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

БАКАЛАВРСКАЯ РАБОТА

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

В повседневной жизни для решения многих задач часто используются случайные графы. Случайные графы нашли практическое применение во всех областях, где нужно смоделировать сложные сети. Имеется большое число моделей случайных графов, отражающих разнообразные типы сложных сетей в различных областях. Случайные графы применяются при моделировании и анализе биологических и социальных систем, сетей, а также при решении многих задач класса NP.

Случайные графы впервые определены венгерскими математиками П. Эрдёшем и А. Реньи в книге 1959 года «On Random Graphs» [1] и независимо американским математиком, Э. Гильбертом, в его статье «Random graphs» [2].

Случайный граф — общий термин для обозначения вероятностного распределения графов [3]. Их можно описать просто распределением вероятности или случайным процессом, создающим эти графы.

Теория случайных графов находится на стыке комбинаторики, теории графов и теории вероятностей. В основе ее лежит глубокая идея о том, что мощные инструменты современной теории вероятностей должны поспособствовать более верному осознанию природы графа, призваны помочь решению многих комбинаторных и теоретико-графовых задач [4].

С математической точки зрения случайные графы необходимы для ответа на вопрос о свойствах типичных графов. Для этого используются модели Эрдёша—Реньи, Барабаши—Альберт, модель триадного замыкания, модель Бьянкони-Барабаши и другие. Все модели основываются на различных свойствах социальных сетей.

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

Существует множество различных моделей построения случайных графов. Рассмотрим некоторые из них.

1.1 Модель Эрдеша—Ренье

Модель Эрдеша—Ренье является одной из первых моделей случайного графа. Граф построенный по этой модели представляет собой совокупность множества вершин $V = \{1, ..., n\}$ и множества рёбер E, состоящего из рёбер полного графа K_n построенного на множестве V, выбранных по схеме Бернулли. Таким образом образуется случайный граф $G = (V, E)$. Формально выражаясь, мы имеем вероятностное пространство

$$G(n, p) = (\Omega_n, F_n, P_{n,p}),$$

в котором: n — количество вершин, p — вероятность появления нового ребра, F_n — сигма-множество,Ω_n — множество возможных рёбер ($|\Omega_n \ | \ = \ 2^n$), $P_{n,p}\ (G) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$ — вероятностная мера. Таким образом, в модели Эрдеша—Реньи каждое ребро независимо от других ребер входит в случайный граф с вероятностью p. Модель Эрдеша—Реньи на данный момент является самой изученной моделью случайных графов [5].

1.2 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 [6]. 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.

1.3 Модель Боллобаша—Риордана

Одной из наиболее удачных и часто используемых моделей предпочтительного присоединения является модель Боллобаша—Риордана. Существуют две основных и, по сути, совпадающих модификации этой модели. В одной дается динамическое, а в другой статическое описание случайности [7].

В динамической модификации при добавлении n-ной вершины проводятся n новых рёбер, при этом рёбра могут быть кратными, а также петлями. При создании графа с единственной вершиной проводится петля в этой точке [5]. Таким образом вероятность появления ребра $(n,\ i),\ \ i \in [0,\ n-1]$ равна $\frac{deg_i}{2n-1}$, где deg_i — степень вершины i.

Статическая модель (LCD-модель) основывается на объекте называемом линейной хордовой диаграммой (LCD). Для построения данного объекта требуется зафиксировать на оси абсцисс $2n$ точек $1, \ldots, 2n$, разбить их на пары и соединить элементы каждой пары дугой, лежащей в верхней полуплоскости. Количество различных диаграмм равно

$$l_n = \frac{(2n)!}{2^n n!}.$$

По каждой диаграмме строится граф с n вершинами и n ребрами по следующему алгоритму:

— Идти слева направо по оси абсцисс пока не встретится правый конец какой-либо дуги, пусть позиция этой точки равна i_k

— Последовательность $i_{k-1}+1, i_k$ объявляется списком смежности для k-той вершины, $i_0 = 0$

— Если $k < n$, k увеличивается на 1, переход на шаг (1).

При построении модели LCD случайно выбирается одна из возможных LCD и вероятность каждой диаграммы равной $\frac{1}{l_n}$, где l_n – общее число диаграмм. Графы построенные по такой модели имеют те же свойства, что и графы построенные по динамической модификации схемы Боллобаша—Риордана.

1.4 Модель Чунг-Лу

Пусть нам задано некоторое конечное множество вершин $V = v_1, \ldots, v_n$ и степень каждой вершины d_i, $i = \overline{1, n}$. Генерация графа $G = (V, E)$ происходит следующим образом:

— формируем множество L, состоящее из $i \cdot d$ копий $i \cdot v$ для каждого i от 1 до n;

— задаем случайные паросочетания на множестве L;

— для вершин u и v из Vs количество ребер в графе G , соединяющее их, равно числу паросочетаний между копиями u и v в L [8].

Сгенерированный таким образом граф соответствует степенной модели $P(a, b)$, описывающей графы, для которых:

$$|\{v|\deg_v = \ x\}| = \frac{e^\alpha}{e^\beta}.$$

1.5 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 (A, B) и (A, C), в некоторой социальной сети существует взаимосвязь, то велика вероятность формирования триады из этих трёх вершин, т.е., велика вероятность связи (B, C) [9]. 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.

1.6 Индекс дружбы

С момента появления социальных сетей — Facebook, Vkontakte, LiveJournal, Instagram, LinkedIn, MySpace и т. д. прошло не так много времени, но они уже плотно вошли в повседневную жизнь многих людей.

Опросы показывают, что 76% пользователей Интернета в России (по данным агентства PRT на январь 2014) и примерно 73% жителей Соединенных Штатов являются активными пользователями социальных сетей, и эта цифра растет [10].

В современном обществе социальные сети становятся огромной базой информации, которую ученые и работодатели все чаще привлекают для решения конкретных задач, будь то научное исследование или оценка кандидата на определенную должность.

Научный интерес к изучению пользователей социальных сетей стремительно растет. На данный момент накоплено большое количество эмпирического материала в отношении характеристик пользователей социальных сетей, который требует систематизации и осмысления.

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 [11].

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 соседей0 вершины v_i:

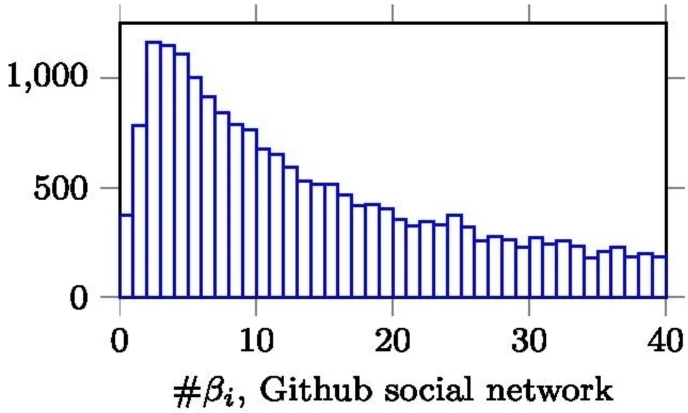
$$\alpha_i(t) = \frac{s_i(t)}{deg_i(t)},$$

а индекс дружбы $\beta_i(t)$ определяется как отношение средней степени соседей v_i к степени самой v_i:

$$\beta_i(t) = \frac{\alpha_i(t)}{deg_i(t)} = \frac{s_i(t)}{deg_i^2(t)} = \frac{\sum_{j:(v_i,v_j)\in E(t)} deg_j(t)}{deg_i^2(t)}.$$

Таким образом, если средняя степень соседей больше степени v_i и парадокс дружбы выполняется, то $\beta_i(t) > 1$ [12].

Для примера, социальные сети Facebook и Github подтверждают парадокс дружбы, что показано на Рис. 1 [13]. Здесь на оси Oy отложено количество узлов сети, для которых индекс дружбы β_i попадает в диапазон, отложенный на оси Ox. Как видим, значительное большинство вершин графов имеют значение индекса дружбы большее единицы.



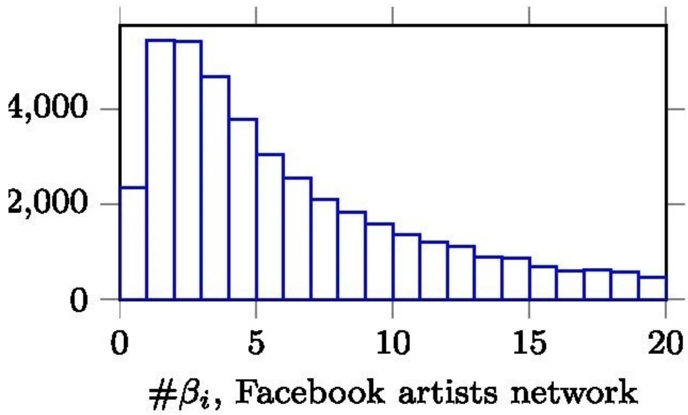


Рисунок 1 – Распределения индекса дружбы в сети телефонных звонков и сети доставки Amazon

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 *Erdös, P.* On random graphs i / P. Erdös, A. Rényi // *Publicationes Mathematicae Debrecen*. — 1959. — Vol. 6. — Pp. 290–297.

2 *Gilbert, E. N.* Random graphs / E. N. Gilbert // *Annals of Mathematical Statistics*. — 1959. — Vol. 30, no. 4. — Pp. 1141–1144.

3 *Blum, A.* Foundations of data science / A. Blum, J. E. Hopcroft, R. Kannan. — 2020.

4 *Bollobas, B.* Random Graphs / B. Bollobas. Cambridge Studies in Advanced Mathematics. — 2 edition. — Cambridge University Press, 2001.

5 *Райгородский, А. М.* Модели случайных графов / А. М. Райгородский // *Труды МФТИ*. — 2010. — Т. 2, № 4. — С. 130 – 140.

6 *Albert, R.* Statistical mechanics of complex networks / R. Albert, A.-L. Barabasi // *Reviews of Modern Physics*. — 2002. — Vol. 74, no. 1. — Pp. 47 – 98.

7 *Райгородский, А. М.* Модели случайных графов и их применени / А. М. Райгородский. — Москва, М.: Издательство МЦНМО, 2011.

8 *Берновски, М.* Случайные графы, модели и генераторы безмасштабных графов. / М. Берновски, Н. Кузюрин // *Труды Института системного программирования РАН*. — 2012. — Т. 22, № 4. — С. 419 – 432.

9 *David, E.* Networks, Crowds, and Markets: Reasoning About a Highly Connected World / E. David, K. Jon. — USA: Cambridge University Press, 2010.

10 Пользователи социальных сетей: современные исследования [Электронный ресурс]. — URL: https://psychojournal.ru/article/887-polzovateli-socialnyh-setey-sovremennye-issledovaniya.html (Дата обращения 10.03.2023). Загл. с экр. Яз. рус.

11 *Feld, S. L.* Why your friends have more friends than you do / S. L. Feld // *American Journal of Sociology*. — 1991. — Vol. 96. — Pp. 1464 – 1477.

12 *Pal, S.* A study on the friendship paradox quantitative analysis and relationship with assortative mixing / S. Pal, F. Yu, Y. Novick, A. Swami, A. Bar-Noy // *Applied Network Science*. — 09 2019. — Vol. 4. — Pp. 71 – 118.

13 *Sidorov, S.* Friendship paradox in growth networks: analytical and empirical analysis / S. Sidorov, S. Mironov, A. Grigoriev // *Appl Netw Sci*. — 2023. — Vol. 74, no. 6. — Pp. 47 – 51.