

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

Кафедра информатики и программирования

АНАЛИЗ СОЦИАЛЬНЫХ СЕТЕЙ С ПОМОЩЬЮ ГРАФОВ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 441 группы

направления 02.03.03 Математическое обеспечение и администрирование
информационных систем

факультета компьютерных наук и информационных технологий

Сурковой Кристины Алексеевны

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

М.В. Огнева

Заведующий кафедрой
к.ф.-м.н., доцент

М.В. Огнева

Саратов 2023

ВВЕДЕНИЕ

Анализ социальных сетей представляет собой область анализа данных, в которой используется теория графов для понимания социальных структур. Например, пользователей социальной сети можно представить вершинами (узлами) графа, связав знакомых людей ребрами. Эту идею можно применить не только для вычисления самого популярного блогера или его поста, но и для выявления неформальных лидеров в крупных организациях. Направленные графы могут показывать авторитет и уровень полномочий, а также силу влияния отдельного человека на принятие управленческих решений [1].

Отдельное направление в науке о сетях отведено случайным сетям [2].

Цель бакалаврской работы – исследование и сравнительный анализ характеристик моделей случайных сетей и реальных социальных сетей.

Поставленная цель определила **следующие задачи**:

1. Дать основные понятия и определения.
2. Рассмотреть модели случайных графов, такие как модель Эрдеша-Реньи, Барабаши-Альберт и Боллобаши-Риордана.
3. Разработать генератор графов на основе разных моделей графов.
4. Подобрать примеры реальных сетей из открытых источников.
5. Проанализировать и сравнить свойства полученных моделей с реальными сетями.

Методологические основы анализа социальных сетей с помощью графов представлены в работах А.Н.Вичугова [1], С.Сидорова [2], Дж.Уоттса [3], Н.Ф.Гусарова [4]. Основы и преимущества некоторых технологий анализа сетей представлены в работах К.М.Ушакова [5] и А.М.Райгородского [6].

Теоретическая значимость бакалаврской работы. Произведен сравнительный анализ характеристик реальных и моделируемых сетей. Полученный результат демонстрирует поведение моделей случайных графов размерности аналогичной реальным данным.

Практическая значимость бакалаврской работы. Создано приложение для генерации моделей графов и их обработки, которое может быть применимо при комплексном анализе реальных и случайных сетей.

Структура и объём работы. Бакалаврская работа состоит из введения, 3 разделов, заключения, списка использованных источников и 6 приложений. Общий объем работы – 86 страниц, из них 56 страниц – основное содержание, включая 21 рисунок, цифровой носитель в качестве приложения, список использованных источников информации – 24 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Основные понятия и определения» посвящен изучению ключевых терминов, связанных с графами. В нем представлены основные характеристики графа, такие как диаметр, средняя длина кратчайшего пути и степень вершины. Также изучены понятия коэффициента ассортативности, все виды коэффициента кластеризации и мер центральности, определения плотности и взаимности сети. Дополнительно рассмотрено распределение степеней узлов в сложных сетях, поскольку большинство узлов в них имеют низкую степень, но в то же время встречаются узлы с высокой степенью. Так, распределение степеней отличается от того, которое возникло бы в совершенно произвольных сетях.

Информация из данного раздела применяется и в теоретической, и в практической части.

Второй раздел «Обзор математических моделей социальных сетей» посвящен изучению принципов построения моделей случайных графов. Существует много видов моделей, генерирующих случайные графы. Их можно разделить на несколько основных классов:

- Модели случайных графов (модель Эрдёша – Реньи).
- Простейшие модели безмасштабных сетей (модель Боллобаша – Риордана, модель копирования и др.).
- Более гибкие модели безмасштабных сетей (модель Чунг-Лу, модель Янсона–Лучака).

В работе рассматриваются три модели: Эрдеша-Реньи, Боллобаша-Риордана и Барабаша-Альберта.

Одной из характеристик социальных сетей являются специфические законы распределения степеней узлов. В большинстве сетей преобладают узлы с низкой степенью, но в то же время имеются узлы высокой степени,

что является типичной особенностью социальных сетей и необходимым условием для определения сети как сложной. Поэтому в работе по каждой модели было рассмотрено распределение степеней узлов. Выяснилось, что модель Эрдеша-Реньи обладает дискретным распределением, что отличает ее от реальных сетей.

Третий раздел «Практическая часть» посвящен разработке программного обеспечения для выполнения экспериментов по моделированию случайных и реальных сетей. Для реализации использовался язык C# на платформе .Net. Изначально был создан основной класс `Graph` с необходимыми полями. Доступны вычисления степени вершины, диаметра, медианы графа. После чего были добавлены методы получения коэффициентов ассортативности и кластеризации, плотности и мер центральности для одной вершины графа, чтобы в дальнейшем проводить сравнительный анализ графиков распределения. Описанный функционал применим как к смоделированным сетям так и к реальным.

Для генерации моделей случайных графов первоначально были определены соответствующие требования. Далее создан модуль генерации случайных сетей с заданными параметрами, которые зависят от конкретной модели генерации случайных сетей. Построение модели Эрдеша-Реньи зависит от вероятности p и задействует лишь один метод основного класса. А модели Боллобаши-Риордана и Барабаши-Альберта хоть и не принимают никаких параметров, но задействуют 3 и 4 функции соответственно. Построенная сеть обрабатывается — вычисляются значения выбранных коэффициентов. Каждый результат вычисления данных величин для отдельного эксперимента записывался в файл. После чего все значения по характеристикам вершин переносятся на график распределения, а минимальные, максимальные и средние значения заносятся в таблицу. Имея описанные визуальные отображения, проще

проводить дальнейший анализ. В итоге получилось, что независимо от размера графа модель Эрдеша-Реньи обладает низкой плотностью, высоким коэффициентом кластеризации и близким к нулю коэффициентом ассортативности. А модели Боллобаши-Риордана и Барабаши-Альберта, наоборот, имеют зависимость от размера сети. С ростом количества вершин в графах увеличивается коэффициент кластеризации. Однако размер не влияет на меры центральности, плотность и ассортативность.

Полученные результаты были сопоставлены с аналогичными вычислениями для трех реальных социальных сетей, загруженных из онлайн хранилищ данных о сетях, соответствующих взаимодействию пользователей в реальных системах. Размеры сетей составляют 1876, 3330 и 553 вершин. В среднем данные первого и второго социальных графов обладают низкой плотностью, коэффициентом кластеризации выше его среднего значения и близким к нулю коэффициентом ассортативности. Третий граф наоборот имеет высокую плотность, высокий коэффициент кластеризации и близкий к нулю коэффициент ассортативности.

В результате стало известно, что нельзя сделать однозначного вывода, что какая-то из моделей явно копирует реальную сеть. Модели схожи с реальными данными лишь по некоторым характеристикам. Однако, учитывая локальные параметры, модель Боллобаши-Риордана оказалась ближе к реальной сети чем другие. Аналогично настоящим сообществам смоделированная была не ассортативна, обладала коэффициентом кластеризации выше среднего значения, низкой плотностью и мерами центральности ниже среднего.

Таким образом, было разработано полноценное программное обеспечения для генерации и обработки случайных и реальных сетей. Работа с приложением велась из интегрированной среды разработки. Визуализация результатов осуществлялась через png графики. Дополнительно созданы таблицы с минимальными, максимальными и

средними значениями характеристик. На основе этого были сделаны выводы по сравнительному анализу.

ЗАКЛЮЧЕНИЕ

В процессе достижения цели бакалаврской работы были выполнены все поставленные задачи, а именно:

- Изучены основные понятия и определения.
- Разработана программа для генерации моделей случайных графов и вычисления характеристик социального графа.
- Построены социальные сети, основанные на открытых данных из онлайн сервисов.
- Проведена серия экспериментов на реальных сетях и графах, полученных из моделей и обобщены полученные результаты.

Также было проведено сравнение характеристик полученных по моделям графов и реальным сетям, в результате которого выявлено, что модели не полностью повторяют поведение реальных сетей, а лишь отражают некоторые из их особенностей. Однако, модель Боллобаши-Риордана наиболее полно описывает ряд характеристик, таких как распределение степеней вершин, ассортативность и взаимность сети, что делает ее более подходящей для моделирования реальных сетей, чем другие модели. Представленное в бакалаврской работе исследование позволяет проследить взаимосвязи между параметрами различных моделей графов, что в свою очередь может быть применено в анализе реальных сетей.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Вичугова А.Н. Анализ социальных связей с Graph Data Science // Москва – Школа больших данных. – 2022. – С.58.
- 2 С. П. Сидоров, С. В. Миронов Анализ динамики локальных характеристик в сложных сетях // Саратов: Издательство Саратовского университета. – 2022. – С.104.
- 3 Watts, D. J.; Strogatz, S. H. Collective dynamics of 'small-world' networks // Nature. — 1998.
- 4 Гусаров Н.Ф. Анализ социальных сетей. Основные понятия и метрики // Санкт-Петербург – Университет ИТМО – 2016. – С.123.
- 5 Ушаков К.М., Куксо Е.Н. Возможности сетевого анализа для исследований в образовании // Народное образование. – 2015. – С.10
- 6 А.М.Райгородский Математические модели Интернета // Квант – № 4 – 2012 – С.96.
- 7 Erdos-Renyi random graphs [Электронный ресурс] URL: https://www.ndsu.edu/pubweb/~novozhil/Teaching/767%20Data/chapter_3.pdf (Дата обращения: 26.04.2023)
- 8 Albert R., Barabasi A. Statistical mechanics of complex networks // REVIEWS OF MODERN PHYSICS. – 2002. – № 74 – С. 47-97
- 9 Щербакова Н.Г. Меры центральности в сетях // Проблемы информатики. – 2015. – С.28.
- 10 Kemper A. Valuation of Network Effects in Software Markets: A Complex Networks Approach // Springer Science & Business Media. – 2009. – С.309
- 11 Grigoriev A., Sidorov S., Mironov S., Malinskii I., “Average Nearest Neighbor Degree and Its Distribution in Social Networks”, Digital Transformation and Global Society. DTGS 2021, Communications in Computer

and Information Science, 1503, eds. Alexandrov D.A. et al., Springer, Cham, 2022, 36-50