МИНОБРНАУКИ РОССИИ

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

Кафедра теории функций и стохастического анализа

МОДЕЛЬ БАРАБАШИ-АЛЬБЕРТ С ФИКСИРОВАННЫМ И СЛУЧАЙНЫМ ЧИСЛОМ ДОБАВЛЯЕМЫХ РЕБЕР

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 412 группы
направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета
Хромов Данила Александрович

Научный руководитель
д. ф.-м. н., доцент ______________ С. П. Сидоров

Заведующий кафедрой
д. ф.-м. н., доцент ______________ С. П. Сидоров

Саратов 2022

Теория графов – область дискретной математики, особенностью которой является геометрический подход к изучению различных объектов.

Граф – основной объект этой теории. Это – система, состоящая из множества ребер и множества вершин. Вершны – это обозначения какого – то объекта (например, города или автобусной остановки).

Ребра обозначают связи между объектами (например, дороги или стоимость перевозки из одного пункта в другой, или длина дороги)

Степень вершины – количество ребер, выходящих из этой вершины. *Случайный граф* – граф, на котором задано распределение вероятностей.

Обычно под случайным графом понимается целый класс (множество) графов. Каждый конкретный граф из этого класса называется реализация случайного графа.

Случайные графы оказываются полезными при построении моделей сетей связи, которые подвержены случайным изменениям.

В естественно возникающих сетях(коммуникационных, природных, сетях цитирования) широко распространенны безмасштабные сети.

Безмасштабная сеть – это граф, степени вершин в котором распределены по степенному закону, т.е. доля вершин со степенью k примерно равна $k^{-\gamma}$ Например, если из вершины выходит 2 ребра, то число таких вершин равно $\frac{1}{2}$ или $\frac{1}{4}$ и т.д. Теория графов- основная часть теории науки о сетях. Это научная область, которая изучает компьютерные сети: коммуникационные, биологические, социальные и другие.

Кроме теории графов эта наука заимствует теорию и методы из информатики из статистики.

Польза науки о сетях заключается в том, что она даёт возможность прогнозировать исследуемые явления.

Цель данной работы - изучения безлимитных сетей, а именно - модели Барабаши-Альберт, а также её построение и анализ. Эта модель предназначена для генерации случайных безмаштабных сетей с использованием принципа предпочитаемых соединений. В работе она будет представленна в двух вариантах: с фиксированным числом добавляемых ребер на итерации и со случайным, распределенным по Гауссовскому или Пуассоновскому распределениям.

Математическая модель - одно из основных понятий прикладной математики. В математической модели реальные объекты и связи между ними описываются в виде математических соотношений. А затем математическая модель исследуется, изучаются её свойства и эта информация применяется в работе с реальными объектами.

В приложении приводится программа численной реализации алгоритма и численные результаты.

Основное содержание работы. Работа состоит из введения, трех разделов, заключения и приложения.

Во введении указывается область науки, к которой относится работа, цели работы , практическая ценность, проводимые исследование.

В разделе 1 описаны общие теоретические основы математического моделирования, как основного инструмента прикладной математики.

Раздел 2 посвящен моделям сложных безмасштабных сетей. Основное внимание уделено модели Барабаши-Альберт. Модель Барабаши-Альберт - это модель генерации случайных безлимитных сетей с использованием предпочтительного присоединения.

Модель Барабаши Альберт

Описание модели

Пусть $m \in \mathbb{N}$ есть фиксированное натуральное число (это будет параметр модели, равный количеству рёбер, присоединяемых на каждой из итераций построения сети).

В начальный момент времени $t = m$, $G_m = \{V_m, E_m\}$ есть полный граф с $|V_t| = m$ и $|E_t| = m(m+1)/2$.

Согласно модели Барабаши—Альберт, граф G_{t+1} получается из графа G_t (в каждый дискретный момент времени $t+1 = m+1, m+2, \ldots$) следующим образом:

— один узел v_{t+1} присоединяется к графу, т.е., $V_{t+1} = V_t \cup \{v_{t+1}\}$;

— добавляются m_{t+1} рёбер, $m_{t+1} \leqslant m$, соединяющих узел v_{t+1} с m_{t+1} уже существующими узлами; каждое из этих рёбер появляется в результате 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). Мы проводим m таких независимых испытаний. Если случайная величина ξ^{t+1} принимает одно и то же значение i в двух или более повторениях на итерации, то добавляется только одно ребро (в графе отсутствуют мульти-ребра).

Предварительный анализ

Для получения стохастических соотношений, описывающих динамику интересующих нас процессов во времени, мы введём вспомогательные индикаторные переменные, которые характеризуют случайные события, связанные с тем, будет ли узел v_i выбран новым узлом (переменная $\xi_{i,l}^{t+1}$), или будет ли выбран один из соседей узла v_i (переменная $\eta_{i,l}^{t+1}$):

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

— Пусть случайная переменная равна $\eta_{i,l}^{t+1} = 1$, 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.

Для простоты мы предполагаем, что все m рёбер на итерации $t+1$ проводятся одновременно и независимо друг от друга. Однако в этом случае возможно, что узел v_i будет выбран два или более раз. Вероятность того, что узел v_i будет выбран k раз во время итерации, пропорциональна $\left(\frac{d_i(t)}{2mt}\right)^k$. Хотя это значение не равно нулю, оно на порядок меньше вероятности выбора узла v_i один раз в серии экспериментов m, и поэтому мы можем исключить эти случаи из анализа.

Для дальнейшего удобства будем писать $\xi_i^{t+1} = \sum_{l=1}^{m} \xi_{i,l}^{t+1}$ и $\eta_i^{t+1} = \sum_{l=1}^{m} \eta_{i,l}^{t+1}$.

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

$$\mathbb{E}(\xi_i^{t+1}) = m\frac{d_i(t)}{2mt} = \frac{d_i(t)}{2t} \tag{1}$$

и

$$\mathbb{E}(\eta_i^{t+1}) = m \sum_{j:(v_j,v_i)\in E_t} \frac{d_j(t)}{2mt} = m\frac{1}{2mt}s_i(t) = \frac{s_i(t)}{2t}. \tag{2}$$

В результате присоединения каждого из m рёбер значения $s_i(t)$ может увеличиваться только в трех случаях:

— новый узел v_{t+1} (степени m) присоединяется к узлу v_i на итерации $t+1$, и таким образом, значение величины $s_i(t)$ увеличится на m, в то время как величина $d_i(t)$ увеличится на 1: $d_i(t+1) = d_i(t) + 1$, *а*);

— новый узел v_{t+1} присоединяется к одному из соседей узла v_i посредством одного из своих m ребер, и тогда значение величины $s_i(t)$ вырастет на 1, в то время как величина $d_i(t)$ останется без изменений: $d_i(t+1) = d_i(t)$, *б*);

— новый узел v_{t+1} присоединится одним из своих m ребер с одним из соседей узла v_i, в то время как одно из оставшихся ребер связывает его с узлом v_i. Тогда суммарная степень всех соседей узла v_i увеличится на $m+1$ и степень узла v_i вырастет на 1, *в*).

Следует отметить, что вероятность третьего случая на порядок ближе к нулю по сравнению с вероятностями первых двух случаев, поэтому мы исключаем этот случай из анализа, чтобы не загромождать вывод уравнений.

Динамика степени узла

Покажем, что динамика степени каждого узла во времени описывается степенным законом

$$\mathbb{E}(d_i(t)) = m\left(\frac{t}{i}\right)^{\beta} \tag{3}$$

с одной и той же экспонентой $\beta = \frac{1}{2}$ для всех узлов.

Обозначим $\Delta d_i(t+1) := d_i(t+1) - d_i(t)$. С учётом проведённого анализа, получаем следующее стохастическое уравнение

$$\Delta d_i(t+1) = (d_i(t)+1)\xi_i^{t+1} + d_i(t)(1-\xi_i^{t+1}) - d_i(t) = \xi_i^{t+1}.$$

Тогда, используя (1), имеем

$$\mathbb{E}(\Delta d_i(t+1)|G_t) = \frac{d_i(t)}{2t}.$$

Этому разностному уравнению соответствует дифференциальное уравнение

$$\frac{df(t)}{dt} = \frac{f(t)}{2t},$$

где $f(t) = d_i(t)$. Решением уравнения будет $d_i(t) = ct^{\frac{1}{2}}$, где $c = c(i)$ есть некоторая константа, которую можно найти, используя начальное условие $d_i(i) = m$. Таким образом, соотношение (3) установлено.

Распределение степеней вершин

Исследуем распределение степеней при увеличении числа итераций t до бесконечности и покажем, что оно сходится к распределению степенного типа.

Найдём изменение числа узлов степени k после добавления в сеть одного нового узла на итерации t (при этом присоединяются m рёбер, соединяющие этот новый узел с m существующими вершинами).

Обозначим через $p_k(t)$ вероятность того, что случайно выбранная вершина имеет степень k (на итерации t). Пусть $q_t(s,k)$ есть вероятность того, что узел v_s имеет степень k в момент (на итерации) t.

По определению $q_t(s,k)$ можно записать следующие равенства:

$$\frac{1}{t}\sum_{s=1}^{t} q_t(s,k-1) = p_{k-1}(t),\ \frac{1}{t+1}\sum_{s=1}^{t+1} q_{t+1}(s,k) = p_k(t+1),\ \frac{1}{t}\sum_{s=1}^{t} q_t(s,k) = p_k(t). \tag{4}$$

Поскольку каждый новый узел получает степень m, мы можем записать

$$q_{t+1}(t+1,k) = \begin{cases} m, & k = m \\ 0, & k \neq m. \end{cases} \quad .$$

Вершина s может иметь степень k на итерации $t+1$ только в двух случаях:

— узел v_s имеет степень $k-1$ на итерации t и увеличивает ее на 1 на следующей итерации $t+1$;

— узел v_s уже имеет степень k на итерации t, которая не меняется на следующей итерации $t+1$.

Вероятность того, что узел v_s имеет степень $k-1$ на итерации t и этот узел выбран на шаге 2 итерации $t+1$, равна

$$q_t(s,k-1)m\frac{k-1}{2|E(t)|}.$$

Вероятность того, что узел v_s имеет степень k и узел *не выбран* на итерации $t+1$, равна

$$q_t(s,k)\left(1-m\frac{k}{2|E(t)|}\right).$$

Тогда

$$q_{t+1}(s,k)=q_t(s,k-1)m\frac{k-1}{2|E(t)|}+q_t(s,k)\left(1-m\frac{k}{2|E(t)|}\right). \qquad (5)$$

После суммирования левой и правой частей (5) по $s=1,\ldots,|V(t)|$ с использованием приближений $|V(t)|\sim t$, $|E(t)|\sim 2mt$, получаем

$$(t+1)p_k(t+1)-m=\frac{1}{2}(k-1)p_{k-1}(t)+tp_k(t)-\frac{1}{2}kp_k(t). \qquad (6)$$

Чтобы найти стационарное распределение степеней, обозначим через $p_k=p_k(\infty)$ предельную вероятность, тогда имеем

$$(t+1)p_k(t+1)-tp_k(t)\to p_k,\ t\to\infty,$$

так как $(t+1)p_k(t+1)-tp_k(t)\to(t+1)p_k(\infty)-tp_k(\infty)=p_k(\infty)=p_k$.

Таким образом, динамика изменения числа узлов степени k выглядит следующим образом:

$$2p_k=(k-1)p_{k-1}-kp_k. \qquad (7)$$

Приближенное дифференциальное уравнение к уравнению (7) есть

$$\frac{dp(k)}{dk}=-\frac{3p(k)}{k}, \qquad (8)$$

где $p(k)\sim p_k$.

Решением уравнения (8) является

$$p(k)=Ck^{-3},$$

где С есть некоторая постоянная.

Необходимые определения:

Математическим ожиданием случайной дискретной величины называется сумма, заданная её значениями $x_1, x_2 ... x_n$ и значениями вероятностей $p_1, p_2 ... p_n$, соответствующим этим случайным величиным.

$M(X) = \sum\limits_{i}^{n} x_i p_i$

($M(X)$ - среднее значение)

Дисперсией случайной дискретной величины X называется величина $D(X) = \sum\limits_{i}^{n} M(X^2) - (M(X))^2$

Математическое ожидание квадрата отклонения этой величины от её математического ожидания. $D(X)$обозначается еще как $\sigma^2(x)$

Среднеквадратическим отклонением случайной величины X от средних ожидаемых называется величина. $\sigma(X) = \sqrt{\sigma^2(x)} \equiv \sqrt{D(X)}$

В разделе 3 приводится описание алгоритма построение модели Барабаши-Альберт и результаты математического эксперимента.

Алгоритм модели Барабаши-Альберт

Сеть начинается с начальной сетки с m_0 узлами, и степень каждого узла начальной сети должно быть не меньше 1, иначе они будет всегда отделены от остальных частей сети. В каждый момент времени в сеть добавляется новый узел. Каждый новый узел соединяется с существующими узлами вероятностью, пропорциональной числу связей этих узлов. Формально, вероятность p_i того, что новый узел соединится с узлом i равна $p_i = \frac{k_i}{\sum_i k_j}$, где

k_i - степень i-ого узла, а в знаменателе суммируются степени всех узлов.

Алгоритм программной генерации графа

Алгоритм начинается с введения простого интерфейса для взаимодействия с пользователем. В процессе на экран будут выводиться вопросы двух типов: закрытые и со свободным ответом. Также реализована проверка вводимых параметров на корректность. При выполнении алгоритма система запрашивает тип рассматриваемой модели, количество итераций и графов. При выборе фиксированного количества добавляемых ребер(m) мы задаем m вручную, либо оставляем стандартное значение, в противном случае выбираем распределение (распределение Гаусса или распределение Пуассона) с параметрами и случайно генерируем по нему вектор значений m. Генерация графа начинается с одной вершины, и до тех пор, пока возможное количество ребер не станет больше m, каждая новая вершина будет соединяться с той или иной вершиной. Она рассчитывается как отношение степени этой вершины к сумме степеней всех вершин в графе ($p_i = \frac{k_i}{\sum_i k_j}$). Затем, с помощью функции np.random.choice мы по вероятностям случайным образом выбираем одно из значений, а затем проводим искомые ребра к i-ым вершинам. Алгоритм повторяется до тех пор, пока количество вершин не станет равным числу итераций. Также в ходе генерации графа мы чертим график степени первой вершины. Затем мы повторяем алгоритм генерации графа заданное в переменной count количество раз, выстраиваем график среднего значения степени первой вершины. После того как график построен, мы считываем номера итераций, для которых считаем значение среднеквадратичного отклонения по формуле

$S = \sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$, где x_i, $\overline{x}$ - значения точек на графике,

и выводим их.

В приложении приводится код программы на языке Python, реализующий модель Барабаши-Альберт. Приведённая в приложении программа позволяет увидеть результаты моделирования с помощью библиотек по визуализации графов и графиков функций.

Практическая ценность данной работы. При генерации большого числа случайных сетей и наблюдении за изменениями их параметров в динамике, эта работа даёт возможность оптимально прогнозировать реальные сложные сети, например социальные или транспортные, а также выстраивать оптимальную стратегию при их создании. Такая применимость достигается за счет того, что принцип предпочтительного присоединения часто выполняется на практике.

Заключение В данной работе изучены общие принципы построения моделей сложных сетей, построен алгоритм модели Барабаши-Альберт, разработаны программа, приведены графики полученных результатов.

Сравнение результатов в случаях фиксированного и случайного числа добавляемых ребер показывает, что второй случай дает нам больше возможностей применения модели в реальных практических задачах.

Полученные результаты могут представлять интерес при моделировании реальных сетей