

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

Кафедра дискретной математики и информационных технологий

**ИЗУЧЕНИЕ ИЗОМОРФИЗМА СОЦИАЛЬНЫХ ГРАФОВ
МАТРИЧНЫМ МЕТОДОМ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Зыковой Анна-Марии Борисовны

Научный руководитель
доцент, к. ф.-м. н.

Л. Б. Тяпаев

Заведующий кафедрой
доцент, к. ф.-м. н.

Л. Б. Тяпаев

Саратов 2020

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Основная часть	5
ЗАКЛЮЧЕНИЕ	10
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	11

ВВЕДЕНИЕ

Изучение социальных сетей, в том числе тех, которые выстраиваются на основе платформ в Интернете необходимо не только для оптимизации рынка информационных услуг и научного интереса. Изучение социальных образований, структур и законов внутри социальных объединений является основой для разработки концепций машинного обучения, психологического содержания дистанционного и кабинетного обучения и даже систем безопасности самих платформ социальных сетей.

В свою очередь, практически любая задача, связанная с подобием может быть приведена к решению задачи изоморфизма графов. Именно поэтому необходимы эффективные алгоритмы определения сходства структур графов на основе прямых признаков в категориях изоморфизма. Считается, что задача определения изоморфизма без полного перебора подстановок сходства не может быть решена, но прямой алгоритм перебора имеет фрактальную сложность.

Таким образом, актуальность работы определена совокупностью неизученности обеих этих задач.

Задачи и цели:

- обзор методов генерации и анализа случайных графов, описывающих социальные сети;
- разработка программного продукта для облегчения процесса анализа и моделирования случайных графов, описывающих социальные сети;
- проведение сравнения основных алгоритмов генерации случайных графов, описывающих социальные сети;
- обозначить цели, задачи и методы исследований изоморфизма графов, поиска спектра графов;
- формулирование метода оптимизации прямой проверки изоморфизма графов с помощью вычисления инварианты: спектра графов;
- проведение тестов методов на реальном социальном графе и сгенерированных графах.

Выпускная квалификационная работа структурно состоит из введения, пяти глав основной части, заключения. Имеется библиографический список использованных источников и приложения. Общий объем работы –142 страницы, из них 52 страницы – основное содержание, включая заключение.

Первая часть дает определение графа математически приводит понятие изоморфности графа, матричных методов анализа. Делается обзор современного подхода к работе с социальными графами.

Во второй части делается обзор на различные модели построения случайных графов социальных сетей, приводится практический пример построения изображения существующего социального графа отдельной группы на платформе Twitter.

В третьей описываются методы расчета спектра графа и приводится алгоритм, используемый в реализованном программном продукте.

В четвертой и пятой частях приводятся матричные методы анализа изоморфности графов, приводится ряд основных теорем, описывающих свойства собственных значений матриц смежности случайных графов и приводится предложенный на рассмотрение метод установления изоморфизма социальных графов на основе семплирования(алгоритм Метрополиса), кластеризации(алгоритм Маркова) и эвистики(инвариантой является спектр).

1 Основная часть

Под социальной сетью в данной работе понимается ресурс, содержащий данные пользователей и связи между ними, позволяющий пользователю:

- создавать публичный или частично публичный профиль, в котором он может указывать личную информацию
- задавать и поддерживать список других пользователей, которые его интересуют и которые интересуются им.
- просматривать свои и чужие списки связей и посещать профили разных списков
- создавать публичные записи, комментировать записи других пользователей.
- обмениваться с другими пользователями текстовой мультимедийной и иной информацией.

Под социальной связью понимается любая связь между двумя пользователями, произведенная с помощью инструментов ресурса. Социальную сеть, как и любую сеть, можно математически моделировать графом, в котором вершины представляют объекты сети, а ребра — взаимоотношения.

Матрица смежности графа G с конечным числом вершин (пронумерованных числами от 1 до n) — это квадратная матрица A размера n .

Спектром графа G называется вектор (чаще всего отсортированный) собственных значений матрицы смежности.

Результатом данной выпускной квалификационной работы является эвристический алгоритм проверки изоморфизма социальных графов посредством методов случайных блужданий на основе изоспектральных свойств матриц смежности.

Алгоритм подразумевается как оптимизация прямого метода проверки изоморфизма графов, который, как известно факториальной временной сложности в зависимости от количества вершин графа.

Алгоритм предполагает последовательность шагов:

1. Первичная проверка на равное количество вершин и ребер анализируемых графов.
2. Поиск локального или глобального минимума в задаче перестановки строк и столбцов с помощью алгоритма Метрополиса над матрицами смежности.

3. Симметричная кластеризация полученных изоспектральных графов.
4. Расчет спектров каждого кластера графов.
5. Поиск симметричных изоспектральных кластеров.

Идея алгоритма в том, что если два изоспектральных графа имеют одинаковое количество кластеров и каждому кластеру в одном графе можно найти изоспектральный ему кластер в другом графе, эти графы изоморфны.

Ядром алгоритма соответственно являются алгоритм семплирования Метрополиса и алгоритм кластеризации Маркова.

Входными значениями для алгоритма являются спектр и список спектров кластеров одного графа и матрица смежности другого.

До тех пор, пока имеющийся спектр графа не будет равен рассчитываемому спектру полученной матрицы смежности, матрица смежности семплируется в новую перестановкой строк и столбцов с помощью алгоритма Метрополиса. Затем полученная матрица смежности кластеризуется с помощью алгоритма Маркова.

Затем среди полученных кластеров ищутся изоспектральные тем, которые относятся к другому графу. Если хотя бы пара кластеров не подобны, делается вывод, что графы не подобны.

Объектом исследования не случайно выбраны именно социальные графы - основой все еще служит прямой алгоритм проверки изоморфизма графов с помощью установления перестановочного подобия двух матриц одного порядка. Графически метод перестановочного подобия двух матриц выглядит как если есть два одинаковых отображения одного графа в которых вершины пронумерованы на разному и задачей является рассмотреть все возможные варианты перенумерации вершин, пока два графа не будут выглядеть идентично.

Социальные графы, особенно графы, построенные на основе данных с интернет-платформ социальных сетей дают матрицы смежности удовлетворяющих условию марковости. В Марковской сети вероятность события зависит только от текущего состояния сети. Для матриц это означает, что возможность попасть в какую-то вершину зависит только от того, в какой вершине на данной итерации находится алгоритм при случайном блуждании.

Для поиска изоспектральной близости кластеров (а не полного равен-

ства) достаточно, если максимальные значения собственных значений были равны. Это упрощение возможно в силу спектральных свойств неотрицательных матриц, к которым относятся матрицы смежности социальных графов.

Для решения задач расчета инвариант и облегчения построения графиков распределения степеней вершин для различных моделей генерации случайных графов, был разработан программный продукт.

SpectraPictureAnalys (S[pic]trA) - инструмент для анализа графа в рамках дипломной работы "Изучение изоморфизма социальных графов матричным методом выполнен на языке C# в среде разработки Visual Studio 2010. В состав программного продукта так же входят Readme и файлы из примеров. Вес программного продукта 800 КБ. Установочная программа создана с помощью Inno Setup Compiler 5.4.1.

Программа позволяет:

1. генерировать случайные графы по заданным показателям с помощью алгоритмов моделей генерации случайных графов Эрдеша-Реньи, Барабаши-Альберта, Уоттса-Строганца и модели копирования.
2. отрисовывать на плоскости сгенерированный ранее или созданный иным способом и представленный в формате как в Matis граф;
3. строить таблицы смежности графов;
4. отрисовывать на плоскости график распределения степеней узлов ;
5. подсчитывать кластерный коэффициент и спектр случайного графа;
6. проводить кластеризацию графа;
7. восстанавливать по спектру оригинального графа и матрице смежности кластеризованного графа оригинальный граф;
8. проводить тестирования в фоновом режиме.

Пример в данной работе базируется на данных из Twitter. Поскольку любые виды деятельности связанные с обработкой персональных данных в том числе физическими лицами с использованием средств автоматизации должны в соответствии с Федеральным законом от 27.07.2006 N 152-ФЗ (ред. от 31.12.2017) "О персональных данных" для осуществления научной деятельности проводится с согласия субъекта персональных данных на обработку его персональных данных специально для дипломной работы в Twitter был создан приватный список-сообщество, куда мог вступить любой желающий помочь в данной работе и согласный на обработку своих персональных дан-

ных.

Для каждого из участников с помощью Твееру были скачены последние 100 твитов и на основе активностей пользователей списка был создан граф реальной социальной сети. Ограничение анализируемого сообщества списком выборки из окружения фолловеров и фолловеров фолловеров автора выпускной квалификационной работы позволило обойти алгоритмы предотвращения несанкционированного автоматического сбора данных и ограничения нагрузки на инфраструктуру сервиса и разработка или использование систем автоматического сбора данных пользователей с целью построения социального графа не потребовалась.

Искусственно сгенерированные графы были созданы с помощью разработанного автором программного продукта по примитивным моделям: Эрдоша–Реньи(множество вершин определено, а ребра существуют с заданной вероятностью), Барабаши – Альберта(генерация методом предпочтительного присоединения), "копирования"(динамическая генерация из меньшего, заданного регулярного графа) и "малого мира"(модель генерации графа, который обладают эффектом «тесного(малого) мира»). Результатом процесса генерации являлся массив данных о связях, в формате как в программах визуализации графов Metis , Chaco , и других.

Случайным образом было сгенерировано 400 графов (по 100 графов на каждую из моделей) по 200 вершин каждый и для каждого было создано по 100 заведомо изоморфных отображений. Вероятно, в силу малой выборки и простоты алгоритма рандомизации графов, а также примитивности выбранных алгоритмов генерации случайных графов, такая проверка на ложноотрицательные результаты показала, что для графа любого метода генерации проверка изоморфности с помощью действий над матрицами смежности всегда выполняется удачно.

Однако, к сожалению, ложноотрицательный тест над реальным графов, созданным на основе социальной сети Twitter, выявил 0.5 процентную возможность возникновения ложноотрицательных результатов. Над единственным реальным графом было построено 400 "сломаных"графов и два из них остановили процесс выполнения алгоритма Метрополиса для полной матрицы из-за слишком большого количества семплов раньше, чем удовлетворительный результат был сформирован.

Все известные алгоритмы, гарантирующие правильный ответ, являются как минимум экспоненциальными (EXPTIME). И так, если алгоритм Маркова имеет вычислительную сложность $O(N \cdot k)$ и алгоритм Метрополиса так же линейно зависит от размеров и количества генерируемых семплов $O(N \cdot k)$, где k – число кластеров (числа, являющиеся ограниченными сверху константой в реализованном программной продукте).

Так в среднем скорость исполнения блока семплирования изоспектральной образцу матрицы (полученной путем перемешивания вершин из матрицы образца) на AMD FX8350 8-Core 4.2GHz и оперативной памятью 8 Гб для графов в 200 вершин занимала 8,5 минут. Тестирование алгоритма, представленного в выпускной квалификационной работе, было выполнено поблочно, поэтому произвести расчет времени не представляется возможным.

ЗАКЛЮЧЕНИЕ

На основе эвристики, используемой для проверки изоморфизма графов, построен алгоритм решения задачи, представляющей собой цикличное семплирование матриц смежности на основе случайных блужданий.

Разработан программный продукт для облегчения процесса анализа и моделирования графов, а так же проведение сравнения основных алгоритмов генерации случайных графов, описывающих социальные сети и оценки характеристик случайных графов посредством методов имитационного моделирования и нахождение зависимостей этих характеристик от объема графа и параметра распределения степеней вершин.

Дальнейшими перспективны развития являются:

- распараллеливание вычислений на несколько ЭВМ для ускорения работы;
- дальнейший анализ и эксперименты для установления вероятности ложноположительных результатов;
- исследование с целью распространить алгоритм для графов других классов;
- точное вычисление временной сложности алгоритма.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Райгородский А.М. Модели случайных графов и их применения / А.М. Райгородский // Труды Московского физико-технического института - 2010. - №4 - С.130-147
- 2 Берновский М.М. Случайные графы, модели и генераторы безмасштабных графов / М.М. Берновский, Н.Н. Кузюрин // Труды Института системного программирования РАН - 2012. - №22 - С.419-434
- 3 Евин И.А. Введение в теорию сложных сетей / И.А. Евин // Компьютерные исследования и моделирование - 2010. - №2(2) - С.121-141
- 4 Ключарев П.Г., Исседование спектральных свойств социального графа сети livejournal/П.Г. Ключарев , В.О. Честноков, //МГТУ им Н.Э.Баумана. Электрон.журн.- 2013. - №9 - С.391-396
- 5 Ключарев П.Г., Спектральные методы анализа социальных сетей/Ключарев П.Г., Басараб М А. //Машиностроение и компьютерные технологии - 2017. - №5 - С.168-177
- 6 Lutzeyer J. F., Comparing Graph Spectra of Adjacency and Laplacian Matrices/Lutzeyer J. F., Walden A. T., //arxiv: Methodology - 2017. - С.191-202
- 7 Пролубников А. В., Алгоритм поиска приближенного решения задачи проверки изоморфизма подграфов / Пролубников А. В. //Математические структуры и моделирование - 2003. - №1 - С.59-67
- 8 Исхаков Л., Кластерный коэффициент в модели пространственного предпочтительного присоединения/Л. Исхаков , М. Миронов, Л. Прохоренкова //Доклады академии наук - 2018. - №481 - С.10-14
- 9 Лери М. М.,Об ассортативности и кластеризации степенной модели конфигурационных графов/Лери М. М. //Труды Карельского научного центра РАН - 2019. - №7 - С.30–35
- 10 Задорожный В. Н.,Точная теория графа Барабаши-Альберт/Задорожный В. Н. //Омский научный вестник - 2009. - №3 - С.13-18

- 11 P. Diaconis, What Do We Know about the Metropolis Algorithm?*/P. Diaconis //Journal of Computer and System Sciences - 1998. - №57 - С.20-36
- 12 Александр Тумачек, Применение QR алгоритма для расчёта собственных структур корреляционных матриц на платформе ADSP TS201 Фирмы Analog Devices/Александр Тумачек //Современная Электроника : Проектирование И Моделирование - 2009. - С.66-69
- 13 Луценко Е. В., Некоторые проблемы классического кластерного анализа/Луценко Е. В., Коржаков В. Е., //Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. - 2011. - №2
- 14 Шалагин С.В., Нурутдинова А.Р.Сравнительный анализ вычислительной сложности алгоритмов идентификации конечных простых однородных цепей Маркова/ Шалагин С.В., Нурутдинова А.Р. //Вестник Казанского технологического университета. - 2016. - №13
- 15 Вильховский д.э.Определение изоморфности графов/Вильховский д.э., Ефимов с.с. //Вестник омского университета - 2012. - №4 - С.141-143
- 16 Сайфуллина Е.С.,Эвристический подход к проверке изоморфности графов/ Сайфуллина Е.С. //Информационные технологии и нанотехнологии (итнт-2016) - 2016. - С.1063-1066
- 17 Martin W. Liebeck, On Graphs Whose Full Automorphism Group is an Alternative Group or a Finite Classical Group/Martin W. Liebeck, //Proceedings of the London Mathematical Society - 1983. - №3 - С.337-362
- 18 Ключарев П.Г. Спектральные методы анализа социальных сетей/Ключарев П.Г. Басараб М.А, //Наука и Образование.МГТУ им Н.Э.Баумана. Электрон.журн. - 2017. - №5 - С.169-177
- 19 Batura T, Методы анализа компьютерных социальных сетей/Batura T, //Вестник Новосибирского государственного университета. Серия: Информационные технологии - 2012. - №10 - С.13-28
- 20 Володичева М.И., Исследование изоморфизма графов с помощью жордановых форм матриц смежности/ М.И.Володичева, С.Н. Леора //Прикладная дискретная математика - 2018. - №40 - С.87-99

- 21 Беспалов А.А., Матричный метод проверки изоморфизма графов/Беспалов А.А. //Прикладная дискретная математика - 2004. - №3 - С.4-11
- 22 Пролубников А. В., Прямой алгоритм проверки изоморфизма графов/Пролубников А. В. //Компьютерная оптика - 2007. - №31 - С.86-92
- 23 Meyerhenke, Drawing Large Graphs by Multilevel Maxent-Stress Optimization/Meyerhenke, Henning and Llenburg, Martin and Schulz, Christian //arxiv:1506.04383v2 - 2015. - №6
- 24 А.А. Айзенберг, Пространства матриц и выпуклые многогранники/А.А. Айзенберг. - Москва:Факультет компьютерных наук НИУ ВШЭ, 2018.
- 25 Гусарова Н.Ф., Анализ социальных сетей. Основные понятия и метрики/Гусарова Н.Ф. - Санкт-Петербург:Университет ИТМО, 2016.
- 26 Тараканов В. Е., Графов Теория/Тараканов В. Е.// Большая российская энциклопедия. -2007. - №7 - С.644-645
- 27 Bollobes Bela, Random Graphs/ Bollobes Bela - Cambridge University Press,2001.
- 28 Алексеев В.Е., Графы. Модели Вычислений. Структуры Данных/ Алексеев В.Е., Таланов В.А. - Нижний Новгород: Издательство Нижегородского государственного университета,2004.
- 29 Н. С. Бахвалов, Численные методы/ Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков - Москва: Лаборатория знаний - 2008.
- 30 Алексеев В.Е., Графы. Модели вычислений. Структуры данных: Учебник/ Алексеев В.Е., Таланов В.А., - Нижний Новгород: Издательство Нижегородского государственного университета, 2004.
- 31 Karsten Borgwardt, Metropolis Algorithms for Representative Subgraph Sampling/Karsten Borgwardt, Zoubin Ghahraman // Eighth IEEE International Conference on Data Mining - 2008.
- 32 Population, total, // World Development Indicators (WDI) URL: <https://data.worldbank.org/indicator/SP.POP.TOTL> Загл. С экр. Яз. Англ.
- 33 Tweepy Documentation, URL: <http://docs.tweepy.org/en/latest/>