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

Кафедра математической кибернетики и компьютерных наук

ИССЛЕДОВАНИЕ ПАРАДОКСА ДРУЖБЫ В МОДЕЛИ БАРАБАШИ-АЛЬБЕРТ СЛУЧАЙНОГО ГРАФА

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 5 курса 551 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Хасанова Саида Фархадовича

Научный руководитель
Зав.кафедрой, к. ф.-м. н. ____________ С. В. Миронов

Заведующий кафедрой
к. ф.-м. н. ____________ С. В. Миронов

Саратов 2022

ВВЕДЕНИЕ

Актуальность темы исследования: теория случайных графов находится на стыке теории графов и теории вероятностей. Её заметное развитие началось с момента публикации известной статьи Эрдеша и Реньи в 1959 году. Позже выяснилось, что предложенная модель плохо описывает реальные графы, но тем не менее, это дало толчок к исследованию новых математических моделей случайных графов, оставляя данное направление актуальным по сей день.

Существует множество видов моделей, с помощью которых можно построить случайные графы близкие по свойствам к реальным сетям. Обычно их разделяют на несколько основных классов:

- модели случайных графов (модель Барабаши–Альберт);
- простейшие модели безмасштабных сетей (модель Боллобаша–Риордана, модель копирования и др.);
- более гибкие модели безмасштабных сетей (модель Чунг-Лу, модель Янсона–Лучака);
- модель стохастических графов Кронекера.

Под сложными сетями понимают всевозможные сети, которые встречаются в природе: компьютерные, биологические, социальные, экономические, транспортные и так далее. Классический пример такой сети — граф сети Интернет. Вершины графа — это веб-страницы, а ребра — ссылки между ними. С появлением сети Интернет началось интенсивное изучение сложных сетей. Было замечено, что все эти сети обладают некоторыми общими свойствами: малый диаметр, степенной закон распределения степеней вершин, кластерная структура и другие.

Цель выпускной квалификационной работы: проведение исследования работы одной из моделей случайных графов. В ходе работы были поставлены и решены следующие **задачи**:

- разработать программу для построения выбранной модели случайного графа;
- для выбранного количества случайных графов исследовать некоторые локальные характеристики и отобразить полученные данные;
- для выбранного количества случайных графов определить выполняется ли парадокс дружбы.

Объём и структура выпускной квалификационной работы: работа состоит из введения, трёх глав, заключения, списка литературы и трёх приложений. Общий объём выпускной квалификационной работы 43 страницы.

Структура глав работы:

1. Постановка задачи
    1) Основные понятия и задачи
    2) Детали реализации
2. Модели случайных графов
    1) Основные понятия теории графов
    2) Случайные графы
    3) Модель Эрдеша-Реньи
    4) Модель Барабаши-Альберт
    5) Модель Боллобаша-Риордана
    6) Модель копирования
    7) Модель Янсона-Лучака
    8) Графы Кронекера
    9) Парадокс дружбы
3. Описание процесса разработки и анализ результатов
    1) Разработка модели Барабаши-Альберт
    2) Результаты запуска

1 Основное содержание работы

Во введении обоснована актуальность темы исследования, сформулированы цели и задачи работы.

Глава 1 содержит подробный список задач, основные понятия, требующие для выполнения задач и анализа результатов их выполнения, а также описание инструментов, использовавшихся во время разработки программы. В данной главе перечислены следующие характеристики случайных графов:

— d_i - степень вершины i или количество смежных с ней вершин;

— s_i - сумма степеней смежных с i вершин;

— α_i - средняя степень смежных вершин

$$\alpha_i = \frac{s_i}{d_i};$$

— β_i - индекс дружбы вершины i

$$\beta_i = \frac{\alpha_i}{d_i}.$$

Значение β_i является важным свойством, с помощью которого можно выявить выполняется ли парадокс дружбы в построенном графе. Узнать это можно используя формулу

$$\overline{\beta} = \frac{\sum\limits_{i=1}^{n} \beta_i}{n},$$

где n - количество вершин. Если результат вычисления $\overline{\beta} > 1$, значит парадокс дружбы выполняется.

Все вышеперечисленные характеристики случайного графа были использованы для исследования работы модели Барабаши-Альберт. Анализ каждой из них приведён в главе 3. Для этого необходимо было определить более подробный список задач для выполнения:

— разработать программу для построения выбранной модели случайного графа;

— определить следующие параметры: m_0 - начальное количество вершин, m – количество вершин в графе;

— для n графов через каждые h новых вершин вычислить средние значения свойств d_i, s_i, α_i, β_i для $i =$ 5, 50 и 100. Полученные данные отобразить на графиках;

- для n графов через каждые h новых вершин определить выполняется ли парадокс дружбы;
- определить выполняется ли в среднем парадокс дружбы по n графам.

В заключении главы описаны детали реализации: разработка программы построения случайного графа была осуществлена с помощью языка программирования Python версии 3.10; модуля random из стандартной библиотеки, который реализует генераторы псевдослучайных чисел, а именно функция choices, возвращающая список элементов размером k, выбранных из заданной последовательности; сторонней библиотеки matplotlib.pyplot, в основном предназначенной для интерактивных графиков и простых случаев программной генерации графиков.

Глава 2 посвящена литературному обзору по теме выпускной квалификационной работы. В начале данной главы рассмотрены основные понятия теории графов, необходимые для понимания дальнейшего анализа работы, такие как граф, вершина, ребро, ориентированный граф, неориентированный граф, список смежности, степень вершины, связность. В качестве примера был представлен граф из пяти вершин и соответствующий ему список смежности.

Продолжает данную главу теория случайных графов. В этой части главы сказано о том, что понятие случайного графа связано со взглядом на всё множество графов как на вероятное пространство $S =< K, Q >$, где K — множество рассматриваемых графов, Q — вероятностный механизм порождения (или случайного выбора) графа из множества K.

Случайные графы нашли практическое применение во всех областях, где нужно смоделировать сложные сети – известно большое число случайных моделей графов, отражающих разнообразные типы сложных сетей в различных областях. В математическом контексте термин случайный граф означает почти всегда модель случайных графов Эрдеша-Реньи. В других контекстах любая модель графов означает случайный граф.

Следующая часть главы описывает различные существующие модели случайных графов:

- модель Эрдеша-Реньи;
- модель Барабаши-Альберт;
- модель Боллобаша-Риордана;

— модель копирования;
— модель Янсона-Лучака;
— графы Кронекера.

В модели Эрдеша-Реньи случайный граф — это случайный элемент со значениями во множестве

$$\Omega_n = \{G = (V_n, E) : V_n = \{1, 2, \dots, n\}\},$$

которое состоит из всех n-вершинных графов без петель, кратных ребер и ориентации, так что $|\Omega_n| = 2^{C_n^2}$. Распределение здесь биномиальное, т.е.

$$P(G) = p^{|E|}(1-p)^{C_n^2 - |E|}, p = p(n) \in [0, 1].$$

Другими словами, можно считать, что каждое ребро случайного графа проводится независимо от всех остальных ребер с одной и той же вероятностью p, величина которой может меняться с течением времени n.

Модель Барабаши-Альберт используется в практической части работы. Мотивацией этой модели послужили растущие сети, которые развиваются со временем и которые нельзя описать моделью Эрдеша–Реньи. Её идея заключается в том, чтобы модернизировать модель случайных сетей в направлении описания динамики развития:
— в начальный момент времени $(t = 0)$ есть m несвязанных узлов;
— на каждом шаге $(t = 1, 2, 3 \dots)$ будем добавлять новый узел с m ребрами;
— хотя количество связей, с которым приходит новый узел в граф, фиксировано (m), к какому именно узлу он присоединяется – выбирается случайным образом, т.е. рост идет случайно.

Формально, вероятность p_i того, что новый узел соединится с узлом i равна

$$p_i = \frac{k_i}{\sum_j k_j}, \tag{1}$$

где k_i - степень вершины i, а в знаменателе суммируются степени всех существующих вершин. Из этого следует, что наиболее связанные вершины, как правило, накапливают ещё больше связей, тогда как вершины с небольшим числом связей вряд ли будут выбраны для присоединения новых узлов.

Заключение второй главы подробнее описывает явление парадокса дружбы: в любой сети средняя степень (т.е. количество соседей) соседа узла строго больше средней степени узлов в сети в целом. Применительно к сетям дружбы это означает, что в среднем у ваших друзей больше друзей, чем у вас. В данной части главы упоминается о проводимой демонстрации парадокса на примере одной из реальных сетей и приводится доказательство того, что среднее значение Δ_i по всем вершин больше нуля, где Δ_i – разница между средним значением степени соседей и собственной степени любой заданной вершины. Далее в текст главы упоминается о ранее определённой характеристике под названием «индекс дружбы» (β_i), которая является отношением средней степени смежных с i вершин к степени вершины i. В рамках выпускной квалификационной работы эта характеристика является важной, так как с её помощью можно определить, является ли выбранная вершина, в соответствии с парадоксом дружбы, превосходящей по популярности своих соседей, или она является одной из вершин, для которых парадокс не выполняется, поскольку она более популярна, чем ее соседи. Также она позволяет определить «расхождение» по сравнению со случаем, в котором не существует парадокса дружбы: если значения β_i превышают единицу, значит вершина проявляет парадокс, а если значения β_i ближе к единице, значит вершина проявляет парадокс в меньшей степени.

Глава 3 содержит описание процесса разработки и анализ результатов. В начале данной главы подробно описан процесс разработки программы для вычисления характеристик случайных графов, включая исходный код программы. Он состоит из трёх частей:

1. **Класс Graph** - разработка класса, реализующего простой граф. Класс состоит из конструктора, который создаёт сильно связный граф с заданным количеством вершин и метода, добавляющего новую вершину в граф по заданному списку смежных вершин. Также данный класс хранит и обновляет информацию о количестве вершин и рёбер созданного графа в соответствующих полях, которые были необходимы для вычислений. Из приведённого кода конструктора класса можно заметить, что у вершин нет имён. Вместо них здесь используются значения соответствующих индексов в списке и для новой вершины задаётся значение следующего индекса в качестве названия
2. **Класс BA_model** - разработка класса-наследника, реализующего модель

Барабаши-Альберт, которая была за основу случайного графа. Он состоит из конструктора, переопределённого метода для добавления вершины, который реализует принцип предпочтительного присоединения, а также методов, вычисляющих характеристики d_i, s_i, α_i, β_i, $\overline{\beta}$, которые были описаны в главе 2. При создании объекта есть возможность задать кол-во связей с новой вершиной. Это этого параметра зависит размер созданного графа: по умолчанию он будет равен параметру m но, если $m < 2$, то начальный граф будет состоять из 2 вершин.

3. **Файл main.py** - разработка основной программы для генерации случайных графов и всех необходимых вычислений, которые требовались для выполнения работы. В постановке задачи были определены некоторые параметры программы. Некоторые из них изменялись в процессе выполнения работы, поэтому они были вынесены в отдельный логический блок с заданными первоначальными значениями. В отдельную функцию были вынесены основные вычисления, требующиеся для выполнения работы. В ней генерируются новые вершины для каждого графа и высчитываются свойства в определённые промежутки времени. Выполнение данной программы печатает значение $\overline{\beta}$ в консоль и выводит окно с 4 графиками, отражающими зависимость каждой из ранее перечисленных характеристик от времени t для заданных вершин.

Далее в главе описан анализ результатов запуска программы в зависимости от различных входных параметров, которые изменялись в процессе выполнения работы.

Изменяемые параметры:

— m - количество связей с новой вершиной;
— n_g - количество случайных графов;
— n_v - конечное количество вершин в случайном графе.

Некоторые параметры были определены на начальном этапе разработки и больше не изменялись в процессе выполнения работы.

Неизменяемые параметры:

— $h = 250$ - шаг;
— $v_i = [5, 50, 100]$ - тестируемые вершины.

Описание анализа каждого запуска сопровождалось графиками, которые формировались в результате исполнения программы.

Первый запуск был осуществлён со следующими параметрами:

- $m = 5$;
- $n_g = 10$;
- $n_v = 1000$.

Результаты показали, что графики значения степени вершины, суммы степеней соседей и средней степени соседей со ростом графа только увеличиваются. Так, например, вершина 5 на графике суммы степеней соседей росла со значения примерно равного 700 при $t = 250$ до значения примерно равного 1700 при $t = 1000$.

График значений индекса дружбы показал что, парадокс дружбы выполняется для двух из выбранных вершин до $t = 600$, а после при добавлении новых вершин значение индекса дружбы вершины 50 становится меньше единицы. Средне значение $\overline{\beta}$ в данном тесте было равным 2.8, из чего следует выполнимость парадокса дружбы в среднем по 10 графам.

Второй запуск был осуществлён со следующими параметрами:

- $m = 5$;
- $n_g = 100$;
- $n_v = 10000$.

Ранее было замечено, что при увеличении t степень вершины только растёт. В данном тесте получается то же самое. На этом тесте вершина 5 на графике суммы степеней соседей росла со значения примерно равного 700 при $t = 250$ до значения чуть меньше чем 7000 при $t = 10000$. Помимо этого было замечено, что вершины, которые были добавлены на ранних этапах, имеют преимущество в связях с новыми вершинами. Например, вершина 5, которая была добавлена при создании графа, имела гораздо большие значения, чем вершины 50 и 100, которые были добавлены позже. Это и есть так называемый принцип предпочтительного присоединения.

Результаты второго запуска также показали, что графики степеней вершин и суммы степеней соседей строятся одинаково при различных параметрах. Это означает, что программа корректно вычисляет значения характеристик d_i и s_i. Из этого следует, что значения характеристик α_i и β_i также корректно вычисляются, так как зависят от первых двух.

График значений индекса дружбы показал как при увеличении количества вершин индекс дружбы становится меньше. Но среднее значение $\overline{\beta}$ было равным 3.7, из чего можно сделать вывод, что в среднем для 100 случайных графов парадокс дружбы выполняется.

В первых двух тестах изменялись только количество вершин и количество графов. При следующих запусках был изменён параметр m.

Третий запуск был осуществлён со следующими параметрами:

— $m = 3$;
— $n_g = 100$;
— $n_v = 10000$.

Результаты запуска показали, что при изменении параметра m графики степени вершины и суммы степеней соседей, как и ожидалось выглядят примерно так же, как при $m = 5$. Отличие в том, что значения этих свойств уменьшились. Например, вершина 5 на графике суммы степеней соседей достигала значения примерно равного 2500, в то время как та же вершина при $m = 5$ достигала значения свыше 6000.

Значения графика индекса дружбы в данном тесте для двух вершин оставались большими единицы на протяжении всего времени, а среднее значение $\overline{\beta}$ было равным 4.6.

Четвёртый запуск был осуществлён со следующими параметрами:

— $m = 1$;
— $n_g = 100$;
— $n_v = 10000$.

При сравнении результатов предыдущих тестов, а конкретно значений характеристик d_i и s_i, было замечено, что они значительно уменьшились, что и следовало ожидать.

График значений индекса для всех вершин оказался больше единицы на протяжении всего времени, а среднее значение $\overline{\beta}$ увеличилось до 9.9.

Последний запуск был осуществлён со следующими параметрами:

— $m = 10$;
— $n_g = 100$;
— $n_v = 10000$.

В данном тесте параметр m был увеличен в два раза по сравнению с изначальным значением. Как и ожидалось, графики в данном тесте строились практически так же, как при $m = 5$. Различие лишь в том, что при $m = 10$ значения свойств имеют большие значения. Также было замечено, что значения средней степени соседей у выбранных вершин на каждом шаге были очень близки друг к другу. Среднее значение $\overline{\beta}$ в данном тесте было равным 3.

Проанализировав результаты выполнения программы были сделаны выводы: как и ожидалось, при большем значении m значения характеристик d_i, s_i и α_i достигают больших значений. В то время как значения β_i для выбранных вершин становятся меньше единицы при большом значении параметра m, но всё же для всех проведенных тестов среднее значение $\overline{\beta}$ было больше единицы.

ЗАКЛЮЧЕНИЕ

В заключении выпускной квалификационной работы подводятся итоги проведённого исследования:

- разработана программа генерации случайных графов на языке программирования Python;
- проведены вычисления средних значений свойств каждого графа в выбранных итерациях для трёх вершин;
- вычислено среднее значение свойства $\overline{\beta}$;
- отображены графики зависимости 4 свойств графа от времени t для каждой из выбранных вершин.

Результаты выполнения задач показали как изменяется генерация выбранной модели случайных графов при различных параметрах, а также продемонстрировали выполнимость парадокса дружбы для всех построенных графов.