Количество опубликованных постов

X4 – Количество опубликованных репостов

X5 – Количество опубликованных комментариев.

Выходными данными нейросетевой модели являлись:

D1 – принимает значение 1, если пользователь является Постером и 0, если нет.

D2 – принимает значение 1, если пользователь является Репостером и 0, если нет.

D3 – принимает значение 1, если пользователь является Комментатором и 0, если нет.

D4 – принимает значение 1, если пользователь является Универсалом и 0, если нет.

D5 – принимает значение 1, если пользователь является Пассивным участником и 0, если нет.

### 2.4.2. Синтезирование датасета

Как было указано выше, для создания датасета, состоящего из 2000 примеров, на основе разметки реальных данных, необходимо было бы собрать данные о 200 тысячах пользователей, вовлеченных в протестные движения в фазе мобилизации пользователей на протестные акции. Если на разметку одного профиля пользователя тратить примерно 10 минут, то на разметку 200 000 пользователей потребовалось бы 2 000 000 минут или 33000 человеко/часов. Затраты на создание такого датасета могли бы составить приблизительно 20 млн. руб. В виду отсутствия указанных ресурсов и во избежание рисков разглашения

конфиденциальной информации было принято решение искусственно синтезировать датасет на основе экспертных знаний о предметной области.

Обычно эксперт, глубоко разбирающийся в своей предметной области, может без особого труда вспомнить, спрогнозировать, придумать наиболее характерные варианты поведения моделируемого объекта, процесса, явления. Для эксперта обычно не вызывает затруднений приведение примеров, так называемых, крайних случаев. Например, эксперт-врач может назвать параметры больного с ярко выраженными симптомами какого-либо заболевания и безошибочно поставить диагноз этого заболевания. Эксперт-инженер может, опираясь на свой собственный опыт спрогнозировать момент выхода из строя какого-либо технического объекта, изучив параметры его текущего состояния. Эксперт-металлург может достаточно точно спрогнозировать механические характеристики выплавляемого им металла и привести примеры удачных и неудачных плавок, и т.д.

Используя множества таких экспертных данных в качестве обучающих примеров, можно создать нейронную сеть, моделирующую знания эксперта, заложенные в эти данные.

Основная трудность применения этого способа состоит в том, что эксперт, как правило, не в состоянии предоставить достаточно большое количество таких примеров, необходимое для качественного обучения нейронных сетей. Для преодоления этой проблемы можно было бы рекомендовать метод добавления шума в обучающие множества [21, 94]. Однако, более эффективно нам представляется попросить эксперта привести в создаваемых им примерах не конкретные значения входных параметров, а указывать допустимые диапазоны, в пределах которых изменения того, или иного параметра, по мнению эксперта, не приведет к существенному изменению выходных параметров. Используя эти диапазоны, не представляет труда случайным образом сгенерировать дополнительные примеры и довести обучающее множество до объема, необходимого для качественного обучения нейронной сети.

Основная задача при формировании датасета состояла в том, чтобы определить диапазоны каждого из входных параметров для каждой роли, а также внести в датасет определенные закономерности. В качестве значений, необходимых для генерации датасета, были использованы диапазоны, полученные в результате анализа материалов, опубликованных пользователями социальной сети ВКонтакте относительно так называемого «Дворца Путина» методом численной оценки (см. табл. 5), а также экспертные знания о данной предметной области в виде закономерностей соотношения входных параметров для каждой из ролей.

Для того, чтобы нейросеть могла хорошо обучиться и качественно классифицировала входные множества, необходимо было обеспечить достаточное количество примеров для каждой роли, поэтому было сгенерировано по 400 примеров для каждой роли (всего 2000 примеров).

Для генерации множества использовалась функция случайного выбора (рандомизации) значения из диапазона значений. Обозначим эту функцию как

$$R(X_{\min}; X_{\max}) \quad (10)$$

В общем виде алгоритм генерации множества представлен на рис. 19.

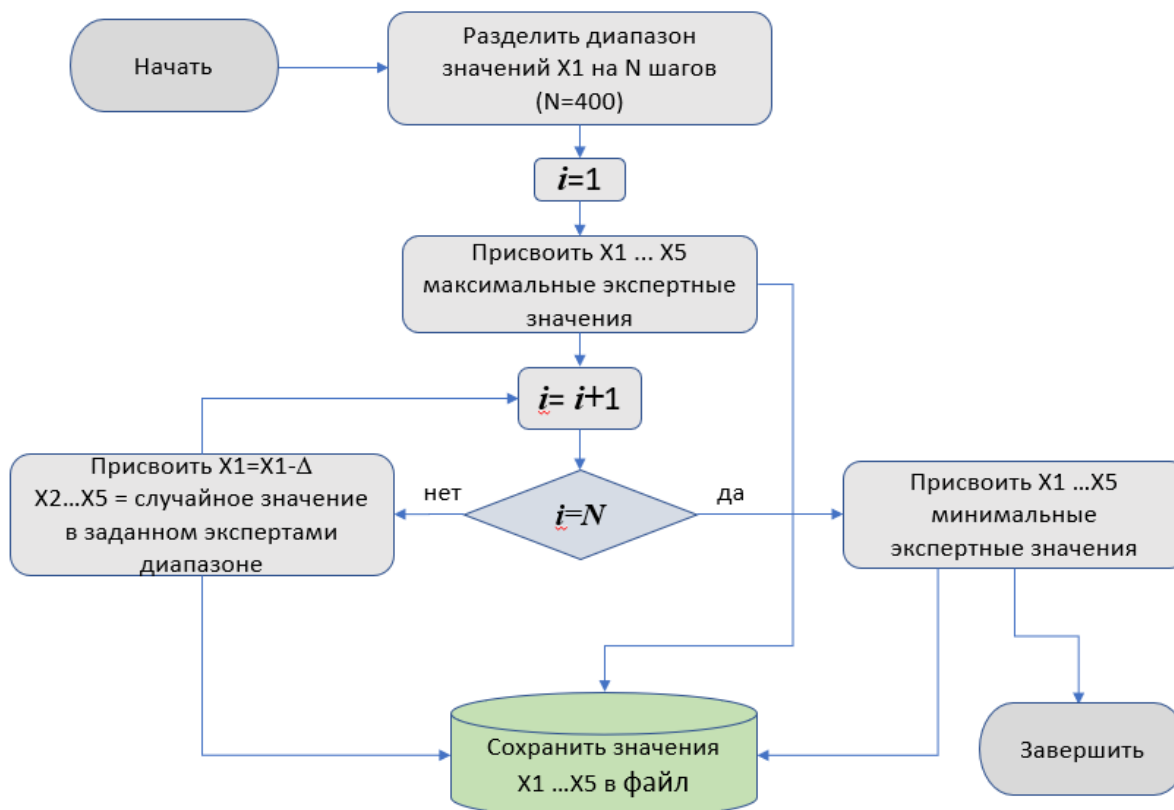


Рисунок 19 - Алгоритм генерации множества на основе экспертных знаний

**Пассивными** считаются участники, которые имеют не высокий рейтинг активности, при этом соотношение между различными видами материалов не имеет значения, главное, чтобы значения не выходили за пределы максимальных и минимальных значений. В таблице 7 представлены формулы, использованные при генерации множества для роли Пассивный участник. Через  $\Delta$  обозначен шаг уменьшения возраста аккаунта.

Таблица 7 – Формулы генерации множества для роли Пассивный участник

№ п/п	X1	X2	X3	X4	X5
1	2	3	4	5	6
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, опубликованных пользователем  $i$ ,

$r_i$  – количество репостов, опубликованных пользователем  $i$ ,

$k_i$  – количество комментариев, опубликованных пользователем  $i$ ,

тогда общее количество материалов этого пользователя  $m_i$  можно выразить как

$$m_i = p_i + r_i + k_i \quad (11)$$

На основании представлений экспертов, полученных при анализе поведения пользователей, вовлеченных в протестные движения и мобилизацию других пользователей на участие в протестных акциях, 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:

$$p_i \geq \alpha m_i, r_i \leq \beta m_i, k_i \leq \gamma m_i, \quad (12)$$

где  $\alpha = 0,6$ ,  $\beta = 0,37$ ,  $\gamma = 0,03$ . В этом случае значения  $r_i$  и  $k_i$  можно выразить как

$$r_i \leq \delta p_i, k_i \leq \varepsilon p_i \quad (13)$$

где  $\delta = \frac{\beta}{\alpha}$  и  $\varepsilon = \frac{\gamma}{\alpha}$ .

Используя эти соотношения и граничные значения для данной роли из таблицы 1, представим набор формул для генерации множества роли Постер (см. табл.8).

Таблица 8 – Формулы генерации множества для роли Постер

№ п/п	X1	X2	X3	X4	X5
1	2	3	4	5	6
1	$X1_{\max}$	$X2_{\max}$	$X3_{\max}$	$R(0; \delta X3)$	$R(0; \varepsilon X3)$
2	$X1_{\max} - \Delta$	$R(X2_{\min}; X2_{\max})$	$R(X3_{\min}; X3_{\max})$	$R(0; \delta X3)$	$R(0; \varepsilon X3)$
3	$X1_{\max} - 2\Delta$	$R(X2_{\min}; X2_{\max})$	$R(X3_{\min}; X3_{\max})$	$R(0; \delta X3)$	$R(0; \varepsilon X3)$
..	..	..	..	..	..
N	$X1_{\min}$	$X2_{\min}$	$X3_{\min}$	$X4_{\min}$	$X5_{\min}$

**Репостером** является пользователь, количество репостов которого, составляет не менее 60% от всех опубликованных им материалов, количество постов – не более 37% и количество комментариев – не более 3%. То есть пользователь должен удовлетворять следующим условиям:

$$r_i \geq \alpha m_i, p_i \leq \beta m_i, k_i \leq \gamma m_i \quad (14)$$

где  $\alpha = 0,6, \beta = 0,37, \gamma = 0,03$ . В этом случае значения  $p_i$  и  $k_i$  можно выразить как

$$p_i \leq \delta r_i, k_i \leq \varepsilon r_i \quad (15)$$

где  $\delta = \frac{\beta}{\alpha}$  и  $\varepsilon = \frac{\gamma}{\alpha}$ .

Набор формул для генерации множества роли Репостер представлен в таблице 9.

Таблица 9 – Формулы генерации множества для роли Репостер

№ п/п	X1	X2	X3	X4	X5
1	2	3	4	5	6
1	$X1_{\max}$	$X2_{\max}$	$R(0; \delta X4)$	$X4_{\max}$	$R(0; \varepsilon X4)$
2	$X1_{\max} - \Delta$	$R(X2_{\min}; X2_{\max})$	$R(0; \delta X4)$	$R(X4_{\min}; X4_{\max})$	$R(0; \varepsilon X4)$
3	$X1_{\max} - 2\Delta$	$R(X2_{\min}; X2_{\max})$	$R(0; \delta X4)$	$R(X4_{\min}; X4_{\max})$	$R(0; \varepsilon X4)$
..	..	..	..	..	..
N	$X1_{\min}$	$X2_{\min}$	$X3_{\min}$	$X4_{\min}$	$X5_{\min}$

**Комментатором** является пользователь, количество комментариев которого, составляет не менее 60% от всех опубликованных им материалов, количество постов – не более 10% и количество репостов – не более 30%. То есть пользователь должен удовлетворять следующим условиям:

$$k_i \geq \alpha m_i, p_i \leq \beta m_i, r_i \leq \gamma m_i, \quad (16)$$

где  $\alpha = 0,6, \beta = 0,1, \gamma = 0,3$ . В этом случае значения  $p_i$  и  $r_i$  можно выразить как

$$p_i \leq \delta k_i, r_i \leq \varepsilon k_i, \quad (17)$$

где  $\delta = \frac{\beta}{\alpha}$  и  $\varepsilon = \frac{\gamma}{\alpha}$ .

Набор формул для генерации множества роли Комментатор представлен в таблице 10.

Таблица 10 – Формулы генерации множества для роли Комментатор

№ п/п	X1	X2	X3	X4	X5
1	2	3	4	5	6
1	$X1_{\max}$	$X2_{\max}$	$R(0; \delta X5)$	$R(0; \varepsilon X5)$	$X5_{\max}$
2	$X1_{\max} - \Delta$	$R(X2_{\min}; X2_{\max})$	$R(0; \delta X5)$	$R(0; \varepsilon X5)$	$R(X5_{\min}; X5_{\max})$
3	$X1_{\max} - 2\Delta$	$R(X2_{\min}; X2_{\max})$	$R(0; \delta X5)$	$R(0; \varepsilon X5)$	$R(X5_{\min}; X5_{\max})$
..	..	..	..	..	..
N	$X1_{\min}$	$X2_{\min}$	$X3_{\min}$	$X4_{\min}$	$X5_{\min}$

**Универсалом** является пользователь, который не удовлетворяет условиям (12), (14) и (16). Фактически Универсал, это активный пользователь, не вошедший в число Постеров, Репостеров или Комментаторов. Используя это представление и граничные значения для данной роли из таблицы 5, представим набор формул для генерации множества роли Универсал (см. табл.6).

Множества примеров для каждой роли, полученные в результате работы алгоритма, были объединены, перемешаны и разбиты на две части:

Обучающее – 1700 примеров,

Тестовое – 300 примеров.

### 2.4.3. Проектирование нейронной сети

Подготовленный датасет был использован для обучения и тестирования нейросетевой модели на платформе Нейросимулятор 5.0 Nsim5sc [95] (доступ [www.LbAi.ru](http://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:

$$E = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - y_n)^2}{N}}}{\max(d_n) - \min(d_n)} 100\% \quad (18)$$

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 11.

Таблица 11 – Результат тестирования нейросетевой модели классификатора ролей пользователей на базе Нейросимулятора Nsim5sc

№ п/п	№ роли	Наименование роли	Ошибка %
1	2	3	4
1	Y1	Постер	10,3%
2	Y2	Репостер	10,2%
3	Y3	Комментатор	2,9%
4	Y4	Универсал	16,5%
5	Y5	Пассивный участник	6,3%

Поскольку обучающие множества были синтезированы на основе набора определенных комбинаций параметров, назначенных экспертом для каждой социальной роли, нейросетевое исследование предметной области выглядит как иллюстрация закономерностей, заложенных экспертом в нейронную сеть. Для исследования зависимости роли пользователя от количества постов использовался случайно выбранный пользователь, параметры которого представлены в таблице 12.

Таблица 12 – Параметры пользователя для прогноза зависимости выходных параметров от количества постов.

№ п/п	Наименование параметра	X1	X2	X3	X4	X5
1	2	3	4	5	6	7
1	Значение параметра	2050	2368529	64	9	6



Результат нейросетевого сценарного прогноза, выполненного для данного пользователя представлен на рисунке 20.

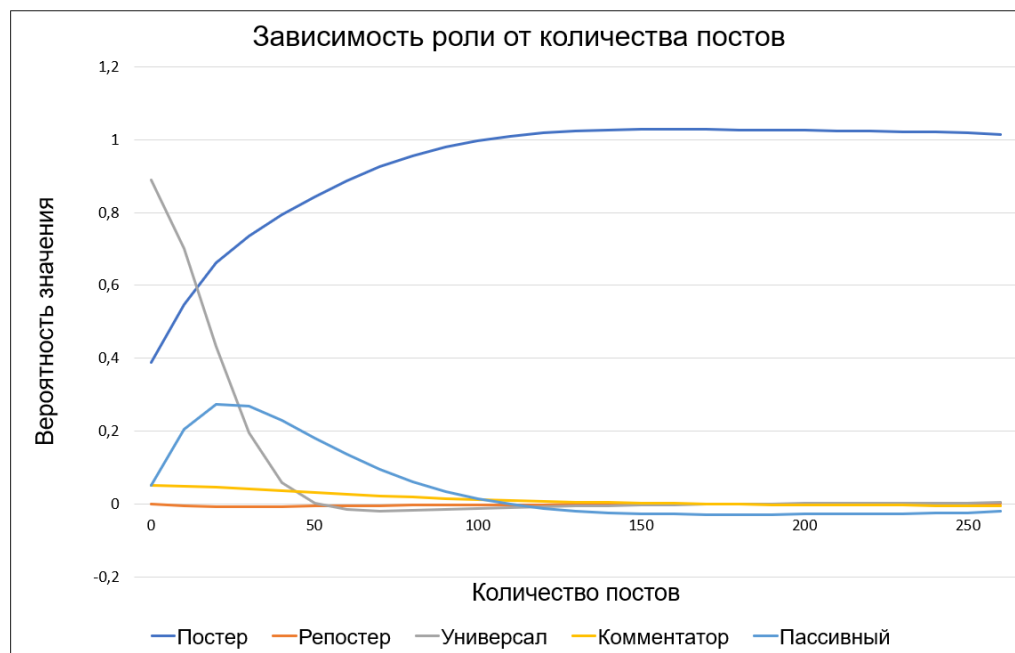


Рисунок 20 - Распределение вероятностей принадлежности пользователя к определенному классу в зависимости от количества постов

Как следует из рисунка 20, при увеличении количества постов вероятность принадлежности пользователя к классу «Постер» сначала резко возрастает, а затем, достигнув максимального значения 100% перестает изменяться. Вероятность принадлежности пользователя к роли «Репостер», «Комментатор» и «Универсал» с увеличением количества постов падает. Таким образом, можно констатировать, что синтетический датасет адекватно отражает заложенные в него зависимости.

Кроме того, 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 последних 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. Результаты классификации были сопоставлены с результатами выявления ролей пользователей, полученными методом численной оценки. В таблице 13 представлены результаты сравнения трех основных наиболее ярко выраженных классов Постер, Репостер и Пассивный участник.

Таблица 13 – Корреляция между результатами метода численной оценки и нейросетевого метода классификации

№ п/п	Коэффициент корреляции Пирсона между численной оценкой и нейросетевым методом классификации	Значение
1	2	3
1	Постер	0,97
2	Репостер	0,96
3	Пассивный	0,98

Как следует из таблицы 13, результаты, полученные методом численной оценки и методом экспертной нейросетевой классификации очень близки и оба этих метода оказались применимы для решения задачи классификации ролей пользователей.

## 2.5. Выявление ролей пользователей методом кластеризации

Другим методом, который часто применяется для выделения групп схожих пользователей, является кластеризация. Для выполнения кластеризации были использованы алгоритмы KMeans, спектральной кластеризации, а также иерархической кластеризации, для которых возможно задание количества

искомых кластеров. Целью кластеризации было выявление ролей Постер, Репостер и Пассивный участник, то есть количество искомых кластеров равно 3. В качестве входных данных для кластеризации были выбраны те же параметры X1 – X5, что и для нейросетевой классификации.

В таблице 14 представлены значения корреляции между результатами численной оценки и различными методами кластеризации для активных ролей Постер и Репостер.

Таблица 14 – Корреляция между результатами численной оценки и методами кластеризации для ролей Постер и Репостер

№ п/п	Метод кластеризации	Корреляция между численной оценкой и методами кластеризации <i>Постер</i>	Корреляция между численной оценкой и методами кластеризации <i>Репостер</i>
1	2	3	4
1	КMeans	0,20	0,20
2	Спектральная кластеризация	0,04	0,04
3	Иерархическая кластеризация	0,22	0,23

Как следует из таблицы 14, классические методы кластеризации при попытках решения задачи выявления искомых ролей пользователей положительного результата не дали.

## 2.6. Выявление наиболее влиятельных пользователей в каждой роли

Использование методов выявления наиболее влиятельных пользователей и выявления ролей пользователей, описанных в главах 2.3 и 2.4, позволяет разделять множество пользователей на активных и пассивных, а в среде активных выделять тех пользователей, кто играет роли Постер, Репостер, Комментатор, Универсал во временной социальной сети. Однако, количество активных пользователей может быть достаточно велико, а уровень

информационного влияния у них может сильно отличаться. Поэтому, для того чтобы выявить только наиболее влиятельных пользователей в каждой роли, для каждого пользователя было вычислено значение ПУВ. В результате сочетания методов выявления наиболее влиятельных пользователей и ролей пользователей был получен список наиболее влиятельных пользователей в каждой роли (см. табл. 15).

Таблица 15 – Статистические данные о наиболее влиятельных пользователях в каждой роли

№ п/п	Параметр	«Жыве Беларусь»		«Дворец Путина»	
1	2	3		4	
1	Кол-во вовлеченных пользователей всего, в том числе:	28 447		35 502	
2	Количество пассивных пользователей	27 190		30 655	
3		Всего	Наиболее влиятельных	Всего	Наиболее влиятельных
4	Количество Постеров	81	49	531	109
5	Количество Репостеров	74	25	1050	58
6	Количество Комментаторов	521	30	132	17
7	Количество Универсалов	581	21	3134	47

В контексте противодействия деструктивным воздействиям на социальные сети, направленным на мобилизацию пользователей для участия в протестных акциях, наибольший интерес представляют наиболее влиятельные Постеры, от

которых зависит количество контента, внедренного в сеть, и Репостеры, которые способствуют максимальному охвату аудитории. В меньшей степени это касается Универсалов, так как они могут как внедрять контент, так и распространять его. То есть в рамках инфоповода «Живе Беларусь» из 28,4 тысяч пользователей оперативный интерес могут представлять не более 100 человек, а по инфоповоду «Дворец Путина» из 35,5 тысяч пользователей – не более 200. Такое существенное снижение количества пользователей, подлежащих контролю, позволяет резко повысить эффективность работы правоохранительных органов.

## **2.7. Выводы по главе 2**

Результаты выявления ролей пользователей социальных сетей, полученные методом численной оценки и нейро-сетевым методом практически полностью совпали, в то время как кластеризация оказалась не применима для решения нашей задачи. Аналитические исследования профилей самых влиятельных и популярных Постеров и Репостеров показали, что большая их часть является ярыми противниками действующей власти, активно продвигающими протестную повестку, имеющими высокое влияние в социальной сети ВКонтакте. Это подтверждает, что были получены два одинаково эффективных метода для классификации ролей пользователей.

Использование метода численной оценки рекомендуется при проведении разовых аналитических исследований массивов пользователей в неустоявшихся или временных социальных явлениях. Нейросетевой метод классификации в большей степени подходит, если социальное явление имеет устойчивый характер, а эксперты могут предсказать адекватные значения диапазонов и соотношений параметров предметной области. Важная особенность нейросетевого метода классификации состоит в том, что этот метод может быть использован в потоковых онлайн приложениях оценки параметров пользователей. Такие приложения могут использоваться при подготовке

информации для принятия решения о блокировке, подавления активности или постановке на мониторинг ключевых пользователей в социальной сети. Также удачным может быть использование метода численной оценки для выявления диапазонов значений в предметной области в совокупности с экспертными знаниями для создания онлайн приложений, основанных на нейросетевом методе выявления ролей.

Использование искусственно синтезированных данных позволило сократить затраты на создание датасета приблизительно в 1000 раз и предотвратить риск утечки большого количества конфиденциальных данных.

Синтетические данные для обучения и тестирования нейросетевой модели были зарегистрированы в качестве базы данных для ЭВМ [53] и доступны для использования по запросу [96]. Структура и параметры базы данных, а также пример данных в датасете представлены в приложениях 5 – 7.

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 количества реальных примеров, полученных классическими способами, недостаточно для качественного обучения нейросети, как это сделано в работах [85,91,92].

### ГЛАВА 3. ВЫЯВЛЕНИЕ МОСТОВ В КЛАСТЕРНЫХ СЕТЯХ

Как было сказано выше, для определения социальных связей пользователей, вовлеченных в публикационную активность по инфоповодам «Жыве Беларусь» и «Дворец Путина» были собраны данные об их друзьях и друзьях их друзей. Для каждого пользователя был рассчитан уровень публикационной активности, который учитывал количество постов, репостов и комментариев. В результате ранжирования по уровню публикационной активности был составлен рейтинг активности пользователей. Для всех пользователей были построены графы их социальных связей. При этом учитывались следующие условия:

- пользователь включается в граф, если он является другом члена рейтинга активности или является другом любого из его друзей (максимальное расстояние до целевого пользователя на графе равно двум);
- пользователь, чей уровень активности равен нулю, включается в граф только в том случае, если он является другом не менее, чем двух пользователей из рейтинга активности.

Примеры файлов, содержащих графовую информацию представлены в приложениях 3 и 4.

На рисунке 21 показаны примеры графов таких сетей, где вершины графа, это пользователи, а ребра – связи между ними. Цветами обозначен уровень публикационной активности пользователей в социальной сети, то есть количество любого вида материалов целевой тематики, опубликованных пользователем. Красный цвет показывает максимальный уровень публикационной активности, а серый – отсутствие активности.

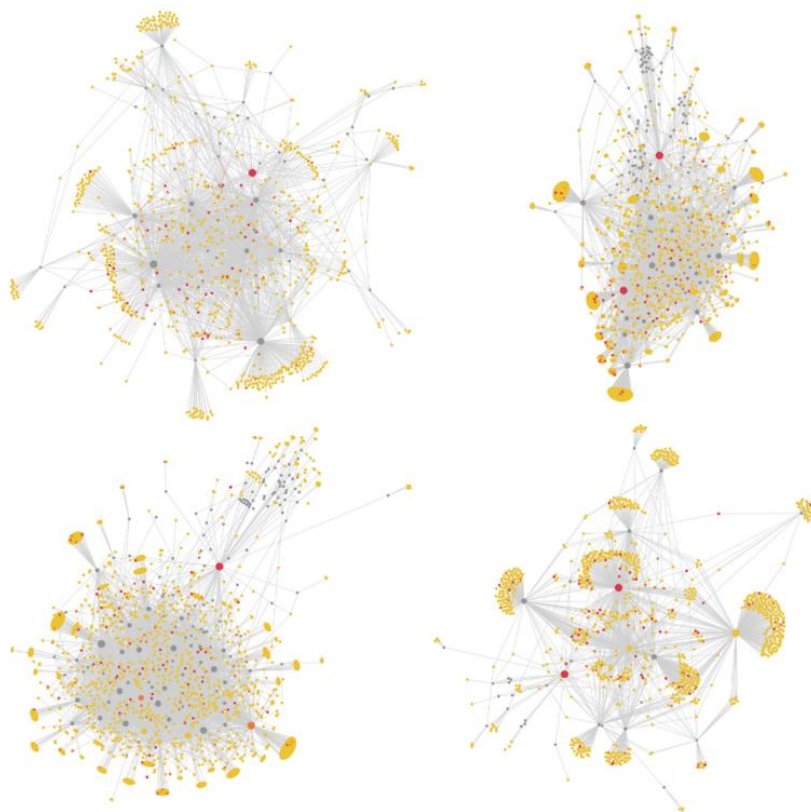


Рисунок 21 - Примеры графов социальных связей наиболее активных пользователей

Как видно из рисунка 21, все графы имеют ярко выраженную кластерную структуру в виде ядра с большим количеством перекрестных связей между пользователями и изолированными кластерами, которые соединены с ядром через одного пользователя, выполняющего роль «моста» между кластером и ядром.

Поскольку виды публикационной активности пользователей в кластерах аналогичны пользователям ядра, логично предположить, что сумма уровней активности каждого узла в кластере может быть выше уровня активности любого узла в ядре, а узел, соединяющий кластер с ядром, возможно будет иметь больший вклад в общий уровень активности сети, чем любой узел в ядре социальной сети. Анализ профилей пользователей, выполняющих роль мостов в тематике протестной активности, показал следующие результаты:

- пользователи в кластерах зачастую лишь частично разделяют мнение пользователей ядра к объектам обсуждения;
- пользователи в кластерах часто объединены одной тематикой;



- предпочтения, интересы, и политические взгляды (для сетей протестной активности) пользователей в разных кластерах могут отличаться (они могут принадлежать к разным политическим партиям или движениям), однако данные пользователи разделяют противостояние действующей власти;
- как правило, пользователи, соединяющий кластер с ядром, выступают в роли модераторов сообщества;
- такие пользователи имеют связи друг с другом и образуют подструктуру в графах социальных связей, за счет чего могут координировать свои действия, вовлекая в исследуемое социальное явление разные категории пользователей, возможно, не согласных с общей точкой зрения ядра сети по некоторым вопросам.

Таким образом, выявление пользователей, выполняющих роль мостов в социальных сетях, имеет огромное значение для противодействия протестным движениям и управления параметрами распространения вирусной и деструктивной информации в социальных сетях.

В представленных на рисунке 21 графах мостом, является узел, 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.

На рисунке 22 представлен фрагмент типичного узла, выполняющего роль моста.

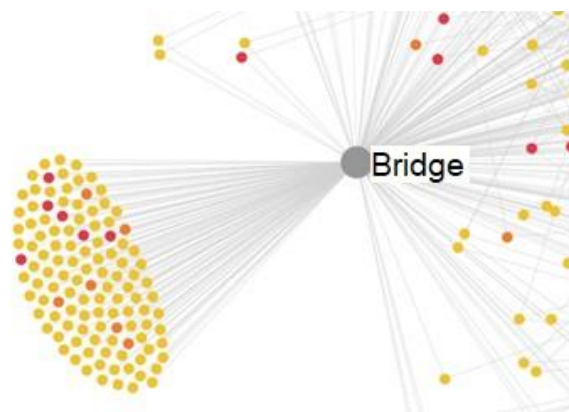


Рисунок 22 – Фрагмент графа с узлом, являющимся мостом

Для оценки уровня влияния мостов в социальной сети необходимо было подобрать или разработать методику, которая однозначно выявляет мосты в кластерных сетях, а также определяет степень влияния мостов на общий уровень активности сети.

### 3.1. Существующие методы выявления мостов в кластерных сетях

Для выявления характеристик узлов в сетях используются различные метрики центральности, описанные в обзорной работе [35]. Самой известной метрикой центральности, характеризующей коммуникативные способности узлов, является впервые введенная [30] центральность по промежуточности, которая соответствует тому, с какой вероятностью данный узел будет связывать множество пар узлов в графе. Поскольку каждый узел в исследуемых сетях имел определенный уровень активности, необходимо было учитывать вес каждого узла в сети. Для вычисления центральности по промежуточности для взвешенных сетей можно было использовать методики, предложенные в работах [97–101]. Однако, центральность по промежуточности, не важно с учетом веса или без, может выявить уровень коммуникативной способности узла сети, но не может точно определить действительно ли данный узел является мостом, так как

узлы с высоким значением центральности по промежуточности могут находиться как в центре ядра сети, так и на периферии сети, являясь мостами.

Влиятельные узлы, по мнению [24], всегда действуют как «мост» между сообществами и существуют в рамках пересечения сообществ. Авторы предлагают использовать для определения таких влиятельных узлов локальный центральный метод, который предполагает, что чем больше сообществ принадлежит узлу, тем большее влияние он оказывает. В работах [25, 26] для определения мостов предлагаются метрики «центральность передачи» и «модульная центральность», однако значение центральности передачи может быть высоким как у вершин ядра, так и у мостов, поэтому ее смысл не сильно отличается от промежуточности, а в модульной центральности мостом являются узлы, соединяющие сообщества, тогда как мы исследовали узлы между сообществами и ядром сети, то есть понятие мост в данном контексте имело иной смысл.

Метод, успешно определяющий мосты, представлен в работе [27], в которой авторы ввели понятие «Мостовая центральность». Эта метрика более точно выявляет мосты, однако она работает только в небольших разреженных сетях с большим количеством колен, так как в ее основе лежит идея о том, что для выявления мостов необходимо отбрасывать значение связей с узлами, находящимися в непосредственной близости узла, то есть связи первого колена графа. Поскольку, в исходном определении моста пользователи кластера связаны только с мостом, то они не могут быть учтены в расчете данной метрики, что было неприемлемо для решения поставленной задачи. Наиболее близкая метрика для решения поставленной задачи, это «Центральность вклада», предложенная [28], суть которой состоит в том, что центральность узла пропорциональна сумме центральностей узлов в его окрестности, взвешенной по их вкладам. Чтобы проиллюстрировать суть данного метода, рассмотрим сеть на рисунке 23.

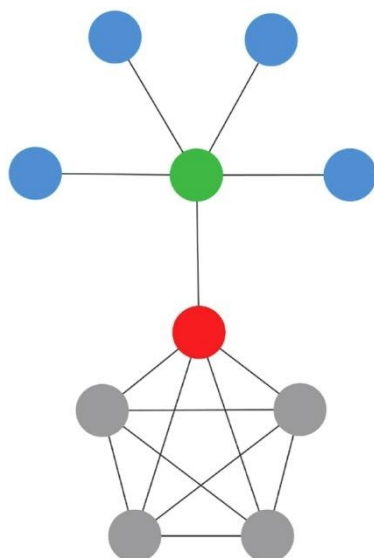


Рисунок 23 – Схема сети узлов с разными вкладами

Для красного узла информация или центральность, которые приносят конкретный серый узел, являются низкими, потому что красный узел может получить доступ к остальным серым узлам напрямую, без какого-либо посредника (это потому, что у них почти одинаковые соседи). Однако, зеленый узел важен для красного узла, потому что без него красный узел не мог бы получить доступ к синим узлам.

Центральность вклада действительно наиболее близко определяет вклады узлов в уровень активности, но она не гарантирует однозначное определение моста, что в нашем случае являлось необходимым условием.

Как видим, все представленные выше метрики в той или иной степени могли бы определить уровень коммуникативной способности узла, но не могут точно определить является ли данный узел мостом в нашем понимании.

### 3.2. Метод выявления мостов в ярко выраженных кластерных сетях

Будем говорить, что все пользователи, имеющие публикационную активность по тематике заданного социального явления, а также их социальные связи составляют «временную социальную сеть», порожденную данным

социальным явлением, а сумма уровней активности всех пользователей составляет общий уровень активности временной социальной сети.

Поскольку количество пользователей в различных кластерах и уровень их активности различны, то мосты могут иметь различный уровень влияния. Пусть степень влияния моста на общий уровень публикационной активности временной социальной сети определяется как суммарный уровень активности кластера, который соединен с ядром через мост. Согласно приведенному выше определению моста, узлы кластера должны быть связаны только с мостом и не должны иметь связи между собой. Рассмотрим граф, представленный на рисунке 24, и определим какой из узлов на графе является мостом в контексте предложенного определения, а также вычислим вес его кластера.

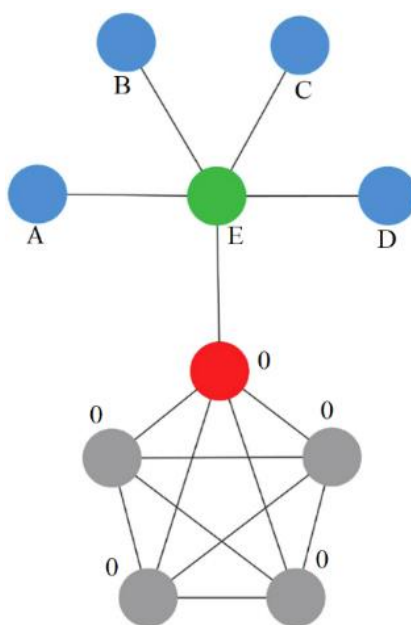


Рисунок 24 – Схема сети, содержащей мост

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.

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 дает значение весов этих узлов. Вес зеленого узла будет равен сумме весов синих узлов  $E=A+B+C+D$ .

Как видим, мы получили одно ненулевое значение для зеленого узла во всей сети, которое точно определяет наличие искомого моста и его вклад в общий уровень активности временной социальной сети, равный значению  $E$ .

С учетом того, что в предлагаемом методе существенную роль играет вес каждого узла графа, для определения степени влияния моста на общий уровень активности был предложен термин «Центральность по взвешенному вкладу».

Будем говорить, что центральность по взвешенному вкладу, это вклад узла социальной сети в общий уровень публикационной активности, равный сумме активности каждого узла кластера, подключаемого узлом к ядру сети, отнесенная к общему уровню активности сети. Другими словами, центральность по взвешенному вкладу моста это вес кластера, подключаемого мостом к ядру, отнесенный к общему весу сети.

Значение центральности по взвешенному вкладу = 0, если узел не является мостом, и  $> 0$ , если узел является мостом. Самый влиятельный узел в роли моста имеет максимальное значение центральности по взвешенному вкладу для данной сети. Алгоритм идентификации мостов и вычисления центральности по взвешенному вкладу представлен на рис. 25.

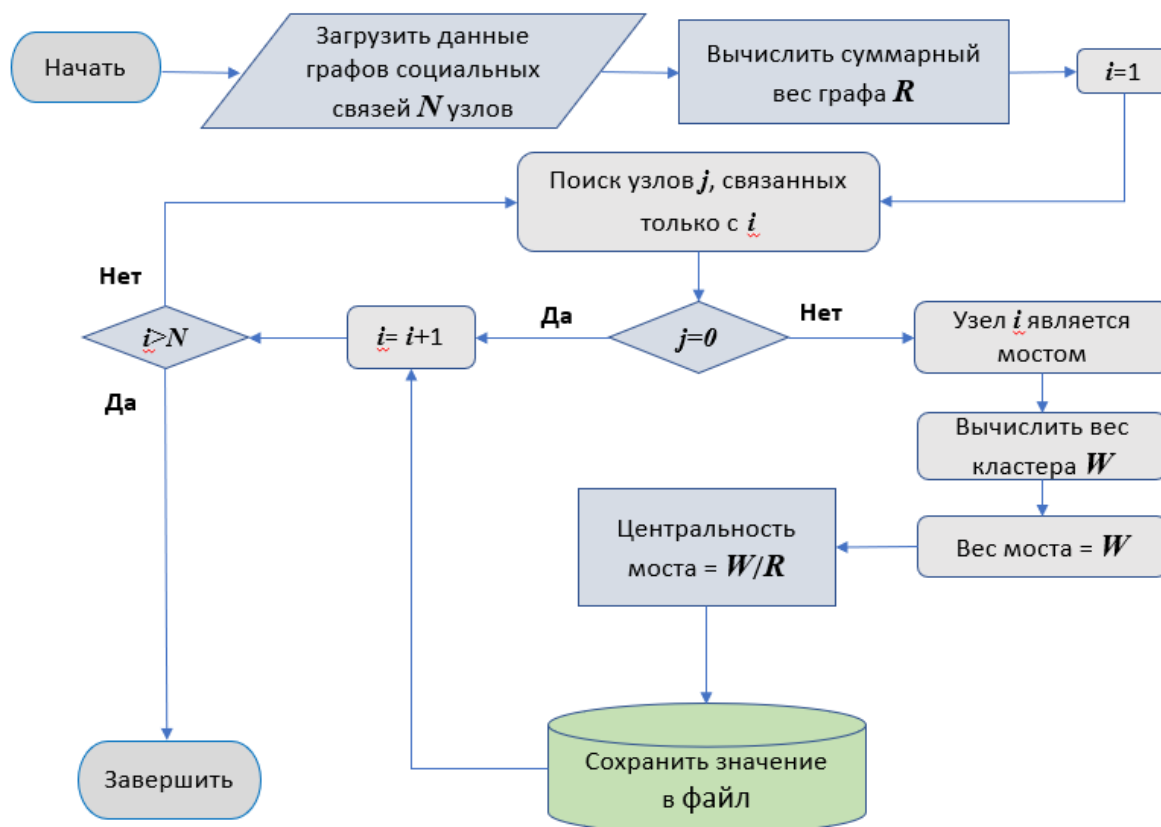


Рисунок 25 – Алгоритм идентификации мостов и вычисления центральности по взвешенному в кластерной сети

Введем обозначения для формального описания предлагаемой методики. Пусть  $G = (U, F)$  – граф, состоящий из множества  $U$  пользователей и множества  $F$  неупорядоченных пар различных элементов множества  $U$ , отражающего дружественные связи между пользователями (рёбра графа). Если пользователи  $u$  и  $v$  являются друзьями, то есть образуют связь  $f \in F$ , будем писать  $f = (uv) = (vu)$ . Множество друзей пользователя  $u \in U$  обозначим  $F(u) = \{v \in U : (uv) \in F\}$ . Тогда степень узла, т. е. число друзей пользователя  $u \in U$ , естественно обозначить  $|F(u)|$ .

Множество пользователей, связанных только с данным пользователем  $u$ , назовем соседями пользователя  $u \in U$  и обозначим  $S(u)$ . Тогда:

$$S(u) = \{v \in U : (uv) \in F, |F(v)| = 1\} = \{v \in U : |F(v)| = 1\} \cap F(u) \quad (19)$$

Если уровень активности пользователя  $u \in U$  обозначить как  $r(u)$ , то суммарный уровень активности некоторого подмножества пользователей  $V \subset U$  будет вычисляться по формуле:

$$R(V) = \sum_{u \in V} r(u) \quad (20)$$

Пользуясь формулой (20), получаем формулу вычисления веса кластера, подключаемого мостом, произвольного пользователя графа:

$$W(u) = R(S(u)) = \sum_{w \in S(u)} r(w) \quad (21)$$

Центральность по взвешенному вкладу определяется как отношение веса кластера, соединенного мостом с ядром сети, к общему уровню активности всех кластеров в сети. Общий уровень активности всех кластеров в сети может быть рассчитан как

$$R = \sum_{u \in V} W(u) \quad (22)$$

Тогда, центральность по взвешенному вкладу можно выразить как

$$C_{wc}(u) = \frac{W(u)}{R}. \quad (23)$$

Код функции на языке Python3, используемой для вычисления мостов в составе программы вычисления мостов в кластерных сетях [52], представлен в приложении 8.

### 3.2.1. Оценка эффективности метода выявления мостов

Для определения уровня влияния мостов 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. В таблице 16 показано насколько уменьшился вес графов в целом и суммарный вес вершин, входящих в кластеры.



Таблица 16 – Изменение веса графа в результате удаления из графа 10 мостов с их вершинами

№ п/п	№ графа	Изменение веса графа	Изменение веса кластеров
1	2	3	4
1	1	-32,7%	-67,5%
2	2	-53,0%	-88,6%
3	3	-44,1%	-65,9%
4	4	-92,0%	-94,5%
5	5	-97,3%	-99,7%
6	6	-72,1%	-92,6%
7	7	-21,9%	-61,0%
8	8	-27,2%	-68,3%
9	9	-48,5%	-72,3%
10	10	-83,4%	-98,9%
11		<i>Среднее значение</i>	<i>Среднее значение</i>
12		<b>-57,2%</b>	<b>-80,9%</b>

Как видим, при удалении из графов 10 самых влиятельных мостов, суммарный вес графа или суммарный уровень активности пользователей в графе снижается в среднем на 57,8%, что показывает высокий уровень влияния мостов. При этом суммарный вес кластеров снижается на 80,9%, что соответствует тому, какую роль играют 10 самых влиятельных мостов в расширении сети. Из этого можно сделать вывод, что узлы сети, выявленные как мосты, действительно вносят существенный вклад в общий уровень активности сети. Сравнение результатов, которые получились при использовании центральности по промежуточности и центральности по вкладу представлены в таблице 17.

Таблица 17 – Изменение метрик центральности в результате удаления из графа 10 мостов с их вершинами

№ п/п	Метрика	Изменение веса графа	Изменение веса кластеров
1	2	3	4
1	Центральность по промежуточности	-54,4%	-76,0%
2	Центральность по вкладу	-52,7%	-66,4%
3	Центральность по взвешенному вкладу	-57,2%	-80,9%

Из представленных данных видно, что при удалении из графов 10 самых влиятельных узлов вместе со всеми их вершинами, общий вес графа снижается примерно одинаково. Это говорит о том, что все три метрики одинаково эффективно выявляют коммуникативные способности влиятельных узлов в сети. В тоже время, изменение веса кластеров заметно больше при использовании центральности по взвешенному вкладу. Это объясняется тем, что центральность по промежуточности и центральность по вкладу выявляют наиболее коммуникативные узлы сети, в том числе и мосты, в отличие от центральности по взвешенному вкладу, которая выявляет только мосты. А поскольку удаление мостов дает наибольший вклад в снижение общего уровня активности сети, то влияние мостов больше, чем у любых других узлов сети.

С другой стороны, отличия в изменении веса графа для разных методик является незначительным. Однако, какие именно узлы учитываются при расчете предложенных метрик центральности, не известно. Для того чтобы ответить на этот вопрос, достаточно выявить уровень корреляции между значениями различных метрик центральности для каждого узла. Для этого для каждой вершины графа были выполнены расчеты центральности по промежуточности и центральности по вкладу, а также сортировка вершин в порядке убывания значений центральности по взвешенному вкладу, промежуточности и

центральности по вкладу. Фрагмент таблицы результатов сравнения представлен в таблице 18.

Таблица 18 – Результат сравнения метрик для определения мостов

№ п/п	Центральность по взвешенному вкладу	Центральность по промежуточности	Центральность по вкладу
1	2	3	4
1	0,22813836	0,184517382	0,171341258
2	0,094830937	0,095150554	0,124428047
3	0,058297707	0,106259885	0,241967659
4	0,056743101	0,142721966	0,272410681
5	0,049358725	0,099964616	0,042964031
6	0,046638166	0,056979675	0,111459583
7	0,038865138	0,007516593	0,008515794
8	0,037310532	0,048190741	0,215037669
9	0,03381267	0,087922315	0,128106399
10	0,031480762	0,07894859	0,204574996
11	0,030314808	0,0555527	0,01115609
12	0,023319083	0,049015734	0,153465642
13	0,022930431	0,038278612	0,198719105
14	0,016712009	0,024024563	0,005532817
15	0,015546055	0,012520585	0,002904175
16	0,01127089	0,045379783	0,195821616
17	0,01127089	0,030091145	0,090063158
18	...	...	...
19	Коэффициент корреляции	0,62	0,45

Для наглядности рассмотрим результаты сравнения центральности по взвешенному вкладу с другими метрик центральности вершин в виде графика, представленного на рисунке 26.



Рисунок 26 – Сравнение центральности по взвешенному вкладу с другими метриками

Как видно из рис.26, центральность по взвешенному вкладу и центральность по промежуточности наиболее близки между собой. То есть наиболее влиятельные мосты имеют высокие значения центральности по промежуточности. В то же время, центральность по вкладу повторяет общую закономерность, но имеет очень большой разброс значений. Кроме того, видно, что мосты с низкими значения центральности по взвешенному вкладу имеют достаточно высокие значения центральности по вкладу. Отсортируем те же данные по промежуточности и представим результат на рис.27.



Рисунок 27 – Сравнение центральности по промежуточности с другими метриками

На графике (рис.27) видно, что вершина, имеющая максимальное значение центральности по промежуточности, не является мостом. Также на графике встречается много нулевых значений центральности по взвешенному вкладу в средней и правой части графика. Это означает, что такие узлы не являются мостами. В тоже время, в левой части графика, не считая первого значения, все мосты имеют высокой значение центральности по промежуточности.

Для того, чтобы соотнести центральность по вкладу и центральность по взвешенному вкладу, отсортируем данные по значениям центральности по вкладу и представим в виде графика на рисунке 28.



Рисунок 28 – Сравнение центральности по вкладу с другими метриками

Из этого графика (рис.28) видно, что при высоких значениях центральности по вкладу довольно много вершин не является мостами по нашему определению, а узлы, не являющиеся мостами, имеют высокие значения центральности по вкладу.

Таким образом, центральность по взвешенному вкладу не противоречит центральности по промежуточности, что вполне объяснимо, так как обе эти метрики характеризуют коммуникативные способности узлов. Уровень их корреляции составляет 0,62. В тоже время, центральность по вкладу и центральность по взвешенному вкладу явно имеют дело с разными типами узлов при уровне взаимной корреляции 0,45.

Самое главное отличие сравниваемых метрик состоит в том, что центральность по взвешенному вкладу однозначно определяет мосты, тогда как промежуточность и центральность по вкладу дают вероятностную характеристику и определяют центральность любых типов узлов, часть из которых мостами не являются.

Для того, чтобы убедиться, что данные выводы справедливы и для других кластерных графов, выполним те же операции для 10 различных графов и рассмотрим корреляцию между метриками, представленную в таблице 19.

Таблица 19 - Корреляция между различными метриками центральности для 10 графов.

№ п/п	Центральность по промежуточности -Центральность по взвешенному вкладу	Центральность по вкладу - Центральность по взвешенному вкладу
1	2	3
1	0,618296614	0,453709388
2	0,900542062	0,670132895
3	0,230838433	0,276205477
4	0,920268705	0,737462024
5	0,976781061	0,871017217
6	0,94954293	0,734470215
7	0,900449697	0,516278165
8	0,908968882	0,520256952
9	0,649853573	0,419937569
10	0,941472856	0,694827395
11	<i>Среднее значение</i>	<i>Среднее значение</i>
12	<b>0,799701481</b>	<b>0,58942973</b>

Из таблицы 19 следует, что центральность по взвешенному вкладу не только однозначно выявляет мосты, но и наиболее эффективно выявляет узлы сети, обладающие наибольшей коммуникативной способностью. Таким образом, можно утверждать, что центральность на взвешенном вкладе наиболее эффективно решает задачу по выявлению мостов по сравнению с другими представленными метриками.

### 3.3. Идентификация наиболее влиятельных мостов

При сборе данных для расчета значения ПУВ, как было сказано выше, был выполнен сбор информации о социальных связях вовлеченных пользователей и построен двухколенный граф социальных связей. Количество связей в графах по каждому инфоповоду исчисляется в миллиардах. Метод выявления вовлеченных пользователей, основанный на уровне публикационной активности, не считает вовлеченными большую часть мостов, так как они не публикуют контент по целевой тематике, но как было показано имеют важное значение для расширения охвата аудитории и повышения общего уровня публикационной активности сети. Для выявления мостов из данного графа был применен алгоритм вычисления центральности по взвешенному вкладу для всех узлов сети и таким образом были выявлены узлы сети, которые играют роль мостов. Следует обратить внимание, что мост, это тоже роль, но в отличие от ролей, основанных на интерпретации паттернов поведения пользователей, мосты определяются исключительно на основе их структурных характеристик, то есть на основе теоретико-графового подхода. В результате выполнения алгоритма в графе социальных связей пользователей, вовлеченных в тематику инфоповода «Дворец Путина», было выявлено 21261 мост. Очевидно, что не все мосты одинаково влиятельны. Большая часть мостов подключают к ядру небольшое количество пользователей с очень низкой активностью, соответственно вес таких небольших кластеров близок к нулю. Также как и для других ролей, необходимо было выявить наиболее влиятельные мосты, воздействие на которые могло бы существенно снизить общий вес сети, охват аудитории и уровень публикационной активности.

Влиятельность мостов рассчитывалась на основе веса кластера, подключаемого мостом к ядру сети. Вес кластера, это сумма количества материалов целевой тематики, опубликованных пользователями в кластере, который стоит за мостом. Чем больше суммарный уровень активности пользователей в кластере, тем более влиятельным считался мост. Далее был



построен рейтинг мостов по убыванию веса и взяты первые  $N$  мостов, суммарный вес кластеров которых составлял половину общего веса всех кластеров. Методика расчета половины массы рассчитывалась по формуле (6) описана в главе 1.7. Далее был построен граф связей этих мостов между собой и отброшены те мосты, у которых количество связей с другими мостами меньше 2. Полученный таким образом список мостов был оценен по весу кластера и по количеству связей с другими мостами путем построения диаграммы Венна. Таким образом был получен перечень наиболее влиятельных мостов, в состав которых вошли мосты с наибольшим количеством связей с другими мостами и наибольшим весом кластера (Супер-мосты). Результаты выявления НВП для всех типов ролей по инфоповоду «Дворец Путина» представлены в таблице 20.

Таблица 20 – Статистические данные по ролям пользователей и НВП для инфоповода «Дворец Путина»

<b>№ п/п</b>	<b>Параметр</b>	<b>Всего</b>	<b>Наиболее влиятельных</b>
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
1	Количество мостов	21 261	102
2	Кол-во вовлеченных пользователей, в том числе:	35 502	231
3	Количество Постеров	531	109
4	Количество Репостеров	1 050	58
5	Количество Комментаторов	132	17
6	Количество Универсалов	3 134	47
7	Количество пассивных пользователей	30 655	
8	Итого	56 763	333

Как следует из таблицы 20, из более чем 21 тысячи выявленных мостов, только 102 моста являются наиболее влиятельными, что существенно снижает трудозатраты на аналитические исследования и повышает эффективность работы сотрудников правоохранительных органов.

### 3.4. Моделирование структуры временной протестной социальной сети

Как следует из таблицы 20, из более чем 56 тысяч пользователей, наиболее влиятельными являются всего лишь 333 пользователя, из которых наиболее влиятельные мосты (Супер-мосты) составляют 30,6%. Рассмотрим структуры графов наиболее влиятельных пользователей (Супер-пользователей) и Супер-мостов, представленных на рисунках 30 и 31, где уровень влиятельности узла соответствует размеру точки в графе.

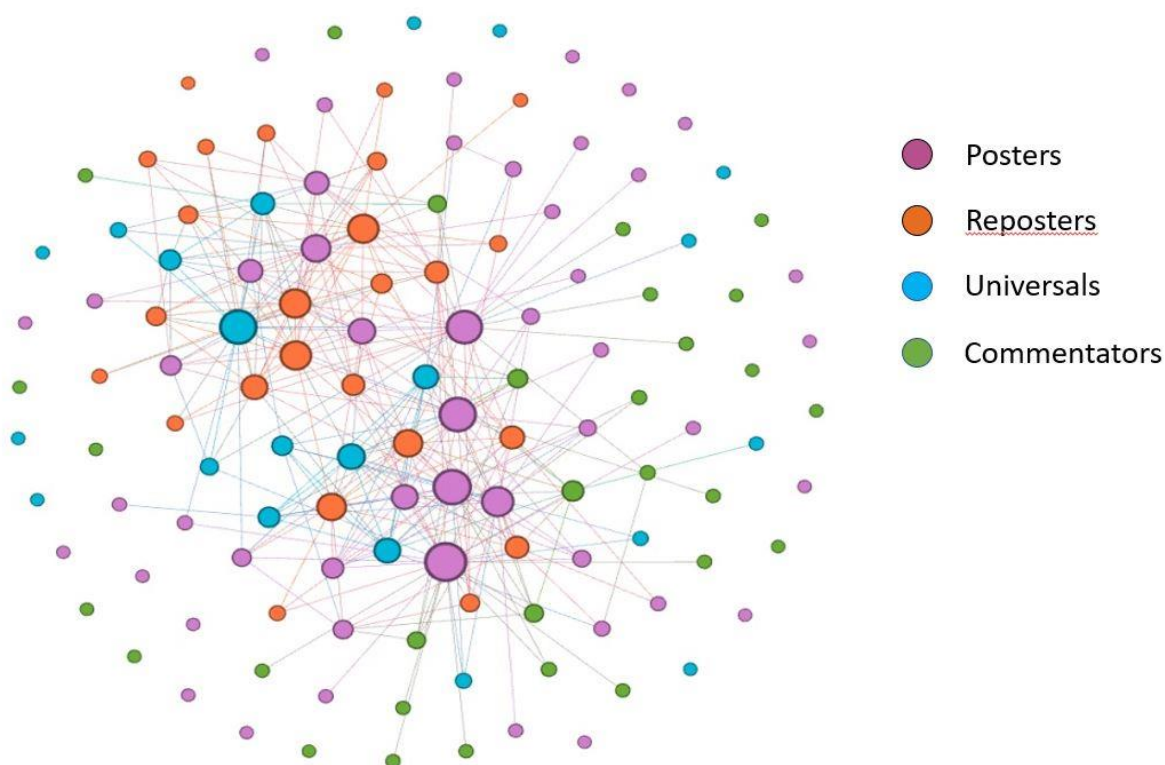


Рисунок 29 – Структура сети наиболее влиятельных Постеров, Репостеров, Комментаторов и Универсалов

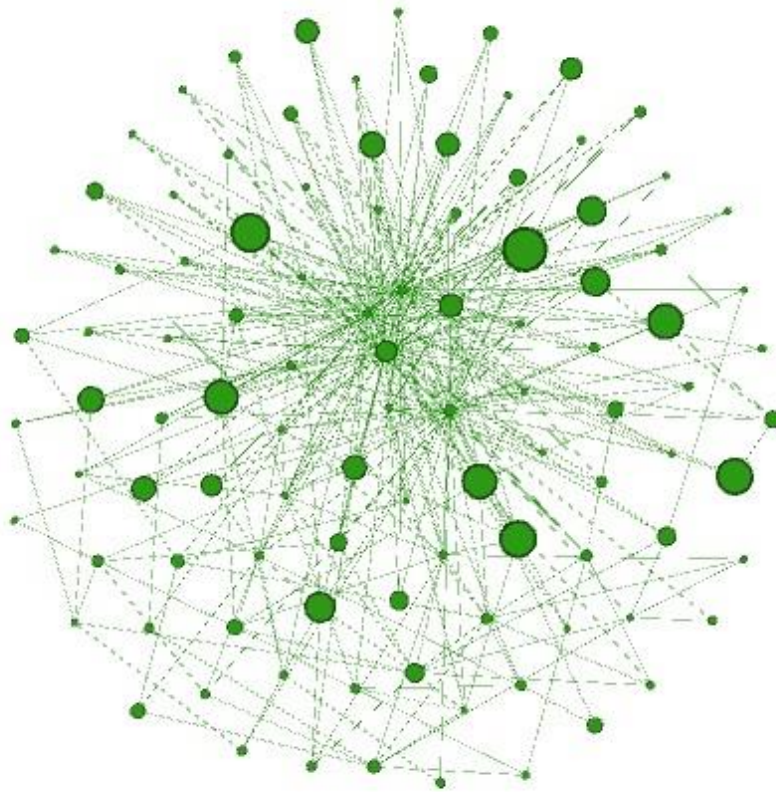


Рисунок 30 – Структура сети наиболее влиятельных Мостов

Как видно из рисунка 29, структура графа Супер-Пользователей содержит большое количество узлов, не связанных или слабо связанных с другими узлами сети. В тоже время граф Супер-Мостов (рис. 30) является полно-связным, где все узлы связаны с другими узлами графа. Сводный граф социальных связей Супер-Пользователей и Супер-Мостов представлен на рисунке 31.

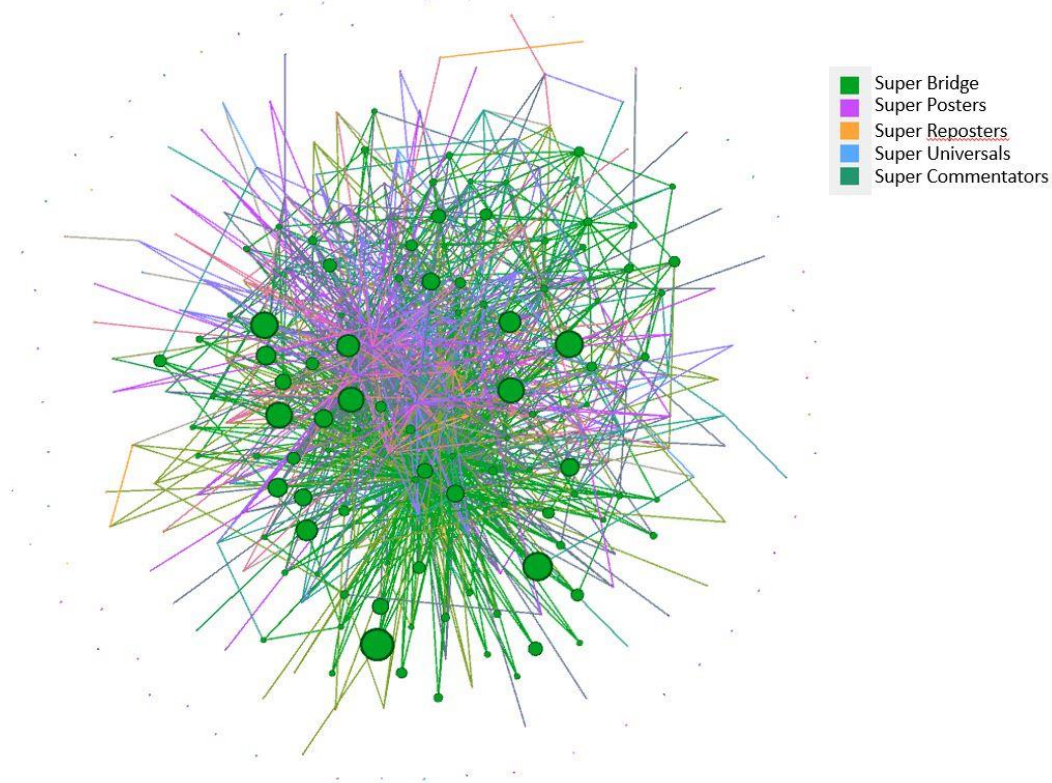


Рисунок 31 – Сводный граф связей Супер-Пользователей и Супер-Мостов

На основании анализа количественных характеристик и структуры сети протестного движения были получены следующие результаты:

- Пользователей, выполняющих роль Мостов приблизительно столько же, сколько всех остальных типов ролей пользователей вместе взятых.
- Супер-Мостов больше, чем Супер-Пользователей каждой отдельной роли, за исключением Супер-Постеров.
- Мосты более связаны между собой, в отличие от других ролей.
- Супер-Мосты образуют полно-связную сеть между собой, к которой подключены другие Супер-Пользователи.

Изучение сводного графа (см.рис. 31) показывает, что сеть Супер-Мостов является основной полно-связной структурной тканью, на которой крепятся Супер-Пользователи, в большей или меньшей степени связанные друг с другом. То есть, некоторые Супер-Пользователи могут быть не связаны с другими Супер-

Пользователями, но они обязательно связаны с сетью с Супер-Мостов и через них с другими пользователями в сети.

Несмотря на то, что исследование проводилось на основе данных пользователей социальной сети ВКонтакте, его результаты применимы также для Facebook и других подобных социальных сетей. В представленной модели структуры пропаганды протестного движения Постеры, Репостеры и Универсалы отвечают за внедрение контента в социальную сеть. Поскольку количество Супер-Пользователей невелико, а их сеть является слабо связанной, блокирующее воздействие на самых влиятельных пользователей может существенно снизить параметры распространения контента. В тоже время мосты, представляя из себя обширную полно-связную сеть, являются гораздо более устойчивой структурой. Следовательно, если в момент вброса контента в сеть заблокировать супер-пользователей, то, параметры распространения контента по сети изменятся значительно, а если контент уже вброшен, то через сеть мостов, устойчивую к разрушению, он будет распространен через Репостеров и Универсалов как в ядре сети, так и в кластерах, которые стоят за мостами.

Таким образом, значение мостов очень велико, так как они обеспечивают устойчивость сети к разрушающему воздействию и способствуют максимально быстрому росту охвата аудитории.

Представляет интерес построение модели поведения Постеров, Репостеров, Комментаторов, Универсалов и Мостов в процессе информационного воздействия на социальную сеть и подтверждение изложенных выше выводов на основе методов, с помощью которых возможно получение численных оценок изменения параметров распространения контента при изменении конфигурации сети.

В качестве результатов моделирования необходимо рассматривать получение рекомендаций по наиболее эффективному воздействию на сеть для предотвращения распространения деструктивного контента протестного характера.

### 3.5. Выводы по главе 3

Главная особенность метрики «Центральность по взвешенному вкладу» состоит в том, что она однозначно определяет является ли узел мостом в описанных ранее сетевых конфигурациях.

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 способом снижения уровня охвата аудитории и активности временной социальной сети.

Оценка уровня информационного влияния мостов осуществляется путем ранжирования узлов по метрике центральности по взвешенному вкладу. Алгоритм вычисления мостов в кластерных сетях на основе метрики центральности по взвешенному вкладу реализован в «Программе вычисления мостов в кластерных сетях» (Свидетельство о государственной регистрации программы для ЭВМ № 2021616086 от 16 апреля 2021г.).



## ГЛАВА 4. ВНЕДРЕНИЕ МНОГОФАКТОРНОГО АНАЛИЗА ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ

### 4.1. Прототип системы выявления наиболее влиятельных пользователей, мостов и определения уровня их информационного влияния

Внедрение многофакторного анализа в производственный процесс компании «СЕУСЛАБ» было выполнено в два этапа: прототипирование и разработка приложения для промышленной эксплуатации. Прототипирование выполнялось с использованием нейросетевого пакета Нейросимулятор 5.0, пакета визуального программирования Orange Data Mining, табличного процессора MS Office 2017 и графического пакета Gephi. На рисунке 33

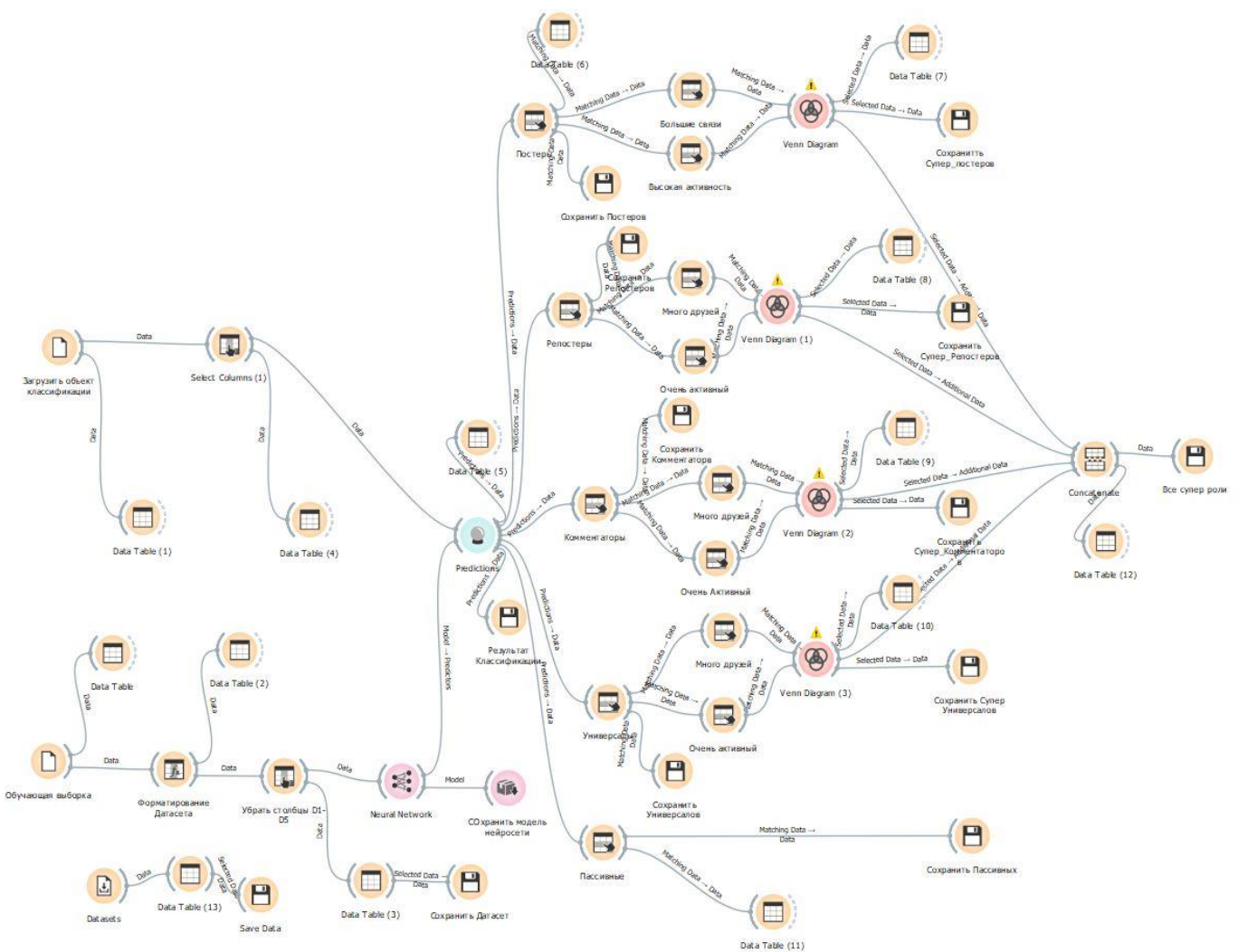


Рисунок 32 – Сценарий нейросетевой классификации и выявления наиболее влиятельных пользователей в каждой роли

представлен сценарий нейросетевой классификации пользователей и выявления наиболее влиятельных пользователей в каждой роли с использованием пакета визуального программирования Orange Data Mining. С его помощью загружается файл с данными о пользователях и файл обучающей выборки. Далее выполняется предобработка файла датасета и нейросетевая классификация объекта классификации. После чего в каждой роли отбираются самые активные пользователи и пользователи с большим количеством связей и затем методом диаграммы Венна выявляются Супер-Постеры, Супер-Репостеры, Супер-Комментаторы и Супер-Универсалы.

Использование данного пакета очень удобно, так как не требует знания языков программирования и позволяет оперативно вносить изменения в сценарий. Программа вычисления мостов в кластерных сетях была реализована на языке Python 3. Результаты вычисления мостов передавались для дальнейшей обработки в Excel файлах.

Далее с помощью пакета Gephi выполнялся графовый анализ структуры выявленных наиболее влиятельных пользователей и наиболее влиятельных мостов. Пример анализа графа наиболее влиятельных пользователей представлен на рисунке 33.



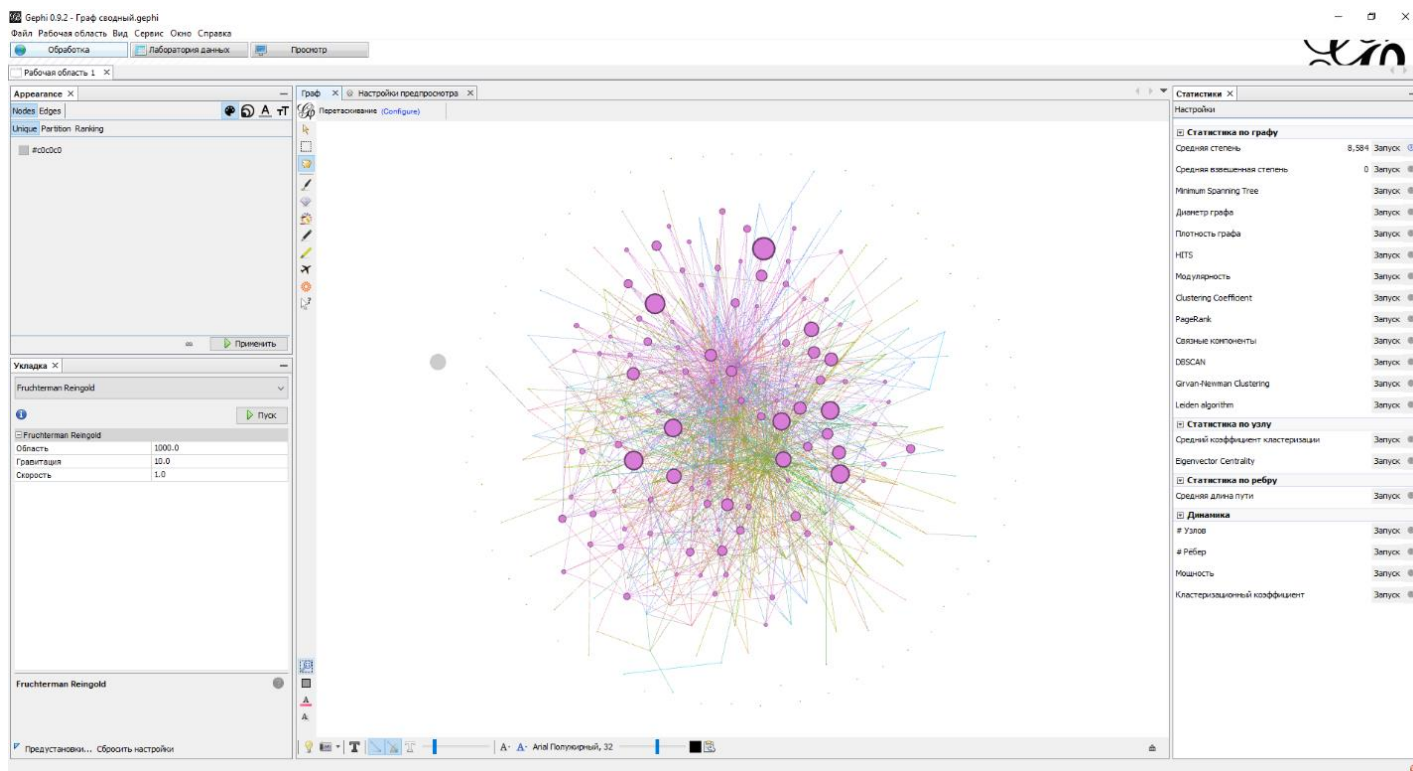


Рисунок 33 – Анализ сводного графа наиболее влиятельных пользователей

После проверки результатов работы прототипа путем проведения аналитических исследований на предмет релевантности выявленных пользователей, уровня их вовлеченности в целевую тематику и оценки уровня их информационного влияния, было принято решение о создании приложения, пригодного к промышленной эксплуатации.

#### 4.2. Приложение «Сервис анализа распространения контента»

На основе предложенных в данном исследовании методик выявления ролей пользователей и уровня их информационного влияния в компании СЕУСЛАБ было разработано приложение «Сервис анализа распространения контента». Данное приложение реализовано в виде веб сервиса, доступного пользователям по подписке. Ниже представлены основные функциональные требования и требования к выходным данным, получаемым в результате работы приложения.

#### **4.2.1. Сервис анализа распространения контента. Функциональные требования**

Сервис должен выполнять поиск публикаций и обсуждений, соответствующих заданному информационному поводу, совершенных пользователями, отвечающими определенной целевой аудитории.

- 1) Для выполнения поиска Сервис должен позволять описывать следующие входные данные:
  - Перечень ключевых слов и словосочетаний по теме, которая описывает информационный повод;
  - Временной интервал поиска публикаций.
- 2) Дополнительно Сервис должен позволять осуществлять фильтрацию публикаций по параметрам авторов этих публикаций, определяющих целевую аудиторию искомых публикаций:
  - География целевой аудитории, с указанием страны, федерального округа (только для России), региона (край, область, штат) и города;
  - Пол целевой аудитории (мужской / женский);
  - Возраст целевой аудитории в виде интервала с указанием неотрицательных целых чисел.
- 3) Сервис должен предоставлять выбор осуществления поиска следующих видов публикаций:
  - Пост (в т.ч. содержащий вложение в виде видеозаписи, аудиозаписи, ссылки, фотографии);
  - Видеозапись;
  - Комментарий;
  - Обсуждение.
- 4) Сервис должен устанавливать связи распространения между публикациями, соответствующими заданному информационному поводу.

Примечание. Связь репоста между постами может быть построена без учета промежуточных связей. Это значит, что любые репосты относятся к

оригинальному посту, который был опубликован раньше всех репостов. Такое ограничение связано с форматом данных, предоставляемых социальной сетью ВКонтакте.

- 5) Сервис должен предоставлять выбор поиска связей распространения следующих типов:
  - Репоста (между двумя постами, между постом и видео);
  - Комментирования (между двумя комментариями, между постом и комментарием, между видео и комментарием, между обсуждением и комментарием).
- 6) Сервис должен предоставлять возможность восстановления путей распространения постом через связь репоста на основе социальных связей с указанием вероятности осуществления репоста от одного пользователя к другому.
- 7) Сервис должен предоставлять возможность определения пользователей, которые являются мостами, согласно метрике центральности по взвешенному вкладу на основе социального графа пользователей, связанных с авторами публикаций, соответствующих заданному информационному поводу.
- 8) Сервис должен предоставлять возможность определения ключевых мостов следующими методами:
  - Определение первых  $N$  пользователей с наибольшим значением метрики центральности по взвешенному вкладу;
  - Определение пользователей с наибольшим значением метрики центральности по взвешенному вкладу согласно квантилю;
  - Определение пользователей с наибольшим значением метрики центральности по взвешенному вкладу согласно коэффициенту отсечки.

- 9) Сервис должен предоставлять возможность определения типов социальных связей, на основе которых строится граф для определения мостов:
- Дружбы;
  - Подписки.
- 10) Сервис должен предоставлять возможность определения пользователей, которые являются постерами, репостерами, комментаторами и универсалами, согласно степени активности пользователей, являющихся авторами публикаций, соответствующих заданному информационному поводу.
- 11) Сервис должен предоставлять возможность определения ключевых пользователей перечисленных ролей следующими методами:
- Определение первых  $N$  пользователей с наибольшим значением степени потенциального влияния пользователя;
  - Определение пользователей с наибольшим значением степени потенциального влияния пользователя согласно квантилю;
  - Определение пользователей с наибольшим значением степени потенциального влияния пользователя согласно коэффициенту отсечки.
- 12) Сервис должен предоставлять возможность определения ключевых пользователей по ролям следующими способами:
- В отдельности для каждой роли;
  - В общем списке, содержащем всех постеров, репостеров, комментаторов и универсалов.

Место Сервиса в общей архитектуре системы SEUS обозначено на рисунке 34 (диаграмма компонентов в нотации UML 2.x).

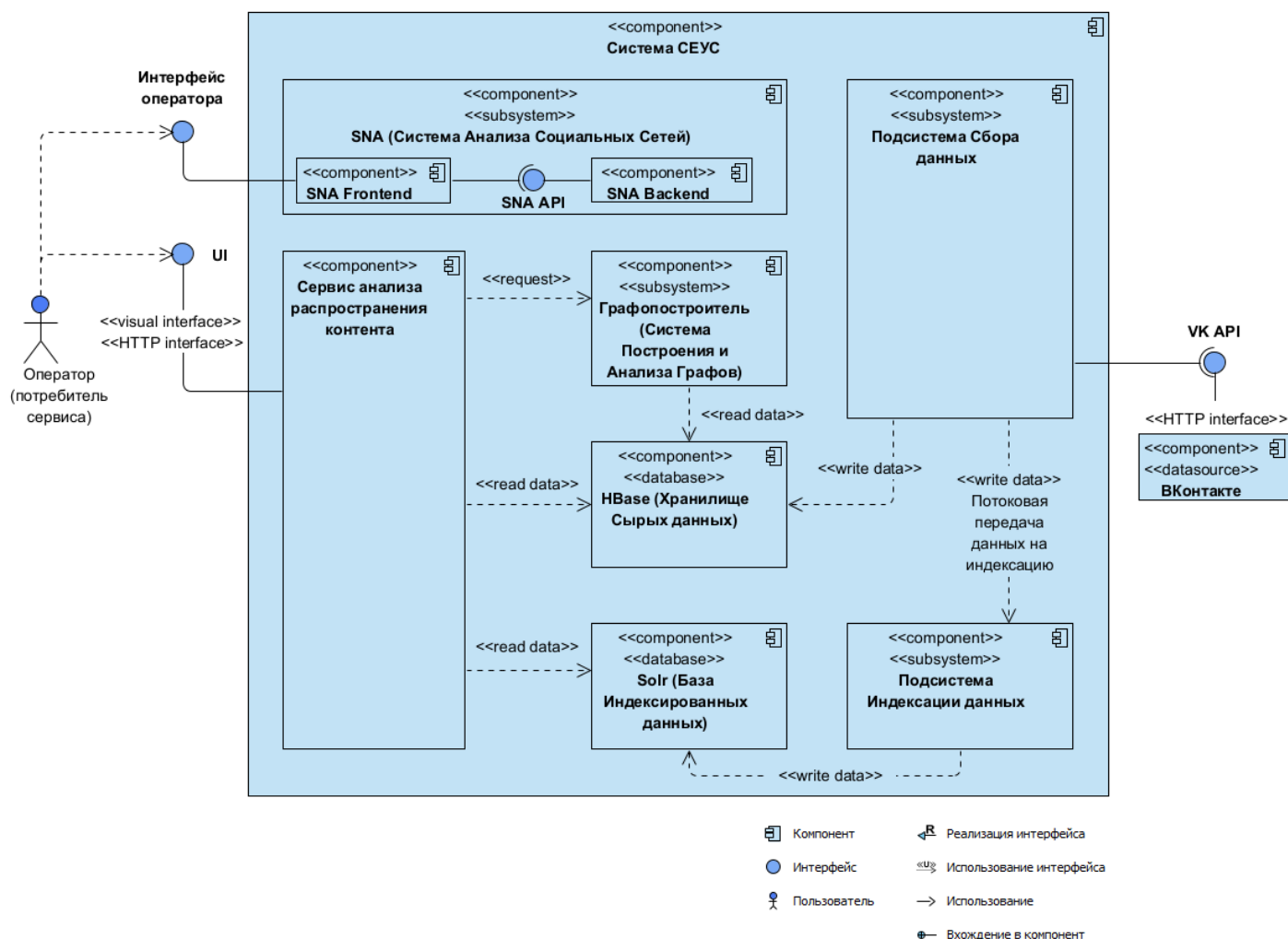


Рисунок 34 – Архитектура сервиса анализа распространения контента

#### 4.2.2. Сервис анализа распространения контента. Требования к выходным данным

- 1) Сервис должен предоставлять возможность экспорта списка публикаций (коллекции), полученных в процессе поиска в формате файла \*.csv, с указанием следующих атрибутов:
  - Уникальный идентификатор публикации в пределах списка (используется для определения связей распространения между публикациями);

- Текстовое содержание публикации (не указывается для обсуждений);
  - Ссылка на публикацию (или обсуждение);
  - Дата и время публикации (или создания обсуждения);
  - Тип публикации;
  - Ссылка на автора публикации (или создателя обсуждения);
  - Общее количество репостов публикации (только для постов, не являющихся репостами, и видео), представленных в коллекции;
  - Общее количество прямых комментариев к публикации или обсуждению, представленных в коллекции;
  - Общее количество всех комментариев (включая все промежуточные комментарии, оставленные к комментариям публикации) к публикации или обсуждению, представленных в коллекции.
- 2) Сервис должен предоставлять возможность экспорта информации о связях распространения между публикациями в формате \*.csv с указанием следующих атрибутов:
- Уникальный идентификатор (или ссылка) публикации или обсуждения, являющегося началом связи;
  - Уникальный идентификатор (или ссылка) публикации, являющейся концом связи;
  - Тип связи;
  - Время образования связи (т. е. время совершения репоста или написания комментария);
  - Оценка вероятности связи (в случае восстановления промежуточных репостов).
- 3) Сервис должен предоставлять возможность экспорта списка пользователей (выборки), полученных в процессе анализа в формате файла \*.csv с указанием следующих атрибутов:

- Идентификатор пользователя в социальной сети;
  - Ссылка на пользователя в формате `https://vk.com/id<Идентификатор_пользователя>`;
  - Имя;
  - Фамилия;
  - Количество друзей, представленных в выборке;
  - Количество подписчиков, представленных в выборке;
  - Количество пользователей, на которых подписан данный пользователь и которые представлены в выборке;
  - Количество публикаций, найденных на странице данного пользователя;
  - Количество опубликованных пользователем постов среди найденных публикаций;
  - Количество опубликованных пользователем видеозаписей среди найденных публикаций;
  - Количество совершенных репостов среди найденных публикаций;
  - Количество совершенных комментариев среди найденных публикаций;
  - Количество созданных обсуждений среди найденных публикаций.
- 4) Сервис должен предоставлять возможность экспорта списков ключевых пользователей, по ролям постеров, репостеров, комментаторов и универсалов в ранжированном виде относительно убывания степени потенциального влияния пользователя в форматах общего файла \*.csv (содержащего пользователей всех ролей) или отдельных файлов \*.csv классифицированных по перечисленным ролям с указанием следующих атрибутов:
- Идентификатор пользователя (или ссылка);
  - Значение степени потенциального влияния пользователя;
  - Роль пользователя (в случае экспорта общего файла).

5) Сервис должен предоставлять возможность экспорта списков ключевых пользователей, являющихся мостами, в ранжированном виде относительно убывания центральности по взвешенному вкладу с указанием следующих атрибутов:

- Идентификатор пользователя;
- Значение центральности по взвешенному вкладу.

б) Сервис должен предоставлять возможность экспорта информации о социальных связях между пользователями, найденных в процессе выявления мостов в формате файла \*.csv с указанием следующих атрибутов:

- Идентификатор пользователя (или ссылка), являющегося началом связи;
- Идентификатор пользователя (или ссылка), являющегося концом связи;
- Тип связи.

Таким образом, Сервис должен поддерживать экспорт 10 различных файлов формата \*.csv:

- Коллекция публикаций и обсуждений;
- 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.



Модель отношений между экспортируемыми из Сервиса сущностями приведена на рисунке 35 (ER-диаграмма в нотации Crow's foot).

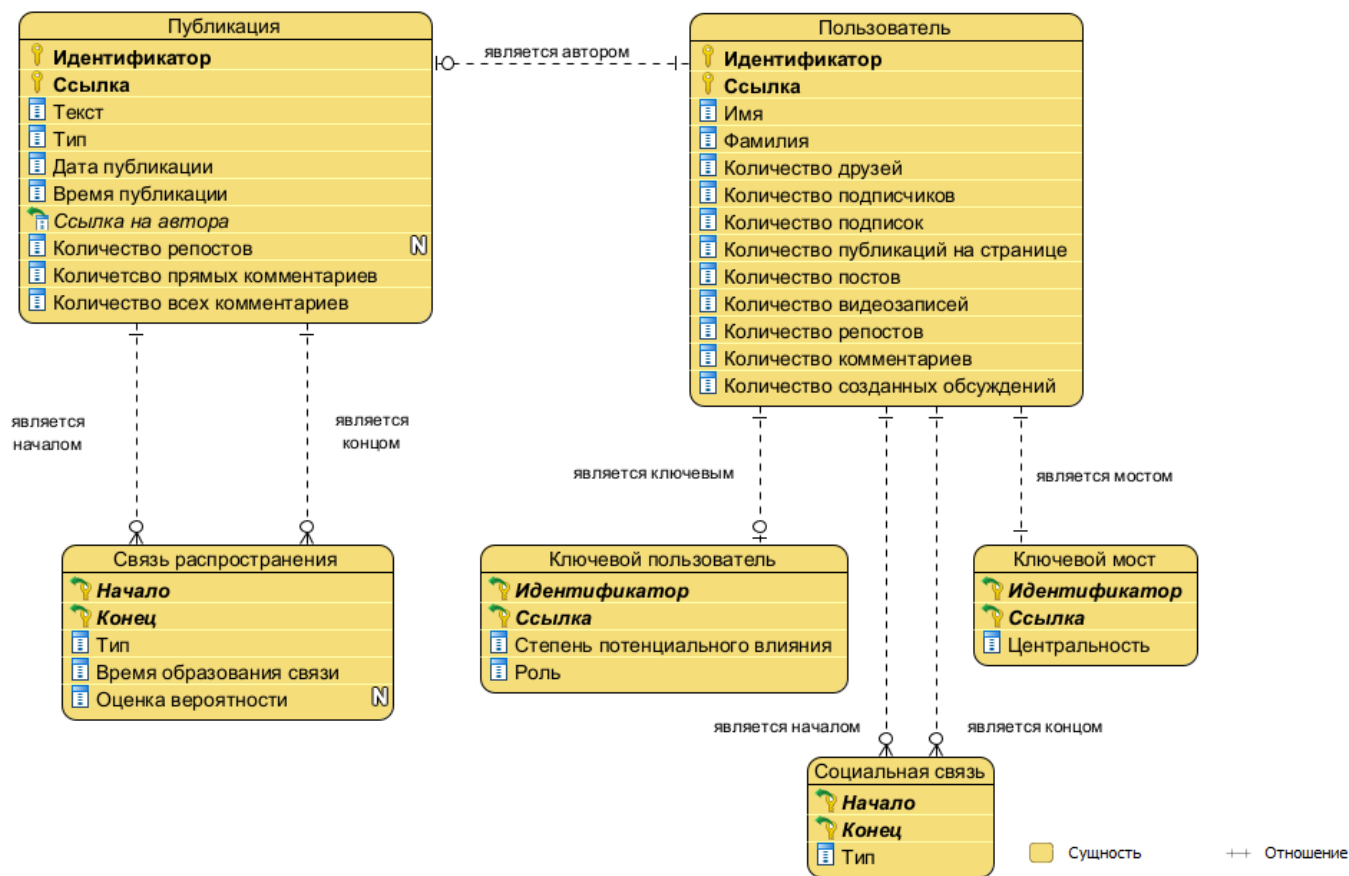


Рисунок 35 - Логическая модель выходных данных

### 4.3. Оценка эффективности метода многофакторного анализа пользователей социальных сетей

Для оценки эффективности предложенного метода было выполнено сравнение трудозатрат, необходимых для выявления наиболее влиятельных пользователей до и после внедрения метода. До внедрения результатов данного исследования в поисковой системе “SEUS”, наиболее влиятельные пользователи выявлялись только на основе анализа их публикационной активности. Однако, как было показано в главе 1.4, наиболее активные пользователи не обязательно обладали наибольшим уровнем влияния, поэтому для выявления НВП аналитикам необходимо было проанализировать всех наиболее активных пользователей для того, чтобы выявить среди них наиболее влиятельных. Таким

образом можно сравнить количество профилей пользователей, которые необходимо было исследовать аналитикам до и после внедрения. Результаты сравнения представлены в таблице 21.

Таблица 21 – Результаты сравнения количества профилей, подлежащих исследованию до и после внедрения метода

№ п/п	Параметр	Жыве Беларусь		Дворец Путина	
		До внедрения	После внедрения	До внедрения	После внедрения
1	2	3		4	
1		До внедрения	После внедрения	До внедрения	После внедрения
2	Количество Постеров	197	49	521	109
3	Количество Репостеров	235	25	1050	58
4	Количество Комментаторов	555	30	132	17
5	Количество Универсалов	392	25	649	36
6	Итого	1379	129	2362	220
7	Среднее значение	344,8	32,3	590,5	55,0
8	Соотношение количества профилей		10,7		10,7
9	Соотношение количества профилей в одной роли		4,0		4,0

#### 4.4. Выводы по главе 4

Использование стандартных пакетов позволило существенно сократить затраты и время на создание прототипа программного приложения. По результатам, полученным с помощью прототипа, были проведены аналитические исследования, которые показали высокий уровень релевантности выявленных пользователей и их ролей, что являлось целью настоящего исследования. Методики, разработанные в данном исследовании, реализованы в виде «Сервиса анализа распространения контента», предназначенного для промышленной эксплуатации.

Применение предложенного метода выявления наиболее влиятельных пользователей на основе вычисления ПУВ позволило сократить необходимое количество исследуемых профилей пользователей в 10,7 раза, таким образом увеличив производительность труда экспертов аналитиков. Кроме того, поскольку в число наиболее влиятельных пользователей входят пользователи, выполняющие различные роли, проведение аналитиками более детальных исследований, учитывающих структуру воздействия на сеть благодаря внедрению метода выявления ролей пользователей позволило дополнительно сократить трудозатраты аналитиков в 4 раза.

Внедрение метода выявления мостов на основе вычисления центральности по взвешенному вкладу позволило выявлять новый класс ключевых узлов, отвечающих за скорость и широту охвата аудитории, которые до внедрения этого метода в принципе не могли быть выявлены поисковой системой “SEUS”, так как они не проявляют публикационной активности, хотя и обладают уникальными структурными характеристиками.

## ЗАКЛЮЧЕНИЕ

Был проведен анализ существующих методов определения уровня информационного влияния пользователей в социальных сетях и их социальных ролей с целью определения возможности выявления элементов структуры информационного воздействия на социальные сети. В результате было определено, что для выявления наиболее влиятельных пользователей особенно выделяются алгоритмы, основанные на учете публикационной активности пользователей и количестве их социальных связей. Основным препятствием существующих методов является их алгоритмическая сложность и высокие требования к затратам вычислительных ресурсов. Для повышения производительности вычисления уровня информационного влияния был предложен метод вычисления потенциального уровня влияния, определяющий максимально возможное количество актов донесения информации до пользователей социальных сетей.

Предложенный метод выявления наиболее влиятельных пользователей имеет преимущество над существующими методами не только по простоте и скорости, но и по уровню релевантности выявленных пользователей, что снижает требования к объему подробных аналитических исследований при проверке специалистами выявленных пользователей с оперативной и политологической точек зрения.

Также было определено, что для выявления социальных ролей пользователей в социальных сетях в контексте противодействия деструктивным воздействиям протестного характера существуют только теоретические подходы, но не существует устоявшейся методики. Подходы, основанные на кластеризации, показали очень низкий уровень релевантности. Более релевантными являются подходы, основанные на нейросетевой классификации, но для их реализации необходимы очень высокие затраты ресурсов для подготовки датасетов, используемых для обучения нейросетей. Поэтому для выявления ролей пользователей был предложен метод создания датасета на

основе синтетических данных, алгоритм создания которых учитывает знания экспертов в данной предметной области. Предложенный метод, при высоком уровне его релевантности, позволяет отказаться от разметки сотен тысяч профилей реальных пользователей, что снижает затраты на создание датасета в тысячи раз, и исключает риск утечки персональных данных реальных пользователей социальных сетей.

Помимо ролей пользователей, определенных на основании интерпретации паттернов их поведения в социальных сетях в ходе исследования были также выявлены пользователи, не имеющие публикационной активности, но играющие важную роль в расширении охвата протестной аудитории в виде мостов между ядром сети и сообществами, модераторами которых они являются. Было определено, что все существующие методики выявляют мосты с определенной долей вероятности, но не существует методик, точно определяющих является ли исследуемый узел социальной сети мостом. Поэтому был предложен метод вычисления мостов в кластерных сетях на основании оригинальной метрики центральности по взвешенному вкладу. Основным преимуществом этого метода является то, что он позволяет однозначно определять, является исследуемый узел мостом или нет. Кроме того, использование центральности по взвешенному вкладу позволяет определять уровень информационного влияния каждого моста и выявлять наиболее влиятельные мосты в сети.

Сопоставление графов наиболее влиятельных пользователей и наиболее влиятельных мостов позволило выявить структуру воздействия на социальную сеть и позволило определить элементы структуры, наиболее пригодные для воздействия в контексте противодействия деструктивным влияниям протестного характера в социальных сетях.

Эксперименты по применению предложенных методов были проведены при проведении аналитических исследований пользователей социальной сети ВКонтакте, вовлеченных в протестное движение и активно участвующих в публикационной активности по информационным поводам, связанным с выборами президента республики Беларусь (инфоповод «Жыве Беларусь») и

обсуждением так называемого дворца Путина (инфоповод «Дворец Путина»), в результате чего были выявлены основные акторы, участвующие в создании и распространении информации, а также были получены сведения о заговоривших следах иностранного вмешательства.

На основании результатов, полученных в ходе экспериментов, было разработано программное приложение «Сервис анализа распространения контента», внедрение которого позволило решить следующий ряд проблем:

- Выявление ролей пользователей социальных сетей на основе паттернов их поведения;
- Выявление пользователей, выполняющих роль мостов между сообществами и ядром временной социальной сети;
- Определение уровня информационного влияния пользователей с учетом социальной роли пользователя;
- Выявление структуры деструктивного воздействия на социальные сети;
- Повышение производительности труда специалистов при проведении подробных аналитических исследований за счет снижения количества целевых пользователей и повышения их релевантности.

Данные, получаемые в результате применения Сервиса анализа распространения контента, позволяют выявлять структуру воздействия на социальную сеть, элементы этой структуры, а также признаки целенаправленного воздействия на социальную сеть, поэтому могут быть использованы в качестве инструмента интеллектуальной поддержки принятия решения по блокировке, контролю, оперативному и другим видам воздействий на пользователей социальных сетей в контексте противодействия деструктивным воздействиям на социальные сети.

**Основные результаты и выводы.** В рамках проведенного исследования были получены следующие результаты, обладающие научной и практической значимостью:

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

2. Предложенный метод идентификации пользователей, выполняющих роль мостов между ядром социальной сети и изолированными кластерами пользователей на основе расчета метрики центральности по взвешенному вкладу, позволил выявить пользователей, играющих решающую 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 наиболее влиятельных пользователей на основе учета публикационной активности пользователей и количества их социальных связей позволил повысить релевантность выявления наиболее влиятельных пользователей 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. 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.

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

**Перспективы дальнейшего развития.** Применение многофакторного анализа ролей пользователей и уровня их влияния на социальную сеть позволяет выявлять элементы структуры целенаправленного воздействия на сеть. Дальнейшие исследования планируются в части выявления паттернов поведения

таких структур, цепочек распространения информации в пределах структур и установления закономерностей с целью разработки методики выявления признаков целенаправленного воздействия на социальную сеть, прогнозирования поведения воздействующей структуры и подготовки управляющей информации для наиболее эффективного противодействия деструктивным воздействиям на социальную сеть.



## СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

**НВП** – наиболее влиятельный пользователь;

**ПУВ** – потенциальный уровень влияния;

**СЕТЬ** – социальная сеть;

**СОЦИАЛЬНАЯ СЕТЬ** – Социальные средства интернет-коммуникаций;

**ПОЛЬЗОВАТЕЛЬ** – Пользователь социальной сети. Одному пользователю соответствует один аккаунт социальной сети;

**ПУБЛИКАЦИОННАЯ АКТИВНОСТЬ** – Количество опубликованных материалов в заданный промежуток времени;

**РЕЙТИНГ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ** – Результат сортировки публикационной активности пользователей по убыванию;

**ВОВЛЕЧЕННЫЙ ПОЛЬЗОВАТЕЛЬ** – Пользователь, проявляющий публикационную активность по заданной тематике;

**СООБЩЕСТВО** – Группа людей, объединенных одними взглядами, идеями, интересами или событиями, являющихся пользователями одной и той же социальной сети;

**КЛАСТЕР** – Сообщество пользователей, которое может рассматриваться как самостоятельная единица, обладающая определёнными свойствами.

**СПИСОК ЛИТЕРАТУРЫ**

1. Gleave, E. A Conceptual and Operational Definition of “Social Role” in Online Community / E. Gleave, H.T. Welser, T.M. Lento, M.A. Smith // 42nd Hawaii International Conference on System Sciences. IEEE (2009). – 2009. – DOI: 10.1109/HICSS.2009.6.
2. Füller, J. User Roles and Contributions in Innovation-Contest Communities / J. Füller, K. Hutter, J. Hautz, K. Matzler // Journal of Management Information Systems. – 2014. – Vol. 31. – P. 273–308. – DOI: 10.2753/MIS0742-1222310111.
3. Arazy, O. Functional Roles and Career Paths in Wikipedia / O. Arazy, F. Ortega, O. Nov, L. Yeo, A. Balila // Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, New York, NY, USA (2015). – 2015. – P. 1092–1105. – DOI: 10.1145/2675133.2675257.
4. Welser, H.T. Finding social roles in Wikipedia / H.T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, M. Smith // Proceedings of the 2011 iConference on - iConference '11. ACM Press, New York, New York, USA (2011). – 2011. – P. 122–129. – DOI: 10.1145/1940761.1940778.
5. Brandtzaeg, P.B. A typology of social networking sites users / P.B. Brandtzaeg, J. Heim // International Journal of Web Based Communities. – 2011. – Vol. – 7. – P. 28–51. – DOI: 10.1504/IJWBC.2011.038124.
6. Cicek, M. Social media marketing: exploring the user typology in Turkey / M. Cicek, rem E. Erdoğan // International Journal of Technology Marketing. – 2013. – Vol. – 8. – P. 254–271 – DOI: 10.1504/IJTMKT.2013.055343.
7. Arularasan, A.N. Identification and classification of best spreader in the domain of interest over the social networks / A.N. Arularasan, A. Suresh, K. Seerangan // Cluster Computing. – 2019. – Vol. 22. – P. 4035–4045. – DOI: 10.1007/s10586-018-2616-y.

8. Ruas, P.H.B. Identification and characterisation of Facebook user profiles considering interaction aspects / P.H.B. Ruas, A.D. Machado, M.C. Silva, M.R.G. Meireles, A.M.P. Cardoso, L.E. Zárata, C.N. Nobre // Behaviour & Information Technology. – 2019. – Vol. 38. – P. 858–872. – DOI: 10.1080/0144929X.2019.1566498.

9. Jabłońska, M.R. Artificial neural networks for predicting social comparison effects among female Instagram users / M.R. Jabłońska, R. Zajdel // PLOS ONE. – 2020. – Vol. 15. – DOI: 10.1371/journal.pone.0229354.

10. Lin, H. User-level psychological stress detection from social media using deep neural network / H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J. Huang, L. Cai, L. Feng // Proceedings of the 22nd ACM international conference on Multimedia. ACM, New York, NY, USA (2014). – 2014. – P. 507–516. – DOI: 10.1145/2647868.2654945.

11. Segalin, C. What your Facebook Profile Picture Reveals about your Personality / C. Segalin, F. Celli, L. Polonio, M. Kosinski, D. Stillwell, N. Sebe, M. Cristani, B. Lepri // Proceedings of the 25th ACM international conference on Multimedia. ACM, New York, NY, USA (2017). – 2017. – P. 460–468. – DOI: 10.1145/3123266.3123331.

12. Wijenayake, P. Automated Detection of Social Roles in Online Communities using Deep Learning / P. Wijenayake, D. de Silva, D. Alahakoon, S. Kirigeegamage // Proceedings of the 3rd International Conference on Software Engineering and Information Management. ACM, New York, NY, USA (2020). – 2020. – P. 63–68. – DOI: 10.1145/3378936.3378973.

13. Matsumoto, K. Classification of Emoji Categories from Tweet Based on Deep Neural Networks / K. Matsumoto, M. Yoshida, K. Kita // Proceedings of the 2nd International Conference on Natural Language Processing and Information Retrieval – NLPIR. ACM Press, New York, New York, USA (2018). – 2018. – P. 17–25. – DOI: 10.1145/3278293.3278306.

14. Sunghwan, M.K. Detecting Social Roles in Twitter / M.K. Sunghwan, W. Stephen, P. Cecile // Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media, Austin, TX. – 2016. – P. 34–40.

15. Dilmegani, G. Synthetic Data Use Cases & Applications in 2021. Top 20 / G. Dilmegani // URL: <https://research.aimultiple.com/synthetic-data-use-cases/> (дата обращения 15.07.2021).

16. Dilmegani, G. Synthetic Data Generation: Techniques, Best Practices & Tools / G. Dilmegani // URL: <https://research.aimultiple.com/synthetic-data-generation/> (дата обращения 29.07.2021).

17. Dilmegani, G. The Ultimate Guide to Synthetic Data in 2021 / G. Dilmegani // URL: <https://research.aimultiple.com/synthetic-data/> (дата обращения 29.07.2021).

18. Dilmegani G. The Ultimate Guide to Synthetic Data: Uses, Benefits & Tools / G. Dilmegani // URL: <https://research.aimultiple.com/synthetic-data/> (дата обращения 29.07.2021).

19. Nikolenko, S.I. Synthetic Data for Deep Learning. / S.I. Nikolenko // Springer International Publishing, Cham. – 2021. – DOI: 10.1007/978-3-030-75178-4.

20. Ясницкий Л.Н. О возможностях применения методов искусственного интеллекта в сфере туризма / Л.Н. Ясницкий, А.С. Бржевская, Ф.М. Черепанов // Сервис plus. – 2010. – № 4. С. 111–115.

21. Румянцев, М.А. Повышение качества обобщения нейронной сети с помощью добавления шума в обучающую выборку на примере задачи определения прочности бетона по составу входящих в него компонентов / М.А. Румянцев, Л.Н. Ясницкий // Часть I. Сборник статей по материалам Четвертой всероссийской научно-практической конференции, проводимой в рамках Пермского естественнонаучного форума «Математика и глобальные вызовы XXI века». – 2019 – С. 57–62.

22. Ясницкий Л.Н. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта / Л.Н. Ясницкий, В.И. Грацилев, Ю.С. Куляшова, Ф.М. Черепанов // Вестник Пермского университета. Серия: Философия. Психология. Социология. – 2015. – № 1. – С. 61–71.

23. Ясницкий Л.Н. Использование методов искусственного интеллекта в изучении личности серийных убийц / Л.Н. Ясницкий, С.В. Ваулева,

Д.Н. Сафонова, Ф.М. Черепанов // Криминологический журнал Байкальского государственного университета экономики и права. – 2015. – № 9. – С. 423–430.

24. Wei, H. Identifying influential nodes based on network representation learning in complex networks / H. Wei, Z. Pan, G. Hu, L. Zhang, H. Yang, X. Li, X. Zhou // PLOS ONE. – 2018. – Vol. 13. – DOI: 10.1371/journal.pone.0200091.

25. Zhang, Q. Link transmission centrality in large-scale social networks / Q. Zhang, M. Karsai, A. Vespignani // EPJ Data Science. – 2018. – Vol. 7. – 33. – DOI: 10.1140/epjds/s13688-018-0162-8.

26. Ghalmane, Z. Centrality in modular networks / Z. Ghalmane, M. El Hassouni, C. Cherifi, H. Cherifi // EPJ Data Science. – 2019. – Vol. 8. – 15. – DOI: 10.1140/epjds/s13688-019-0195-7.

27. Jensen, P. Detecting global bridges in networks / P. Jensen, M. Morini, M. Karsai, T. Venturini, M. Jacomy, J.-P. Cointet, P. Merckle, E. Fleury // IMA Journal of Complex Networks. – 2015. P. 1–14.

28. Alvarez-Socorro, A.J. Eigencentrality based on dissimilarity measures reveals central nodes in complex networks / A.J. Alvarez-Socorro, G.C. Herrera-Almarza, L.A. González-Díaz // Scientific Reports 17095. – 2015. – Vol. 5. – DOI: 10.1038/srep17095.

29. Freeman, L.C. A Set of Measures of Centrality Based on Betweenness / L.C. Freeman // Sociometry. – 1977. – Vol. 40 – DOI: 10.2307/3033543.

30. Freeman, L.C. Centrality in social networks conceptual clarification / L.C. Freeman // Social Networks. – 1978. – Vol. 1. – P. 215–239. – DOI: 10.1016/0378-8733(78)90021-7.

31. Carrington, P.J. Models and Methods in Social Network Analysis / P.J. Carrington, J. Scott, S. Wasserman // Cambridge University Press, Cambridge. – 2005.

32. Castells, M. Networks of Outrage and Hope. Social Movements in the Internet Age / M. Castells // Polity, Cambridge. – 2012.

33. Faris, D.M. Dissent and Revolution in a Digital Age / D.M. Faris // I.B.Tauris. – 2013. – DOI: 10.5040/9780755607839.

34. Gerbaudo, P. Tweets and the Streets. Social Media and Contemporary Activism / P. Gerbaudo // Pluto Books, London. – 2012.

35. Lü, L. Vital nodes identification in complex networks / L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, T. Zhou // Physics Reports. – 2016. – Vol. 650. – P. 1–63. – DOI: 10.1016/j.physrep.2016.06.007.

36. Chen, D. Identifying influential nodes in complex networks / D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, T. Zhou // Physica A: Statistical Mechanics and its Applications. – 2012. – Vol. 391. – P. 1777–1787. – DOI: 10.1016/j.physa.2011.09.017.

37. Губанов, Д.А. Об определении влияния пользователей и мета-пользователей онлайн-социальной сети на основе акциональной идеи / Д.А. Губанов, А.Г. Чхартишвили // XIII ВСЕРОССИЙСКАЯ ШКОЛА-КОНФЕРЕНЦИЯ МОЛОДЫХ УЧЕНЫХ «УПРАВЛЕНИЕ БОЛЬШИМИ СИСТЕМАМИ». Учреждение Российской академии наук Институт проблем управления РАН, Москва. – 2016. – С. 342–351.

38. Губанов, Д.А. Влияемость пользователей и мета-пользователей социальной сети / Д.А. Губанов, А.Г. Чхартишвили // Проблемы управления. – 2016. – С. 12–17.

39. Губанов, Д.А. Акционная модель влияния пользователей социальной сети / Д.А. Губанов, А.Г. Чхартишвили // Проблемы управления. – 2014 – № 4. – С. 20–25.

40. Губанов, Д.А. Об определении влияния пользователей и мета-пользователей онлайн-социальной сети на основе акциональной модели / Д.А. Губанов, А.Г. Чхартишвили // Управление большими системами (УБС'2016) Самара, 05–09 сентября 2016 года. Самара. – 2016. – С. 342–351.

41. Торопов, Б.А. Максимизация конкурентного влияния в социальных сетях на основе игры Вороного / Б.А. Торопов // Моделирование, оптимизация и информационные технологии. – 2019 – № 7.

42. Торопов, Б.А. Теоретико-графовая модель анализа общественной опасности распространения экстремистской информации / Б.А. ТОРОПОВ,

И.В. ГУРЛЕВ, В.Б. КОРОБКО // Труды Академии управления МВД России. – 2017. – № 4.

43. Торопов Б.А. Модели террористических сетей и теоретико-игровой подход к оценке центральности их участников / Б.А. Торопов, З.И. Тагиров // Вопросы безопасности. – 2016. – № 6. – DOI: 10.7256/2409-7543.2016.6.21436.

44. 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.

45. Рабчевский, А.Н. 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.

46. 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.

47. 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 / 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.

48. Rabchevsky A. N. Modelling the structure of protest movement advocacy in social media using graph and neural network analysis / 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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. опубл. 15.07.2021; Бюл. №7. – 1 с.

54. Tindall, D.B. From metaphors to mechanisms: Critical issues in networks and social movements research / D.B. Tindall // Social Networks. – 2007. – Vol. 29. – P. 160–168. – DOI: 10.1016/j.socnet.2006.07.001.

55. Bennett, W.L. The logic of connective action / W.L. Bennett, A. Segerberg // Information, Communication & Society. – 2012. – Vol. 15. P. 739–768. – DOI: 10.1080/1369118X.2012.670661.



56. Juris, J.S. Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation / J.S. Juris // *American Ethnologist*. – 2012. – Vol. 39. – P. 259–279. – DOI: 10.1111/j.1548-1425.2012.01362.x.

57. Соколов, А.В. Интернет-аспекты протестной активности в современной России / А.В. Соколов, А.А. Фролов // *Известия Алтайского государственного университета*. – 2012 – № 4–2. С. 280–283.

58. Градосельская, Г.В. Информационные волны в социальных сетях: проблематизация, определение, механизмы распространения / Г.В. Градосельская, Т.Е. Щеглова, И.А. Карпов // *Управление развитием крупномасштабных систем (MLSD'2018)*. – 2018. – С. 509–512.

59. Bavelas, A. A Mathematical Model for Group Structures / A. Bavelas // *Human Organization*. – 1948. – Vol. 7. – DOI: 10.17730/humo.7.3.f4033344851gl053.

60. Bavelas, A. Communication Patterns in Task-Oriented Groups / A. Bavelas // *J Acoust Soc Am*. – 1950. – Vol. 22. – DOI: 10.1121/1.1906679.

61. Shimbel, A. Structural parameters of communication networks / A. Shimbel // *The Bulletin of Mathematical Biophysics*. – 1953. – Vol. 15. – DOI: 10.1007/BF02476438.

62. Nash-Williams, C.St.J.A. Applications of Graph Theory to Group Structure / C.St.J.A. Nash-Williams, C. Flament // *The Mathematical Gazette*. – 1966. – Vol. 50. – DOI: 10.2307/3614862.

63. Beauchamp, M.A. An improved index of centrality / M.A. Beauchamp // *Behavioral Science*. – 1965. – Vol. 10. – DOI 10.1002/bs.3830100205.

64. Sabidussi, G. The centrality index of a graph / G. Sabidussi // *Psychometrika*. – 1966. – Vol. 31. – DOI: 10.1007/BF02289527.

65. Stephenson, K. Rethinking centrality: Methods and examples / K. Stephenson, M. Zelen // *Social Networks*. – 1989. – Vol. 11. – P. 1–37. – DOI: 10.1016/0378-8733(89)90016-6.

66. Comin, C.H. Identifying the starting point of a spreading process in complex networks / C.H. Comin, L. da Fontoura Costa // *Physical Review E*. 84, 056105-1-056105–6. – 2011. – DOI: 10.1103/PhysRevE.84.056105.
67. Kitsak, M. Identification of influential spreaders in complex networks / M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse // *Nature Physics*. – 2010. – Vol. 6. – P. 888–893. – DOI: 10.1038/nphys1746.
68. Bae, J. Identifying and ranking influential spreaders in complex networks by neighborhood coreness / J. Bae, S. Kim // *Physica A: Statistical Mechanics and its Applications*. – 2014. – Vol. 395. – DOI: 10.1016/j.physa.2013.10.047.
69. Hou, B. Identifying all-around nodes for spreading dynamics in complex networks / B. Hou, Y. Yao, D. Liao // *Physica A: Statistical Mechanics and its Applications*. – 2012. – Vol. 391 – DOI: 10.1016/j.physa.2012.02.033.
70. Basaras, P. Detecting Influential Spreaders in Complex, Dynamic Networks / P. Basaras, D. Katsaros, L. Tassiulas // *Computer (Long Beach Calif)*. – 2013. – Vol. 46. – DOI: 10.1109/MC.2013.75.
71. Zeng, A. Ranking spreaders by decomposing complex networks / A. Zeng, C.-J. Zhang // *Physics Letters A*. – 2013. – Vol. 377. – DOI: 10.1016/j.physleta.2013.02.039.
72. Liu, J.-G. Ranking the spreading influence in complex networks / J.-G. Liu, Z.-M. Ren, Q. Guo // *Physica A: Statistical Mechanics and its Applications*. – 2013. – Vol. 392. – DOI: 10.1016/j.physa.2013.04.037.
73. Kleinberg, J.M. Authoritative sources in a hyperlinked environment / J.M. Kleinberg // *Journal of the ACM*. – 1999. – Vol. 46. – P. 604–632. – DOI: 10.1145/324133.324140.
74. Brin, S. The Anatomy of a Large-Scale Hypertextual Web Search Engine / S. Brin, L. Page // *Computer Science Department, Stanford University, Stanford, CA* 94305. – 1998.
75. Lü, L. Leaders in Social Networks, the Delicious Case / L. Lü, Y.-C. Zhang, C.H. Yeung, T. Zhou // *PLoS ONE*. – 2011. – Vol. – 6. – DOI: 10.1371/journal.pone.0021202.

76. Sukhodolov, A. Mathematical Methods in Law Enforcement: Counteracting Extremism on Social Media / A. Sukhodolov, A. Lebedev, B. Toropov, A. Babkin, V. Spasennikov // Russian Journal of Criminology. – 2018. – Vol. 12 – DOI: 10.17150/2500-4255.2018.12(4).468-475.

77. Мазалов, В.В. Модифицированное значение Майерсона для определения центральности вершин графа / В.В. Мазалов, В.А. Хитрая // Математическая Теория Игр и ее Приложения. – 2019. – № 11. С. 19–39.

78. Чхартишвили, А.Г. Об измерении влиятельности в социальных сетях / А.Г. Чхартишвили // В.Н. Бурков (ed.) ТЕОРИЯ АКТИВНЫХ СИСТЕМ. Материалы международной научно-практической конференции. Институт проблем управления им. В.А. Трапезникова РАН, Москва. – 2014. – С. 219–220.

79. Свидетельство о регистрации программы для ЭВМ RU 2019665357 Программа для расчета влияния и влиятельности пользователей социальных сетей на основе акциональной модели / Губанов, Д.А. опубл. 22.11.2019.

80. Ишков, Д.А. Автоматизация поиска распространителей деструктивного контента в социальных сетях / Д.А. Ишков, Е.Р. Нежелский, М.Н. Степанов // Информация и безопасность. – 2019. – № 22. – С. 256–259.

81. Блок, М. Социальный капитал: к обобщению понятия / М. Блок, Н.А. Головин // Вестник Санкт-Петербургского университета. Серия 12. Психология. Социология. Педагогика. – 2015. – № 4. – С. 99–111.

82. Butler, B.S. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures / Brian S. Butler // Information Systems Research. – 2001. – Vol. 12.

83. Cummings, J.N. The quality of online social relationships / J.N. Cummings, B. Butler, R. Kraut // Communications of the ACM. – 2002. – Vol. 45. – P. 103–108 – DOI: 10.1145/514236.514242.

84. Cheung, C.M.K. A theoretical model of intentional social action in online social networks / C.M.K. Cheung, M.K.O. Lee // Decision Support Systems. – 2010. – Vol. 49. – P. 24–30. – DOI: 10.1016/j.dss.2009.12.006.

85. Ganley, D. The ties that bind: Social network principles in online communities / D. Ganley, C. Lampe // *Decision Support Systems*. – 2009. – Vol. 47. – P. 266–274. – DOI: 10.1016/j.dss.2009.02.013.

86. Heidemann, J. Identifying Key Users in Online Social Networks: A PageRank Based Approach / J. Heidemann, M. Klier, F. Probst // *ICIS 2010 Proceedings - Thirty First International Conference on Information Systems*. – 2010. – P. 79.

87. Trusov, M. Determining Influential Users in Internet Social Networks / M. Trusov, A. Vodapati, R. Bucklin // *Journal of Marketing Research*. – 2009. – DOI: 10.2139/ssrn.1479689.

88. Модестов, С.А. Социальные сети как театр информационного противоборства в условиях современной “гибридной” войны / С.А. Модестов, Д.А. Никитин, Е.А. Рабчевский // *Вестник Академии военных наук*. – 2019. – № 3. – С. 20–25.

89. <https://www.seuslab.ru/seus>.

90. Doran, D. On the Discovery of Social Roles in Large Scale Social Systems / D. Doran // *Social Network Analysis And Mining*. – 2015. – Vol. 5. – 49-undefined. – DOI: 10.1007/s13278-015-0290-0.

91. Akar, E. User Roles and Contribution Patterns in Online Communities: A Managerial Perspective / E. Akar, S. Mardikyan // *SAGE Open*. – 2018. – Vol. 8. – DOI: 10.1177/2158244018794773.

92. Wijenayake, P. Automated Detection of Social Roles in Online Communities using Deep Learning / P. Wijenayake, D. de Silva, D. Alahakoon, S. Kirigeegamage // *Proceedings of the 3rd International Conference on Software Engineering and Information Management*. ACM, New York, NY, USA (2020). – 2020. – DOI: 10.1145/3378936.3378973.

93. Lauterbach, A. The Artificial Intelligence Imperative: A Practical Roadmap for Business / A. Lauterbach, A. Bonime-Blanc, I. Bremmer // *ABC-CLIO*. – 2018.

94. Осовский, С. Нейронные сети для обработки информации / С. Осовский // *Пер. с польского*. – М.: Финансы и статистика. – 2002.

95. Свидетельство о государственной регистрации программы для ЭВМ № 2014618208. Нейросимулятор 5.0 / Ф.М. Черепанов, Л.Н. Ясницкий. Опубликовано 12.07.2014 г.

96. SEUSLAB: Страница запроса датсета. URL: [https://seuslab.ru/registered\\_db/2021621533?lang=en](https://seuslab.ru/registered_db/2021621533?lang=en)

97. Opsahl, T. Node centrality in weighted networks: Generalizing degree and shortest paths / T. Opsahl, F. Agneessens, J. Skvoretz // *Social Networks*. – 2010. – Vol. 32. – P. 245–251 – DOI: 10.1016/j.socnet.2010.03.006.

98. Кузнецов, Е.Н. Анализ структуры сетевых взаимодействий: контекстно-зависимые меры центральности / Е.Н. Кузнецов // *Управление большими системами*. М.: ИПУ РАН. – 2019. – С. 57–82.

99. Wang, H. Betweenness centrality in a weighted network / H. Wang, J.M. Hernandez, P. van Mieghem // *Physical Review E*. 77. – 2008. – DOI: 10.1103/PhysRevE.77.046105.

100. Van Mieghem, P. Influence of the link weight structure on the shortest path / P. van Mieghem, S. van Langen // *Physical Review E*. 71. – 2005. – DOI: 10.1103/PhysRevE.71.056113.

101. Levandowsky, M. Distance between Sets / M. Levandowsky, D. Winter // *Nature*. – 1971 – Vol. 234. – P. 34–35 – DOI: 10.1038/234034a0.

## ПРИЛОЖЕНИЕ 1

### Пример файла выборки коллекции материалов

Ссылка на материал	Дата публикации	Время публикации	Вид материала	Актор	Ссылка на профиль актора
<a href="https://vk.com/">https://vk.com/.....</a>	30.01.2021	13:12:22	Вложение (Ссылка)	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	31.01.2021	06:13:56	Комментарий к посту/репосту	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	20.01.2021	00:43:39	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	25.01.2021	17:25:31	Комментарий к посту/репосту	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	23.01.2021	11:35:42	Вложение (Ссылка)	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	29.01.2021	19:43:06	Комментарий к посту/репосту	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	28.01.2021	07:08:58	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	22.01.2021	20:15:57	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	23.01.2021	21:55:38	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	23.01.2021	21:55:31	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	22.01.2021	20:15:51	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	11:58:11	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	27.01.2021	00:55:33	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	28.01.2021	15:48:36	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	27.01.2021	21:48:30	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	08:13:31	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	25.01.2021	22:51:45	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	09:32:22	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	14:21:08	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	25.01.2021	23:12:18	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	25.01.2021	22:51:26	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	18:08:22	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	15:27:10	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	25.01.2021	22:39:17	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	27.01.2021	00:55:33	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	29.01.2021	17:01:41	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	14:20:48	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	31.01.2021	16:42:21	Комментарий к посту/репосту	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	31.01.2021	14:38:16	Комментарий к посту/репосту	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	01:02:56	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	31.01.2021	15:38:10	Пост	Пользователь	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	18:08:26	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	27.01.2021	04:44:32	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	15:12:03	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	25.01.2021	22:39:16	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>
<a href="https://vk.com/">https://vk.com/.....</a>	26.01.2021	01:24:55	Репост	Группа	<a href="https://vk.com/">https://vk.com/....</a>

## ПРИЛОЖЕНИЕ 2

### Пример файла рейтинга активности пользователей

№	Общее кол-во друзей	Посты	Комментарии	Репосты	Кол-во материалов в выборке
1	24	8	0	160	168
2	0	94	0		94
3	246	90	0		90
4	953	58	0		58
5	944832	46	7		53
6	2704	3	48		51
7	493	0	47		47
8	16720244	46	0		46
9	519595	5	1	40	46
10	451	42	0		42
11	19763	6	36		42
12	378	42	0		42
13	2490	1	37	3	41
14	0	36	3		39
15	6925	39	0		39
16	54	38	0		38
17	2922884	7	11	19	37
18	1270369	23	0	13	36
19	1	0	36		36
20	3572894	33	0	2	35
21	88945	35	0		35
22	428809	34	0	1	35
23	2932	34	0		34
24	13822	34	0		34
25	7358	33	0		33
26	40	32	0		32
27	19363092	17	1	13	31
28	6	31	0		31
29	148	30	0		30
30	69	0	30		30
31	411	0	29		29
32	9	0	29		29
33	796	0	28		28
34	172	28	0		28
35	1616803	28	0		28

### ПРИЛОЖЕНИЕ 3

#### Пример файла формата CSV, содержащего информацию о вершинах графа, то есть о пользователях социальной сети

id,Тип вершины,Названия вершин,Цвет,Страна,Город,Родной город,Друзья,Подписчики,Группы,Роль
https://vk.com/id.....,Пользователь,Алек.... Абра.....,белый,Бѣлоруссия,Минскъ,,,,,Reposter
https://vk.com/ id.....,Группа, \_(ʘ)_/,белый,Россия,Санктъ-Петербургъ,,,,,Kommentator
https://vk.com/ id.....,Пользователь,Юра Па.....,белый,Бѣлоруссия,Минскъ,,,,,Poster
https://vk.com/ id.....,Пользователь,Але.... Дзьяч.....,белый,Бѣлоруссия,Могилевъ,,,,,Poster
https://vk.com/ id.....,Группа,☀ Развитие образования в России ☀,белый,Россия,Москва,,,,,Poster
https://vk.com/ id.....,Пользователь,Вален.... Мозо.....,белый,Россия,Санктъ-Петербургъ,,,,,Universal
https://vk.com/ id.....,Пользователь,Васи.... Мусі.....,белый,Бѣлоруссия,Минскъ,,,,,Universal
https://vk.com/ id.....,Группа,★ ЭЛИЗ... ★,белый,Россия,Нижній Новгородъ,,,,,Poster
https://vk.com/ id.....,Группа,Newstube,белый,Россия,Москва,,,,,Poster
https://vk.com/ id.....,Группа,Trib..... Беларусь,белый,Бѣлоруссия,Минскъ,,,,,Poster



## ПРИЛОЖЕНИЕ 4

### Пример файла формата CSV, содержащего информацию о ребрах графа, то есть о друзьях пользователей

source,target,Тип связи
https://vk.com/id.....,https://vk.com/ id.....,Подписчики
https://vk.com/ id.....,https://vk.com/ id.....,Друзья
https://vk.com/id4943.....,https://vk.com/id2342.....,Друзья
https://vk.com/id4943.....,https://vk.com/id4705.....,Друзья
https://vk.com/id4943.....,https://vk.com/id9588.....,Друзья
https://vk.com/id4943.....,https://vk.com/id11216.....,Друзья
https://vk.com/id4943.....,https://vk.com/id14170.....,Друзья
https://vk.com/id4943.....,https://vk.com/id19360.....,Друзья
https://vk.com/id4943.....,https://vk.com/id23972.....,Друзья
https://vk.com/id4943.....,https://vk.com/id37097.....,Друзья
https://vk.com/id4943.....,https://vk.com/id49879.....,Друзья
https://vk.com/id2022.....,https://vk.com/id16871.....,Подписчики
https://vk.com/id682.....,https://vk.com/id14170.....,Подписчики
https://vk.com/club10571.....,https://vk.com/id1470.....,Участник группы
https://vk.com/club10571.....,https://vk.com/id1087.....,Участник группы
https://vk.com/club10571.....,https://vk.com/id13891.....,Участник группы

## ПРИЛОЖЕНИЕ 5

## Диаграмма структуры базы данных для обучения нейросетевой модели

Dashboard Properties SQL Statistics Dependencies Dependents datasets\_db/p...

1M MM

datasets\_db/postgres@datasets\_local\_server

👁	
🔹 datasets_schema	
📄 dataset_table	
📄 comments_count integer	
📄 edges_count integer	
📄 is_commentator boolean	
📄 is_poster boolean	
📄 is_reposter boolean	
📄 is_universal boolean	
📄 posts_count integer	
📄 profile_age_days integer	
📄 reposts_count integer	
🔑 fact_id integer	
📄 is_passive boolean	

## ПРИЛОЖЕНИЕ 6

### Свойства базы данных для обучения нейросетевой модели

Statistics	Value
Backends	1
Xact committed	14640
Xact rolled back	22
Blocks read	708
Blocks hit	306616
Tuples returned	3303965
Tuples fetched	82213
Tuples inserted	4278
Tuples updated	20
Tuples deleted	398
Last statistics reset	2021-06-18 11:32:38.65585+03
Tablespace conflicts	0
Lock conflicts	0
Snapshot conflicts	0
Bufferpin conflicts	0
Deadlock conflicts	0
Temporary files	0
Size of temporary files	0 bytes
Deadlocks	0
Block read time	0
Block write time	0
Size	8189 kB

## ПРИЛОЖЕНИЕ 7

### Пример содержания базы данных для обучения нейросетевой модели

datasets\_db/postgres@datasets\_local\_server

Query Editor Query History

```

1 SELECT *
2 FROM datasets_schema.dataset_table

```

Data Output Explain Messages Notifications

	comments_count integer	edges_count integer	is_commentator boolean	is_poster boolean	is_reposter boolean	is_universal boolean	posts_count integer	profile_age_days integer	reposts_count integer	fact_id [PK] integer	is_passive boolean
1	11	30597141	false	false	false	true	3	953	8	1	false
2	1	20056165	false	true	false	false	49	1642	2	2	false
3	2	27241064	false	true	false	false	34	1782	5	3	false
4	27	22552610	true	false	false	false	0	4497	8	4	false
5	10	15252218	false	false	false	true	7	3173	17	5	false
6	0	27341247	false	false	false	false	0	5060	1	6	true
7	1	27115237	false	false	true	false	14	4521	91	7	false
8	0	1329587	false	false	false	false	1	3960	0	8	true
9	48	17054299	true	false	false	false	0	3525	13	9	false
10	5	19543267	false	true	false	false	39	1812	10	10	false
11	9	12872887	false	false	false	true	14	3221	14	11	false
12	48	16223743	true	false	false	false	2	2781	6	12	false
13	7	22212806	false	true	false	false	61	562	16	13	false
14	1	30566187	false	false	false	false	0	2541	1	14	true
15	24	20896430	true	false	false	false	0	1053	4	15	false

## ПРИЛОЖЕНИЕ 8

**Код функции, выполняющей вычисление метрики центральности по взвешенному вкладу**

```
def calculate_weighted_contribution centrality(graph, rating):
centrality = { }
    R = 0 # R = 0 # accumulated weight value of all clusters
    for user, friends in graph.items():
        c = 0 # accumulative value of cluster weight to be connected by
user
        # accumulate rating by user's friends
        for friend in friends:
            # if a user's friend is linked in the column only, add their
            rating
            if len(graph[friend]) == 1:
                c = c + rating[friend]
            centrality[user] = c
    R = R + c
    for user in centrality.keys():
        centrality[user] = centrality[user] / R
return centrality
```

## ПРИЛОЖЕНИЕ 9

## Свидетельство о регистрации базы данных для классификации ролей пользователей социальных сетей

РОССИЙСКАЯ ФЕДЕРАЦИЯ



## СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2021621533

База данных для классификации ролей пользователей социальных сетей

Правообладатель: *Общество с ограниченной ответственностью "СЕУСЛАБ" (RU)*Авторы: *Рабчевский Андрей Николаевич (RU), Заякин Виктор Сергеевич (RU)*

Заявка № 2021621400

Дата поступления 07 июля 2021 г.

Дата государственной регистрации

в Реестре баз данных 15 июля 2021 г.

Руководитель Федеральной службы  
по интеллектуальной собственности

Г.П. Ивлиев





## ПРИЛОЖЕНИЕ 10

## Свидетельство о регистрации программы вычисления мостов в кластерных сетях

РОССИЙСКАЯ ФЕДЕРАЦИЯ



## СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2021616086

## Программа вычисления мостов в кластерных сетях

Правообладатель: *Общество с ограниченной ответственностью "СЕУСЛАБ" (RU)*Авторы: *Рабчевский Андрей Николаевич (RU), Заякин Виктор Сергеевич (RU)*

Заявка № 2021615157

Дата поступления 13 апреля 2021 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 16 апреля 2021 г.

*Руководитель Федеральной службы  
по интеллектуальной собственности*

Г.П. Ивлиев





## ПРИЛОЖЕНИЕ 11

**СЕУСЛАБ**

ООО "СЕУСЛАБ"  
ИНН 5903097803, КПП 590501001  
ОГРН 1125903002675  
614066, Пермский край, г. Пермь  
ул. Шоссе Космонавтов 111 к.3  
тел.: 8-(342)-207-11-70  
[info@seuslab.ru](mailto:info@seuslab.ru), [www.seuslab.ru](http://www.seuslab.ru)

## АКТ

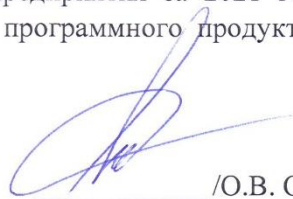
о внедрении научных результатов диссертационной работы  
Рабчевского Андрея Николаевича,  
представленной на соискание учёной степени  
кандидата технических наук

Настоящим актом подтверждается, что практические результаты диссертационной работы Рабчевского А.Н., посвящённой разработке метода многофакторного анализа ролей пользователей социальных сетей, внедрены и являются частью программного продукта «Сервис анализа распространения контента» компании ООО «СЕУСЛАБ». Предложенные методы выявления ролей пользователей и уровня их информационного влияния, модели и алгоритмы интеллектуальной поддержки использованы при разработке модуля поддержки принятия решений по противодействию деструктивным воздействиям в социальных сетях в составе «Поисковой системы SEUS» компании ООО «СЕУСЛАБ», предоставляющей неисключительные права на использование программы сотрудникам органов государственной власти.

Основные показатели эффективности внедрения:

- сокращение трудозатрат на обработку информации со стороны пользователей «Поисковой системы SEUS» в деятельности по выявлению структуры деструктивных воздействий на социальную сеть;
- повышение качества принимаемых решений пользователей «Поисковой системы SEUS» по противодействию деструктивным воздействиям в социальных сетях за счет получения дополнительной информации о структуре деструктивных воздействий на социальную сеть;
- повышение конкурентоспособности «Поисковой системы SEUS» на рынке систем анализа социальных сетей;
- получение дополнительной выручки предприятия за 2021 год за счет поступлений оплаты лицензий внедренного программного продукта «Сервис анализа распространения контента» на 12%.

Директор по стратегическому управлению  
ООО «СЕУСЛАБ»  
«10» декабря 2021 г.



/О.В. Олейников/



**ПРИЛОЖЕНИЕ 12**

УТВЕРЖДАЮ

Проректор по учебной работе

ФГБОУ ВО «Пермский

государственный национальный  
исследовательский университет»

доктор физ.-мат. наук, доцент

 / С.О. Макаров /

АКТ

о внедрении научных результатов,  
полученных Рабчевским Андреем Николаевичем  
при выполнении диссертационной работы  
на соискание учёной степени кандидата технических наук

Комиссия в составе:

председатель: декан механико-математического факультета ПГНИУ,  
к.т.н. Кузнецов А.Г.

члены комиссии: заведующий кафедрой прикладной математики и  
информатики, д.ф.-м.н., профессор Русаков С.В.,  
профессор кафедры прикладной математики и  
информатики, д.т.н, профессор Ясницкий Л.Н.

составила настоящий акт о том, что основные теоретические положения и практические результаты диссертационной работы Рабчевского А.Н. используются в образовательном процессе, реализуемом для бакалавров направления подготовки 01.03.02 «Прикладная математика и информатика».

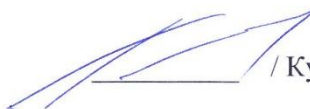
Предложенный в диссертационной работе экспертный способ создания синтетических данных для обучения нейронных сетей рассматривается в ходе изучения дисциплины «Методы искусственного интеллекта».

Эффект от внедрения результатов диссертационной работы заключается в повышении уровня освоения профессиональных компетенций и их компонентов (знаний, умений, владений) в области построения интеллектуальных систем и математических моделей сложных систем, что

соответствует требованиям Федеральных государственных образовательных стандартов высшего образования.

Итоги внедрения результатов диссертационной работы обсуждались на заседании кафедры прикладной математики и информатики 19.10.2021, протокол № 02.

Председатель комиссии:  
Декан механико-математического  
факультета ПГНИУ, к.т.н.



/ Кузнецов А.Г. /

Члены комиссии:  
заведующий кафедрой прикладной  
математики и информатики, д.ф.-м.н.,  
профессор



/ Русаков С.В. /

профессор кафедры прикладной  
математики и информатики, д.т.н,  
профессор



/ Ясницкий Л.Н. /

« 19 » сентября 2022 г.