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 МОДЕЛИ СЛУЧАЙНОГО ГРАФА БАРАБАШИ-АЛЬБЕРТ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студентки 4 курса 451 группы
направления 09.03.04 — Программная инженерия
факультета КНиИТ
Медведевой Татьяны Витальевны

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 Модели растущих сетей ... 5

1.1 Обозначения и определения ... 6

1.2 Характеристики случайной величины ... 7

1.3 Переменные, характеризующие случайные события ... 7

1.4 Описание модели ... 8

2 Реализация модели Барабаши-Альберт ... 9

3 Экспериментальная часть ... 10

3.1 Степень узла ... 10

ЗАКЛЮЧЕНИЕ ... 12

ВВЕДЕНИЕ

С ростом количества данных и развитием технологий обработки информации все большую популярность приобретают исследования в области теории графов и сетевых моделей. Для описания сложных и распределенных систем, таких как социальные, экономические, биологические и технические сетевые системы, широко используются растущие сети.

Растущие сети обладают свойством самоорганизации, то есть процесс их развития происходит без централизованного управления. Они характеризуются постоянным добавлением новых вершин и связей, что приводит к формированию скейлинговой структуры: в них присутствуют выделенные вершины с высокой степенью, окруженные по большей части вершинами с низкой степенью.

Одной из самых изучаемых моделей растущих сетей является модель Барабаши-Альберт. Создание модели было мотивировано социальными сетями. Социальное сети — это сети, полученные в результате изучения отношений в социальных системах. Простым примером социальной сети является сеть друзей. Например, можно построить графическое представление сети дружбы, где вершины представляют людей, а ребра указывают на дружбу между людьми. Вершина i и вершина j соединены ребром, если и только если люди i и j считаются друзьями.

Модель построения случайного графа Барабаши-Альберт использовалась в качестве модели для социальных сетей. Эта модель была предложена в [**?**] с целью эмулировать рост таких сетей, как сети цитирования и Всемирная паутина.

Цель данной работы — провести исследование динамики локальных характеристик вершин и распределения локальных характеристик среди вершин с одинаковой степенью графа, построенного по модели Барабаши-Альберт.

Полученные результаты позволят более глубоко понять организацию подобных сетей и сравнить их с теоретическими ожиданиями.

В связи с поставленной целью возникают следующие задачи:

- исследование различных случайных графов, их свойств и характеристик вершин;
- исследование поведения реальных случайных сетей, их динамики и распределения;
- построение модели сети на языке программирования Python с использо-

ванием многопоточности;

- сбор и анализ статистических данных, полученных из модели;
- изучение распределения характеристик вершин графа относительно количества степеней вершин;
- исследование динамики характеристик вершин графа с изменением во времени.

1 Модели растущих сетей

В этом разделе будет исследовано понятие растущих сетей, кратко описаны различные модели: Эрдёша-Реньи, Барабаши-Альберт, Боллобаша-Риордана, копирования.

Понятие растущих сетей появилось в связи с необходимостью моделирования реальных активно развивающихся сетей. Таким образом, оно представляет собой тип сетей, которые увеличиваются во времени в связи с появлением новых вершин и ребер. Такие сети существует в различных областях: социальных, биологических, технических и так далее.

Свойства таких сетей представляют большой интерес. Так как зная их и имея возможность предсказать их поведение в будущем, можно контролировать реальные сети. Характеристики структуры, динамики сети очень важны, однако одним из наиболее важных свойств является распределение степеней узлов, ведь именно оно отличает растущую сеть от случайной.

Представим краткое описание различных моделей существующих сетей.

— **Модель Эрдёша-Реньи** представляет собой описание случайного графа, который создается без какой либо заданной структуры. Модель представлена 1959 году Паулем Эрдёшем и Альфредом Реньи. Именно она стала отправной точкой в активном исследовании случайных графов. В модели Эрдёша-Реньи образуется из n узлов и m ребер, где каждое ребро добавляется независимо друг от друга с некоторой вероятностью p. Модель стала важным открытие для понимания основных свойств случайных графов, однако она не учитывает много важных свойств реальных сетей, в том числе степенной закон.

— **Модель Барабаши-Альберт** является модификацией вышеописанной модели. В ней новые вершины присоединяются к графу с вероятностью, пропорциональной степени уже существующих вершин. Модель была предложена Альбертом-Ласло Барабаши и Реки Альберт в 1999 году. Эта модель значительно соответствует реальным сетям, например Интернет или социальной сети, где больше всего влияния у вершин с наибольшим количество соседей.

— **Модель Боллобаша-Риордана** была опубликована Белу Боллобашем и Дэвидом Риорданом в 2001 году. Каждая новая вершина такой модели имеет фиксированную вероятность, которая не меняется со временем и не зависит от степени самой вершины. Таким образом, в отличие от модели Барабаши-Альберт, в Боллобаша-Риордана вероятность связывания равномерно распределена внутри графа.

— **Модель копирования** использует простой детерминированный процесс для роста сети. Он заключается в копировании уже существующих узлов и ребер, при том новые узлы строятся на основе старых, вероятность выбора которых также зависит от степени вершины. Она была предложена Альберто-Ласло Барабаши и Рекки Майхила в 1999 году и может использоваться для объяснения свойств реальных сетей.

1.1 Обозначения и определения

В этом разделе определены характеристики узла в момент времени t, на основе которых будет строиться анализ сети.

Описание начинается с состояния сложной сети на итерации t. Оно представляется в виде графа G_t, определенного как пара (V_t, E_t), где $V_t = v_1, \ldots, v_t$ — набор вершин, представляющих собой узлы сети и $E_t = (v_i, v_j)|v_i, v_j \in V_t$ — соответствующие связи узлов сети. Из этого следует, что если $(v_i, v_j) \in E_t$, то вершины v_i и v_j являются соседями в данной сети на момент времени t. Индекс i определен как итерация, на которой появилась вершина v_i

Были определены основные компоненты сети и их состояния. Далее характеризуются непосредственно характеристики узлов сети. Первой является степень узла v_i на итерации t. Обозначим её как $d_i(t)$. Тогда суммарная степень всех соседей узла на момент времени t будет равна:

$$s_i(t) = \sum_{j:(v_i,v_j)\in E_t} d_i(t).$$

Средняя степень всех соседей узла v_i ($\alpha_i(t)$) является отношением суммы степеней соседей узла v_i к количеству его соседей:

$$\alpha_i(t) = \frac{s_i(t)}{d_i(t)}.$$

Согласно статье [**?**] индекс дружбы узла v_i вычисляется следующим отношением:

$$\beta_i(t) = \frac{\alpha_i(t)}{d_i(t)} = \frac{s_i(t)}{d_i^2(t)}.$$

Последней характеристикой узла v_i является коэффициент локальной кластеризации:

$$\theta_i(t) = \frac{2S_i(t)}{d_i(t)(d_i(t)-1)},$$

где $S_i(t)$ - количество треугольников, в которых узел v_i является вершиной.

Описанные ранее характеристики узла v_i могут изменяться при каждой итерации. Это зависит от того, присоединилась ли какая-либо новая вершина к v_i или к его соседу или нет.

1.2 Характеристики случайной величины

Для изучения поведения и закономерностей в графе, построенном по модели Барабаши-Альберт, потребуются характеристики различных случайных величин (например, степень вершины или индекс дружбы и т. д.). Первой такой величиной является математическое ожидание. Фактически, эта характеристика является средним значением величины и вычисляется по следующей формуле:

$$\mathbb{E}(x) = \frac{\sum_{i=0}^{n} x_i}{n},$$

где x_i некоторая случайная величина.

Для нахождения дисперсии случайной величины или стандартного отклонения σ применяются следующие расчеты:

$$D(x) = \mathbb{E}(x^2) - \mathbb{E}(x)^2,$$

$$\sigma(x) = \sqrt{D(x)}.$$

Дисперсия случайной величины характеризует рассеяние, разбросанность случайной величины около ее математического ожидания.

Для неотрицательной случайно величины x в качестве характеристики степени ее случайности применяют коэффициент вариации:

$$C(x) = \frac{\sigma}{\mathbb{E}(x)}.$$

1.3 Переменные, характеризующие случайные события

Для дальнейшего анализа стохастических соотношений потребуются вспомогательные переменные, характеризующие случайные события. Эти случайных события связаны с тем, будет ли узел v_i выбран новым узлом для связи (переменная $\xi_{i,l}^{t+1}$) или будет выбран один из соседей узла v_i (переменная $\eta_{i,l}^{t+1}$):

— пусть случайная переменная $\xi_{i,l}^{t+1} = 1$, если узел v_i выбран узлом v_{t+1} для присоединения на итерации $t+1$ посредством одного из ребер $l \in 1, \ldots, m$ и пусть она равна $\xi_{i,l}^{t+1} = 0$, в противном случае;

— Пусть случайная переменная $\eta_{i,l}^{t+1} = 1$, если новый узел v_{t+1} присоединяется к одному из узлов, уже связанных с узлом v_i, используя одну из созданных m ссылок $l \in 1, \ldots, m$ и пусть $\eta_{i,l}^{t+1} = 0$ в противном случае.

В дальнейшем для удобства определено:

$$\xi_{i,l}^{t+1} = \sum_{l=1}^{m} \xi_{i,l}^{t+1} \text{ и } \eta_{i,l}^{t+1} = \sum_{l=1}^{m} \eta_{i,l}^{t+1}.$$

Так как узел v_i имеет m шансов быть выбранным на итерации $t+1$ с вероятностью, пропорциональной его степени $d_i(t)$, получаем

$$\mathbb{E}(\xi_{i,l}^{t+1}) = m\frac{d_i(t)}{2mt} = \frac{d_i(t)}{2t}$$

и

$$\mathbb{E}(\eta_{i,l}^{t+1}) = m \sum_{j:(v_j,v_i)\in E_t} \frac{d_i(t)}{2mt} = m\frac{1}{2mt}s_i(t) = \frac{s_i(t)}{2t}.$$

1.4 Описание модели

Рассмотрим описание модели, представленное в книге [**?**].

Пусть в начальный момент времени существует полный граф $G_m = V_m, E_m$. Размер графа определен параметром $m \in \mathbb{N}$, который характеризует количество вершин, к которым присоединяется новая вершина в каждый момент времени t. Таким образом, количество вершин у полного графа равно $|V_m| = m$, а количество ребер $|E_m| = \frac{m(m+1)}{2}$.

Построение нового графа $G(t+1)$ в момент времени $t+1$ из графа $G(t)$ происходит путем добавления новой вершины $v_{t+1} : V_t \cup \{v_{t+1}\}$ к множеству вершин V_t. Далее к этой вершине добавляются m ребер. Они соединяют узел v_{t+1} с m уже существующими узлами. Выбор соседних вершин происходит с помощью реализации дискретной случайной величины ξ^{t+1}, которая принимает значение i с вероятностью $P(\xi^{t+1} = i) = \frac{d_i(t)}{2mt}$. Если $\xi^{t+1} = i$ то к графу добавляется ребро (v_{t+1}, v_i). Проводится m таких испытаний.

2 Реализация модели Барабаши-Альберт

В данной главе описана практическая часть работы. Здесь представлена реализация графа, построенного по модели Барабаши-Альберт на языке Python, а так же описан подсчет характеристик вершины. Кроме того в коде реализовано многопоточное построение 100 графов из 200 000 вершин и вывод данных в .txt файлы для дальнейшего их использования при построении графиков. Так же реализовано построение одного графа из 100 000 вершин, в котором производится анализ распределения локальных характеристик среди вершин с одинаковой степенью.

Многопоточная реализация представляет собой файл .py, который запускается в командной строке, тем самым начинается выполнение кода. Все остальные решения написаны в среде Jupyter Notebook, запуск кода производится в соответствие с этой средой.

3 Экспериментальная часть

В данной части будет проводится анализ распределения локальных характеристик среди вершин с одинаковой степенью и динамика локальных характеристик вершин графа при изменении момента t.

Реализация первого анализа состоит в построении одного графа на 100 000 вершин при двух параметрах $m = 3$ и $m = 4$. Далее локальные характеристики вершин группируются по степеням вершин и высчитываются характеристики внутри каждой подгруппы. Так же строится логарифмическая функция, которая позволяет замечать более очевидные зависимости.

Для второго анализа потребуется построить 100 графов из 200 000 вершин каждый. На основе него высчитываются характеристики локальных характеристик для 4-ёх вершин каждый 2000 момент времени t.

3.1 Степень узла

Соблюдается степенной закон в растущих сетях, в том числе в Барабаши-Альберт, согласно которому:

$$P(k) \sim k^{-\lambda},$$

где $P(k)$ — вероятность того, что степень вершины равна k, а λ — показатель степенного закона, который обычно, в реальных сетях, находится в диапазоне от 2 до 3.

Вероятность того, что новая вершина подсоединится к вершине степени k, пропорциональна ее степени k. Таким образом, вероятность того, что новая вершина будет иметь степень k, пропорциональна сумме степеней всех вершин, уже имеющих степень k:

$$P(k) \sim k \sum_{j<k} P(j),$$

где $\sum_{j<k} P(j)$ — это сумма вероятностей всех вершин с меньшей степенью, чем k.

Используя формулу вероятности:

$$\sum_{j<k} P(j) = \sum_{j<k} \frac{P(k)}{kj^{-\lambda}},$$

получаем степенной закон:

$$P(k) \sim k^{-\lambda}.$$

Таким образом, можно заметить, что степенное распределение степеней вершин в графе, реализованном на основе модели Барабаши-Альберт, является результатом процесса присоединения новых вершин к существующим, за счет чего сильно связанные вершины получают высокие степени вершин, а более слабые связи не получают никаких связей. Этот процесс приводит к сильной корреляции между степенями вершин, которая в конечном итоге приводит к степенному распределению $P(k)$.

ЗАКЛЮЧЕНИЕ

Исследование динамики и распределения растущих сетей является важным шагом к пониманию поведения реальных сетей. Изучение построенных моделей, поведение и структура которых похожа на, например, социальные или технические сети, помогает предсказывать развитие событий в реальной жизни. Кроме того, они могут стать основой для созданий новой работающей структуры на практике, и благодаря проводимым исследованием, ее развитие будет полностью предсказуемым.

Исследуемая сеть Барабаши-Альберт является хорошей основой для дальнейшего прогресса в сфере растущих сетей. Работа с экспериментальной частью показала, что в ней соблюдается степенной закон, что является одной из самых главных закономерностей поведения реальных растущих сетей. Однако она не может полностью описывать их поведение, это заметно по динамике индекса дружбы, который в отличие от реальных сетей, уменьшается со временем.

В работе были проведены исследования динамики таких локальных характеристик сети Барабаши-Альберт, как степень вершин, сумма степеней соседей вершины, средняя степень соседей, индекс дружбы и коэффициент локальной кластеризации, а так же изучены распределение этих локальных характеристик среди вершин с одинаковой степенью.

Полученные результаты могут быть полезны для дальнейших исследований в области теории графов и сетевых моделей, а также для практического применения в различных областях, описывающих социальные, экономические, биологические и технические сетевые системы.