

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 ИЗМЕНЧИВОСТИ ЛОКАЛЬНЫХ  
ХАРАКТЕРИСТИК В КОНЕЧНЫХ СОЦИАЛЬНЫХ ГРАФАХ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 5 курса 551 группы  
направления 09.03.04 — Программная инженерия  
факультета КНиИТ  
Февралева Алексея Валерьевича

Научный руководитель  
зав. каф., к. ф.-м. н.

\_\_\_\_\_

С. В. Миронов

Заведующий кафедрой  
к. ф.-м. н.

\_\_\_\_\_

С. В. Миронов

Саратов 2023

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	3
1 Описание теоретической части .....	5
2 Описание практической части .....	6
ЗАКЛЮЧЕНИЕ .....	11

## ВВЕДЕНИЕ

Предметом исследования работы на базе которой написан данный реферат является изучение поведения статистических характеристик в реальных социальных графах, тенденций их изменений и выявление замеченных закономерностей. Также в работе проводится сравнение реальных социальных графов и графов построенных с помощью алгоритмов построения случайных графов. В частности в работе рассмотрены такие модели построения случайных графов, как Барабаши — Альберт, Боллобаша — Риордана и модель копирования. Актуальность выбранной темы обусловлена постоянным ростом социальных связей между людьми, что приводит к закономерному росту социальных графов, которые является математическим отображением этих связей. В случае данной работы мы рассматриваем эти связи в рамках сети Интернет, а именно в рамках некоторых популярных социальных платформ. Сложные сети представляют собой графы состоящие из колоссального количества узлов, связанных между собой в зависимости от взаимодействия социальных единиц, которые эти узлы представляют. В связи с этим важно исследовать средства, которые помогли бы спрогнозировать те или иные явления и взаимодействия в таких огромных сетях. Одним из таких средств являются модели случайных графов, которые довольно близко могут имитировать рост социальных платформ. Параметром на котором делается упор в работе является изучение, так называемого парадокса дружбы. Парадокс дружбы является важной характеристикой, которая присуща всем реальным социальным графам, подробнее о его сути описано в теоретической части работы. Вершины социальных графов представляют собой какую-либо социальную единицу, это может быть отдельно взятый человек, платформа, группа людей или сайт. Связями между этими вершинами будут различные социальные взаимодействия между соответствующими социальными единицами. В рамках рассмотренных платформ, это может быть как дружба в рамках одной платформы, так и например цитирование одного автора другого в рамках другой. Основной целью работы является изучение социальных графов, их параметров, на примере такого феномена, как парадокс дружбы и возможности воссоздания приближенных к реальным социальным графам случайных графов, опираясь на полученные знания возникновения исследуемого парадокса. Также одной из целей работы является поднять интерес людей, к такой важной теме как развитие глобальных

сетей, и заинтересовать читателей в исследовании и открытии новых средств моделирования таких сложных сетей. Основные задачи, которые были поставлены на выпускную квалификационную работу:

- поиск источников по выбранной теме;
- изучение найденных материалов;
- поиск наборов данных для практической части работы;
- создание алгоритмов для обработки наборов данных;
- написание программной реализации моделей построения случайных графов;
- построение графов с использованием написанных алгоритмов;
- изучение поведения локальных характеристик на полученных графиках;
- формулирование выводов по выполненной работе на основе полученных результатов.

## 1 Описание теоретической части

В теоретической части работы, которая носит общее название «Модели построения случайных графов» представлена базовая информация для понимания основ на которых состоит проводимое исследование. В данном разделе описана начальная информация по теории графов, случайных графов, исследуемых параметрах и, принципах и алгоритмах случайных графов применяемых в практической части работы.

В работе рассмотрены не простые графы, а социальные графы. Социальный граф — это граф, вершины которого представляют собой социальные единицы, такие как профиль человека, сообщества, медиа-контент и так далее, а ребра в таких графах характеризуют социальное взаимодействие между ними. Если говорить более конкретно в работе рассматриваются неявные социальные графы, это такие графы, где явного указания взаимодействия между вершинами нет, и в зависимости от набора данных, мы можем понимать под «дружбой» между вершинами совершенно разные взаимодействия. В социальных сетях такое взаимодействие может быть как дружба, так и взаимное упоминание, а например на платформах по хранению научных работ, цитирование одним автором работ другого. Также в работе приведено описание трех алгоритмов построения случайных графов, не считая описания некоторых их реализаций, по типу статической или динамической в рамках одной модели Боллобаши-Риордана. В основе всех моделей построения случайных графов стоит принцип, суть которого, заключается в том что некая вершина, соединяется с другой вершиной с какой-то вероятностью, либо статической, заданной заранее, либо динамической, зависящей от какого-то стороннего параметра или вычисления. Большинство моделей используют так называемое предпочтительное присоединение, которое подразумевает, что соединение с вершиной зависит от величины какой-то ее характеристики, в большинстве случаев под этой характеристикой понимается степень вершины. В случае же с моделями со статической вероятностью, в основном подбираются такие параметры вероятности, которые выдают результат, наиболее приближенный к результатам реальных социальных графов. По исследуемым параметрам в работе приведена основная информация про парадокс дружбы, такой параметр как индекс дружбы, дисперсию, коэффициент вариации и математическое ожидание.

## 2 Описание практической части

Практическая часть работы под названием «Изучение локальных характеристик» начинается с описание программных реализаций, применяемых для поиска искомых характеристик. Для программной реализации всей практической части работы использовался язык Python ver. 3.11.0. Сперва в работе представлены программные реализации двух алгоритмов, которые использовались для получения и структурирования наборов данных взятых с Интернет источников. В рамках работы были взяты пять наборов данных, с пяти известных социальных платформ: Twitter, Facebook, Vkontakte, dblp, Github. Наборы брались из общедоступных открытых источников, таких как Google Dataset Search, Kaggle и dblp. Для того что бы обработать взятые наборы и перевести их в форму графа, были написаны два алгоритма, один из которых взаимодействовал с наборами, которые были представлены в виде таблиц с данными, а другой для наборов представленных в виде словарей. В рамках этих алгоритмов были задействованы такие библиотеки языка Python, как NetworkX и Pandas. После создания графов, требуется получить локальные характеристики, которые нужны нам в рамках выбранного исследования, а именно нас интересуют: распределение степеней в графе, количество социальных единиц в рамках каждой степенной группы, средняя степень соседей вершины и тот же параметр у ее соседей, индекс дружбы у вершины, у всего графа в целом, и средний индекс дружбы в каждой отдельно взятой степенной группы, дисперсия в рамках каждой из степенной групп и коэффициент вариации в тех же степенных группах. В основном все способы для нахождения данных параметров представлены в теоретической части, в разделе «Исследуемые параметры». Для визуального отображения полученных данных и результатов расчетов были построены графики с помощью написанной функции и средств библиотеки Plotly. Следующим пунктом в практической части работы идет раздел «Построение графов на основе моделей случайных графов». В данном разделе на основе описанных в теоретической части работы алгоритмах, написана программная реализация трех моделей случайных графов. Конкретно собственная программная реализация была написана для алгоритма Боллобаши-Риордана и модели копирования. Для реализации модели Барабаши-Альберт, использовалась библиотека NetworkX и функция входящая в нее *barabasi\_albert\_graph*. После прохождения через алгоритм, на выходе получается граф, на подобие

того, который был получен из реальных наборов данных, после чего все также производится подсчет параметров, и строятся графики для визуализации. После описания программных реализаций идет раздел «Результаты исследования» в котором демонстрируются полученные графики параметров и содержатся их описание и разбор. Что в итоге складывается в определенные выводы по каждой из рассмотренных платформ или моделей.

В социальном графе сети Twitter преобладают вершины с низкой степенью, разброс степеней соседей которых достаточно велик, члены одной и той же группы степеней, могут иметь значительные различия в их коэффициентах дружбы. Также на графиках можно заметить определенное количество хаотичных связей. Пользователи в основном, не закрыты в рамках малых социальных групп и присутствует свойство диассортативности. Большинство друзей пользователей, имеют довольно большие степени соседей, в следствии чего значение индекса дружбы в основном превышает 1, и общий процент возникновения парадокса дружбы близок к результатам известных исследований и равен 93,22%.

Под «дружбой» в рамках социальных графов, может пониматься практически любое социальное взаимодействие между пользователями, в случае с dblp, под этим взаимодействием понималось цитирование одним пользователем работы другого. Как и в графиках Twitter наблюдается численное превосходство вершин с малой степенью, которые занимают большинство выборки. График степеней соседей разительно отличается от такого у Twitter, где большую часть графика средняя степень друзей превосходила степень вершины, здесь же только несколько первых групп вершин, соответствуют сказанному, после чего практически до конца сохраняется обратная зависимость. Можно утверждать, что данный набор данных подвержен свойству ассортативности. Все графики для данной платформы, так или иначе изменяются согласно степенному закону. Максимальные значения индекса дружбы в данном наборе данных, в разы меньше, чем те которые мы наблюдали на графиках Twitter. Это можно списать на еще один фактор, влияющий на индекс, а конкретно на специфику платформы. Графики дисперсии и коэффициента вариации, показывают схожее с аналогичными графиками Twitter поведение, с единственной разницей в более низких значениях дисперсии.

Следующей рассмотренной в практической части платформой был Github -

крупнейший веб-сервис для хостинга IT-проектов и их совместной разработки. В данном наборе данных, под «дружбой» понималось обоюдное отслеживание пользователей друг друга, и совместная работа, хотя бы над одним проектом (коммиты в рамках одного проекта). Снова наблюдается большинство за вершинами с малой степенью. Данная платформа отличается очень высоким показателем средней степени соседей, что может заставить нас предположить, что это также связано со спецификой платформы, а степень растет к примеру за счет крупных компаний, с открытым доступом к своим репозиториям, либо репозиториям известного и часто применяемого алгоритма. Имеется довольно большая дисперсия среди групп с низкой степенью. Изменения на графиках происходят по правилам степенного закона.

Следующие две рассмотренные платформы Facebook и VKontakte являются полноценными социальными сетями, что является их отличительной особенностью. Благодаря чему графики данных платформ показали совместные сходства друг с другом, но при этом существенные отличия от других рассмотренных платформ. Под «дружбой» в рамках графов этих платформ, будем понимать прямое значение этого слова. Первое что можно выделить, это распространенность «хаотичных значений». Количественно также наблюдалось преобладание вершин с низкой степенью, как и в прошлых рассмотренных графах. График степеней соседей довольно сильно отличается от прошлых. Оба графа показывают наличие свойства ассортативности, намного более ярко выраженного, в отличие от графика dblp. Как правило в социальных сетях мы предпочитаем дружить с людьми, близкого к нашему круга общения, из за чего разница между степенями нас, и наших друзей не так и велика. Но несмотря на такую разницу в степенях, индекс дружбы по прежнему в основном большинстве положительный, что говорят высокие значения вероятности возникновения парадокса в данных графах. Графики дисперсии и коэффициента вариации показывают наиболее низкие параметры из рассмотренных платформ, что символизирует о большей однородности индекса дружбы в рамках одной степенной группы.

Что касается исследования графов построенных на основе моделей случайных, все три алгоритма показали довольно похожие результаты, наиболее приближенные к результатам графов Twitter и Github. Для каждого из используемых алгоритмов был выбран оптимальный набор начальных параметров,

для получения наиболее приближенного к реальным сетям результатам. На графиках степенного распределения все также наблюдается количественное доминирование вершин с низкой степенью, в частности со стартовой заданной степенью. Графики степеней соседей и индекса дружбы, довольно сильно похожи на таковы у ранее указанных платформ Twitter и Github. Наблюдается свойство диассортативности. Другими словами, вершины с малой степенью в основном связываются с вершинами с намного большей степенью, в основном это происходит благодаря предпочтительному присоединению, используемому в алгоритмах. Из за чего вероятность возникновения парадокса дружбы в данных графах также высока, и примерно равняется 90% для всех трех алгоритмов. Дисперсия и коэффициент вариации частично варьируется от модели к модели, но в основном показывает схожие значения с рассмотренными реальными сетями.

Что мы имеем в итоге: парадокс дружбы отлично проявляет себя не только в безразмерных графах, но его также отчетливо можно наблюдать и в не самых больших конечных социальных графах, показывающих лишь срез от глобальных графов сетевых ресурсов. Во многих источниках и исследованиях на тему изучения парадокса дружбы можно найти информацию, что парадокс в социальных графах возникает для вершин в промежутке от 82% до 98% от общего числа элементов. Эта информация хорошо сочетается с полученными в ходе нашего исследования результатами, все проведенные эксперименты дали результаты возникновения парадокса в этих границах. Также все графы показали четкую концентрацию большинства вершин в группах с малой степенью. Почти все исследуемые социальные графы, показывали признаки изменения согласно степенному закону, но с оговорками, в частности графы социальных сетей, больше других подвержены хаотичному разбросу значений. На графиках глобальных социальных сетей, таких как Vkontakte и Facebook, а также частично в dblp, можно заметить возникновения свойства ассортативности, которое показывает что узлы чаще связываются с узлами похожими на них, в частности это касается значения степени вершины. Графы других же платформ показывают диассортативность. Дисперсия и коэффициент вариации в графах построенных на основе реальных платформ имели зависимость от значения индекса дружбы, и убывали по похожим траекториям, но стабилизировались ближе к высоким значениям степеней. В графах же построенных с помощью

моделей случайных графов, имеется похожая зависимость, но в них данные параметры медленно стремились к 0. Основной сложностью и фактором ограниченности данного исследования можно выделить различия в рассматриваемых наборах данных, по количеству исследуемых вершин и связей, что проистекает от редкого и ограниченного сбора статистических данных таких социальных платформ.

## ЗАКЛЮЧЕНИЕ

В заключении работы подводится итог о проделанной работе и общих выводах полученных от анализа построенных в практической части работы графиках. Численность социальных связей в мире растет, соизмеримо с ростом числа населения. Не малую долю из эти связей в нашу эпоху составляют различные взаимодействия в рамках сети Интернет, которая представляет собой громадную практически безразмерную, постоянно растущую сеть. В той же сети Интернет, существует бесконечное множество более маленьких сетей, но которые в тоже время все равно огромны. И их количество, и рост отдельно взятых также постоянно увеличивается. Сети развились до таких размеров, когда их исследование стало иметь весомые трудности, сюда можно отнести как огромные массивы данных, которые становится сложнее собирать, хранить и обрабатывать, так и сложность доступа к таким данным для стороннего исследователя. Одной из целей работы было предложить читателю альтернативный способ изучения глобальных сетей, с использование вместо настоящих сетей моделей случайных графов. Показать их актуальность и приближенность к реальным сетевым графам. В ходе исследования были получены выводы о том, что модели случайных графов довольно близки к имитации реальных сетей, в основном эта заслуга того, что социальные графы, которые были основой исследования изменяются в рамках степенного закона. Да в них встречаются хаотические значения, которые сложно получить в рамках случайных моделей, но в их сути графы довольно близки. Парадокс дружбы выступавший основным исследуемым феноменом является одним из тех важных кирпичиков, которые требуется учитывать, что бы добиться идеальной модели, которая смогла бы воспроизвести реальную сеть. Все графы построенные на основе реальных наборов данных и с помощью моделей случайных графов показали высокие значения вероятности возникновения парадокса, что соответствует ранее проводимыми исследованиями на данную тематику. На полученных графиках хорошо прослеживается тенденция на количественное преобладание вершин с малой степенью, в то время как вершины «хабы», имеющие высокие значения степени, составляют лишь малую часть от всех вершин. Из особенностей было выявлено то, что графики социальных сетей, таких как Facebook и Vkontakte, предрасположены к меньшей разнице между степенью вершины и степенью соседей, что может указывать на то, что в рамках этих

платформ пользователи общаются в кругах людей с близкой к ним популярностью, вершины с большой степенью, скорее всего будут взаимодействовать с подобными, а вершины с малой степенью предпочитают «дружить» с вершинами степень которых хоть и больше чем у них, но не настолько сильно, как например в графиках Twitter и Github. Графики Twitter и Github показали, что в них пользователи с малой степенью предпочитают «дружить» с намного более известными людьми чем они сами. Если учитывать специфику платформ, то в случае этих сервисов правильнее будет понимать, не дружбу как таковую, а скорее заинтересованность пользователя, с точки зрения опыта или мыслей. Графики dblp стоят особняком и в плане взаимодействия между групп степеней, выдают среднее значение между графиками двух предыдущих пар, немного перевешивая в сторону графиков социальных сетей. В случае же случайных графов, их результаты наиболее близко совпадают с результатами Twitter и Github. Это в основном вся информация полученная в ходе исследования, касающаяся взаимодействия между степенными группами. Остальные же графики, не имеют больших отличий и в основном все убывают с увлечением степени узла. Это касается и индекса дружбы, и дисперсии, и коэффициента вариации. В основном различия заключаются в максимальных и минимальных значениях, которые те или иные величины достигают. К чему можно подвести в конце, теория графов, а конкретно изучение социальных графов, обширная и важная тема, область применения которой просто колоссальна. В рамках работы было рассмотрены лишь три модели случайных графов, а таких моделей уже насчитывается десятки, так что всегда есть куда развернуться и что изучить. С ростом глобальных сетей потребность в изучении способов их исследования только растет, из за чего работа над данным направлением, совмещившем в себе аспекты как математики, так и социологии, очень актуальна в современном мире.