

101. ПРИМЕНЕНИЕ ТЕОРИИ ГРАФОВ В АНАЛИЗЕ СОЦИАЛЬНЫХ СЕТЕЙ

Шилов Б.В. студент гр.477602; Якжик А.Д. студент гр.477602

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Русина Н.В. – ст. преп., каф. ЭИ

Аннотация: Социальные сети стали неотъемлемой частью современного общества, формируя сложные системы взаимодействий между людьми. Анализ этих объёмных и комплексных структур требует математического аппарата, способного описать связи и закономерности между их элементами. Теория графов предоставляет инструменты для моделирования и изучения социальных сетей, что делает её применение особенно востребованным в социологии, маркетинге, IT, криминалистике и других областях.

Ключевые слова: граф, социальные сети, узел графа, маршрут, кластеризация, сообщества, центральность.

Краткий обзор применимости теории графов

Теория графов позволяет формализовывать социальные связи, выявлять ключевых участников, находить сообщества и анализировать устойчивость коммуникационной сети. Эти методы используются в рекомендательных системах, маркетинговых исследованиях и борьбе с дезинформацией [1].

Основные понятия теории графов

Граф — это совокупность вершин (узлов) и рёбер (связей между ними). При работе с социальными сетями, вершина может представлять пользователя, а ребро — их взаимодействие. Степенью вершины является количество рёбер, связанных с ней.

К основным типам графов относят [2]:

- Ориентированные, в которых рёбра имеют направление (например, подписки в Twitter);
- Неориентированные, имеющие двусторонние связи (дружба в Facebook);
- Взвешенные, где рёбрам присвоены числовые значения;
- Мультиграфы, в которых допускается наличие несколько рёбер между одной парой вершин.

Представления графов

Существует два основных способа представления графов:

- Матрица смежности (квадратная матрица, где элемент $a[i][j]$ указывает на наличие ребра между вершинами i и j);

- Списки смежности, где для каждой вершины существует список связей этой вершины с другими.

Применение графовых методов в исследовании онлайн-сообществ

Графовые модели служат эффективным средством для изучения взаимодействий в цифровых сообществах. Традиционно участники сети здесь отображаются как узлы графа, а их отношения — как связи. В зависимости от типа взаимодействия граф может быть направленным или ненаправленным. Так, в Twitter или Instagram подписка формирует направленную связь, так как не требует взаимности, а дружеские связи в Facebook или ВКонтакте являются ненаправленными, поскольку устанавливаются по согласию обеих сторон [3].

Методы исследования структуры графов

Исследование структуры графов представляет собой важное направление в анализе социальных сетей, позволяющее выявлять закономерности взаимодействий между участниками, оценивать степень связности, а также обнаруживать локальные и глобальные особенности сети. Для этого применяется широкий набор алгоритмических подходов, обеспечивающих эффективную обработку больших графовых данных. Одним из базовых инструментов являются алгоритмы обхода графа, применяемые для последовательного посещения вершин и анализа структуры связей. Наиболее известны два классических метода: поиск в ширину (Breadth-First Search, BFS) и поиск в глубину (Depth-First Search, DFS). Алгоритм BFS реализуется через очередь и последовательно исследует все соседние вершины текущего узла до перехода к следующему уровню. Он особенно эффективен в задачах определения минимального расстояния в невзвешенных графах, а также при построении дерева кратчайших путей. Сложность алгоритма составляет $O(V + E)$, где V — число вершин, а E — число рёбер, что делает его применимым для графов большого масштаба [4].

В противоположность этому, DFS исследует путь максимально вглубь, прежде чем вернуться и перейти к следующей вершине. Алгоритм основан на использовании стека, что позволяет выявлять скрытые связи между узлами. Он широко применяется для обнаружения компонент сильной связности, построения временных меток входа и выхода в ориентированных графах, а также в задачах топологической сортировки. При DFS можно идентифицировать мосты и точки сочленения — критические элементы структуры, удаление которых приводит к фрагментации сети.

Кроме базовых алгоритмов обхода, в социально-графовом анализе находят применение более сложные методы, направленные на изучение иерархической структуры сети. К ним относится алгоритм Тарьяна, предназначенный для поиска компонент сильной связности в ориентированных графах, и алгоритм Косарайю, работающий в два прохода и демонстрирующий высокую эффективность на плотных графах [5].

Для выявления циклов в графе используются модифицированные DFS-подходы с отслеживанием посещённых вершин и идентификацией обратных рёбер. Это важно в контексте анализа обратной связи в сетевых взаимодействиях и в изучении устойчивых поведенческих паттернов.

Отдельное направление составляет анализ центральности на уровне обхода, когда при обходе фиксируются показатели, такие как глубина вложенности узлов, количество посещений и порядок достижения. Эти данные затем используются в построении ранговых моделей, позволяющих оценить влияние пользователей внутри сети.

Также широко применяется метод обратного обхода (Reverse BFS/DFS), особенно актуальный для анализа влияния отдельных узлов на более ранние этапы распространения информации. Такой подход эффективен при моделировании ретроспективных сценариев, например, при изучении первоисточников дезинформации или выявлении начальных точек эпидемического распространения контента.

Визуализация результатов обхода, как правило, сопровождается построением деревьев обхода, на основе которых можно оценивать не только структурные особенности, но и динамические параметры сети — например, скорость распространения сообщений или среднюю глубину вовлечённости участников в коммуникацию.

Для более сложных сетей с весовыми характеристиками применяются гибридные методы, сочетающие DFS/BFS с анализом весов рёбер. Например, взвешенный обход с ограничением по глубине используется для изучения локального окружения конкретного узла с учётом интенсивности связей.

В последнее время активно развиваются методы параллельного обхода графов, позволяющие существенно сократить время анализа крупных сетей за счёт распределённой обработки. Такие технологии находят применение при работе с графами, содержащими миллионы и более элементов, в том числе при анализе сетей влияния или корпоративных коммуникаций.

Таким образом, методы исследования структуры графов охватывают широкий спектр алгоритмов, каждый из которых решает специфическую задачу в рамках анализа сетевой информации. Выбор конкретного метода определяется типом графа, характером исследуемых связей, а также масштабом и целью исследования. Совокупность этих подходов составляет методологическую основу для извлечения значимой информации из структурированной графовой модели социальных взаимодействий.

Оптимизация маршрутов во взвешенных графах

Если связи имеют числовые характеристики (например, частоту общения или уровень доверия), для поиска оптимальных путей используется алгоритм Дейкстры — метод вычисления кратчайших маршрутов от стартовой вершины ко всем остальным в графах с неотрицательными весами. В социальных сетях это позволяет оценить «близость» пользователей, основываясь на интенсивности их взаимодействий [6].

Идентификация связанных подграфов

Ключевой задачей является обнаружение компонентов связности — изолированных групп, где каждый участник связан с другими, но отсутствуют связи с внешними узлами. Это помогает:

Обнаружить замкнутые сообщества по интересам или иным признакам;

оценить уровень фрагментации сети [2];

Выявлять барьеры для распространения контента, например, в рекламных кампаниях.

Кластеризация пользовательских групп

Цифровые платформы часто содержат сообщества, сформированные по профессиональным, географическим или другим критериям [3]. Для их автоматического выделения используются следующие методы:

Метод Гирвана-Ньюмана — последовательное удаление связей с высокой промежуточной центральностью (частота участия в кратчайших путях), для разделения графа на кластеры;

Алгоритм Лувена — итеративная оптимизация показателя качества группировки узлов (модулярности), где сообщества объединяются в суперузлы до достижения максимальной плотности внутренних связей [7].

Итог

Графовые алгоритмы превращают массив данных социальных сетей в структурированную модель, раскрывающую скрытые закономерности. С их помощью можно:

Определять ключевых участников через метрики влияния;

Анализировать силу социальных связей;

Планировать стратегии продвижения;

Изучать эволюцию сети во времени.

Такой подход обеспечивает переход от хаотичных данных к логичной системе, на основе которой строятся аналитические выводы для принятия решений [8].

Метрики и характеристики социальных графов

Анализ графов в контексте социальных сетей требует количественной оценки свойств как отдельных узлов, так и структуры графа в целом. Метрики позволяют формализовать такие характеристики, как степень влияния, степень связанности, плотность взаимодействий и устойчивость сети. Выбор конкретных метрик зависит от целей анализа: от выявления ключевых участников до оценки целостности сети или её фрагментации.

Одной из базовых характеристик является степень вершины — число рёбер, инцидентных данной вершине. В ориентированных графах различают входящую и исходящую степень, что позволяет различать активных инициаторов коммуникации и пассивных получателей. Анализ распределения степеней по всем вершинам предоставляет информацию о неоднородности структуры: наличие вершин с высокой степенью указывает на наличие центров активности или лидеров мнений.

Важное место занимает центральность — обобщённое понятие, отражающее уровень значимости вершины в контексте всей сети. Наряду с уже упомянутыми степенной, близостной и посреднической центральностями, применяются более специализированные показатели:

Эксцентриситет вершины — наибольшее расстояние от данной вершины до любой другой, отражающее её максимальную удалённость в сети;

Эго-центральность — локализованный аналог центральности, рассчитываемый на основе непосредственного окружения вершины;

Пейджеранг (PageRank) — метрика, учитывающая не только количество, но и значимость входящих связей. Особенно полезна при оценке авторитетности в сетях, где присутствует направленная динамика, как в блогах, форумах или научных цитированиях;

Собственная центральность (eigenvector centrality) — расширяет представление PageRank, определяя значимость узла через значимость его соседей. Используется для построения моделей «социального капитала» и сетевого влияния.

Для анализа связности всей сети в целом рассчитываются глобальные показатели. Один из ключевых — диаметр графа, представляющий собой максимальное кратчайшее расстояние между двумя вершинами. Он характеризует наихудший сценарий достижения информации в пределах сети. Комплементарной метрикой выступает радиус, то есть минимальное значение эксцентриситета среди всех узлов.

Коэффициент кластеризации (clustering coefficient) измеряет вероятность того, что два соседа вершины также связаны между собой. Высокие значения этого коэффициента указывают на наличие устойчивых локальных сообществ и часто наблюдаются в социальных сетях, где распространены «треугольники доверия» — например, в профессиональных или семейных группах [7].

Плотность графа определяется как отношение числа имеющихся рёбер к максимально возможному их количеству. Эта характеристика позволяет оценить насыщенность связей в сети: высокая плотность говорит о высоком уровне взаимодействий, тогда как разреженные графы часто соответствуют децентрализованным сообществам.

Особое значение в социальных графах имеет модулярность — метрика, отражающая качество кластеризации. Она показывает, насколько хорошо граф разделяется на сообщества, в пределах которых связи плотные, а между которыми — слабые. Модулярность активно используется при автоматическом выявлении сообществ и оценке результатов кластеризации.

Дополнительные характеристики включают:

Средняя длина пути — среднее значение кратчайших расстояний между всеми парами узлов. Позволяет оценить эффективность распространения информации;

Коэффициент ассортативности — измеряет тенденцию вершин связываться с вершинами с аналогичной степенью. В социальных сетях часто наблюдается положительная ассортативность, когда высокоактивные пользователи взаимодействуют преимущественно друг с другом;

Коэффициент устойчивости — параметр, определяющий чувствительность графа к удалению вершин. Используется при оценке устойчивости сетей к атакам или сбоям.

В случае динамического анализа графа во времени используются временные метрики, позволяющие отслеживать изменения характеристик. Например, темп роста степени может свидетельствовать о повышении популярности участника, а динамика центральности — о смене лидеров мнений в сообществе [9].

Поиск сообществ в социальных сетях

Сообщество - группа вершин, плотно связанных внутри себя, но слабо — с остальными.

Методы разбиения

После получения исходных данных, специалисты обрабатывают их и используют методы кластеризации для получения базовых графов.

Примеры применения

Выявление преступных ячеек, ведущих коммуникацию с помощью социальных сетей [12]. После обнаружения подозрительной активности (например, по ключевым словам), следователи могут попытаться выявить других членов группировки, а также изучить структуру ячейки, используя моделирование связей между людьми, выражаемую в виде графов. В начале структура будет крайне неструктурированной, и из-за недостатка информации о действующих и возможных соучастниках группировки будет указывать на связи между людьми лишь косвенно, однако с появлением больших объёмов подтверждённой информации, система будет становиться точнее.

Выявление нескольких других потенциальных членов преступной ячейки может стать основой для будущей работы. Частота и объём сообщений, рассылаемые между контактами, позволит более точно отслеживать деятельность ранее обнаруженной группы, а при грамотном наблюдении — вывести на другие подразделения [10].

Поиск высокопоставленных членов группировок также может осуществляться с помощью анализа сообщений: частота, содержание, время отправки и особая шифровка может сильно выделяться на фоне общения между рядовыми участниками, что может быть невидимо при традиционных методах анализа, но очевидно на графе.

Практическая часть

С помощью Python и библиотеки NetworkX можно построить граф подозреваемых, рассчитать центральность вершин и визуализировать связи для облегчения работы с ними [11].

Интерпретация данных

Анализ и его визуализация может указать на приблизительное число участников ячейки, их активность за различные промежутки времени, структуру группировки и другие ценные данные для правоохранительных органов [12].

Заключение

Теория графов предоставляет мощные инструменты для анализа социальных сетей. В работе рассмотрены основные методы, позволяющие выявлять ключевых участников, сообщества и структурные особенности. Перспективы дальнейших исследований связаны с применением машинного обучения для прогнозирования динамики сетей.

Список использованных источников:

1. Учебное пособие Омельченко В.П. и Куликова А.С. «Теория графов в задачах и примерах» (2021)
2. Монография Барахнина В.Б. и Борисовой Л.Б. «Социальные сети: модели анализа и управления» (2019)
3. Статья Петрова А.А. и Смирновой Е.В. «Алгоритмы выявления сообществ в социальных сетях»
4. Практическое руководство Скворцова А.В. «Анализ данных в Python: от визуализации до машинного обучения» (2022)
5. Учебник Кормена Т., Лейзерсона Ч. и Ривеста Р. «Алгоритмы: построение и анализ» (русское издание, 2022)
6. Статья Иванова Д.К. «Метрики центральности в социальных сетях» (Журнал «Прикладная информатика», 2021)
7. Монография Гусятникова В.Н. «Сетевой анализ: математические модели и приложения» (2018)
8. Вебинар «Анализ социальных сетей с помощью Python» (платформа Stepik, 2023)
9. Статья Козлова М.И. и Федоровой Т.П. «Применение алгоритма Лувена для кластеризации пользователей Twitter» (eLibrary, 2020)
10. Учебник Черкасова Е.А. «Математические методы в социологии» (2020)
11. Руководство Сидорова П.Р. «Визуализация графов в Gephi и NetworkX» (платформа Habr, 2022)
12. Монография под редакцией Новикова Д.А. «Динамика социальных сетей: от структуры к прогнозированию» (2021)

UDC 519.17

APPLICATION OF GRAPH THEORY IN THE ANALYSIS OF SOCIAL NETWORKS

Shilov B.V. student gr.477602; Yakzhik A.D. student gr.477602

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Rusina N. V. - Senior Lecturer of the Department of EI

Abstract: Social networks have become an integral part of modern society, forming complex systems of interactions between people. The analysis of these three-dimensional and complex structures requires a mathematical apparatus capable of describing the relationships and patterns between their elements. Graph theory provides tools for modeling and studying social networks, which makes its application particularly in demand in sociology, marketing, IT, criminology and other fields.

Keywords: graph, social networks, graph node, route, clustering, communities, centrality.