

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**МОДЕЛИРОВАНИЕ ДИНАМИКИ ХАРАКТЕРИСТИК УЗЛОВ В
СЛОЖНЫХ СЕТЯХ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 2 курса 273 группы
направления 02.04.03 — Математическое обеспечение и администрирование
информационных систем
факультета КНиИТ
Козырева Юрия Дмитриевича

Научный руководитель

зав. каф., к. ф.-м. н., доцент

С. В. Миронов

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2025

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Теоретические сведения	4
1.1 Случайные графы	4
1.2 Метрики	4
2 Модификации индекса дружбы	6
3 Экспериментальная проверка гипотезы	7
3.1 Проверка на моделях растущих сетей	7
3.2 Проверка на конфигурационных моделях	8
3.3 Проверка на реальных сетях	8
4 Анализ результатов	9
ЗАКЛЮЧЕНИЕ	14

ВВЕДЕНИЕ

В повседневной жизни для решения многих задач часто используются случайные графы. Случайные графы нашли практическое применение во всех областях, где нужно смоделировать сложные сети. Имеется большое число моделей случайных графов, отражающих разнообразные типы сложных сетей в различных областях. Случайные графы применяются при моделировании и анализе биологических и социальных систем, сетей, а также при решении многих задач класса NP.

Случайный граф — общий термин для обозначения вероятностного распределения графов. Их можно описать просто распределением вероятности или случайным процессом, создающим эти графы.

С математической точки зрения случайные графы необходимы для ответа на вопрос о свойствах типичных графов. Для этого используются модели Эрдёша—Реньи, Барабаши—Альберт, модель триадного замыкания, модель Бьянкони-Барабаши и другие. Все модели основываются на различных свойствах социальных сетей.

При практическом применении случайных графов для моделирования систем, содержащих миллионы вершин необходима возможность определять пригодность той или иной модели для рассматриваемой симуляции, не прибегая к необходимости построения модели в полную величину. Для этого необходима локальная метрика позволяющая сравнивать графы разных размеров.

Цель настоящей работы — предложить потенциальный вариант локальной масштабируемой метрики графа и проанализировать его поведение. Для достижения этой цели необходимо решить следующие задачи:

- проанализировать существующие масштабируемые метрики графов;
- формализовать локальную масштабируемую метрику графов;
- реализовать предложенную метрику;
- провести анализ поведения предложенной метрики в случайных и реальных графах.

1 Теоретические сведения

1.1 Случайные графы

Случайные графы широко применяются во многих сферах человеческой деятельности.

Некоторые протоколы передачи данных в компьютерных сетях основаны на различных свойствах случайных графов и в данный момент активно ведутся исследования по усовершенствованию этих алгоритмов. Например, Алессандро Больоло и Кристель Сирокки в своей статье «Topological network features determine convergence rate of distributed average algorithms» изучают топологию и различные метрики графов в компьютерных сетях и Gossip-алгоритмах для разработки более эффективной замены для Gossip-протоколов. Gossip — это группа протоколов в одноранговой компьютерной коммуникации, в которых распространение информации идёт способом, схожим с образом распространения эпидемий, и сводящимся к тому, что каждый или некоторые из узлов могут передавать обновляемые данные известным этому узлу соседям. Основными преимуществами данных алгоритмов являются их надёжность и отказоустойчивость, однако передача большого количества избыточных данных приводит к с лишком высокой нагрузке на оборудование.

1.2 Метрики

С популяризацией темы графов, графовой аналитики и их визуализации, возрастает необходимость измерять, сравнивать, выявлять зависимости и различные иерархии.

Для этого необходимы некие численные характеристики графов. Графовая метрика – это функция, которая возвращает действительное число, характеризующее граф согласно применяемым мерам.

Метрики на графах часто применяются в различных областях включая медицину.

Метрики делятся на:

- глобальные - принимают на вход данные о всех вершинах графа и описывают весь граф целиком;
- локальные - характеризуют некую окрестность заданной вершины.

В магистерской работе рассматриваются локальные метрики графа, в силу того что они позволяют более точно описать внутреннюю структуру графа и

позволяют ускорить процесс анализа, так как могут вычисляться параллельно.

К ним относятся: сумма степеней соседей, средняя степень соседей, ANND, neighborhood-based bridge node centrality tuple (NBNC), индекс дружбы и многие другие.

Однако предельное поведение распределения индекса дружбы по степеням вершин, при размерности графа стремящейся к бесконечности, зависит от γ и отличается для следующих двух случаев: $\gamma > 2$ и $1 < \gamma \leq 2$. Где параметр γ отражает коэффициент степенной функции $f(x) = cx^{-\gamma-1}$, в соответствии с которой распределены степени узлов сети. Если показатель γ больше 2, то для сети, полученной по конфигурационной модели, предельное распределение будет существовать. Однако, если показатель находится в диапазоне от 1 до 2, то распределение сходится к случайной величине.

Значения индекса дружбы и ANND масштабируются относительно размера сети t . Следовательно, эти метрики не являются подходящим показателем для сравнения сетей разного размера. Поэтому было бы полезно предложить меру для оценки парадокса дружбы, которая не зависит от размера сети.

Существует несколько подходов к построению масштабируемых метрик. Среди них можно выделить следующие использование химических расстояний, рассмотрение сетевых мотивов, создание модели глубокого обучения, использование ранговых метрик.

2 Модификации индекса дружбы

В данной работе предлагается использование аналогичного подхода для масштабирования и нормализации индекса дружбы, для формирования масштабируемой локальной метрики. ANND ($\Phi_n(k)$) можно рассматривать как глобальный аналог глобальной метрики средней суммы соседей вершины ($\alpha_i(t)$):

$$\begin{aligned}\Phi(k) &= 1_{\{f_n(k)>0\}} \frac{\sum_{l>0} h_n(k, l)l}{f_n^*(k)} = 1_{\{f_n(k)>0\}} \frac{\sum_{l>0} (\frac{1}{L_n} \sum_{i \rightarrow j} 1_{\{deg_i(t)=k, deg_j(t)=l\}})l}{\frac{nk f_n(k)}{L_n}} = \\ &= 1_{\{f_n(k)>0\}} \frac{\frac{\sum_{l>0} (\sum_{i \rightarrow j} 1_{\{deg_i(t)=k, deg_j(t)=l\}})l}{L_n}}{\frac{nk f_n(k)}{L_n}}.\end{aligned}$$

Предложенная метрика получила обозначение RI (Ranged Friendship Index) и, соответственно, первый вариант обозначен как

$$RI1_i(t) = \beta_i(t)L_i(t) = \frac{\sum_{j:(v_i, v_j) \in E(t)} deg_j(t)}{deg_i^2(t) \sum_{j:(v_i, v_j) \in E(t)} deg_j(t)},$$

а второй:

$$RI2_i(t) = \beta_i(t)L_n(t) = \frac{\sum_{j:(v_i, v_j) \in E(t)} deg_j(t)}{deg_i^2(t) \sum_{i=1}^n deg_i(t)}$$

3 Экспериментальная проверка гипотезы

3.1 Проверка на моделях растущих сетей

В ходе выполнения работы были проведены эксперименты моделирующие поведение стандартного индекса дружбы и обоих предложенных его модификаций, в сетях случайных графов Барабаши-Альберт и SCM размера 10 000 вершин. Для получения более усреднённых результатов каждый эксперимент проводится 10 раз и в качестве результата берётся их среднее арифметическое.

Результаты экспериментов представлены в виде графиков распределения, с разбиением на 50 групп, перечисленных метрик в различных графах на различных этапах построения сети.

На Рис. 1 можно увидеть результаты описанных экспериментов.

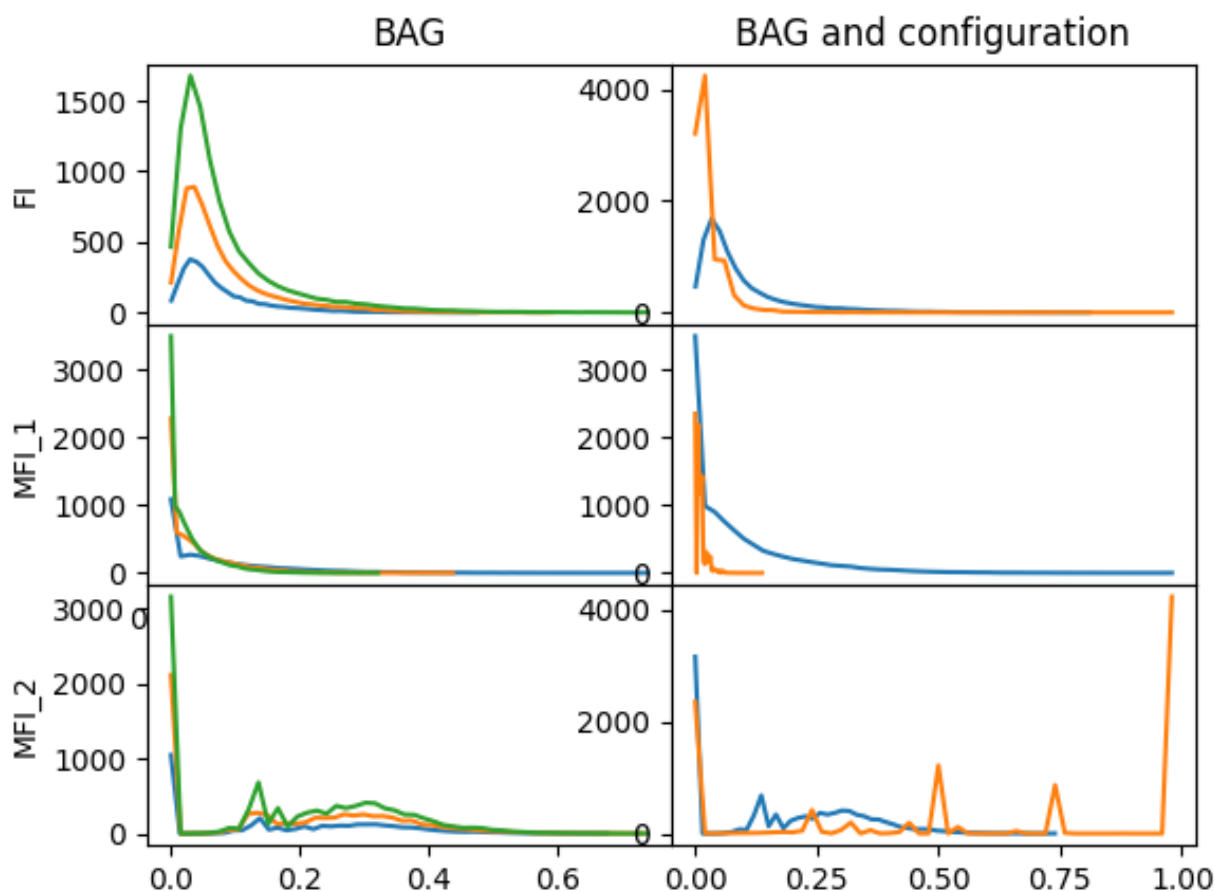


Рисунок 1 – Сравнение поведения $\beta_i(t)$, $RI1_i(t)$ и $RI2_i(t)$

Также был проведён эксперимент для анализа динамики среднего индекса дружбы, $RI1_i$ и $RI2_i$ в сети модели Барабаши-Альберт. Результаты этого эксперимента также были усреднены по 10 попыткам. Графики изменений средних

значений перечисленных величин можно увидеть на Рис. 2.

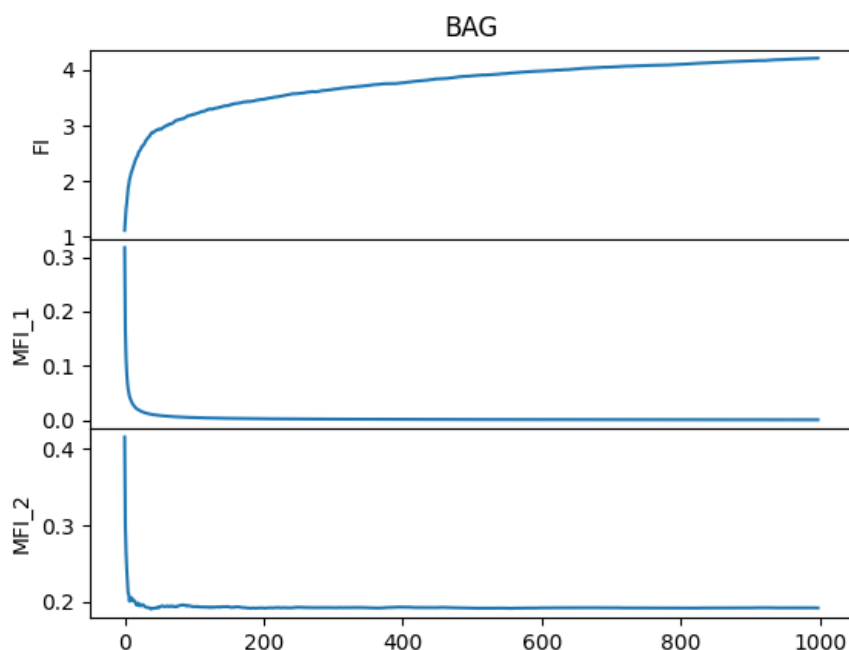


Рисунок 2 – Сравнение динамики изменения $\beta_i(t)$, $RI1_i(t)$ и $RI2_i(t)$

3.2 Проверка на конфигурационных моделях

Также были построены графики количественного распределения величин и распределения среднего значения величин по степени вершины для индекса дружбы и ранжированного индекса дружбы в графах построенных в соответствии с конфигурационной моделью. Конфигурационная модель построения случайного графа подразумевает создание неориентированного графа, в котором степени вершин соответствуют значениям, содержащимся в векторе, являющемся параметром данной модели. В ходе экспериментов использовались последовательности сгенерированные в соответствии с различными распределениями перечисленными в таблице 1.

3.3 Проверка на реальных сетях

Помимо прочего в ходе работы были построены графики количественного распределения величин и распределения среднего значения величин по степени вершины для индекса дружбы и ранжированного индекса дружбы в графах представляющих сети из реального мира. В ходе экспериментов использовались сети перечисленные в таблице 1. Рассмотренные графы можно разделить на три категории: социальные сети, сети цитирования и веб-графы.

4 Анализ результатов

В ходе проведения экспериментов были построены усреднённые графики распределения $\beta_i(t)$, $RI1_i(t)$ и $RI2_i(t)$ в графах построенных по модели Барабаши—Альберт с параметром $m = 5$, графики строились трижды, при значениях n принадлежащих множеству $n \in \{3333, 6666, 10\,000\}$. Они представлены в левом столбце Рис. 1. На основании этих графиков можно сделать вывод о том, что $RI1_i(t) = \beta_i(t)L_i(t)$ является единственной метрикой из рассмотренных, распределение которой остаётся прежним при росте сети, в то время как индекс дружбы и, в меньшей степени, $RI2_i(t) = \beta_i(t)L_n(t)$ увеличиваются при изменении размера сети.

В правом столбце Рис. 1 расположено сравнение графиков распределения соответствующих метрик в графах одинакового размера построенных по разным моделям: синие графики отражают распределение метрик в сети Барабаши—Альберт при $n = 10\,000$, а оранжевые графики — распределение в графе конфигурационной модели (SCW) построенной по последовательности Парето при значении параметра $\lambda = 2.5$ и $n = 10\,000$. Эти графики указывают на то, что не смотря на способность масштабирования при росте сети $RI1_i(t)$ позволяет различать графы построенные по разным моделям и, соответственно, обладающих различными свойствами.

Из графиков динамики средних значений $\beta_i(t)$, $RI1_i(t)$ и $RI2_i(t)$ представленных на Рис. 2 можно сделать вывод, о том что в отличии от бесконечно растущего индекса дружбы, средние значения обеих предложенных метрик весьма быстро стабилизируются после короткого периода быстрого изменения среднего значения на начальных итерациях генерации графа. Это можно объяснить, тем что на начальных итерациях структура графа модели Барабаши—Альберт непредсказуема и претерпевает значительные изменения при каждом добавлении новой вершины, однако со временем модель стабилизируется и её поведение становится всё более и более предсказуемым.

Также из Рис. 3, на котором изображены результаты того же эксперимента на логарифмической шкале, можно заметить, что график $RI1_i(t)$ представляет собой нисходящую прямую линию, что указывает на то что среднее значение данной метрики изменяется по обратному степенному закону.

В таблице 1 содержатся угловые коэффициенты линейных аппроксимаций графиков количественного и степенного распределения индекса и ранжирован-

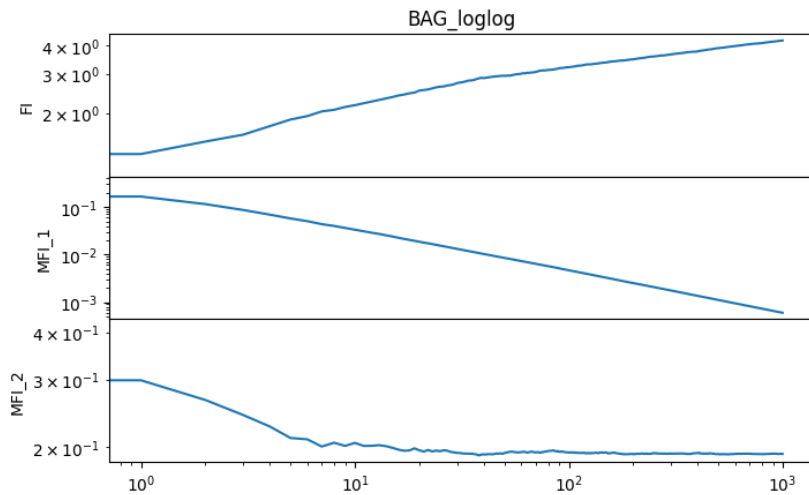


Рисунок 3 – Сравнение динамики изменения $\beta_i(t)$, $RI1_i(t)$ и $RI2_i(t)$ на логарифмической шкале

ного индекса дружбы в графах сетей из реального мира и графах, построенных в соответствии с конфигурационной моделью по различным распределениям. Так как все графики были построены на логарифмической шкале, то прямая линия в этих координатах отражает степенную функцию

$$y = bx^a,$$

в таблице отражены значения параметры a данной функции. Запись N/A в ячейке таблицы означает, что для соответствующего графика произвести линейную аппроксимацию оказалось невозможным.

Из таблицы можно увидеть, что ранжированный индекс дружбы, также как и индекс дружбы позволяет различать графы обладающие разными свойствами. Например, различные конфигурационные графы построены по нескольким различающимся последовательностям, что подразумевает различие внутренней структуры. И угловые коэффициенты для графов построенным по конфигурационной модели в значительной степени различаются: в одних случаях они положительные, в других отрицательные, а их значения принадлежат весьма широкому диапазону.

В свою очередь реальные сети были выбраны из трёх разных категорий:

- социальные сети;
- сети цитирования;
- веб-графы.

Вероятно, графы принадлежащие к одной категории должны обладать схожей структурой и свойствами, в то время как параметры сетей из разных категорий могут различаться.

Графы для LiveJournal, YouTube, Facebook и Twitter можно отнести к социальным сетям - они все отражают множество пользователей ресурса связанных отношением дружбы. За редким исключением, коэффициенты наклона для этих графов не отличаются более чем на 0.5. Аналогично в сетях цитирования среди патентов и научных работ в области физики высоких энергий также наблюдается заметное сходство. Однако в случае веб-графов статей в Википедии и поисковых результатов Google параметры степенной функции для количественного распределения ранжированного индекса дружбы различаются в различной степени.

В свою очередь таблица 2 отражают динамику исследуемых значений в конфигурационных моделях. По признаку степенного коэффициента распределения степеней вершин (γ), рассмотренные конфигурационные модели можно разделить на три группы:

- $1 < \gamma \leq 2$ (распределения Пуассона и Парето);
- $\gamma > 2$ (экспоненциальное, геометрическое и гамма-распределения)
- распределение степеней вершин не подчиняется степенному закону (биномиальное распределение).

Для первой группы характерно изменение индекса дружбы, для второй и третьей - сохранение визуального сходства графиков для индекса дружбы, не смотря на то что его распределения не соответствуют степенному закону. Однако во всех группах количественное распределение ранжированного индекса дружбы не меняет степенного коэффициента, а распределение среднего значения ранжированного индекса дружбы абсолютно не изменяется при увеличении размера графа. Эта закономерность указывает, что, в проведённых экспериментах ранжированный индекс дружбы проявляет свойство масштабируемости.

Таблица 1 – Сводная таблица результатов экспериментов

Сеть	Степенной коэффициент количественно- го распределения индекса дружбы	Степенной коэффициент количественно- го распределения ранжированного индекса дружбы	Степенной коэффициент распределения индекса дружбы по степеням вершин	Степенной коэффициент распределения ранжированного индекса дружбы по степеням вершин
распределение Пуассона	N/A	-10	-10	-10
распределение Парето	-4	-4	-5	-4
экспоненциальное распределением	N/A	-8	-8	-8
гамма-распределение	N/A	N/A	-11	-11
геометрическим распределение	N/A	N/A	-11	-10
биномиальное распределение	-12	N/A	N/A	N/A
социальная сеть LiveJournal	-3	-3	-3	-3
граф друзей среди пользователей видео-хостинга YouTube	-2	-3.5	-3.5	-3.5
сеть пользователей Facebook	-1.5	-1.5	-2	-2
граф взаимодействия пользователей соцсети Google+	-1.5	-2	-2	2
сеть пользователей Twitter	-4	-3	-3	-3
сеть цитирования патентов	-3.5	-3.5	-4	-4
сеть цитирования научных статей в области физики высоких энергий	-2	-3	-3	-3.5
сеть цитирования научных статей в области теоретической физики высоких энергий	-2	-4	-4	-4
веб-граф ссылок между статьями Википедии	-3	-6	-4	-4
веб-граф перекрёстных ссылок между сайтами в поисковой выдаче Google	-2	-4	-4	-4

Таблица 2 – Сводная таблица динамики результатов экспериментов над конфигурационными моделями

Сеть	γ	Степенной коэффициент количественно- го распределения индекса дружбы	Степенной коэффициент количественно- го распределения ранжированного индекса дружбы	Степенной коэффициент распределения индекса дружбы по степеням вершин	Степенной коэффициент распределения ранжированного индекса дружбы по степеням вершин
распределение Пуассона	2	изменяется	не изменяется	изменяется	не изменяется
распределение Парето	2	не изменяется	не изменяется	не изменяется	не изменяется
экспоненциальное распределением	6	изменяется	не изменяется	изменяется	не изменяется
гамма- распределение	6	изменяется	не изменяется	изменяется	не изменяется
геометрическим распределение	6	изменяется	не изменяется	изменяется	не изменяется
биномиальное распределение	N/A	изменяется	не изменяется	изменяется	не изменяется

ЗАКЛЮЧЕНИЕ

В ходе выполнения дипломной работы были рассмотрены различные локальные и глобальные метрики графов. А также был предложен вариант масштабируемой локальной метрики, потенциально позволяющей сравнивать графы разного размера и были проведены эксперименты подтверждающие данное свойство этой метрики, в целом наборе различных моделей построения случайных графов включая модели растущих графов и конфигурационных моделей, а также сетях из реального мира.

Однако, не смотря экспериментальное подтверждение масштабируемости метрики $RI1_i(t)$, наличие этого свойства не было доказано математически, а следовательно, без дополнительных исследований невозможно утверждать, что ранжированный индекс дружбы является масштабируемым.