

Моделирование социальных сетей: визуализация и расчет метрик

М. А. Решетников, Е. Б. Замятин

Пермский государственный национальный исследовательский университет (ПГНИУ), Пермь, Россия

Аннотация. В статье рассматриваются методы визуализации и алгоритмы вычисления метрик в ходе моделирования социальных сетей. Реализованы алгоритмы поиска наиболее влиятельных узлов социальной сети. Проведено моделирование процесса распространения информации на примере модели SIR: с помощью вычисленных метрик центральности выявлены ключевые узлы наиболее влиятельных пользователей и проведено исследование, которое определяет влияние изменения количества этих узлов на скорость распространения сообщений в сети. Результаты моделирования иллюстрируются диаграммами и графами, полученными с помощью системы моделирования Triad.Net.

Ключевые слова: социальные медиа, визуализация графов, метрики центральности, моделирование распространения информации.

Simulation of social media: visualization and metrics calculation

M. A. Reshetnikov, E. B. Zamyatina

Perm State National Research University, Perm, Russia

Abstract. The paper discusses visualization methods and algorithms for calculating metrics during the simulation of social networks. Algorithms for searching for the most influential nodes of a social network are implemented. The process of information dissemination is simulated using the SIR model as an example: using the calculated centrality metrics, key nodes of the most influential users are identified, and a study is conducted that determines the impact of changing the number of these nodes on the speed of message dissemination in the network. The simulation results are illustrated by diagrams and graphs obtained using the Triad.Net simulation system.

Keywords: social media, graph visualization, centrality metrics, information diffusion modeling.

Введение

Социальные сети прочно вошли в повседневную жизнь. Современный человек потребляет огромное количество информации из социальных сетей, таких как Twitter, Вконтакте, Facebook, Одноклассники и т.д. В интернете они являются средством быстрого распространения информации, позволяют людям объединяться в группы по интересам и могут влиять на общественное мнение.

На основе данных из социальных сетей можно понять, кого из политиков поддерживает народ, одобряют ли люди действия и позицию властей. Проанализировав структуру социальных сетей, можно выявить, как необходимо распространять информацию, чтобы она имела наибольший охват аудитории. Таким образом, социальные сети можно успешно использовать и для проведения маркетинговых операций. Благодаря анализу контента, который потребляет пользователь, можно сказать, какая реклама будет ему интересна, а какая нет [1–3].

Социальные сети также являются средством массовой информации, где каждый пользователь может распространять новости от своего имени или имени сообщества, из-за чего возникает необходимость дополнительной проверки новостей на их достоверность [4].

По сути дела, исследование социальных сетей и поведение пользователей в сети является актуальным, поскольку необходимо решить задачу скорейшего распространения нужной информации (для маркетологов, например). Важной является и обратная задача: выявление и сдерживание распространения нежелательной информации [4–6].

При решении этих задач используют методы Social Mining [7]. Основными задачами Social Network Mining являются: (а) анализ информационных потоков: выявление наиболее «весомых» узлов в социальной сети, а именно людей, через которых проходит большой поток информации и чье мнение оказывает существенное влияние на другие узлы (применяется в таких сферах, как интернет-маркетинг, реклама и безопасность пользователей); (б) персонификация предложений: фильтрация контента, потребляемого пользователем, в соответствии с интересами этого пользователя. Например, показ рекламы автомобилей человеку, который собирается покупать машину в ближайшее время, повышает удовлетворенность пользователя от использования ресурса, а реклама становится более эффективной; (в) поиск аномалий, компьютерных ботов и мошенников: необходимо вовремя выявлять пользователей, чьи действия негативно влияют на социальную сеть.

При изучении социальных сетей зачастую не учитывают временные характеристики объектов исследования в социальных сетях. В настоящей статье рассматривается динамическое моделирование социальных сетей, учитываются моменты выполнения тех или иных событий в социальной сети, их последовательность. Моделирование выполняется с использованием программной системы Triad.Net [8]. Особенности реализации системы имитационного моделирования Triad.Net рассматриваются ниже.

Очень часто при изучении социальных сетей используют виртуальные сети, синтетические данные, модели случайных графов [9], модели процессов распространения информации. Модели распространения также основываются на теории графов: пользователи – узлы (вершины) графа, а связи между ними – ребра графа. Существующие модели распространения информации разделяются на следующие категории: независимые (а) каскадные модели, (б) линейные пороговые модели, (в) модели эпидемии. При этом последние различаются по количеству возможных состояний: модели эпидемии, описывающие процесс распространения на основе трех состояний узлов (SIR), эпидемиологические модели, включающие четыре состояния, а также модели распространения эпидемии, подразумевающие наличие множества состояний.

Наряду с виртуальными социальными сетями исследуют характеристики пользователей и каналы связей между ними в реальных социальных сетях [10, 11].

В моделях социальных медиа часто требуется выявить наиболее влиятельных участников и проследить пути распространения информации в социальной сети. Для этого используют алгоритмы, определяющие метрики центральности, в частности метрику PageRank [12], и биоинспирированные алгоритмы для поиска лидеров мнений. В настоящем исследовании были выбраны биоинспирированные алгоритмы Firefly и Grey Wolf Optimization (GWO) [13, 14].

Одновременно важна качественная визуализация структуры сети и результатов моделирования, поскольку наглядное представление графа упрощает анализ сложных связей между узлами.

Особенностью работы является создание инструмента, позволяющего не только отображать сеть, но и активно вмешиваться в её структуру для мгновенной оценки последствий на моделях распространения (SIR). Для этого используется онтологический подход к хранению данных, обеспечивающий семантическую согласованность и связь с предметной областью и подход для выявления наиболее влиятельных участников через биоинспирированные алгоритмы.

В статье представлены расширения модуля TriadNSim (модуль, предназначенный для исследования социальных медиа в системе моделирования Triad.Net), интегрирующего методы поиска лидеров мнений, моделирования распространения информации и визуализации на графе социальной сети.

Постановка задачи

Цель исследования – разработка прикладного инструмента для вычисления метрик центральности и визуализации результатов моделирования социальных сетей. В рамках этой цели ставились следующие задачи:

- исследовать методы определения влиятельных узлов в сети, включая классические метрики центральности и биоинспирированные алгоритмы оптимизации;
- реализовать алгоритмы PageRank, Firefly, GWO для оценки влияния участников;
- выполнить визуализацию социальной сети в виде графа, где параметры узлов (цвет, размер, шрифт) отображают вычисленные метрики;
- разработать пользовательский интерфейс для управления визуализацией, выбором алгоритма и анализа результатов;
- обеспечить доступ к онтологии с данными о пользователях и сообществах с помощью возможностей Triad.Net;
- провести экспериментальное моделирование информационных процессов (SIR-модель распространения новости) в социальной сети из 20 пользователей и проанализировать влияние выявленных лидеров на время распространения информации.

Теория

Задачи поиска ключевых узлов и оптимизации информационных потоков важны для понимания динамики социальных процессов, таких как вирусное распространение новостей или воздействие на общественное мнение. Для их решения используются методы анализа графов, позволяющие

количественно оценивать значимость узлов и устойчивость сети. Классические метрики центральности служат фундаментальным инструментом для ранжирования узлов по их структурной важности. Эти показатели позволяют обнаружить узлы с большим потенциалом распространения информации и понять роль каждого участника в сети.

Наряду с метриками центральности применяются биоинспирированные алгоритмы оптимизации, способные выявлять оптимальные стратегии распространения и выбор лидирующих узлов. Алгоритм светлячков (Firefly) и алгоритм серых волков (GWO) имитируют поведение природных систем поиска оптимальных решений и могут быть использованы для подбора набора узлов, обеспечивающих максимальный охват сети. Сочетание метрик центральности и метаэвристических алгоритмов обеспечивает мощный арсенал методов для анализа социальных сетей и моделирования информационных процессов. В последующих подпунктах будут рассмотрены теоретические основы некоторых упомянутых метрик и биоинспирированных алгоритмов.

- *Алгоритм PageRank*

PageRank — алгоритм ранжирования узлов графа, основанный на вероятностной модели перехода между вершинами. Узел считается тем более значимым, чем больше на него ссылаются другие значимые узлы. В контексте социальной сети это позволяет выявить участников, на которых «ориентируются» другие, что может означать авторитет или популярность узла.

- *Алгоритм светлячков (Firefly Algorithm)*

Метаэвристика – основана на поведении светлячков, у которых яркость соответствует качеству решения. В социальной сети каждый узел интерпретируется как светлячок, чья яркость зависит от метрик центральности. Узлы перемещаются (влияют друг на друга) в сторону более ярких, тем самым алгоритм итеративно усиливает влияние лидеров. На рисунке 1 представлена блок-схема алгоритма светлячков.

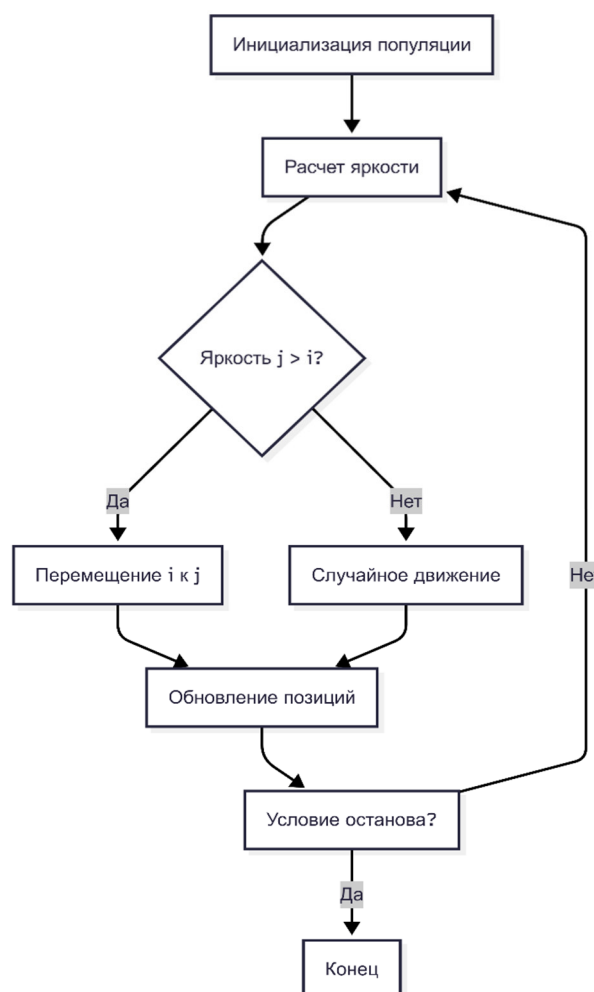


Рисунок 1. Блок-схема алгоритма светлячков

- *Алгоритм серых волков (GWO)*

Метаэвристический алгоритм моделирует иерархическую охоту стаи волков, блок-схема которого представлена на рисунке 2. Узлы-волки стремятся к «добыче» (оптимальному решению), следуя за тремя главными (альфа, бета, дельта). GWO показал лучшую устойчивость в динамических сетях по сравнению с Firefly.

- *SIR-модель распространения информации*

Модель SIR (Susceptible–Infected–Recovered) адаптирована под задачу информационного распространения. Агенты (узлы) могут быть в одном из трёх состояний:

- 1) S — уязвимы к получению информации,
- 2) I — обладают информацией и могут передавать её другим,
- 3) R — вышли из процесса распространения.

Передача информации осуществляется с вероятностью β , переход в состояние R — с вероятностью γ . Эта модель позволяет эмулировать динамику распространения новостей, слухов или вирусного контента в социальных сетях.

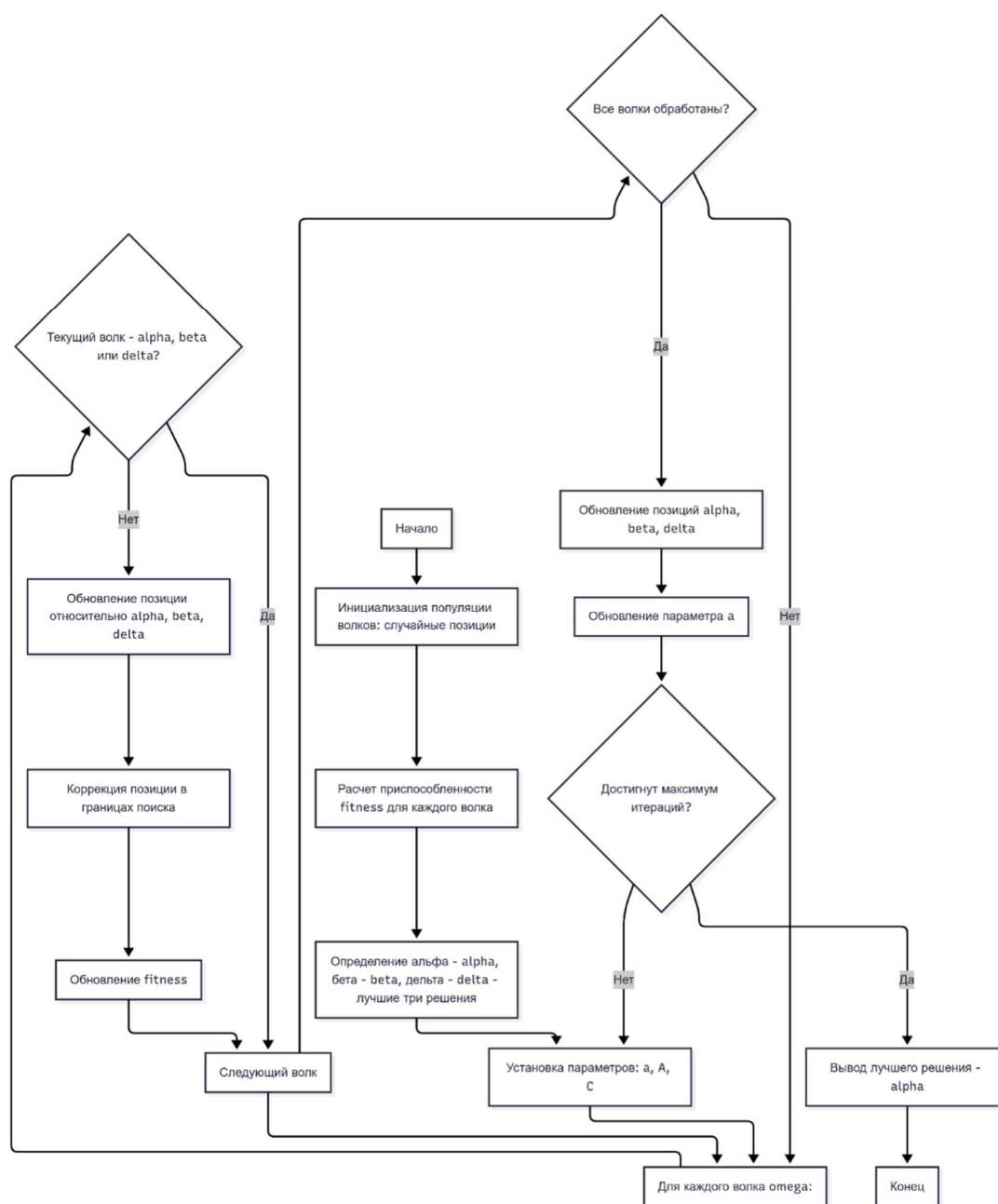


Рисунок 2. Блок-схема алгоритма серых волков

• Система моделирования *Triad.Net*

Triad.Net — это программная среда имитационного моделирования, предназначенная для построения, анализа и визуализации сложных динамических систем. Она широко применяется в задачах, связанных с анализом поведения программных агентов, алгоритмов маршрутизации в компьютерных сетях, информационными процессами в социальных сетях и т.д.

Для загрузки данных в онтологию используется программный агент, реализующий сбор информации из социальной сети «ВКонтакте» в реальном времени. Агент отслеживает действия пользователей и связи между сущностями, сохраняя их в OWL-онтологию через RabbitMQ-очереди. Онтология отражает актуальное состояние сети, позволяет Triad.Net визуализировать не только топологию, но и семантические паттерны, пример ее структуры представлен на рисунке 3. Онтология включает в себя:

- динамические атрибуты пользователей (пол, возраст, интересы);
- историю действий (лента новостей, просмотры постов);
- граф связей с весами (частота взаимодействий).

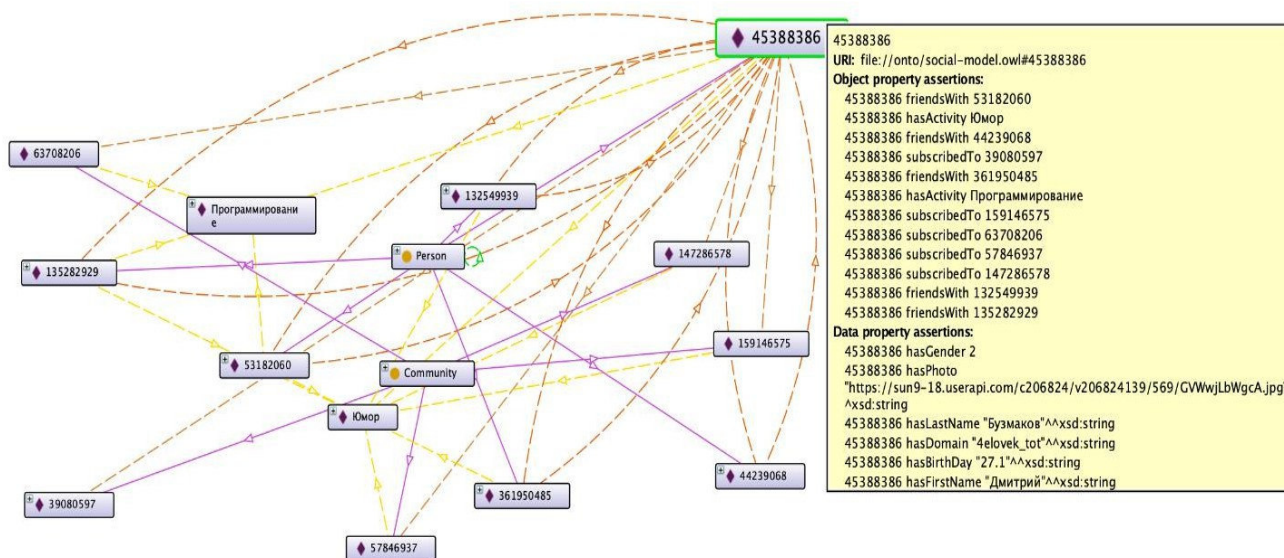


Рисунок 3. Онтология пользователя социальной сети

В Triad.Net принято трехуровневое представление имитационной модели: $M = (STR, ROUT, MES)$, где STR – слой структур, ROUT – слой рутин, MES – слой сообщений.

Слой структур представляет собой совокупность объектов, взаимодействующих друг с другом посредством посылки сообщений. Основа представления слоя структур – графы. Дуги графа определяют связи между объектами. Имитационная модель имеет иерархическое представление. Отдельные объекты, представляющие вершины графа, могут быть расшифрованы подграфом более низкого уровня и т.д. Для представления социальной сети вершины графа – это пользователи, сообщества; дуги – связи между пользователями, сообществами и т.д.

Для описания структуры имитационной модели используют графовые константы. Графовая константа – это граф некоторого стандартного вида. Например, кольцо, путь, дерево, полный граф и т.д. Количество вершин графа задают с помощью параметров. В Triad.Net реализованы несколько моделей случайных графов, такие как случайный граф Эрдеша–Реньи, граф Боллобаши–Риордана, модель П. Бакли и Д. Остгуса и модель копирования. Известно, что случайные графы позволяют исследовать виртуальные социальные сети.

Слой структур позволяет выполнять операции над графами. Для исследования характеристик построенного графа в слое структур реализованы процедуры, которые позволяют определять количество вершин в графе, диаметр, степень вершины, диаметр графа и т.д. Кроме того, реализованы процедуры для исследования социальных сетей.

Объекты структуры действуют по определённом сценарию, который описывают с помощью рутины. Рутинa представляет собой последовательность событий e_i , планирующих друг друга. Выполнение события сопровождается изменением состояния объекта. Состояние объекта определяется значениями переменных рутины. Таким образом, система имитации является событийно ориентированной. Совокупность рутин определяет слой рутин ROUT.

Слой сообщений (MES) предназначен для описания сообщений сложной структуры. Система моделирования Triad реализована таким образом, что пользователю необязательно описывать все слои. Так, если возникает необходимость в исследовании структурных особенностей модели, то можно описать только слой структур.

Одним из ключевых преимуществ Triad.Net является наличие графического интерфейса, позволяющего исследователю формировать модель наглядно — в виде графа, где узлы и связи отображаются с учётом заданных параметров. Пользователь может интерактивно задавать правила поведения агентов, настраивать свойства элементов системы и наблюдать за динамикой процессов в реальном времени.

Особенно стоит уделить внимание поддержке интеграции с онтологиями — формальными описаниями предметной области, содержащими информацию о сущностях (например, пользователях, сообществах, событиях) и их отношениях. Использование онтологий позволяет автоматически загружать и структурировать данные из реальной социальной сети [15]. Онтологическая модель обеспечивает согласованное представление информации и упрощает анализ семантических связей в социальной сети.

Таким образом, Triad.Net представляет собой мощный инструмент для комплексного моделирования, анализа и визуализации социальных сетей, обеспечивая удобство работы и гибкость при настройке модели. Интерфейс Triad.Net представлен на рисунке 4.

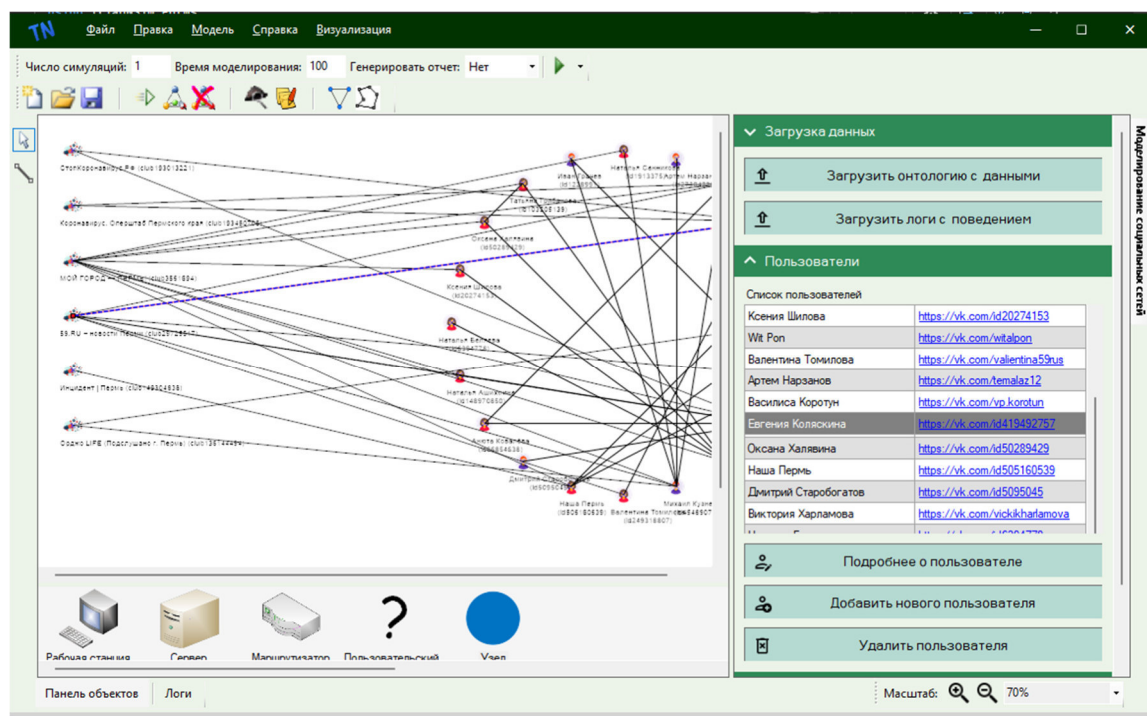


Рисунок 4. Интерфейс Triad.Net

На основе анализа работ [16, 17] был реализован модуль графического представления социальных сетей и визуализации результатов данных по распространению информации в сети. Интерфейс модуля для работы с социальными сетями представлен на рисунке 5.

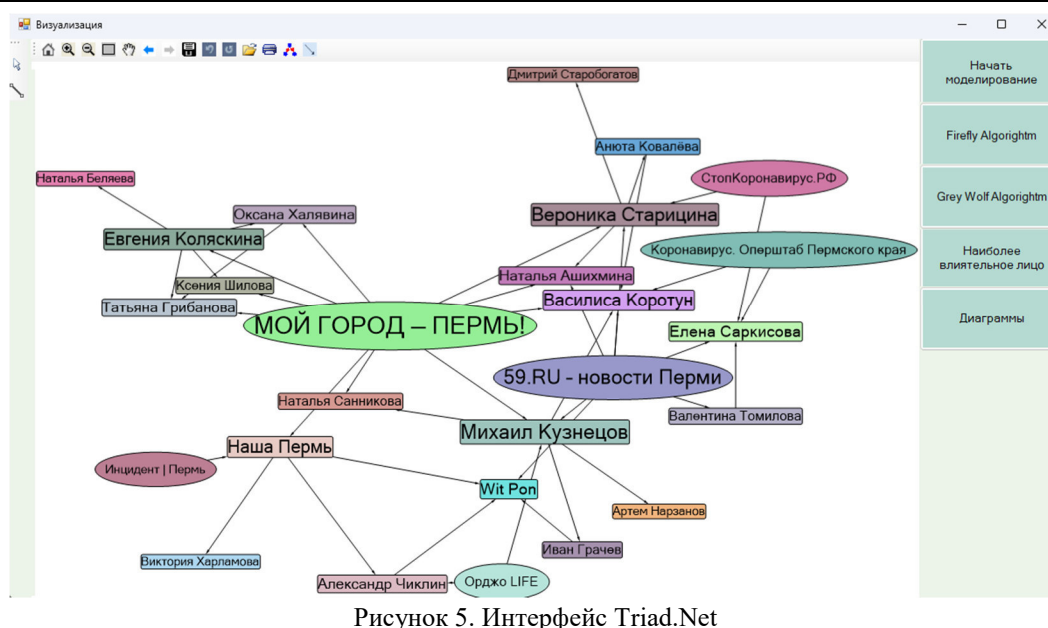


Рисунок 5. Интерфейс Triad.Net

Результаты экспериментов

Для практической оценки методов рассмотрен следующий эксперимент. Модельная социальная сеть содержит 20 пользователей, среди которых заданы группы с разными связями (участники групп «Мой город – Пермь!», «59.RU – новости Перми» и др.). На рисунке 5 показано графическое представление такой сети.

С помощью SIR-модели эмулируется однократная вспышка новости. Сначала определяются три наиболее влиятельных узла сети средствами алгоритмов: наибольшая метрика выдается сообществу «МОЙ ГОРОД – ПЕРМЬ!», а также узлам «59.RU – новости Перми» и «Михаил Кузнецов» (по совокупности мер). Затем два самых активных лидера (за исключением стартового «Мой город – Пермь!», который изначально является источником новости) искусственно удаляются из сети. После этого модель повторно запускается на модифицированной сети без этих узлов. На рисунках 6, 7 показана сравнительная диаграмма времени распространения информации до и после удаления лидеров. Видно, что время затухания новости заметно сократилось, а часть узлов вовсе не получила информацию из-за разрыва ключевых связей. Это свидетельствует о существенном влиянии выявленных лидеров на динамику распространения.

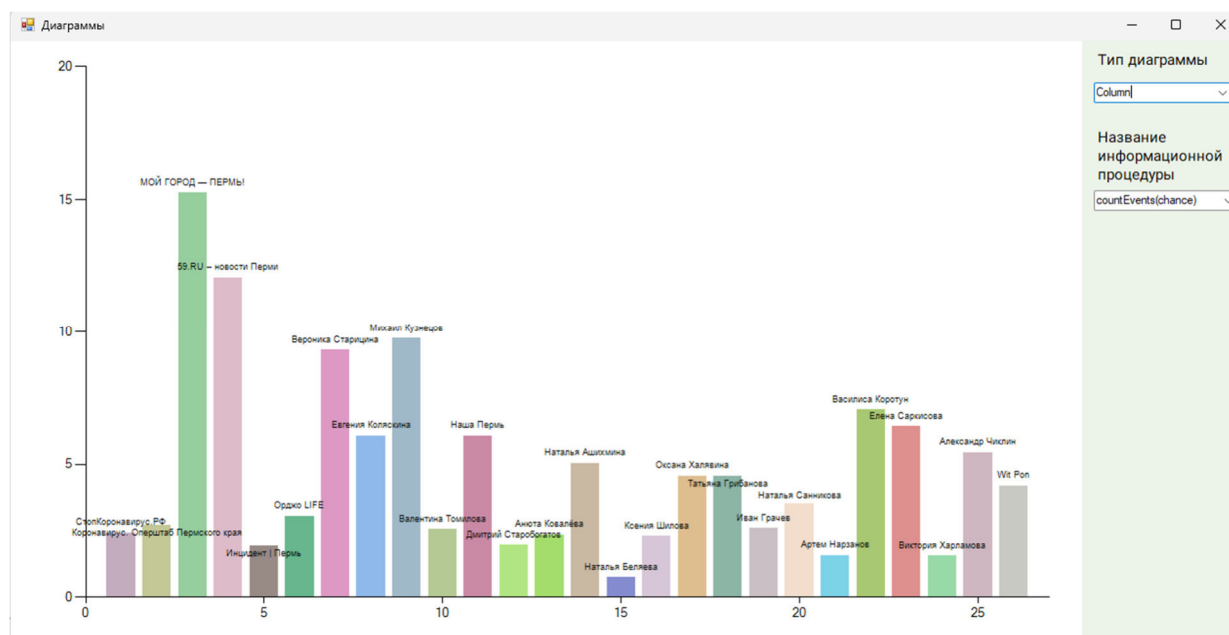


Рисунок 6. Столбчатая диаграмма времени распространения информации

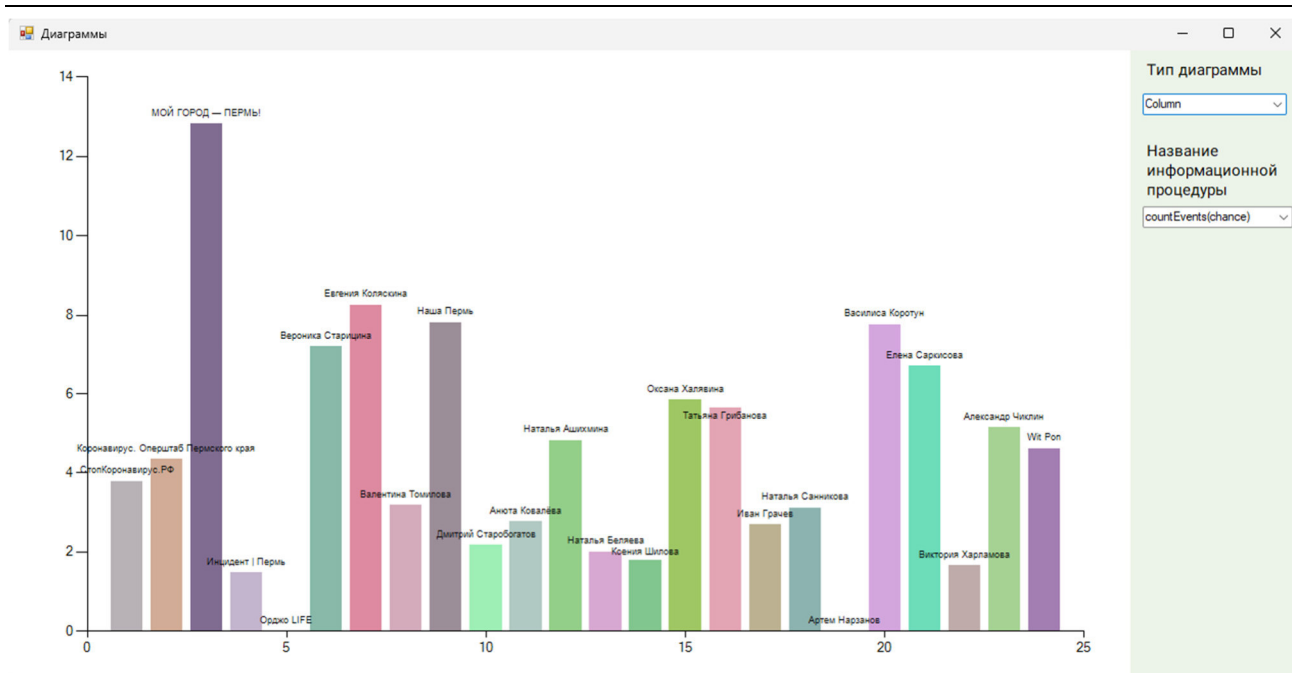


Рисунок 7. Столбчатая диаграмма времени распространения информации после удаления влиятельных пользователей

Обсуждение результатов

Результаты эксперимента подтверждают эффективность реализованных методов. После удаления наиболее влиятельных узлов ключевые связи в сети нарушаются, что приводит к замедлению и фрагментации информационного потока. Выявление таких узлов позволяет моделировать сценарии управления распространением, например, влияние на маркетинговые кампании. Отметим также, что алгоритмы Firefly и GWO в рамках тестовой сети дали схожий набор лидеров, но GWO обеспечивает более стабильную сходимость при увеличении размеров сети. Наглядное отображение социальной сети и диаграмм существенно упрощает анализ: цветовое кодирование влиятельных узлов и диаграммы времени распространения позволяют быстро оценить ключевые результаты моделирования.

Заключение

Были разработаны прикладные средства для вычисления метрик центральности и визуализации результатов моделирования социальных сетей. В частности, в платформу TriadNSim добавлены инструменты:

- Редактирование сети – добавление/удаление вершин и рёбер, интеграция с онтологическими данными.
- Методы поиска лидеров – внедрены алгоритмы Firefly и GWO для нахождения наиболее влиятельных узлов.
- Средства визуализации – автоматическая раскладка графа и построение диаграмм (включая цветовые метки узлов и графики времени распространения).

Реализованная система предоставляет уникальные возможности для контролируемых экспериментов:

- интерактивное вмешательство в структуру сети (удаление/изоляция узлов) с синхронизированной визуализацией последствий в реальном времени;
- глубокая интеграция с онтологиями, обогащающая узлы атрибутами (тип пользователя, тематика сообщества) для контекстной интерпретации;
- адаптивная визуализация, где цвет и размер узлов кодируют метрики влияния (PageRank, Firefly, GWO), а диаграммы отображают динамику распространения информации;
- цикл анализа «метрики → вмешательство → визуализация → интерпретация», недоступный в статических инструментах.

Применение этих инструментов позволяет проводить быстрый анализ структуры сети и динамики информационных процессов. Например, можно «отключить» ключевых пользователей в эксперименте и наблюдать последствия для распространения новости, что важно для оптимизации маркетинговых стратегий или исследования моделирования пандемий в социальных сетях. Разработанный модуль визуализации объединяет вычисление метрик и наглядное представление данных, что повышает эффективность исследования социальных сетей.

Проведённые исследования показали, что поставленные цели достигнуты. В дальнейшем планируется расширить функциональные возможности разрабатываемого модуля: добавить новые алгоритмы анализа, улучшить интерактивность визуализации и исследовать применение метода в реальных больших сетях.

Список литературы

1. Грибанова Е. Б., Логвин И. Н., Ширенков И. В. Алгоритм оценки маркетинговых мероприятий онлайн-социальной сети "ВКонтакте" на основе каскадной модели распространения информации. Доклады Томского государственного университета систем управления и радиоэлектроники. 2018. Т. 21, вып. 3. С. 69-74.
2. Методика выявления каналов распространения информации в социальных сетях / А. А. Проноза, Л. А. Виткова, А. А. Чечулин, И. В. Котенко, Д. В. Сахаров // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2018. Т. 14, вып. 4. С. 362–377.
3. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели влияния в социальных сетях // УБС. 2009. № 27. С. 205–281.
4. Абрамов М.В., Тулупьев А.Л., Сулейманов А.А. Задачи анализа защищенности пользователей от социоинженерных атак: построение социального графа по сведениям из социальных сетей // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18, № 2. С. 313–321.
5. Дмитриев И.П., Замятина Е.Б. Опыт моделирования действий по предотвращению распространению запрещенной информации в социальной сети // Прикладная информатика. 2019. Т. 14, № 1 (79). С. 82–95.
6. Zimin I., Zamyatina E. Strategies of the Social Network Immunization: An Experience of an Investigation by Simulation Tools // 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT). IEEE, 2020. P. 1–6.
7. Давыдов А. А. Системная социология: Social Networks Mining. М.: ИС РАН, 2009.
8. Mikov A., Zamyatina E., Germanova D. Program Tools for Dynamic Investigation of Social Networks // Supplementary Proceedings of the 5th International Conference on Analysis of Images, Social Networks and Texts (AIST-SUP 2016), Yekaterinburg, Russia, April 7-9, 2016 / ed. by D. I. Ignatov. Vol. 1710. Aachen : CEUR Workshop Proceedings, 2016. P. 224-236.
9. Райгородский А.М. Модели случайных графов и их применение // Труды Московского физико-технического института (национальный исследовательский университет). 2010. С. 130–140.
10. A simulation-based approach to analyze the information diffusion in Microblogging Online Social Network / M. Gatti, A. P. Appel, C. Nogueira dos Santos, C. Z. Pinhanez // Proceedings of Winter Simulation Conf. Piscataway, New Jersey, 2013. P. 1685–1696. Usanin A., Zimin I., Zamyatina E. Study of Strategies for Disseminating Information in Social Networks Using Simulation Tools // Analysis of Images, Social Networks and Texts: 9th International Conference, AIST 2020, Skolkovo, Moscow, Russia, October 15–16, 2020, Revised Selected Papers Vol. 12602. Springer, 2021. P. 303–315.
11. Zamyatina E., Mikov A., Lanin V. Automation of Simulation Steps using Ontological Approach // Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018). Volume 2: KEOD / Ed. by D. Aveiro, J. Dietz, J. Filipe. Portugal: SCITEPRESS – Science and Technology Publications, Lda., 2018. P. 223-230.
12. Нуньес-Иглесиас Х. Элегантный SciPy / в. д. Уолт, Х. Дэшноу. Москва: ДМК Пресс, 2018.
13. Sebin J., P.V Paul. A survey on identification of influential users in social media networks using bio inspired algorithms. // International Conference on Machine Learning and Data Engineering. 2023.
14. Интеллектуальные системы: модели и методы метаэвристической оптимизации: монография / Л. А. Гладков, Ю. А. Кравченко, В. В. Курейчик, С. И. Родзин. Чебоксары, 2024.
15. An Approach to Developing Data Visualization Tools Based on Domain Specific Modeling / A. D. Dzheiranian, I. D. Ermakov, K. A. Proskuryakov, L. N. Lyadova // Scientific Vizualisation. 2024. Vol. 16, No. 4. P. 82–101.
16. Li Y., Chan J. Mixed emotion extraction analysis and visualization of social // Data & Knowledge Engineering. 2013. P. 82–92.
17. Chien-Po L., Chien Y., Yu-His L. Progress and future directions for research on social media // Telematics and Informatics. 2023.