

Федеральное государственное бюджетное образовательное учреждение
высшего образования

Санкт-Петербургский государственный университет

Критическая масса в социальных сетях

Направление 01.04.02 Прикладная математика и информатика
Магистерская программа
Математическое и информационное обеспечение экономической
деятельности

Гусева Алина Михайловна

Научный руководитель:

доцент, к.ф. - м.н.

Лежнина Елена Александровна

Санкт-Петербург

2020 г.

Содержание

Введение	3
Постановка задачи	4
Обзор литературы	5
Глава 1. Математическая модель	9
1.1. Поиск критической вероятности	9
1.2. Модель поведения сети выше перколяционного порога	13
1.3. Модель поведения сети ниже критической вероятности	15
Глава 2. Численное моделирование	17
Выводы	24
Заключение	25
Список литературы	26
Приложение 1	29
Приложение 2	30
Приложение 3	31

Введение

В современном мире социальные сети плотно вошли в нашу жизнь. Сложно представить себе человека, который не зарегистрирован или не пользовался бы ими каждый день. Сети позволяют налаживать не только коммуникацию между людьми, но и на своих площадках развивать рекламные и торговые компании. На сегодняшний день социальные сети являются самым посещаемым ресурсом в интернете.

Социальная сеть – социальная структура, состоящая из группы узлов или социальных объектов и связями между ними. Этими объектами могут быть люди или группа людей, сообщества, организации. В буквальном смысле, социальная сеть – это сообщество людей, объединенных одинаковыми интересами, общими социальными связями, каким-либо делом для взаимодействия между собой.

В настоящее время около 3,8 миллиардов людей [1] зарегистрированы в различных социальных сетях. Поскольку каждый пользователь добровольно публикует информацию о себе, о своих взглядах и предпочтениях, социальные сети представляют богатый источник данных для анализа и исследования поведения групп людей. Большую роль в данных исследованиях играют структурные свойства социальной сети. С одной стороны чем больше у пользователя связей, тем больше возможностей через свое окружение повлиять на мнение всей сети. Но, с другой стороны, тем больше уязвимость к чужому мнению.

Постановка задачи

На сегодняшний день существует большое множество работ, посвященных исследованию влияния в социальных сетях. Математические модели, которые показывают критические переходы были объединены в одну группу - перколяционные модели. Такие модели активно используются в физике. В теории перколяции постановка задачи звучит следующим образом: имеется двумерная решетка с некоторым количеством узлов и связей между ними. Часть узлов занята и может проводить электрический ток. Остальная часть узлов является диэлектриком. Необходимо найти такую минимальную концентрацию занятых узлов, при которой имеется сквозной путь через всю решетку, электрический ток пронизывает всю решетку от одного края до другого.

Касательно социальных сетей, узлами являются пользователи социальной сети, а связями – социальные связи между ними. Будем считать занятыми узлами тех пользователей, которые принадлежат какому-либо движению, имеют одинаковые взгляды, либо обладают схожими интересами. Применяя модель теории перколяции к социальным сетям, задача звучит следующим образом: необходимо найти такую минимальную долю занятых пользователей, при которой оставшиеся узлы социальной сети начнут присоединяться к этому движению.

Другой важной задачей данной работы является моделирование поведения социальной сети в зависимости от порога перколяции. При достижении перколяционного кластера свойства социальной сети резко и скачкообразно меняются: возникает или затухает эпидемия, резко увеличиваются просмотры популярного блогера. В ходе данной работы был рассмотрен численный пример. Основной метод – численное моделирование на решетках или деревьях.

Обзор литературы

На данный момент существует множество работ, посвященных моделированию социальных сетей. В одних из них большое внимание уделяется правилам взаимодействия между агентами, в других – моделям управления данными в социальных сетях.

Одно из направлений исследования социального взаимодействия составляют модели критической массы. Например, модель [17], в которой агенты осуществляют бинарный выбор, а их поведение можно описать одной целевой функцией. Функция полезности агента возрастает с увеличением доли агентов, сделавших такой же выбор. Такие модели получили широкое распространение в физике и эпидемиологии. Одним из ярких примеров является SIR-модель [12]. В этой работе рассматривается совокупность людей, где каждый человек может заразиться инфекцией при контакте с другими людьми. Весь процесс заболевания делится на три стадии – восприимчивость, заражение инфекцией, выход из состояния болезни. Важный вопрос, на который необходимо найти ответ – будет ли инфекция распространяться до тех пор пока все восприимчивые люди не переболеют или инфекция затухнет, не коснувшись части населения. После выздоровления индивидум приобретает иммунитет к болезни. В качестве критической массы выступает плотность населения. Существует расширение этой модели [5]. Совокупность людей разбивают на четыре стадии – здоровые, восприимчивые, больные и выздоровевшие. Выздоровевший становится восприимчивый к болезни через некоторое время. Самый простой пример – заболевание гриппом. Распространение вирусов характерно и для интернет ресурсов. В статье [15] авторы находят компьютерный вирус и среднюю продолжительность жизни инфекции. Определяя динамическую модель распространения, ученые утверждают, что в такой модели отсутствует эпидемиологический порог, а вместе с ним и критическое поведение. Однако основополагающей моделью критической массы в социальной науке является работа [20]. В этой модели рассматривается то, как группа людей принимает решение относительно района проживания. Выбор строится по отношению к своему окружению. В другой работе этого же автора [19]

используется модель пространственного соседства. Рассматривается поведение двух групп людей, которые отличаются по одному признаку. Все они, в зависимости от своих предпочтений, имеют две возможности: либо жить в окружении людей из другой группы, либо перемещаться в то место, где люди из своей группы представлены в большей пропорции. В рамках этой работы были построены кластеры, при изменении размера которых нарушалась структура равновесия. Также существует модель ограниченного окружения [21]. В отличие от предыдущей модели этого автора под окружением понимается сообщество, в котором агенту комфортно находиться. В этой модели при незначительном превышении плотности агентов некоего критического значения существует несколько положений равновесия. В книге [22] несколько примеров, основанных на модели ограниченного соседства. Среди них оказались такие примеры как коллективное поведение, посещение баскетбольных матчей, присоединение к аплодисментам, голосование. В работе [11] автор рассмотрел несколько с другой стороны критическую массу. В отличие от моделей [19], [21], где рассматривались группы людей, здесь автор исследовал устойчивость равновесия в поведении одного типа агентов. Они могут принимать одно из двух решений: действовать или бездействовать. Опираясь на свои предпочтения, агенты действуют так, чтобы максимизировать свою прибыль. Однако в своей работе автор не учитывал случаи, когда некоторые из рассматриваемой совокупности являются друзьями. Именно дружеские связи сильно изменяют поведение агентов. Улучшенную модель реализовал Чуэ в своей работе [8], однако в ней есть свои минусы: не учитываются временные ограничения для перемещения агентов. Несмотря на все ограничения, пороговая модель Грановеттера получила широкое использование в области социального поведения.

В социальных сетях пользователи добавляют друг друга в друзья. Численность пользователей растет достаточно быстро. Закон Меткалфа [14] гласит, что если каждый участник может связаться с каждым, то эффект полезность сети пропорционален квадрату численности пользователей сети. Сетевой эффект соответствует числу возможных связей. Позднее в статье [18] автор сформулировал свой закон, основанный на [14]. Ценность се-

ти возрастает геометрически и $\sim 2^n$. Однако ученые, которые исследуют социальные сети, в своей статье [16] предполагают, что социальная сеть размером n растет пропорционально $n \log(n)$. Эта скорость быстрее, чем линейный рост, но намного медленнее, чем квадратичный рост по закону Меткалфа. В статье [6] авторы вводят неиерархический механизм роста социальной сети, который не определяет строгие правила для присоединения новых пользователей. В результате исследования получают уравнения скорости для эволюции распределения степеней.

С помощью ежедневных взаимодействий, таких как комментирование фотографий или отправка сообщений, возникают небольшие группы людей, изменяющиеся со временем, с одинаковыми интересами. В статье [9] авторы моделируют социальную сеть, предлагая новый подход для анализа взаимоотношений. Исследуя поведения отдельных узлов, ученые приходят к выводу, что существует корреляционное отношение между дружественной связью и присутствием пользователя в определенной группе по интересам. Одним из подходов управления потоками данных в информационных сетях являются модели, основанные на теории перколяции. Такие задачи возникают чаще всего в физике при протекании жидкости или газа. Однако встречались случаи, когда данную теорию применяли и в геологии для определения свойств полезных ископаемых. Слово перколяция впервые употребили в 1957 г. в работе [7]. С помощью перколяционных моделей можно производить информационные расчеты, анализировать распространение эпидемии, исследовать свойства пористых материалов, исследовать надежность сети. В работе [4] используется теория перколяции для обеспечения максимально эффективного использования полосы пропускания для водных видов транспорта. Специфика такой задачи состоит в разнообразии целей управления, также существенное влияние оказывает большая размерность системы. В результате этой работы была выделена возможность использовать теорию перколяции для обеспечения целостной передачи данных, когда система динамически изменяется, переставая носить стационарный характер. Позднее, в работе [23] автор применил теорию протекания для описания информационной проводимости многосвязных семантических сетей. В статье описаны подходы к семантическому пред-

ставлению документа. Перколяционный порог определяет потерю смысла и разбиение текста на отдельные несвязные фрагменты, показано соотношение этого порога и смысловой силы документа. В качестве показателя информационной проводимости используется порог протекания.

На сегодняшний день теория перколяции используется во многих точных науках. Существуют целые сборники с подробным описанием алгоритмов работы. Одним из таких является книга [2]. В первых главах рассматриваются регулярные и стохастические фракталы, а дальше автор переходит к описанию фрактальных структур, возникающих при геометрических фазовых переходах. Именно с использованием фрактальной геометрии и теории перколяции автор описывает работу сложных систем.

Глава 1. Математическая модель

1.1 Поиск критической вероятности

Для моделирования сложных сетей обычно используют модель «классический случайный граф», модель «тесный мир» или модель «безмасштабного» графа. В большинстве случаев число связей в графе имеет распределение Пуассона. Это означает, что существует строгий максимум примерно равный среднему значению. Однако для социальных сетей среднего значения не существует. В социальных сетях небольшое количество узлов имеет очень много друзей или связей, и огромная доля содержит маленькое количество связей. Соответствующее вероятностное распределение подчиняется степенному закону распределения

$$P(q) \sim q^\lambda$$

Степенная функция является признаком безмасштабного графа. В этих графах отсутствует определенное число связей для каждого узла.

В данном исследовании используется перколяционная модель, предложенная в работе [13], которая делает три приближения к реальным социальным сетям:

- Социальная сеть имеет бесконечное число членов
- Существует фиксированное число связей для каждого узла
- В модели нет циклов. Если узел А знает узел В, узел В знает узел С, а узел С знает узел А, то узел А знает узел С.

Данная модель предназначена для обеспечения работоспособного приближения социальной сети. Она даст представление об их структуре и управлении, сохраняя при этом эквивалент критической массы.

Теория перколяции изучает вероятностные модели, демонстрирующие фазовый переход. Данная концепция была разработана на симпозиуме по методам Монте-Карло в 1954 году.

«Предположим, что есть пористый материал, в который наливаем жидкость сверху. Этот материал был смоделирован математически: трехмерная сеть, состоящая из n^3 точек и соединений между двумя соседними отверстиями, которые могут быть открыты, пропуская жидкость, с вероятностью p , либо закрыты с вероятностью $p-1$. Тогда для заданного p , какова вероятность, что жидкость просочится сверху ко дну?» [7]

Вероятность протекания $p = p_c$, критическая вероятность, – вероятность, для которой существует путь от начала до конца в пористом составе.

Аналогично, в данной работе используем решетку Бете для создания модели социальной сети. Эта модель зависит от двух параметров:

1. z – максимальное число связей, друзей, которые может иметь человек
2. p – вероятность того, что любые два человека являются друзьями.

Два параметра (p, z) упрощенно означают то, что в любой локальной области будет пара случайных величин (p, z) . В этой модели не учитывается время, игнорируется динамический процесс, который позволял бы знакомым добавляться в сеть. Данные предложения позволили применить теорию перколяции к социальной сети и получить некоторые результаты.

Имеется некоторая матрица связи пользователей социальной сети. Часть пользователей случайно проводит информацию или принадлежит какому-нибудь движению, а другая часть нет. При вероятности p равной 0 социальная сеть не может проводить никакую информацию. Это обусловлено не только отсутствием связи между пользователями. Невосприимчивость к информации сильно зависит от характеристик пользователей социальной сети. С увеличением концентрации проводящих узлов в социальной сети возникает единый большой кластер из проводящих связей. Поскольку в социальной сети бесконечно много пользователей, этот кластер имеет бесконечно большой размер и называется бесконечным. При возникновении такого кластера, свойства социальной сети, характеризующие ее, резко меняются. Все остальные кластеры, не принадлежащие бесконечному кластеру, называют конечными.

При малых концентрациях, когда $p \ll 1$, узлы либо изолированы друг от друга и не имеют общих связей, либо формируют небольшие груп-

пы – кластеры. Существуют только кластеры небольших размеров. Два узла принадлежат одному кластеру тогда и только тогда когда они связаны цепочкой проводящих соседей. Если же $p = 1$, то все узлы принадлежат одному кластеру. Существует некоторое критическое значение, при котором происходит переход от одного состояния в другое. Это значение p_c называется порогом протекания.

Самым удобным способом демонстрации данной модели будет дерево Кейли или решетка Бете. Рассматриваем большую популяцию пользователей социальной сети. Каждый узел представляет собой пользователя социальной сети, который может быть знаком с другими пользователями этой сети. Ребрами является связь между пользователями. У каждого узла может быть максимально z ребер. Формирование дерева Кейли начинается с центрального узла. Данная модель симметрична в том смысле, что центральным может быть выбран любой узел.

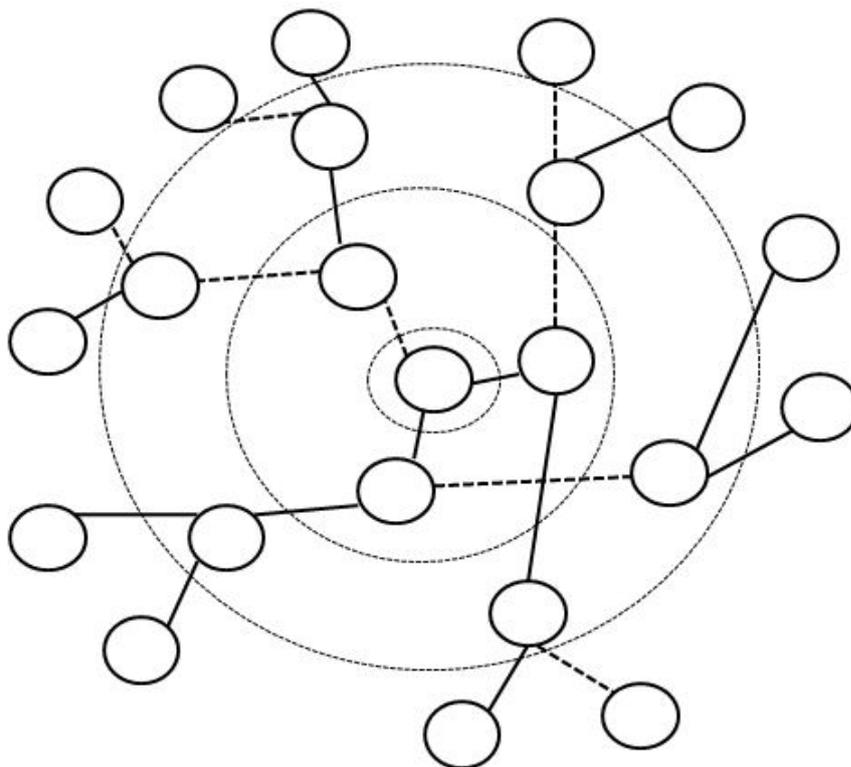


Рис. 1: Модель социальной сети

Каждая связь заканчивается узлом, который, в свою очередь, также является началом для z связей. Причем одна из этих связей соединяет

данный узел с центральным исходным, а дургие $z - 1$ приводят к новым узлам. Этот ветвящийся процесс продолжается снова и снова. Воспользуемся понятием химического расстояния дерева из теории графов. Химическое расстояние в модели – наикратчайшее расстояние от выбранного узла до центрального. Как было замечено выше, в модели предполагается, что нет циклов. Это означает, что любые два узла соединены только одним ребром. Таким образом, на химическом расстоянии равном 1 в сети будет $1 + z$ пользователей. Химическое расстояние может быть бесконечно большим. Выведем рекуррентную формулу. При химическом расстоянии равным двум, максимальное количество пользователей в социальной сети будет $1 + z(1 + (z - 1))$. Раскрывая скобки, получим $1 + z^2$ членов социальной сети. Проводя аналогию, на химическом расстоянии равном r максимальное количество пользователей будет равно

$$1 + z(1 + (z - 1) + (z - 1)^2 + (z - 1)^3 + \dots + (z - 1)^{r-1}) \quad (1)$$

Связные члены имеют тенденцию делиться на кластеры. Кластеры можно воспринимать как личности, которые разделяют интересы. В литературе очень много рассказывается про кластеризацию, но отдельный интерес представляет один большой кластер. В этом исследовании будем называть его гигантским кластером. Он может существовать или не существовать. С математической точки зрения гигантский кластер – это связная компонента всего множества занятых узлов, по которым можно пройти. Стремясь к бесконечности, бесконечным становится и гигантский кластер.

Найдем перколяционный порог протекания на дереве Кейли. Стоя в центральном узле, пытаемся найти бесконечный путь, проходящий через занятые узлы. Сделав один шаг в направлении от этого узла, будет $z - 1$ направление, кроме пути, по которому прибыли. Каждый из этих $z - 1$ направлений приводит к новому соседу, который может оказаться занятым с вероятностью p . Следовательно, в среднем имеем $(z - 1)p$ занятых соседей. Ровно и такое же число возможностей для продолжения пути. Если величина $(z - 1)p$ меньше единицы, то среднее число различных путей, ведущих к бесконечности, уменьшается с каждым шагом в $(z - 1)p$ раз. Таким об-

разом, даже если все z соседей центрального узла тоже заняты и дают z различных возможностей для продолжения бесконечного пути, и даже если z очень велико, при $p < \frac{1}{(z-1)}$ вероятность обнаружения непрерывной цепочки занятых узлов стремится к нулю с увеличением химического расстояния от центрального узла. Поэтому для дерева Кейли с максимальным числом соседей z , получаем:

$$p_c = \frac{1}{z-1} \quad (2)$$

Таким образом, только при $p \geq p_c$, выражаемое формулой 2, появляется гигантский кластер, соответствующий описанию.

1.2 Модель поведения сети выше перколяционного порога

Как только появляется гигантский кластер, возникает интерес к еще одной величине, назовем ее g – вероятность того, что любой член социальной сети подключен к гигантскому кластеру. Вероятность g равна нулю, если гигантского кластера не существует, или другими словами доля занятых узлов меньше критической вероятности для образования гигантского скопления. Вероятность $g \in [0; 1]$. Можно предположить, что данная вероятность используется для нормализованной меры платить. Чем больше g , тем больше вероятность, что произвольный узел социальной сети будет принадлежать гигантскому кластеру. С другой стороны g можно рассматривать как силу сетевого эффекта. Поскольку гигантское скопление определяет интересы социальной сети, то принадлежность к этому скоплению гарантирует связь с остальными узлами с таким же интересом.

Рассмотрим силу сетевого эффекта для модели, изображенной на рисунке 1. Максимальное количество связей у каждого узла равняется трем. Пусть $h = 1 - g$ – вероятность того, что любой произвольный узел не принадлежит гигантскому кластеру через любой путь. Вероятность того, что две ветки этого узла не ведут к гигантскому кластеру равна h^2 . Тогда ph^2 – вероятность того, что узел занят, но ни одна из его двух ветвей не ведет к гигантскому скоплению. Этот сосед является пустым с вероятностью

$1 - p$. В этом случае ни одна ветвь не сможет его соединить с гигантским кластером. Таким образом

$$h = 1 - p + ph^2 \quad (3)$$

вероятность того, что эта фиксированная ветвь не приводит к гигантскому кластеру либо потому, что соединение разорвано уже у первого соседа, либо потому что где-то дальше отсутствует соединение. Уравнение 3 имеет 2 корня: $h = 1$ и $h = \frac{(1-p)}{p}$. Из уравнения 3, выделив член $p - g$, то есть вероятность того, что начало координат не принадлежит гигантскому кластеру ни через одну из его ветвей, получим: $p - g = ph^3$. Выражая отсюда g и подставляя корни уравнения 3, имеем:

$$g = \begin{cases} p(1 - h^3), & \text{если } p < p_c; \\ p(1 - \frac{(1-p)}{p}), & \text{если } p > p_c. \end{cases}$$

Аналогично можно рассмотреть случай, когда $z > 3$. Пусть произвольный узел не принадлежит гигантскому кластеру. Тогда вероятность того, что $z - 1$ ветка не ведут к гигантскому кластеру равна h^{z-1} . Вероятность того, что узел занят, но ни одна из его $z - 1$ ветвей не ведет к гигантскому кластеру равна ph^{z-1} . Таким образом

$$h = 1 - p + ph^{z-1} \quad (4)$$

Формула 4 – вероятность того, что узел не связан с гигантским кластером либо потому что соединение уже разорвано, либо потому что где-то дальше отсутствует соединение. Уравнение 4 имеет $z - 1$ корней. Один из них $h = 1$. Из остальных $z - 2$ корней всего лишь один реальный, оставшиеся – мнимые. Назовем реальный корень $-root_{real}(p)$. Из формулы 4, выделив член $p - g$, получим: $p - g = ph^z$. Подставляя корни уравнения 4 в

формулу, получим:

$$g = \begin{cases} p(1 - h^z), & \text{если } p < p_c; \\ p(1 - (\text{root}_{real}(p))^z), & \text{если } p > p_c. \end{cases} \quad (5)$$

1.3 Модель поведения сети ниже критической вероятности

В этой главе рассматривается поведение сети при показателях $p < p_c$. В таком случае нет гигантского скопления узлов. Мощность перколяционного кластера – отношение узлов, принадлежащих этому кластеру к общему количеству пользователей в сети. Мощность бесконечного кластера обозначается P_∞ . Поскольку в случае $p < p_c$ вероятность протекания равна нулю, то $P_\infty = 0$.

Введем функцию распределения кластеров n_s по размерам или по числу составляющих элементов.

$$n_s(p) = \frac{N_s}{N} \quad (6)$$

где N_s – среднее число кластеров размера s , а N – полное число узлов рассматриваемой сети. Величина sn_s показывает число узлов, принадлежащих кластеру размера s . Очевидно, что

$$\sum_{s=1}^{s=\infty} sn_s(p) = p$$

Другими словами, вероятность того, что узел принадлежит хоть какому-нибудь кластеру совпадает с вероятностью того, что он просто занят.

Значение g – всего лишь мера связности любого члена социальной сети для остальных участников. Ниже критической вероятности нет гигантского скопления, но можно рассмотреть эту ситуацию с точки зрения кластеров с некоторым размером. Чем больше будут размеры кластеров, тем больше будет привлекательность сети, из-за совместного вклада многих маленьких кластеров. Для образования конечного кластера, необходи-

мо, чтобы с двух сторон были не занятые узлы. Тогда вероятность образования кластера размером s будет равна $n_s = p^s(1-p)^2$. Другими словами, эта величина называется плотность s -кластеров и равна числу кластеров отнесенное к общему числу узлов в сети. При $p < 1$ и стремлении s к бесконечности число конечных кластеров уменьшается до нуля.

Теперь можно вычислить средний размер конечного кластера. Заметим, что $\sum_s = sn_s(p)$ – вероятность того, что узел принадлежит какому-либо кластеру, а sn_s – вероятность того, что узел принадлежит s -кластеру. Таким образом,

$$\omega_s = \frac{sn_s}{\sum_s sn_s}$$

вероятность того, что кластер содержит s узлов. Тогда средний размер S кластера вычисляется по формуле:

$$S = \sum_{s=1}^{\infty} \omega_s s = \sum_{s=1}^{\infty} \frac{n_s s^2}{\sum_{s=1}^{\infty} sn_s}$$

Средний размер кластера S обеспечивает надежную меру связности, когда нет гигантского кластера, потому что он измеряет количество контактов, которые имеет узел в социальной сети, непосредственно перед изменением фазы, до появления гигантского скопления. Рассмотрим среднее число групп, содержащих по s членов, перед появлением гигантского скопления. Это число определяется по периметру кластера. Кластер размера s будет окружен $t = (z-2)s + 2$ членами. Поэтому можно вывести формулу:

$$n_{s1}(p) = k_s p^s (1-p)^{2+(z-2)s}$$

где k_s – среднее количество конфигураций для кластера размером s . Однако не существует вывода k_s в открытом виде, поэтому следует изучить отношение $\frac{n_{s1}(p)}{n_{s1}(p_c)}$, где $n_s(p_c)$ – это константа, которая обеспечивает понимание поведения кластеров относительно точки, в которой появляется гигантский кластер.

$$\frac{n_{s1}(p)}{n_{s1}(p_c)} = \left[\frac{1-p}{1-p_c} \right]^2 \left[\left(\frac{p}{p_c} \right) \frac{1-p}{1-p_c} \right]^s$$

Глава 2. Численное моделирование

В данной главе произведены расчеты математической модели и представлены полученные результаты. По статистическим данным [3] около 60 % населения страны ежедневно проявляют какую-либо активность. Одной из самых популярных социальных сетей в России является сеть "ВКонтакте". Как видно из графика 2 основная часть пользователей – молодежь в возрасте от 12 до 24 лет. На гистограмме под процентным соотношением понимается % пользователей, использующих ресурс каждый день от всех участников социальной сети.

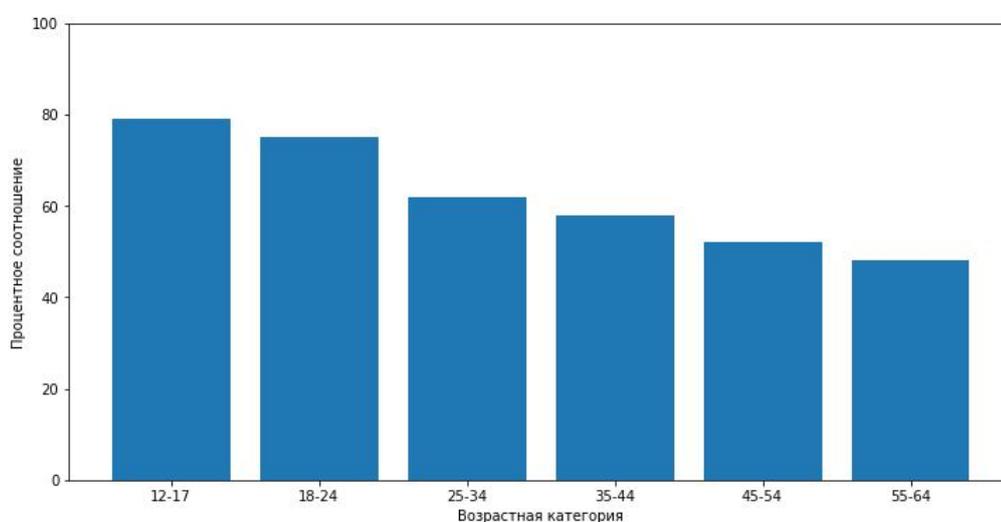


Рис. 2: Зависимость активности пользователя социальной сети от возраста.

Для расчета всех показателей в этом исследовании использовалась самая активная часть пользователей, а именно люди в возрасте от 12 до 24 лет. В качестве товара, для которого рассчитывается критическая масса, выступает научно-популярная книга. Все участники социальной сети представляются в виде узлов. Рассматривается локальная область социальной сети с глубиной дерева равной 2. По статистическим исследованиям человек имеет определенное ограничение на количество связей, которые может поддерживать. В социологии это выражается в виде числа Данбара [10], которое лежит в интервале от 100 до 230 связей. Для определенности в этой работе использовалось среднее арифметическое этих двух чисел равное 165. Таким образом, в данной модели каждый член социальной сети

имеет 165 друзей. Для вычисления общего числа пользователей в локальной области сети использовалась формула 1, где в качестве показателя r выступала максимальная глубина дерева равная 2, а в качестве z использовалось количество друзей каждого пользователя равное 165. Эти данные занесены в таблицу 1. Далее для вычисления критической вероятности применялась формула 2. В модели каждый пользователь имеет 165 друзей. Критическая вероятность равняется $\frac{1}{164}$. Умножая это число на количество пользователей, получается критическая масса, с которой начинает распространяться информация по локальной области.

Таблица 1: Основные показатели

Глубина дерева	Локальное количество участников	Критическая вероятность	Критическая масса
2	27556 человек	$\frac{1}{164}$	168 человек

После вычисления основных показателей рассматривалось два случая:

1. Процент пользователей, читающих одну и ту же научно-популярную книгу в локальной области сети, больше критической вероятности.

В качестве количественной характеристики процент читающих пользователей равен $\frac{1}{100}$. Умножая данный показатель на общее количество пользователей, получилось количество человек, входящих в перколяционный кластер. Такой кластер существует, поскольку рассматривался случай выше критической вероятности. Для вычисления вероятности того, что любой узел из локальной области будет включен в перколяционный кластер была использована формула 4 и формула 5. Все показатели занесены в таблицу 2.

Таблица 2: Основные показатели для первого случая

Процент пользователей	Количество участников	Сила сетевого эффекта g
$\frac{1}{100}$	275 человек	0.004

2. Процент пользователей, читающих одну и ту же научно-популярную книгу в локальной области сети, меньше критической вероятности. Рассматривается случай, близкий к переломному моменту. Этот момент наступает, когда процент пользователей близок к критической вероятности. В качестве количественной характеристики процент читающих пользователей равен $\frac{1}{170}$. Умножая данный показатель на общее количество пользователей, получилось количество человек, читающих книгу. Большого скопления не будет, поскольку рассматривается случай меньше критической вероятности. Все данные занесены в таблицу 3.

Таблица 3: Основные показатели для второго случая

Процент пользователей	Количество участников
$\frac{1}{170}$	162 человека

Для разбиения читающих пользователей на кластеры различного размера была использована формула 6. Под размером кластера понимается суммарное число пользователей, объединенных в некоторой области. Поскольку по условию число читающих пользователей равно 162, то размер кластеров, умноженный на их количество, должен давать суммарное число пользователей. Все данные занесены в таблицу 4.

Таблица 4: Разбиение участников на кластеры в зависимости от размера

Размер кластера	Количество кластеров
1	146
2	8

В рамках работы был рассмотрен численный пример, где известно количество участников и их взаимосвязи. Для исследования поведения графиков ключевых показателей предположим, что значения являются параметрическими. На рисунке 3 представлен график зависимости силы сетевого эффекта от процентного соотношения пользователей, включенных в сеть с значениями, когда максимальное количество связей равно 2, 3, 4. Хорошо видно, что, чем больше вероятность включенности пользователей в сеть, тем сильнее будет сетевой эффект. С увеличением максимального числа связей у каждого пользователя, то есть показателя z , наклон графика становится меньше. Это объясняется тем, что в больших сетях фазовый переход происходит плавнее, чем в сетях с низким показателем числа z .

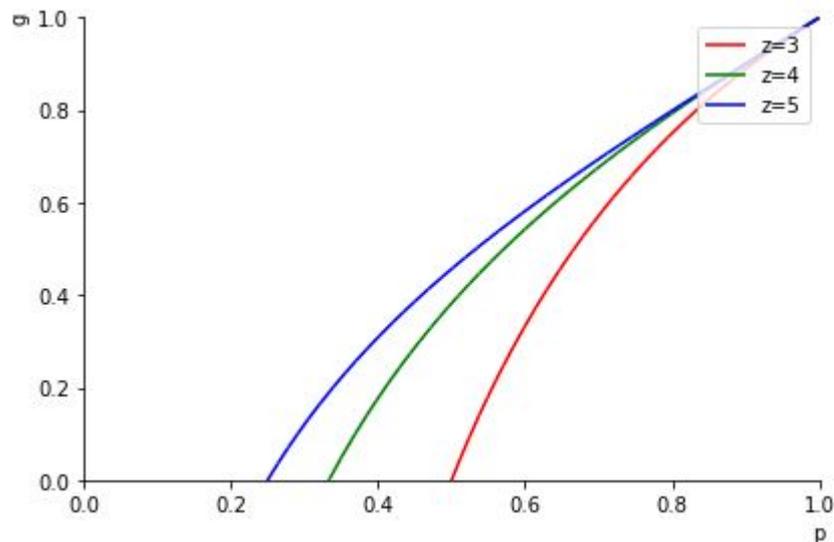


Рис. 3: Зависимость силы сетевого эффекта от пороговой вероятности

Данный случай рассматривается тогда, когда процентное содержание включенных пользователей выше значения критической вероятности, кото-

рая зависит от максимального количества друзей у пользователя. Именно после перехода через критическую вероятность образовывается гигантский кластер. Для понимания того, как будет вести себя график при больших значениях z , было смоделировано трехмерное изображение. Для этого построения использовалась встроенная библиотека python matplotlib и SymPy. Код программы представлен в Приложении 1. Мощность компьютера позволила построить график 4, когда количество связей у каждого пользователя от двух до шести. Даже при таких маленьких значениях параметра z наблюдается тенденция уменьшения наклона графика. При значениях, стремящихся к ∞ , сила сетевого эффекта будет достигать своего максимума тогда, когда значение вероятности равно 1.

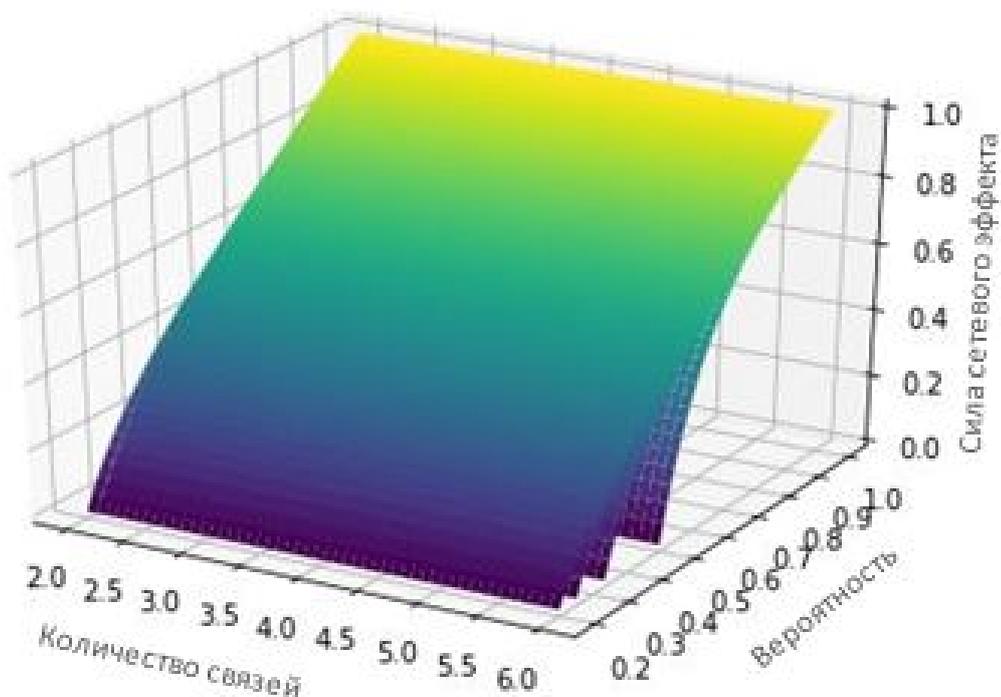


Рис. 4: Зависимость силы сетевого эффекта от количества связей и вероятности включения пользователя в социальную сеть

Другой случай рассмотрен, когда вероятность включения пользователя в сеть ниже критической вероятности. Гигантского перколяционного

кластера не будет, но в этом случае появляются кластеры конечных размеров, изолированные друг от друга. На рисунке 5 построен график образования кластера определенного размера в зависимости от вероятности включения пользователя. Данный график рассмотрен для случая, когда

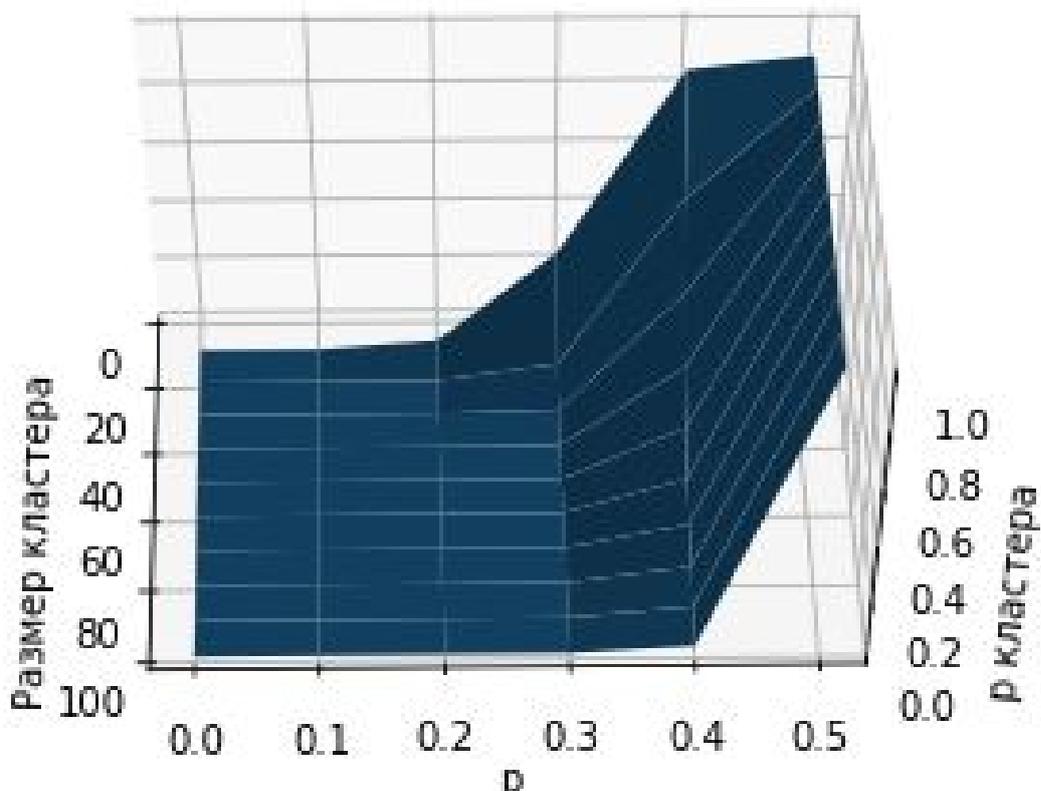


Рис. 5: Вероятность появления кластера определенного размера

критическая вероятность равна $\frac{1}{2}$. Программный код представлен в Приложении 2. Из формулы 2 можно сделать вывод, что максимальное число связей у каждого пользователя равно 3. Размеры кластера лежат в промежутке от 2 до 100. Из графика видно, что для кластеров больших размеров (от 40 до 100) вероятность их появления равна 1, когда сеть имеет критический переход. Действительно, при приближении к критической вероятности узлы будут стараться объединиться в единый перколяционный кластер.

Для размера кластеров от 2 до 20 вероятность их появления будет больше при малых вероятностях включения пользователя в сеть. При стремлении к критической вероятности, вероятность появления таких кла-

стеров в сети будет максимальной. При малых вероятностях включения пользователя в сеть образовывается много кластеров маленького размера, изолированные друг от друга. Чем больше становится включенных пользователей, тем быстрее маленькие кластеры объединяются в кластеры большего размера. На рисунке 4 хорошо видно, что при вероятности близкой

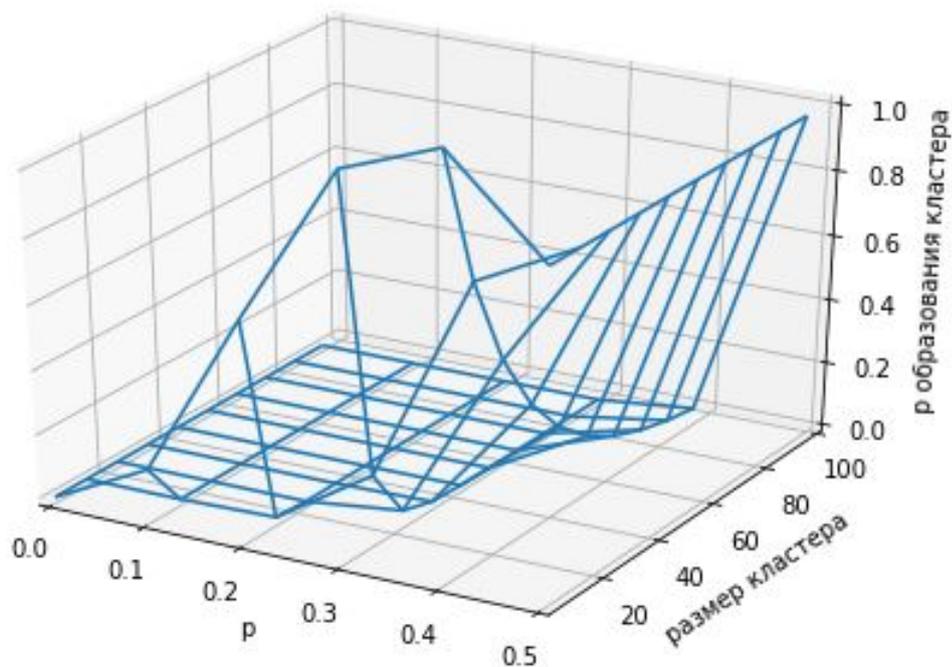


Рис. 6: Вероятность появления кластера определенного размера

к критической в сети будут существовать кластеры любого размера. Данный график был построен в python с использованием библиотек matplotlib и SymPy. Код программы представлен в Приложении 3.

Выводы

Основные результаты, полученные в процессе написания данной работы:

- Изучена и применена теория перколяции для социальной сети
- Выведена формула поиска критической массы пользователей при образовании гигантского кластера
- Выведены формулы поиска ключевых показателей для двух случаев:
 1. Процент занятых узлов во всей сети выше критической вероятности
 2. Процент занятых узлов во всей сети ниже критической вероятности
- Произведен численный расчет всех показателей при определенном количестве друзей
- Построены и проанализированы графики зависимости основных показателей

Заключение

Роль социальных сетей в жизни трудно переоценить. Люди различных возрастов и категорий общаются и обмениваются информацией в сети. Социальные сети давно стали инструментом распространения информации. Воздействуя через них, люди поддаются общественному мнению и подвергают других пользователей на определенные действия.

Однако, в этом исследовании были сделаны предположения, которые расходятся с реальной структурой социальных сетей.

1. Бесконечное число пользователей социальной сети. Это предположение верно для больших сетей, но, если рассматривать сеть, с конечным числом пользователей, то, вероятно, фазовый переход будет постепенным.
2. Фиксированное число связей у каждого пользователя. Это предположение менее реалистично, поскольку число социальных связей, которые может иметь любой человек, может быть переменным. Теорию перколяции с решеткой Бете достаточно трудно использовать, учитывая этот факт. Вероятно, для дальнейших исследований необходимо развивать данную модель, либо использовать другую.
3. В исследуемой модели нет циклов. Данное предположение расходится с реальной структурой социальной сети. Для дальнейших исследований необходимо расширить рассматриваемую модель, чтобы обеспечить эффективное управление социальной сетью.

Список литературы

- [1] Воронкин А.С. Социальные сети: эволюция, структура, анализ. // Образовательные технологии и общество – 2014. – №1. – С. 650-675.
- [2] Иудин Д.И. Фракталы: от простого к сложному / Д.И. Иудин, Е.В. Копосов. – Нижегородский государственный архитектурно-строительный университет – Нижний Новгород: ННГАСУ, 2012. – 200 с.
- [3] Курносова Е. Социальные сети в цифрах [Электронный ресурс] – Режим доступа: <https://mediascope.net/upload/iblock/f97/18.04.2019/Mediascope.pdf> (дата обращения: 26.04.2020)
- [4] Соколов С.С., Некрасова А.А. Исследование возможности применения теории перколяции для управления потоками данных в информационных сетях на транспорте // Вестник Государственного университета морского и речного флота имени адмирала С. О. Макарова. – 2015. – №4. – С. 192-198.
- [5] Bailey N. The Mathematical Theory of Infectious Diseases
- [6] Boccaletti S., Hwang D.-U., Latora V. Growing hierarchical scale-free networks by means of nonhierarchical processes // International Journal of Bifurcation and Chaos. – 2007. – Vol. 17, №7 – P. 2447–2452. and Its Applications. – New York: Hafner Press, 1975.
- [7] Broadbent S.R., Hammersley J.M. Percolation processes I. Crystals and mazes // Mathematical Proceedings of the Cambridge Philosophical Society. – 1957.– Vol. 53, №3. – P. 629-641.
- [8] Chwe M. Structure and Strategy in Collective Action // American Journal of Sociology. – 1999. – Vol. 105, №1. – P. 128–157.
- [9] Du N., Wang H., Faloutsos C. Analysis of large multi-modal social networks: patterns and a generator // In Proceedings of the 2010 European Conference

on Machine learning and knowledge discovery in databases: Part I, ECML PKDD'10 – P. 393– 408. Springer-Verlag.

- [10] Dunbar R., Neocortex size as a constraint on group size in primates // Journal of Human Evolution – 1992. – №2. – P. 469–493.
- [11] Granovetter M. Threshold Models of Collective Behavior // The American Journal of Sociology. – 1987. – Vol. 83, №6. – P. 1420–1443.
- [12] Kempe D., Kleinberg J., Tardos E. Influential Nodes in a Diffusion Model for Social Networks. Automata, Languages and Programming In Automata // Languages and Programming. – 2005. — Vol. 3580. – P. 1127–1138.
- [13] Lesko S.A., Zhukov D.O. Percolation Models of Information Dissemination in Social Networks // 2015 IEEE International Conference on Smart City – 2015. – P. 213-216.
- [14] Metcalfe B. There oughta be a law. New York Times, July 15, 1996.
- [15] Pastor-Satorras R., Vespignani A. Epidemic spreading in finite scale-free networks // Physical Review Letters – 2001. – Vol. 65 – P. 1081–1084.
- [16] Odlyzko, A., Tilly, B. A refutation of Metcalfe's Law and a better estimate for the value of networks and network interconnections // Digital Technology Center – 2005 – P. 148-159.
- [17] Oliver P., Marwell G., Teixeira R. A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action // The American Journal of Sociology. – 1985. – Vol. 91, № 3. – P. 522–556.
- [18] Reed D. P. The law of the pack // Harvard Business Review – 2001 – P. 23–24.
- [19] Schelling T. A process of residential segregation: Neighborhood tipping, in Racial Discrimination // Racial discrimination in economic life – 1972. — P. 157–184.

- [20] Schelling T. Dynamic models of segregation // Journal of Mathematical Sociology. – 1971. – Vol. 1. – P. 143–186.
- [21] Schelling T. Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities // The Journal of Conflict Resolution. — 1973. — Vol. 17, №3. – P. 381–428.
- [22] Schelling T. Micromotives and Macrobehavior. — New York : WW Norton Co, 1978. – 272 p.
- [23] Zhukov D.O., Alyoshkin A.S. Information search model based on the use of percolation properties of semantic networks of texts // Advances in Economics, Business and Management Research. – 2019. – Vol. 81 – P. 205-209.

Приложение 1

```
from sympy import *
from sympy.plotting import plot
from sympy.plotting import plot3d
p = Symbol('p')
h = Symbol('h')
b = []
for z in range(2,6):
    a = solve(p*h**(z)- p + 1 - h, h)
    for i in range(len(a)):
        if (a[i].subs(p,0.9).is_real)\
        and (a[i].subs(p,0.9)>0)\
        and (a[i].subs(p,0.9) != 1):
            b.append(a[i])
p1 = plot(p*(1-(b[0])**2),(p,0,1),show=False,\
        line_color='r',\
        xlabel='p', ylabel='g',\
        axis_center='auto',\
        legend=True)
p2 = plot(p*(1-(b[1])**3),(p,0,1),\
        show=False,line_color='g',\
        legend=True)
p3 = plot(p*(1-(b[2])**4),(p,0,1),\
        show=False,line_color='b',\
        legend=True)
p1[0].line_color='red'
p1[0].label='z=3'
p2[0].line_color='green'
p2[0].label='z=4'
p3[0].line_color='blue'
p3[0].label='z=5'
p1.extend(p2)
```

```

p1.extend(p3)
p1.show()
p = symbols('p')
z = symbols('z')
plot3d((p*(1-(b[0])**3),(h,2,6),(p,1/2,1)),\
        (p*(1-(b[1])**4),(h,2,6),(p,1/3,1)),\
        (p*(1-(b[2])**5),(h,2,6),(p,1/4,1)),\
        (p*(1-(b[3])**6),(h,2,6),(p,1/5,1)))

```

Приложение 2

```

from sympy import *
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
import numpy as np

from mpl_toolkits.mplot3d import axes3d
ax = axes3d.Axes3D(plt.figure())

i = np.arange(0,0.51,0.1)
j = np.arange(3,100,10)
X, Y = np.meshgrid(i, j)
Z = (((1-X)/(1-1/2))**2) * ((1 - 4*(X-1/2)**2)**Y)
ax.plot_wireframe(X,Y,Z, rstride=1, cstride=1)
ax.set_xlabel('p')
ax.set_ylabel('size')
ax.set_zlabel(' _size ')
ax.set_zlim(0, 1)
ax.set_xlim(0, 0.5)
ax.set_ylim(3,100)
plt.show()

```

Приложение 3

```
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
from matplotlib import cm
from matplotlib.ticker import LinearLocator
from matplotlib.ticker import FormatStrFormatter
import numpy as np

fig = plt.figure()
ax = fig.gca(projection='3d')

X = np.arange(0, 0.51, 0.1)
Y = np.arange(3, 100, 10)
X, Y = np.meshgrid(X, Y)
Z = (((1-X)/(1-1/2))**2) * ((1 - 4*(X-1/2)**2)**Y)

surf = ax.plot_surface(X, Y, Z, cmap=cm.coolwarm,
                       linewidth=0.5, antialiased=False)

ax.set_zlim(0, 1, 0.1)
ax.zaxis.set_major_locator(LinearLocator(10))
ax.zaxis.set_major_formatter(FormatStrFormatter('%0.02f'))

fig.colorbar(surf, shrink=0.4, aspect=3)

plt.show()
```