

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 ИНДЕКСА ДРУЖБЫ УЗЛОВ РАСТУЩИХ СЕТЕЙ
ПОСТРОЕННЫХ ПО МОДЕЛЯМ С ПРЕДПОЧТИТЕЛЬНЫМ
ПРИСОЕДИНЕНИЕМ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 411 группы

направления 02.03.02 — Фундаментальная информатика и информационные
технологии

факультета КНиИТ

Козырева Юрия Дмитриевича

Научный руководитель

зав. каф., к. ф.-м. н., доцент

С. В. Миронов

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Теоретические сведения	4
1.1 Модель Барабаша—Альберт	4
1.2 Модель триадного замыкания	4
1.3 Индекс дружбы	5
2 Реализация моделей	6
3 Анализ	7
3.1 Анализ распределения и динамики индекса дружбы в построенных графах	7
3.2 Анализ распределения и динамики индекса дружбы в сети цитирования статей в сфере феноменологии физики высоких энергий ..	8
3.3 Анализ распределения и динамики индекса дружбы в сети форума Reddit	9
3.4 Анализ распределения и динамики индекса дружбы в системе вопросов и ответов AskUbuntu	10
3.5 Анализ распределения и динамики индекса дружбы в системе вопросов и ответов SuperUser	12
ЗАКЛЮЧЕНИЕ	16

ВВЕДЕНИЕ

В повседневной жизни для решения многих задач часто используются случайные графы. Случайные графы нашли практическое применение во всех областях, где нужно смоделировать сложные сети: при моделировании и анализе биологических и социальных систем, сетей, а также при решении многих задач класса NP.

Случайный граф — общий термин для обозначения вероятностного распределения графов. Их можно описать просто распределением вероятности или случайным процессом, создающим эти графы.

В дипломной работе рассматриваются одни из активно используемых и хорошо изученных моделей: модель Барабаши-Альберта и модель триадного замыкания.

Для изучения и анализа моделей случайных графов, анализе социальных явлений и сетей, сообществ и их взаимодействий широко применяется индекс дружбы. Индекс дружбы — один из показателей, используемых в социологии, определяется как отношение средней степени соседей к степени самого объекта.

Целью настоящей работы является анализ индекса дружбы сетей, построенных по модели Барабаши—Альберт. Для достижения этой цели необходимо решить следующие задачи:

- рассмотреть алгоритмы Барабаши—Альберт и триадного замыкания для построения случайного графа;
- реализовать классический алгоритм Барабаши—Альберт, и его модификацию, в которой начальная степень каждого нового узла определяется как случайная величина, заданная пуассоновским распределением;
- реализовать алгоритм триадного замыкания;
- рассмотреть возможность параллельной реализации процесса получения случайного графа и вычисления индекса дружбы;
- организовать серии экспериментов, в которых строятся случайные графы по реализованным моделям;
- провести анализ распределения индекса дружбы построенных графов.

1 Теоретические сведения

1.1 Модель Барабаши—Альберт

Модель Барабаши—Альберт является одной из первых моделей веб-графов. Веб-граф представляет собой ориентированный мульти-граф, вершинами в котором являются какие-либо конкретные структурные единицы в Интернете: речь может идти о страницах, сайтах, хостах, владельцах и пр. Для определенности будем считать, что вершинами веб-графа служат именно сайты. А рёбрами соединяются вершины, между которыми имеются ссылки.

В своей модели А.-Л. Барабаши и Р. Альберт предложили стратегию предпочтительного присоединения. Её основная идея заключается в том, что вероятность присоединения конкретной вершины ребром к новой вершине пропорциональна степени данной вершины. Здесь и далее степенью вершины $v_i \in V$ графа $G = (V, E)$ называется количество вершин, напрямую связанных с данной, т.е. $deg(v_i) = |\{v \in V : (v, v_i) \in E\}|$.

Алгоритм формирования сети по модели Барабаши—Альберт заключается в следующем.

1. Первоначально берется полный граф из m вершин, где m — параметр модели.
2. На каждой итерации роста сети добавляется одна новая вершина, которая соединяется m ребрами с уже имеющимися в соответствии с принципом предпочтительного присоединения.

1.2 Модель триадного замыкания

Помимо стратегии предпочтительного присоединения, модель триадного замыкания использует стратегию формирования триад. Триадное замыкание — свойство социальных систем заключающееся в том, что если между вершинами (А, В) и (А, С), в некоторой социальной сети существует взаимосвязь, то велика вероятность формирования триады из этих трёх вершин, т.е., велика вероятность связи (В, С). В модели триадного замыкания рост сети происходит следующим образом.

1. Первоначально берется полный граф из m вершин, где m — параметр модели.
2. На каждой итерации роста сети добавляется одна новая вершина, которая соединяется m ребрами с уже имеющимися по следующим правилам:

- в соответствии с принципом предпочтительного присоединения выбирается вершина, к которой проводится первое ребро;
- с вероятностью p , где p — параметр модели, выбирается стратегия формирования триады с произвольным соседом вершины, присоединенной первым ребром, или, с вероятностью $(1 - p)$, стратегия предпочтительного присоединения к произвольной вершине графа.

1.3 Индекс дружбы

В социальных сетях часто можно встретить явление именуемое парадоксом дружбы: в среднем друзья любого человека имеют больше друзей, чем он сам. Оно было обнаружено в 1991 году социологом из государственного университета Нью-Йорка Скоттом Фельдом.

Для изучения парадокса дружбы следует ввести несколько обозначений. В момент времени t для вершины v_i в графе $G(t) = (V(t), E(t))$ сумма степеней всех соседей v_i равна: $s_i(t) = \sum_{j:(v_i, v_j) \in E(t)} deg_j(t)$, средняя степень соседей вершины v_i : $\alpha_i(t) = \frac{s_i(t)}{deg_i(t)}$, а индекс дружбы $\beta_i(t)$ определяется как отношение средней степени соседей v_i к степени самой v_i : $\beta_i(t) = \frac{\alpha_i(t)}{deg_i(t)}$.

Таким образом, если средняя степень соседей больше степени v_i и парадокс дружбы выполняется, то $\beta_i(t) > 1$.

2 Реализация моделей

В ходе выполнения дипломной работы были реализованы: стандартная модель Барабаша—Альберт, модель Барабаша—Альберт с пуассоновским распределением начальных степеней и модель триадного замыкания. Все расчёты производились на компьютере с процессором Intel core i5-8265U и 16 ГБ оперативной памяти. Модель реализована на Python 3.9.1. При реализации использовались библиотеки `json`, `multiprocessing`, `random` и `numpy.random`.

Все три реализованные модели представляют собой модели растущего случайного графа с использованием механизма предпочтительного присоединения. В реализованных модификациях на каждом шаге добавляется одна вершина и некоторое количество m рёбер, при этом параметр m и процесс выбора соседей новой вершины зависят от используемой модели. В модели триадного замыкания используется дополнительный параметр p .

В экспериментах перебираются различные модели и различные значения параметров. В каждом эксперименте строится граф из 100000 вершин. Так как получаемые графы случайны, то в ходе эксперимента граф строится десять раз, после чего строится усреднённая гистограмма по диапазонам значений индекса дружбы и график динамики изменений среднего индекса дружбы графа.

В реализации моделей граф описывается парой массивов `degrees` и `neighbours`, хранящие соответственно степень и список номеров соседних вершин для каждой вершины графа. Каждая модель реализована в форме функции `model_name(n, args, funcs, dts)`, где параметр n задаёт размер конечного графа, массив `args` содержит аргументы модели, `funcs` является массивом метрик, которые нужно применить к графу, а `dts` — периодичность применения метрик.

Метрики также представлены функциями вида `metric_name(degrees, neighbours)`, в которых `degrees` представляет собой массив степеней вершин графа, а `neighbours` — массив массивов соседних вершин. В ходе работы были использованы четыре метрики:

`s(degrees, neighbours)` — возвращает массив сумм степеней соседей для вершин графа; `alfa(degrees, neighbours)` — возвращает массив средних степеней соседей для вершин графа; `beta(degrees, neighbours)` — возвращает массив индексов дружбы для вершин графа; `mean_beta(degrees, neighbours)` — возвращает средний индекс дружбы всего графа

3 Анализ

Для того чтобы проверить, что реализованные модели могут выступать в качестве моделей социальных процессов в окружающем мире, нужно показать, что сгенерированные графы обладают некоторыми свойствами реальных сетей. В качестве характеристики для анализа был выбран индекс дружбы $\beta_i(t) = \frac{\alpha_i(t)}{\overline{deg_i(t)}}$. В работе рассматривались распределение индекса дружбы и динамика его среднего значения.

3.1 Анализ распределения и динамики индекса дружбы в построенных графах

На графике, представленном на Рис. 1, изображено распределение индекса дружбы в графе построенном в соответствии с стандартной моделью Барабаши—Альберт с параметром $m = 3$, по этому графику можно сделать вывод о том, что для большинства вершин графа значение индекса дружбы больше единицы. Следовательно в данной модели наблюдается парадокс дружбы.

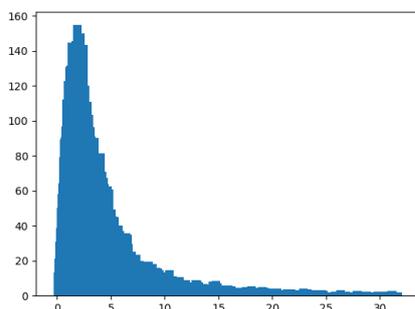


Рисунок 1 – Распределение индекса дружбы в стандартной модели Барабаши—Альберт при $m = 3$

На Рис. 2 выведены те же значения, но в логарифмической шкале. В соответствии с графиком, можно выдвинуть гипотезу, что распределение значений индекса дружбы происходит в соответствии со степенной функцией $y = (\alpha \cdot x)^\gamma$.

Рисунок 3 содержит график в логарифмической шкале динамики среднего индекса дружбы в модели Барабаши—Альберт при $m = 3$ в сравнении с графиком степенной функции $y = (10000 \cdot x)^{0.12}$. На основании данного графика можно сделать предположение, что среднее значение индекса дружбы в рассматриваемых моделях растёт в соответствии со степенной функцией.

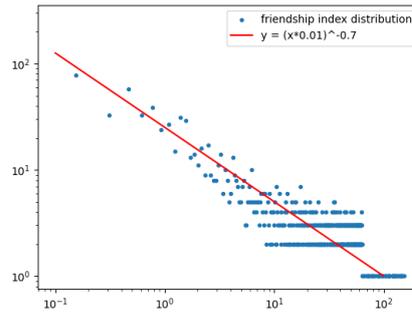


Рисунок 2 – Распределение индекса дружбы в стандартной модели Барабаши— Альберт при $m = 3$, в логарифмической шкале

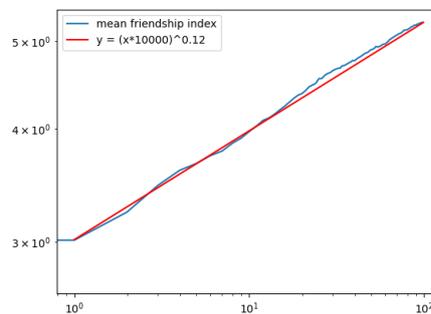


Рисунок 3 – Динамика среднего значения индекса дружбы в стандартной модели Барабаши— Альберт при $m = 3$, в логарифмической шкале

В ходе работы были рассмотрены графы следующих сетей из реального мира: LiveJournal, Twitter, Reddit, Google+, askubuntu, mathoverflow, superuser, сеть цитирования статей в сфере феноменология физики высоких энергий, внутренняя социальная сеть Калифорнийского университета в Ирвайне.

3.2 Анализ распределения и динамики индекса дружбы в сети цитирования статей в сфере феноменологии физики высоких энергий

Набор данных содержит информацию о 34 546 статьях на тему физики высоких энергий, опубликованных с января 1993 по апрель 2003 на arXiv.org. Граф содержит 421 578 рёбер. Датасет был представлен в работе .

На Рис. 4 изображено распределение индекса дружбы в данной сети. Из этого графика можно сделать предположение о том, что индекс дружбы в сети цитирования, как и индекс дружбы рассмотренных случайных графах, распределён по степенному закону $y = (0.0055 \cdot x)^{-2}$.

График динамики индекса дружбы в данной сети представлен на Рис. 5.

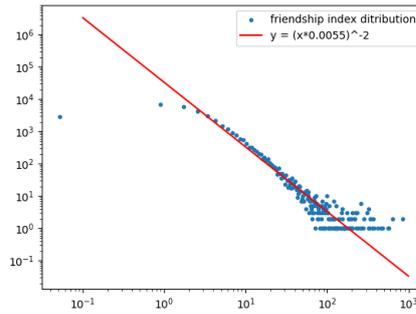


Рисунок 4 – Распределение индекса дружбы в сети цитирования в логарифмической шкале

Можно выдвинуть гипотезу, что среднее значение индекса дружбы в графе растёт в соответствии со степенной функцией $y = (11000 \cdot x)^{0.12}$. Это, в части данного предположения, соответствует росту индекса дружбы в сетях моделей Барабаши— Альберта и триадного замыкания.

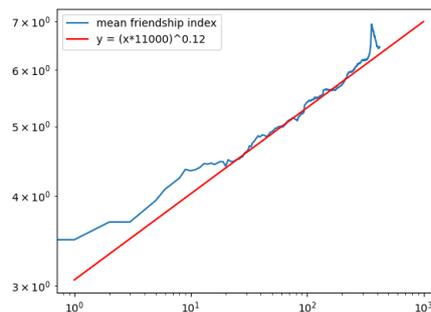


Рисунок 5 – Динамика индекса дружбы в сети цитирования в логарифмической шкале

3.3 Анализ распределения и динамики индекса дружбы в сети форума Reddit

В данном случае рассматривается сеть гиперссылок между субредитами (сообществами на «Reddit») данные собирались на протяжении двух с половиной лет: с января 2014 по апрель 2017. Граф состоит из 55 863 узлов, соединённых 858 490 рёбрами. Данные были представлены в .

На Рис. 6 представлен график распределения индекса дружбы в графе гиперссылок между субредитами. По графику можно сделать предположение, что распределение индекса дружбы в данной сети также соответствует степен-

ной функции, которая в данном случае выглядит следующим образом: $y = (2e^{-6} \cdot x)^{-1}$.

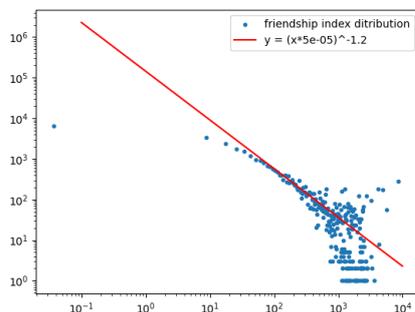


Рисунок 6 – Распределение индекса дружбы в сети форума Reddit в логарифмической шкале

Рис. 7 указывает на сходство графика динамики индекса дружбы на форуме Reddit, с графиком степенной функции $y = (0.03 \cdot x)^{0.8}$.

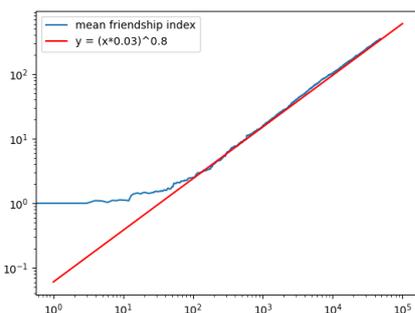


Рисунок 7 – Динамика индекса дружбы в сети форума Reddit в логарифмической шкале

3.4 Анализ распределения и динамики индекса дружбы в системе вопросов и ответов AskUbuntu

Были рассмотрены 4 датасета:

- граф ответов на вопросы, который состоит из 137, 517 вершин, представляющих пользователей, и 280 102 связей между ними, связь между пользователями A и B заключается в том, что пользователь A ответил на вопрос пользователя B ;
- граф цитирования вопросов, содержащий 79 155 узлов и 327 513 рёбер, где направленное ребро проводится между пользователем и автором вопроса, который он процитировал;

- граф цитирования ответов из 75 555 и 356 822 вершин и рёбер, соответственно, ребро проводится от пользователя сославшегося на ответ к пользователю его написавшему;
- и общий граф содержащий всех пользователей и все рёбра всех предыдущих графов, в нём 159 316 узлов и 964, 437 связей.

Данные для этих графов собирались на протяжении более чем семи лет.

Свойства данной сети не полностью соответствуют поведению, изучаемых в дипломной работе моделей: на Рис. 8 можно увидеть, что распределение индекса дружбы плохо соответствует степенной функции и, следовательно, не соответствует построенным моделям.

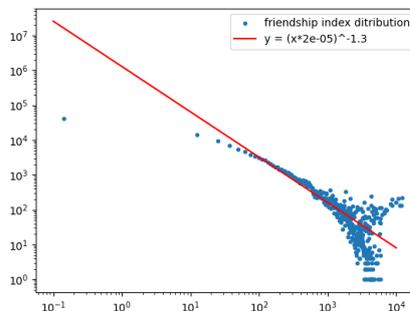


Рисунок 8 – Распределение индекса дружбы в сети AskUbuntu в логарифмической шкале

Однако, среднее значение индекса дружбы, как и в моделях случайных графов, растёт скорее по степенному закону $y = (0.8 \cdot x)^{0.5}$, что показано на графике 9.

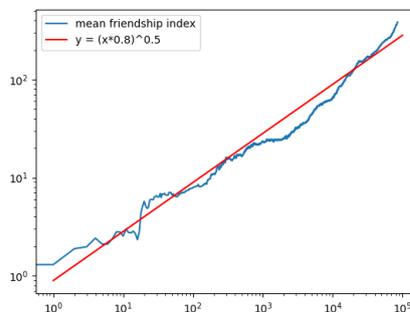


Рисунок 9 – Динамика индекса дружбы в сети askubuntu в логарифмической шкале

Наиболее ярко данные закономерности проявляются в графе сети ответов

на вопросы, графики распределения и динамики индекса дружбы в которой изображены на Рис. 10 и 11, соответственно.

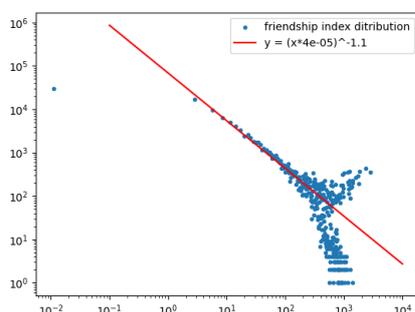


Рисунок 10 – Распределение индекса дружбы в сети ответов на вопросы на AskUbuntu в логарифмической шкале

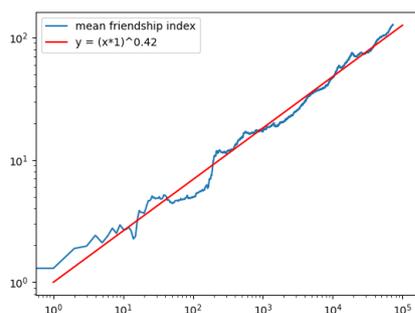


Рисунок 11 – Динамика индекса дружбы в сети ответов на вопросы на AskUbuntu в логарифмической шкале

3.5 Анализ распределения и динамики индекса дружбы в системе вопросов и ответов SuperUser

Для этой сети были выделены также три подсети: сеть ответов на вопросы, сеть цитирования вопросов, сеть цитирования ответов.

Сеть платформы SuperUser не подчиняется закономерностям выявленным в случайных графах в ходе данной работы. На Рис. 12 можно увидеть, что график распределения индекса дружбы скорее плохо совпадает со степенной функцией, также как и график динамики индекса дружбы представленный на Рис. 13.

В сети цитирования вопросов это отклонение только усиливается, что можно наблюдать на графиках 14 и 15.

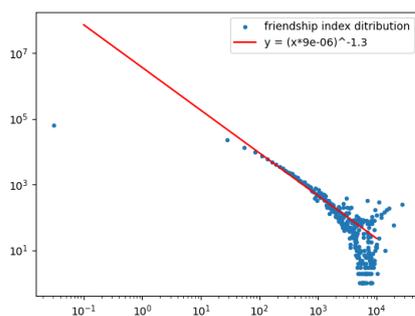


Рисунок 12 – Распределение индекса дружбы в сети SuperUser в логарифмической шкале

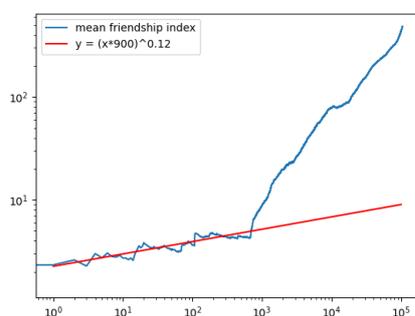


Рисунок 13 – Динамика индекса дружбы в сети SuperUser в логарифмической шкале

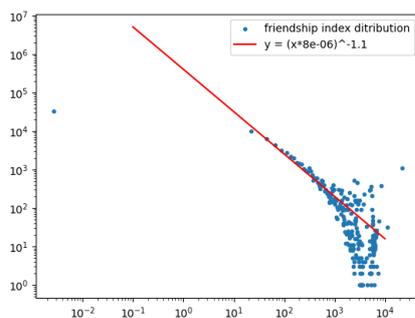


Рисунок 14 – Распределение индекса дружбы в сети цитирования вопросов на SuperUser в логарифмической шкале

В то же время, в подсети ответов на вопросы динамика индекса дружбы всё же, вероятно, подчиняется степенному закону, что можно наблюдать на Рис. 16. В таблице 1 представлен сравнительный анализ всех рассмотренных сетей.

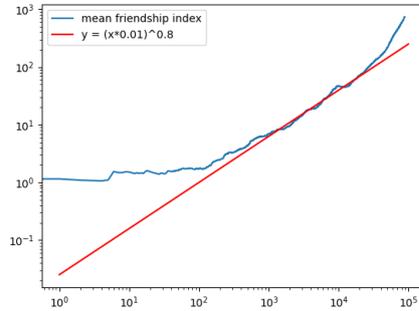


Рисунок 15 – Динамика индекса дружбы в сети цитирования вопросов на SuperUser в логарифмической шкале

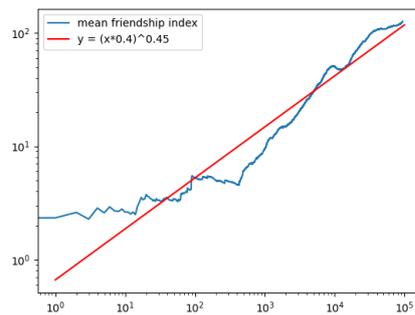


Рисунок 16 – Динамика индекса дружбы в сети ответов на вопросы на SuperUser в логарифмической шкале

Таблица 1 – Сводная таблица результатов экспериментов

Сеть	Функция распределения индекса дружбы	γ_{dist}	Функция динамики индекса дружбы	γ_{dyn}
Модель Барабаши— Альберт при $m = 3$	степенная	-0.7	степенная	0.15
Модель Барабаши— Альберт при $m = 3$	степенная	-0.7	степенная	0.15
Модель Барабаши— Альберт с пуассоновским распределением начальных степеней при $m = 4$	степенная	-0.5	степенная	0.15
Модель Барабаши— Альберт с пуассоновским распределением начальных степеней при $m = 5$	степенная	-0.5	степенная	0.15
Модель Барабаши— Альберт с пуассоновским распределением начальных степеней при $m = 6$	степенная	-0.45	степенная	0.15
Модель триадного замыкания при $m = 3$ и $p = 0.25$	степенная	-0.7	степенная	0.15
Модель триадного замыкания при $m = 3$ и $p = 0.5$	степенная	-0.7	степенная	0.15
Модель триадного замыкания при $m = 3$ и $p = 0.75$	степенная	-0.4	степенная	0.12
Модель триадного замыкания при $m = 5$ и $p = 0.25$	степенная	-0.6	степенная	0.13
Модель триадного замыкания при $m = 5$ и $p = 0.5$	степенная	-0.6	степенная	0.13
Модель триадного замыкания при $m = 5$ и $p = 0.75$	степенная	-0.55	степенная	0.12
Twitter	неизвестная функция	N/A	N/A	N/A
Google+	степенная	-1.7	N/A	N/A
Reddit	степенная	-1.2	степенная	0.8
AskUbuntu	неизвестная функция	N/A	степенная	0.5
MathOverflow	степенная	-1.2	степенная	0.45
SuperUser	неизвестная функция	N/A	экспоненциальная	N/A
Сеть цитирования LiveJournal	степенная	-2	степенная	0.12
Сеть студенческих сообщений	степенная	-2.5	N/A	N/A
	степенная	-0.9	степенная	0.5

ЗАКЛЮЧЕНИЕ

В ходе выполнения дипломной работы были изучены различные модели генерации растущих сетей. Были реализованы модели построения случайных графов:

- стандартная модель Барабаши—Альберт;
- модель Барабаши—Альберт с пуассоновским распределением начальных степеней узлов;
- модель триадного замыкания.

В соответствии с реализованными моделями проведена серия экспериментов в которых строились случайные графы и исследовались динамика среднего значения индекса дружбы в сети и распределение значений индекса дружбы в итоговом графе. В соответствии полученными результатами выдвинута гипотеза, что распределение индекса дружбы вершин итогового графа и рост среднего значения индекса дружбы в ходе формирования сети происходит по степенному закону.

Были исследованы значения распределения индекса дружбы и динамики роста среднего значения индекса дружбы в ряде реальных сетей. Сделан вывод, что для большинства сетей поведение индекса дружбы схоже с его поведением в сетях растущих по исследуемым моделям.

Реализация моделей и эксперименты проводились на языке python с использованием библиотек json, multiprocessing, pyplot, numpy, random.