МИНОБРНАУКИ РОССИИ

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

Кафедра теории функций и стохастического анализа

МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ ЭПИДЕМИЙ НА ГРАФАХ

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 248 группы
направления 01.04.02 — 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 2024

Область математики, к которой относится данная работа - теория графов с привлечением теории и методов информатики и статистики.

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.

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

Цель данной работы - изучение, построение и анализ модели Барабаши-Альберт, а затем построение на ее базе модели распространения эпидемий SI. Модель Барабаши-Альберт предназначена для генерации случайных безмасштабных сетей с использованием принципа предпочтительных присоединений. 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 алгоритмов.

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).

В разделе 3 приводится описание алгоритма программной генерации графа по модели Барабаши-Альберт и результаты математического эксперимента. **В разделе 4** описываются общие проблемы распространения эпидемий в сложных сетях, и изучается модельSI распространения информации.

Цель раздела состоит в изучении и анализе проблем моделирования распространения информации в графах.

Кривая, описывающая диффузию информации в социальных сетях, представляет собой S-образную линию и показывает зависимость между общим количеством пользователей, получивших сообщение по конкретной тематике с течением временем. Эта кривая может быть условно разделена на три участка, которые соответствуют трем этапам диффузионного процесса. Процесс диффузии информации в социальной сети может быть сравнен с эпидемическим распространением болезней.

На первом этапе только небольшая часть пользователей социальной сети обладает определенной информацией, которую часто называют "новаторами". Затем новаторы распространяют эту информацию, публикуя ее на своих страницах в социальных сетях, что делает ее доступной для их подписчиков, в основном для их ближайших контактов. Затем пользователи, которые были заражены данной информацией (тематикой), продолжают распространять информацию дальше по социальной сети, используя свой список контактов и подписчиков. На этом временном отрезке кривая начинает медленно расти от начала участка до его конца.

На втором этапе кривая роста значительно ускоряется и переходит в экспоненциальный режим. На этом этапе происходит наибольший рост числа пользователей, получивших информацию. Увеличение скорости распространения информации обусловлено большим количеством пользователей, которые еще не получили эту информацию.

Затем кривая переходит в режим насыщения, что приводит к замедлению скорости распространения информации. Это происходит из-за того, что большинство пользователей сети уже получили эту информацию и/или потеряли к ней интерес.

Заметим, что аналогичные процессы наблюдаются при распространении инфекционных заболеваний и инноваций. Например, можно рассмотреть рынок мобильной связи в разных странах в качестве инновации. Чаще всего используется логистическая модель, так как она лучше всего подходит для изучения процесса S-образной диффузии мобильных сетей.

На данный момент существует множество различных моделей диффузии, включая модель Басса, модель Гомперца и другие. В последние годы ученые широко применяли эти модели и их вариации для анализа различных явлений . Важным этапом является сравнение нескольких моделей для изучения конкретного набора данных и выявления преимуществ и недостатков каждого метода. Обзор моделей диффузии инноваций, основанный на агентном подходе, можно найти в [13]. В последние годы одной из основных тем стало изучение распространения информации в социальных сетях.

В настоящее время социальные сети стали чрезвычайно популярными благодаря тому, что они предоставляют пользователям огромные объемы данных. Это привело к возникновению множества исследований, посвященных распространению информации в социальных сетях. Ранее проводились экспериментальные исследования или решались системы дифференциальных уравнений для изучения динамики распространения различных процессов. Недавние исследования использовали методы моделирования процессов диффузии информации, такие как уравнения теплопередачи, уравнения реакции-диффузии или уравнения гидродинамики.

К сожалению, существующие модели фокусируются на ранних стадиях распространения информации. В нашем исследовании мы ставим целью создать модель, которая может описывать поведение на более поздних стадиях распространения информации. Предыдущие исследования показали, что сетевая структура играет важную роль при использовании дифференциальных уравнений. Поэтому в нашей работе мы используем модель, которая учитывает сетевую структуру. Мы отказываемся от гипотезы о гомогенном смешении и сосредотачиваемся на диффузии информации в сетях с безмасштабной структурой.

Модель SI

Для анализа диффузии информации, обычно изучаются три основных логистических модели: модели SI, SIS и SIR. Эти модели помогают нам понять основные принципы распространения информации. Все эти модели описывают S-образную кривую, которая отражает процесс распространения информации среди групп пользователей в социальных сетях.

Модель SI является базовой моделью, которая используется для описания распространения информации в социальных сетях. Эту модель впервые использовал Гриличес в своей работе, где он применил логистическую модель для объяснения широкого использования гибридной кукурузы в США. Некоторые ученые рассматривали это исследование как пример использования логистической модели в своих работах.

В соответствии с представленной моделью, пользователи социальных сетей, которые активно участвуют в дискуссии определенной темы, могут быть классифицированы на две основные категории: первая группа - это инициаторы распространения информации, которые публикуют статьи или сообщения, соответствующие заданной тематике и используют уникальные хэштеги, тогда как вторая группа - это те пользователи, которые еще не получили информацию и не принимают участия в ее распространении. Модель SI логистического роста может быть представлена в виде дифференциального уравнения:

$$\frac{dI(t)}{dt} = \beta \langle k \rangle \frac{S(t)I(t)}{N}, \tag{1}$$

где N – количество пользователей в сети, $S(t)$ – количество пользователей, которые не получили новостное сообщение и не обладают информацией относительно данной тематики в момент времени t (такие пользователи будем называть восприимчивыми), $I(t)$ – количество пользователей, которые получили новостное сообщение и продолжают его распространение на заданную тему в момент времени t (таких пользователей будем называть информированными), β – вероятность распространения информации, то есть вероятность того, что пользователь, получивший сообщение, заинтересуется его тематикой и будет передавать его или его производные другим пользователям в сети в единицу времени, k – среднее количество контактов пользователей в социальной сети.

Обозначим через $s = s(t) = S(t)/N$ и $i = i(t) = I(t)/N$ доли восприимчивых и информированных пользователей соответственно в момент времени t. Тогда уравнение (1) может быть переписано следующим образом:

$$\frac{di}{dt} = \beta \langle k \rangle si = \beta \langle k \rangle i(1 - i), \tag{2}$$

где произведение $\beta \langle k \rangle$ является скоростью передачи информации.

Это дифференциальное уравнение первого порядка с начальным условием $i_0 = i(0)$ может быть решено аналитически:

$$i = i(t) = \frac{i_0 e^{\beta \langle k \rangle t}}{1 - i_0 + i_0 e^{\beta \langle k \rangle t}} \tag{3}$$

Уравнение (3) указывает, что:

— В начале процесса доля пользователей, которые получают новостные сообщения, растет экспоненциально. На ранних этапах только восприимчивые пользователи сталкиваются с информацией, поэтому она легко распространяется.

— Характеристическое время, необходимое для того, чтобы доля восприимчивых пользователей достигла значения $1/e$ (приблизительно 36%), определяется выражением:

$$\tau = \frac{1}{\beta \langle k \rangle} \tag{4}$$

Таким образом, значение τ обратно пропорционально скорости распространения информации среди пользователей сети. Соответственно, увеличение количества связей k или параметра β ускоряет распространение информации и сокращает характеристическое время.

— С течением времени пользователь, получивший информацию, передает ее меньшему числу восприимчивых пользователей. Поэтому рост доли i замедляется при больших значениях t (рис. 1). Распространение информации заканчивается, когда все пользователи получили информацию, то есть при $t \to \infty$ доля i становится равной 1, а доля s становится равной 0.

Наконец, в разделе 4 дается алгоритм численного эксперимента и его результаты.

В приложении приводятся коды программ численной реализации построенных моделей. **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. Результаты математического моделирования распространения эпидемий могут быть использованы для других видов распространения информации в социальных сетях.

Заключение В данной работе изучены общие принципы построения моделей сложных сетей, построен алгоритм модели Барабаши-Альберт, алгоритм прогнозирования распространения эпидемий в сложных сетях на примере модели SI , разработаны программы, приведены графики полученных результатов.

Сравнение результатов позволяет сделать вывод, что моделирование распространение эпидемий в сложных сетях приближено дает представление о распространении заболеваний в реальных сетях.

Полученные результаты могут представлять интерес при моделировании реальных сетей и их исследовании.