

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

Кафедра математической кибернетики и компьютерных наук

СРЕДНЯЯ СТЕПЕНЬ СОСЕДЕЙ В СЛОЖНЫХ СЕТЯХ

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студента 2 курса 273 группы
направления 02.04.03 — Математическое обеспечение и администрирование
информационных систем
факультета КНиИТ
Григорьева Алексея Александровича

Научный руководитель

к. ф.-м. н., доцент

С. В. Миронов

Заведующий кафедрой

к. ф.-м. н., доцент

С. В. Миронов

Саратов 2022

ВВЕДЕНИЕ

Наука о сетях занимается анализом сложных систем, которые могут быть представлены в виде сетей с узлами – элементами системы – и связями, которые отражают взаимодействие узлов или представляют процессы, возникающие между узлами. Исследования сетей вносят огромный вклад в развитие научных трудов, информационных технологий, менеджмента. Практическое применение сетей помогает при моделировании, прогнозировании и принятии решений в сложных системах.

В настоящее время сети наблюдаются в различных сферах жизни и научных областях:

- Социальные сети состоят из людей или групп, их отношений, которые могут быть односторонними или двусторонними (представляется ориентированным и неориентированным графом).
- Биологические сети могут возникать при описании связей на молекулярном уровне. Также исследование биологических сетей позволяет предсказывать распространение болезней между людьми, принятие решений во время пандемии.
- В транспортных сетях узлами могут быть транспортные узлы, а взвешенные ребра могут определять плотность потока. Их анализ может помочь сбалансировать поток транспорта.
- Наука о сетях тесно сопряжена с компьютерными сетями и применяется в балансировке нагрузки и оптимизации топологии
- Экономические сети могут быть созданы на основе данных о цене акций на рынке, и корреляции цен между ними. Объектами исследований являются как иностранные, так и российские фондовые рынки.
- Интерес для исследования представляют сети научного сотрудничества ученых, где узлами являются ученые, а ребра представляют факт цитирования одним автором другого или наличие совместно написанной статьи. Анализ таких сетей может выявить актуальные направления науки.

Отдельное направление в науке о сетях отведено случайным сетям. Случайные сети определяются алгоритмами их построения. В отличие от реальных сетей их поведение может быть не только изучено, но и теоретически доказано. При этом случайные сети могут отражать свойства реальных сетей.

Приведением реальной сети к изученной случайной модели раскрывает свойства реальной сети, что упрощает или делает возможным прогнозирование, анализ и принятие решений.

Средняя степень соседей является фундаментальной характеристикой многих сетей. Она непрерывно связана с формированием сообществ, степенными корреляциями. Целью научно-исследовательской работы является комплексный анализ характеристик, основанных на средней степени соседей, в моделях генерации случайных сетей и реальных сетях.

Для выполнения целей были поставлены следующие задачи:

- подбор и изучение материалов по случайным сетям Барабаши—Альберт и тройственного замыкания;
- изучение технических средств для реализации и обработки случайных сетей;
- формулировка требований к разрабатываемому программному обеспечению;
- разработка программного обеспечения для генерации случайных сетей;
- дополнение ПО графическим пользовательским интерфейсом.
- моделирование сетей Барабаши—Альберт и тройственного замыкания и получение распределения случайных величин;
- сбор данных о динамике суммарной степени, средней степени и индекса дружбы, их усреднение для экземпляров синтетических сетей;
- получение распределения случайных величин в реальных социальных сетях;
- внедрение возможности обработки сразу нескольких случайных величин в рамках одного эксперимента над сетями, улучшение работы с программным обеспечением при моделировании огромных сетей.
- подбор растущих реальных сетей из открытых источников;
- улучшение программного обеспечения в части получения распределения случайных величин с использованием метода логарифмической группировки.
- дополнение ПО функционалом обработки растущих реальных сетей.

Методические основы анализа средних степеней соседей в сложных сетях представлены в работах Д. Фишера [1], ван дер Хорна [2], С. Сидорова [3], А–Л. Барабаши [4]. Основы и преимущества некоторых технологий анализа

сетей представлены в работах Е. Платта [5] и К. Харриса [6]. Примеры анализа социальных сетей представлены в работах К. Конга [7] и С. Пенга [8].

Теоретическая значимость магистерской работы. Произведен сравнительный анализ характеристик, основанных на средней степени соседей, в реальных и синтетических сетях. Полученные результаты демонстрирует применимость рассматриваемых моделей с точки зрения моделирования поведения реальных сетей в части средних степеней соседей.

Практическая значимость магистерской работы. Реализовано приложение для генерации и обработки сетей, применимое при комплексном анализе реальных и случайных сетей. Визуализация упрощает взаимодействие пользователя и программного обеспечения.

Структура и объем работы. Магистерская работа состоит из введения, четырех разделов, заключения, списка использованных источников и приложения. Общий объем работы — 81 страница, из них 60 страниц — основное содержание, включая 27 рисунков и 12 таблиц, цифровой носитель в качестве приложения, список использованных источников информации — 30 наименований.

Краткое содержание работы

В первом разделе «Случайные сети» произведен обзор литературы, посвященной случайным сетям. В нем рассматриваются модели генерации случайных сетей, понятия ассортативности и парадокса дружбы, а также такие локальные характеристики сети как средняя степень соседей и индекс дружбы, используемые для измерения степенных корреляций.

Информация из данного раздела применяется и в теоретической, и в практической части работы.

Второй раздел «Программное обеспечение для генерации и обработки случайных сетей» посвящен разработке программного обеспечения для выполнения экспериментов по моделированию случайных и реальных сетей. В начале разработки ПО были определены соответствующие требования, а именно: генерация сетей Барабаши—Альберт и тройственного замыкания, возможность проведения неограниченного числа независимых экспериментов, запись результатов в файл, получение динамики в искусственных сетях, получение степенной корреляции, построение графиков и гистограмм распределений на линейной и логарифмической шкале, генерация совместимой с LaTeX визуализации результатов, считывание сетей из файла, работа с ориентированными сетями, параметризуемость, наличие пользовательского интерфейса, сопровождаемость.

При реализации использовался язык программирования Python с библиотеками для построения графов `networkx`, математических вычислений `numpy` и анализа данных `scipy`. Был создан модуль генерации случайных сетей с заданными параметрами, которые зависят от конкретной модели генерации случайных сетей. Построенная сеть обрабатывается — вычисляются значения выбранных коэффициентов. Доступны вычисления суммарной, средней степени узла, индекса дружбы и средней степени соседей узлов с выбранной степенью. Каждый результат вычисления данных величин для отдельного эксперимента записывается в файл. Были созданы вспомогательные модули для усреднения результатов, полученных для отдельных экземпляров случайных сетей. Более того, сбор данных о локальных характеристиках узлов производится в процессе роста сети. Данный функционал применим как к моделям генерации случайных сетей, так и к реальным растущим сетям.

Поскольку многие случайные величины, основанные на средней сте-

пени соседей, распределены по степенному закону, то разумно привести их к системе координат с логарифмическими осями и воспользоваться линейной регрессией для получения экспоненты степенного закона. При построении графиков случайных величин без дополнительной обработки на логарифмических осях заметно большое количество шумов в низких и высоких степенях. Для получения более точной экспоненциальной зависимости используется метод логарифмической группировки, при котором вершины объединяются в группы, размер которых увеличивается логарифмически с увеличением степеней. Данный метод был реализован в программном обеспечении.

Было разработано полноценное программное обеспечение для генерации и обработки случайных и реальных сетей. Работа с приложением может вестись как из интегрированной среды разработки, так и через предоставленный графический интерфейс приложения. Данное приложение было использовано для получения эмпирических результатов из следующих разделов.

Третий раздел «Распределение средней степени соседей» посвящен анализу ассортативности в случайных и реальных сетях. Одна из наиболее надежных величин измерения ассортативности — средняя степень соседей узлов с данной степенью (ANND), вычисляемая как отношение суммы средних степеней соседей к количеству узлов с выбранной степенью.

Было построено по 500 независимых экземпляров случайных сетей Барабаши—Альберт и тройственного замыкания размером до 25,000 узлов. Вычисление коэффициента ANND позволило сделать выводы о корреляции средней степени соседей и степеней узлов. Было получено, что дисперсия средних степеней стремится к нулю, а узлы, преимущественно, не имеют корреляции. Однако для низких степеней заметно резкое падение коэффициента средней степени соседей.

Полученные результаты были сопоставлены с аналогичными вычислениями для четырех реальных социальных сетей, загруженных из онлайн хранилищ данных о сетях, соответствующих взаимодействию пользователей в реальных системах.

В результате стало известно, что для всех сетей дисперсия средних степеней равномерно убывает по степенному закону, для обеих случайных сетей низкие степени дизассортативны, а высокие степени не коррелируют. Для всех реальных степеней высокие степени также дизассортативны, а поведение для

низких степеней отличается.

Четвертый раздел «Динамика случайных величин в растущих сетях» посвящен анализу средней степени соседей в растущих сетях. Реальные растущие сети отличаются от стационарных тем, что для каждой связи известен момент времени, когда она была добавлена в сеть.

Сначала были предложены методы моделирования растущих сетей. Затем для семи реальных растущих сетей с миллионами связей были вычислены траектории средней степени соседей. Аналогичные действия были произведены для случайных сетей. Было обнаружено, что средние степени узлов непрерывно растут, а дисперсия стремится к константе.

В результате моделирования реальных сетей можно сделать вывод о том, что феномен, наблюдаемый в случайных сетях, не является свойственным реальным сетям. В реальных сетях дисперсия стремится к $+\infty$. Поведение систем, при котором дисперсия стремится к $+\infty$, является типовым для случайных Марковских нестационарных процессов.

Стоит обратить внимание на то, что траектории характеристик для сетей AskUbuntu и SuperUser оказались очень похожими. Обе сети имеют одинаковые механизмы формирования связей, но при этом они построены по независимым данным. Таким образом, можно сделать вывод о том, что реальные сети, имеющие схожие механизмы образования связей, могут иметь одинаковые траектории роста характеристик. Другими словами, одним из ключевых факторов роста характеристик является механизм формирования связей в сети.

ЗАКЛЮЧЕНИЕ

В результате выпускной квалификационной работы было разработано интерактивное программное обеспечение, позволяющее моделировать и анализировать случайные сети. Визуализация результатов осуществляется через png графики и совместимые с LaTeX файлы. Разработанное программное обеспечение было модифицировано в части получения распределения случайных величин. Был внедрен метода логарифмической группировки для устранения шумов в распределениях случайных величин.

С помощью программного обеспечения была проведена оценка динамики случайных величин в синтетических сетях. Для этого были построены случайные сети Барабаши—Альберт и тройственного замыкания, а на основе них получены и усреднены траектории средней степени и индекса дружбы для узлов в случайных сетях. Случайные величины при росте сетей следуют логарифмическому и степенному закону. Были построены социальные сети, основанные на открытых данных из онлайн сервисов. На основе реальных и случайных сетей было получено распределение средней степени соседей. Модели Барабаши—Альберт и тройственного замыкания имеют схожее поведение в части ассортативности.

Было разработано интерактивное программное обеспечение, позволяющее моделировать и анализировать случайные сети. Визуализация результатов осуществляется через png графики и совместимые с LaTeX файлы. Разработанное программное обеспечение было модифицировано в части получения распределения случайных величин. Был внедрен метода логарифмической группировки для устранения шумов в распределениях случайных величин.

С целью анализа динамики средних степеней в реальных сетях были собраны семь растущих сетей, основанных на данных реальных систем, компаний, сервисов. Для каждой сети была проведена оценка динамики случайных величин в реальных растущих сетях. В результате сравнения поведения характеристик средних степеней для случайных и реальных сетей был сделан вывод, что поведение дисперсии средних соседей реальных сетей отличается от рассмотренных моделей.

Результаты научно-исследовательской работы были опубликованы в следующих работах:

- Balash, V. Comparative Analysis of Financial Network Topology for the

Russian, Chinese and US Stock Markets / V. Balash, S. Sidorov, A. Faizliev, A. Grigoriev. – WSEAS transactions on business and economics, 2020. – Vol. 17. – Pp. 120–132.

- Sidorov, S. P. Friendship paradox in growth networks: analytical and empirical analysis / S. P. Sidorov, S. V. Mironov, A. A. Grigoriev // Applied Network Science. – jul 2021. – Vol. 6, no. 1.
- Sidorov, S. Node degree dynamics in complex networks generated in accordance with a modification of the triadic closure model / S. Sidorov, S. Mironov, A. Faizliev, A. Grigoriev // Mathematical modeling and supercomputer technologies. SPRINGER NATURE, 2021. – Vol. 1413. – Pp. 146–153.
- Grigoriev, A. Average nearest neighbor degree and its distribution in social networks / A. Grigoriev, S. Sidorov, S. Mironov, I. Malinskii // Digital Transformation and Global Society. – Springer International Publishing, 2022. – Vol. 1503 – Pp. 36–50.,

а также представлены на конференциях с работами:

- «Средняя степень соседей в социальных сетях» на «студенческой научной конференции» 30 апреля 2021 года на базе «Саратовского национального исследовательского государственного университета им. Н. Г. Чернышевского» в Саратове;
- «Average Nearest Neighbor Degree and Its Distribution in Social Networks» на конференции «DTGS 2021» 23–25 июня 2021 года на базе «университета ИТМО» в Санкт-Петербурге;
- «Analysis of assortativeness in social networks» на конференции «The Sixth Workshop on Computer Modelling in Decision Making (CMDM 2021)» 18–20 ноября 2021 года на базе «Саратовского национального исследовательского государственного университета им. Н.Г.Чернышевского» в Саратове;
- «Trend Stationarity of the Average Degree for a Node’s Neighbors in Growth Networks» на конференции «12th International Conference on Network Analysis NET 2022» 23–25 мая 2022 года на базе «Высшей школы экономики» в Нижнем Новгороде.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Fisher, D. N.* The Perceived Assortativity of Social Networks: Methodological Problems and Solutions / D. N. Fisher, M. J. Silk, D. W. Franks. — Springer International Publishing, 2017. — Pp. 1–19.
- 2 *van der Hoorn, P.* Average nearest neighbor degrees in scale-free networks / P. van der Hoorn // *Internet Mathematics*. — 2018.
- 3 *Sidorov, S.* Surprising behavior of the average degree for a node’s neighbors in growth networks / S. Sidorov, S. Mironov, S. Tyshkevich // *Complex Networks & Their Applications X*. — Cham: Springer International Publishing, 2022. — Vol. 1015. — Pp. 463–474.
- 4 *Barabási, A.-L.* Network science / A.-L. Barabási. — Cambridge: Cambridge University Press, 2016.
- 5 *Platt, E. L.* Network Science with Python and NetworkX Quick Start Guide: Explore and visualize network data effectively / Edward L. Platt / E. L. Platt. — Birmingham: Packt Publishing, 2019.
- 6 Array programming with NumPy / C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser et al. // *Nature*. — 2020. — Vol. 585, no. 7825. — Pp. 357–362.
- 7 *Kong, X.* Academic social networks: Modeling, analysis, mining and applications / X. Kong, Y. Shi, S. Yu, J. Liu, F. Xia // *Journal of Network and Computer Applications*. — 2019. — Vol. 132. — Pp. 86–103.
- 8 *Peng, S.* Influence analysis in social networks: A survey / S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, W. Jia // *Journal of Network and Computer Applications*. — 2018. — Vol. 106. — Pp. 17–32.