

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра социальной информатики

ПОСТРОЕНИЕ СОЦИАЛЬНОГО ГРАФА

(автореферат бакалаврской работы)

студента 5 курса 531 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
Социологического факультета
Торгашина Дениса Игоревича

Научный руководитель
кандидат социологических наук , доцент

_____ К.В. Мохнаткина
подпись, дата

Зав. кафедрой
кандидат социологических наук, доцент

_____ И.Г. Малинский
подпись, дата

Саратов 2022

ВВЕДЕНИЕ

Актуальность темы. В обществе имеется большое количество информации, данную информацию можно изобразить в виде объектов и связей между ними. В качестве примера такими объектами могут быть представлены научные работы, тогда если одна из работ ссылается на другую, то между этими работами образуется связь. Из чего можно заключить, что существующие научные исследования могут быть представлены в виде графа. Аналогичное представление может быть для иных структур из различных областей знаний. Такие как люди и их взаимоотношения, веб-сайты и приложения.

Для похожих структур появляется возможность их представление в виде графа. Примеры, которые будут представлены, обозначаются как социальные графы.

Развитие социальных сетей совершило огромное изменение в области социальных графов. На сегодняшний день количество зарегистрированных аккаунтов в социальной сети Фейсбук насчитывает более 2.5 миллиардов, тем самым объясняя популярность данной площадки для анализа в коммерческих и научных целях. Подобная множественная аудитория позволяет направить интерес на тему различия социальных сетей, на исследование взаимоотношений, связей объектов, а также создавать успешные модели социальных графов.

Различные фирмы позволяют использовать разнообразные средства для классификации по признакам. Таких как национальность, возраст, пол, а также сообществ, образование и медиаконтент. В то время как классификация способствует определению специфики социальных отношений, обнаружить пересечение интересов и направление социальных векторов. Подобная информация окажется познавательной не только для оценки личной аудитории, но и для обнаружения разнообразных зависимостей в общественных отношениях.

По всему миру количество международных сетей насчитывается более 20 штук. В данных сетях различные люди контактируют между собой. При этом основывая свой круг общения с различными интересами. В социальных сетях у каждого пользователя есть свой профиль. В нем доступна определенная информация, которую могут видеть другие пользователи. Например, может быть фамилия, имя, номер телефона. Также значительно количество предпринимателей продвигают свои сервисы в социальных сетях, образуя отдельные сообщества и страницы. В данном случае каждая фирма имеет свою аудиторию людей, и она может рекламировать свою продукцию. Помимо этого, социальные сети являются удобной платформой для проведения социологических исследований.

Граф – это математический объект показывающий множество вершин графа и набора ребер. Граф как объект имеет целый ряд соединений вершин. Из этого следует, что при анализе социальной сети понимаются социальные отношения узлов и связей. Узлы графа являются отдельными элементами внутри сетей, а связи это отношения между ними. Между узлами графа возможно множество видов связей. Различные научные исследования показали, что социальные сети функционируют на многих уровнях и в определении решения задач исполняют основную функцию.

Целью выпускной квалификационной работы является построение социального графа с помощью программных средств.

Сформулированная цель достигается решением, следующим **задач**:

1. Провести теоретический анализ доступных моделей, методов и программных средств, которые будут использованы при построении социального графа;

2. Проанализировать методы анализа социальных сетей на основе теории графов;

3. Построить социальный граф.

Объектом исследования является социальная сеть.

Предметом выступают методы построения социального графа.

Структура работы. Выпускная квалификационная работа структурирована следующим образом: введение, два раздела, заключение и список использованных источников.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Теоретико-методологические подходы к построению графа» описывает анализ и визуализацию сетей. Приводится описание базовых задач анализа социальных сетей, к которым относятся:

- исследование информационных потоков;
- исследование настроения сообщений (эмоциональность);
- исследование и выборка тем (что чаще всего обсуждают);
- исследование изображений.

А способом визуализации сетей является граф.

Также описаны типы взаимоотношений, которые могут являться направленными и ненаправленными. В основном акцентируют два типа связей:

- дружба (в данном случае люди являются, знакомы друг другу);
- интересы (в таком случае у людей имеются совместные интересы или они входят в группу интересов). Эти отношения используются как пример описания людей, их активности и отношений к другим участникам и объектам.

Так же рассмотрены главные методы исследования социальных сетей, базированные на визуализации графов. К которым относятся: Центральность вершин в графе: которая является одним из главных понятий в анализе социальных сетей, которые в свою очередь оценивают важность вершины в графе по сравнению с иными вершинами. Степень центральности: которая является одной из самых первых, а также является самым простым. Определяется как количество связей попадающих на узел, то есть количество связей имеющих узел. В большинстве случаев степень может быть интерпретирована в непосредственном риске узла для захвата информации, через которую протекает сеть. Близость центральности: которая числится в графе средним показателем кратчайшего пути между узлом и остальными

участниками в графе. Показатель «близость центральности» отражает, степень сложности для достижения определенного узла в сети.

Приводятся свойства социального графа:

- Компонент связности – это часть графа, которая является связной. Формально компоненты связности подразумевают набор вершин графа, между которыми существует путь. Понятие связности распространяется на неориентированный граф.

- Среднее расстояние графа. Под расстоянием между вершинами графа понимают число ребер в кратчайшем пути. В социальных сетях чаще всего имеется маленькое расстояние между вершинами. Свойство может проявляться в модели современного общества.

- Распределение степеней узлов. В анализе графов степенью узла называют количество его связей с другими узлами. Это распределение вероятностей во всей сети. Если граф считается ориентированным, то есть ребра имеют направления от одной вершины к другой, то его узлы имеют два значения степени. Исходящая степень представляется как число исходящих ребер, а входящее как число входящих ребер.

- Коэффициент кластеризации. Это средство степени, в которой узлы группируются вместе. В большинстве социальных сетей, узлы склонны создавать связанные группы. Данные коэффициенты показывают определенную характеристику.

В свою очередь рассмотрена одна из самых популярных мер качеств сетей или графов для данной задачи, именуемая как модулярность. Опция была предложена Ньюманом и Гирваном в процессе разработки алгоритма кластеризации графа. Модулярность — равна доле ребер от общего числа которые попадают в данные ребра минус их ожидаемые доли ребер, если бы распределение было случайным.

Достоинства модулярности.

Модулярность просто интерпретируется. Ее значение равно разности между частей ребер в сообществе и предполагаемой доли его связей, в случае

если бы ребра были размещены случайным образом. Модулярность имеет возможность качественно пересчитывать при небольших изменениях в кластерах.

К недостаткам можно отнести следующее:

Функционал не является непрерывным, и задача его оптимизации — целочисленный. Для поиска глобальной оптимизации используют ориентировочные схемы. Какие-то из них действительно могут оптимизировать значение функционала, другие же по значению модулярности выбирают наилучшее найденное решение, то есть без гарантированного оптимального решения

Во втором разделе «Построение социального графа с применением алгоритма выделения сообществ» приведен пример построения социального графа. Для начала была определена задача выделения сообществ, а также приведены и описаны некоторые алгоритмы сообществ.

Промежуточность (Betweenness) — алгоритм на основе коэффициента меры центральности графа, который определяется на количестве кратчайших путей для каждой пары вершин, которая проходит через ребро.

Метод, разработанный Ньюманом и Гирваном. Алгоритм работает по следующей схеме: сначала происходит подсчет коэффициентов центральности на ребрах графа. Далее происходит удаление ребер в порядке очереди с самым большим коэффициентом, а сообществами считаются остальные компоненты связности. Процедура удаления связей завершается, когда достигает своего максимума итогового разбиения

Жадный алгоритм (Fastgreedy) — это метод, который основан на жадной оптимизации функции модулярности. Этот метод состоит в жадной оптимизации модулярности. Сначала сообщества инициализируются в каждой вершине, затем перебираются пары смежных вершин, после чего алгоритм максимизирует значения модулярности путем передвижения вершин в конкретные сообщества.

Многоуровневая очередь (Multilevel) — данный метод, основан на

многоуровневой оптимизации функции модулярности, в котором используется эвристика. Это наиболее продвинутый способ оптимизации того же перечня возможностей, которая более качественна чем алгоритм жадной оптимизации.

Случайные блуждания (Walktrap) — алгоритм, который основан на случайном блуждании. Метод применяет идею о том, что короткие беспорядочные блуждания не приводят к выходу из текущего сообщества. Данный метод основан на методе случайного блуждания, при этом незначительные беспорядочные блуждания не дают возможности выйти из сообщества.

Infomap — метод случайного блуждания, который базируется на понятии информационных потоков в сетях, кодирования и сжатия информации. Данный метод, в случае, как и метод (Walktrap), применяет механизм случайных блужданий. При прохождении каждого шага объект находится на вершине и из соседних вершин он перемещается в вершину беспорядочным образом.

Собственный вектор (Eigenvector) — метод, базирующийся на личных векторах матрицы модулярности, при этом она происходит из матрицы смежности. Также мы получаем Спектральный способ, который будет основан на личных векторах матрицы модулярности.

В конце раздела показан пример построения социального графа. Тестирование работы алгоритма проводилось на сгенерированных графах. Была приведена зависимость времени выполнения алгоритма от средней степени узла.

Далее были приведены примеры анализа реальных подграфов социальных сетей, где вершины — это профили социальной сети, а связи — отношения дружбы между ними. Данные взяты из социальной сети «ВКонтакте». На примерах было видно, что вершины из одного сообщества почти все связаны между собой, но при этом слабо связаны с вершинами из других сообществ. Кроме того, видно, что в сети присутствует несколько объектов, которые имеют связи с вершинами из всех сообществ. Представлено 2 графа, в первом варианте вершины сети, которые размещены случайным образом, во втором варианте с использованием алгоритма выделения

сообществ. Согласно полученной информации можно сделать вывод, что при формировании сообществ «друзей» элементы одного сообщества формировались по окружности между друг другом. Кроме того можно увидеть, что остальные сообщества формируются также относительно других сообществ.

ЗАКЛЮЧЕНИЕ

В данной работе были описаны методы построения анализа социальных сетей. Основными методами анализа социальных сетей прибывают методы визуализации графов.

Также был показан обзор основных методов для выделения сообществ в социальном графе. Показаны и проанализированы модели и алгоритмы визуализации сети в виде графа. При изучении было выявлено, что модели графов используются для преобразования математических данных графа в двумерные геометрические данные, основанную на условном расположении узлов и ребер, автоматического вытягивания графа и реализации его визуализации.

В ходе исследования анализа социальных сетей были приведены примеры графов, как математическим способом, так и визуальным. Показана суть методов исследования на конкретных примерах с описанием их сущности. Было дано понятие социального графа и его свойств.

Рассмотрена оценка критерий качества решений. При его выявлении определяется выбор значений разбиения. В равной мере были описаны и продемонстрированы методы выделения сообществ.

После чего было описано распознавание структуры социальных сетей, которое помогает выявить правильное представление для поиска информации, и была рассмотрена задача по заданному мультиграфу. Исходя, из этого работа алгоритма была описана математически, а также обозначен быстрый метод оптимизации.

В конечном итоге алгоритм использовался на сгенерированных графах. При этом алгоритм был протестирован в работе и были показаны и описаны примеры результатов алгоритма выделения сообществ.