

ФГБОУ ВО “Московский государственный
университет имени М.В. Ломоносова”

На правах рукописи

Прохоренкова Людмила Александровна

**СВОЙСТВА СЛУЧАЙНЫХ ВЕБ-ГРАФОВ, ОСНОВАННЫХ
НА ПРЕДПОЧТИТЕЛЬНОМ ПРИСОЕДИНЕНИИ**

Специальность 01.01.05 — теория вероятностей
и математическая статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2015

Работа выполнена на кафедре математической статистики и случайных процессов механико-математического факультета ФГБОУ ВО “Московский государственный университет имени М.В. Ломоносова”.

Научный руководитель: Райгородский Андрей Михайлович
доктор физико-математических наук, профессор

Официальные оппоненты: Павлов Юрий Леонидович
доктор физико-математических наук, профессор
заведующий лабораторией теории вероятностей и
компьютерной статистики ФГБУН Института при-
кладных математических исследований Карельско-
го научного центра РАН

Карпов Дмитрий Валерьевич
кандидат физико-математических наук
старший научный сотрудник Санкт-Петербургского
отделения Математического института им. В.А.
Стеклова РАН

Ведущая организация: Хабаровское отделение ФГБУН Института
прикладной математики ДВО РАН

Защита состоится 24 июня 2015 года в 16 часов 45 минут на заседании диссер-
тационного совета Д 501.001.85 на базе ФГБОУ ВО “Московский государственный
университет имени М.В. Ломоносова”, по адресу: 119991, Москва, ГСП-1, Ленин-
ские горы, д. 1, МГУ, Механико-математический факультет, аудитория 16-24.

С диссертацией можно ознакомиться в Фундаментальной библиотеке ФГБОУ
ВО “Московский государственный университет имени М.В. Ломоносова” по ад-
ресу: Москва, Ломоносовский проспект, д. 27, сектор А и на сайте механико-
математического факультета <http://mech.math.msu.su/~snark/index.cgi>.

Автореферат разослан мая 2015 года.

Ученый секретарь диссертационного
совета Д 501.001.85 на базе МГУ
доктор физико-математических наук, профессор

В.В. Власов

Общая характеристика работы

Актуальность

Диссертация посвящена изучению моделей сложных сетей и их свойств. Под сложными сетями понимают всевозможные сети, которые встречаются в природе: компьютерные, биологические, социальные, экономические, транспортные и так далее. Классический пример такой сети — граф сети Интернет. Вершины графа — это веб-страницы, а ребра — ссылки между ними. С появлением сети Интернет началось интенсивное изучение сложных сетей. Было замечено, что все эти сети обладают некоторыми общими свойствами: малый диаметр, степенной закон распределения степеней вершин, кластерная структура и другие^{1,2,3,4,5,6}. Возник естественный вопрос: почему сети столь различной природы обладают схожими свойствами? Возможно, в основе всех этих сетей лежит какой-то общий принцип формирования? Так стали появляться первые модели сложных сетей, в том числе вероятностные модели^{7,8,9,10,11,12,13}. Вероятностная модель — это случайный элемент со значениями в множестве графов и некоторым вероятностным распределением на этом множестве.

В 1999 году А.-Л. Барабаши и Р. Альберт¹⁴ предложили моделировать сложные сети с помощью так называемого принципа предпочтительного присоеди-

¹M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

²M.E.J. Newman. Assortative mixing in networks. *Phys. Rev. Letter*, 89, 208701, 2002.

³M.E.J. Newman. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46, N5, 323–351, 2005.

⁴M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

⁵R. Pastor-Satorras, A. Vázquez, A. Vespignani. Dynamical and Correlation Properties of the Internet. *Phys. Rev. Lett.*, 87, N25, 258701, 2001.

⁶D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature* 393, pages 440–442, 1998.

⁷W. Aiello, F. Chung, L. Lu. A random graph model for power law graphs. *Experiment. Math.*, 10, 53–66, 2001.

⁸S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

⁹B. Bollobás, Ch. Borgs, J. Chayes, O. Riordan, Directed scale-free graphs. ACM-SIAM Symposium on Discrete Algorithms, 132–139, 2003.

¹⁰C. Cooper and A. Frieze. A general model of web graphs. *Random Structures and Algorithms*, 22(3):311–335, 2003.

¹¹M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.*, 47:41–72, 2009.

¹²N. Eggemann and S. D. Noble. The clustering coefficient of a scale-free random graph. *Discrete Applied Mathematics*, 159(10):953–965, 2011.

¹³P. Holme and B. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65(2), 2002.

¹⁴R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74:47–97, 2002.

ния. Основная идея их подхода состоит в том, что новые страницы, появляющиеся в Интернете, скорее предпочтут сослаться на уже популярные страницы (то есть на те, на которые ведет много ссылок). Работа Барабаша и Альберт послужила толчком к развитию области вероятностного моделирования сложных сетей. Например, Р. Кумар, П. Рагхаван, С. Раджагопалан, Д. Сивакумар, А. Томкинс и Э. Упфал предложили так называемую модель копирования¹⁵, в которой некоторые ребра, проведенные из новой вершины, “копируют” ребра одной из предшествующих страниц. Еще изучалась конфигурационная модель^{16,17}, которая позволяет построить граф с заданным распределением степеней вершин. Эта модель может применяться для моделирования графа Интернета, если взять подходящее распределение степеней. Обзор перечисленных и многих других моделей можно найти, например, в работах^{18,19,20}. Модели сложных сетей применяются в различных областях: в математике, физике, биологии, области информационных технологий.

Вероятностное моделирование сложных сетей включает в себя две важные задачи. Первая задача — создание новых моделей, которые лучше моделируют свойства реальных сетей, вторая — изучение характеристик уже существующих моделей.

Наиболее распространенная характеристика сложных сетей — это распределение степеней вершин. А именно, если выбрать случайную вершину в сети и посмотреть на ее степень (количество примыкающих ребер), то получим случайную величину, распределение которой интересно изучать. Для большинства наблюдаемых сетей оказалось, что доля вершин степени d убывает как $d^{-\gamma}$ с некоторым параметром γ , который обычно лежит в интервале $(2, 3)$ ^{21,22,23}. Таким образом, принято говорить, что в сложных сетях распределение степеней вершин подчиня-

¹⁵R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. *Proc. 41st Symposium on Foundations of Computer Science*, 2000.

¹⁶M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random structures and algorithms*, 1995.

¹⁷Ю. Л. Павлов. О предельных распределениях степеней вершин условного конфигурационного случайного графа. *Труды Карельского научного центра РАН*, 5, 2012.

¹⁸А. М. Райгородский. Модели случайных графов и их применение. *Труды МФТИ*, 2(4):130–140, 2010.

¹⁹А. М. Райгородский. Модели Интернета. *Интеллект*, 2013.

²⁰B. Bollobás. Mathematical results on scale-free random graphs. *Handbook of Graphs and Networks*, pages 1–34, 2003.

²¹A.-L. Barabási and R. Albert. Emergence of scaling in random network. *Science*, 286(5439):509–512, 1999.

²²A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(16):309–320, 2000.

²³M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. SIGCOMM Conference, 1999.

ется (асимптотически) *степенному закону*.

Еще одной важной характеристикой сложных сетей является PageRank^{24,25,26}. PageRank — это мера важности (популярности, качества) веб-страниц, он активно используется в информационном поиске. Основная идея PageRank заключается в том, что важность вершины определяется важностью (степенью, популярностью) ее соседей. Степень вершины является довольно грубым приближением ее PageRank. Более точное приближение можно получить, если рассматривать так называемую вторую степень вершины, то есть количество путей длины два, ведущих в данную вершину. В данной работе исследуется распределение вторых степеней вершин, а также анализируются k -е входящие степени вершин.

Другой важной характеристикой сети является ее диаметр, то есть максимальная по всем парам вершин длина кратчайшего пути между двумя вершинами. Сложные сети, встречающиеся в природе, имеют малый диаметр (теория 6 рукопожатий)²⁷. Б. Боллобаш и О. Риордан доказали²⁸, что диаметр графов в классической модели предпочтительного присоединения ведет себя асимптотически как $\frac{\ln n}{\ln \ln n}$, где n — количество вершин в графе. В данной работе вводится обобщение понятия диаметра — r -диаметр. Для r -диаметра классических моделей предпочтительного присоединения тоже удастся изучить асимптотическое поведение.

Кроме перечисленных выше свойств, многие сети, встречающиеся в природе, обладают еще и так называемым *свойством устаревания*²⁹. Свойство устаревания отражает тот факт, что в реальных сетях вершины чаще соединены с другими вершинами, близкими к ним по возрасту (времени появления). Классические модели предпочтительного присоединения не обладают свойством устаревания. В данной работе предложена и проанализирована модель, которая обладает этим свойством.

²⁴K. Avrachenkov, D. Lebedev. PageRank of Scale-Free Growing Networks. *Internet Math.*, 3, N2, 207–232, 2006.

²⁵L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the Web. Technical report, Stanford University Database Group, 1998.

²⁶G. Pandurangan, P. Raghavan, E. Upfal. Using PageRank to Characterize Web Structure. *Internet Math.*, 3, N1, 1–20, 2006.

²⁷R. Albert, H. Jeong, and L.-A. Barabási. Diameter of the world-wide web *Nature*, 401, 130–131, 1999.

²⁸B. Bollobás and O. M. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

²⁹D. Lefortier, L. Ostroumova, and E. Samosvat. Evolution of the media web. In *Algorithms and Models for the Web Graph*, pages 80–92. Springer International Publishing, 2013.

Цель работы

1. Обобщить понятие степени вершины и изучить распределение вторых степеней в классических моделях предпочтительного присоединения.
2. Обобщить понятие диаметра графа и изучить асимптотическое поведение r -диаметра в классической модели Боллобаша–Риордана.
3. Предложить вероятностную модель, которая отражает свойство устаревания реальных сетей, а также проанализировать ее свойства.

Научная новизна

Все результаты диссертации являются новыми. Перечислим основные из них:

1. Проанализировано распределение вторых степеней вершин в моделях Бакли–Остхуса и Боллобаша–Риордана. Показано, что это распределение подчиняется (асимптотически) степенному закону [3, 4].
2. Введено понятие k -й входящей степени вершины, вычислено математическое ожидание этой величины для вершин графа в модели Боллобаша–Риордана [1].
3. Введено понятие r -диаметра и найдено его асимптотическое поведение в модели Боллобаша–Риордана [2].
4. Предложена модель, которая сочетает в себе идею предпочтительного присоединения и свойство устаревания. Для этой модели доказано, что при определенных условиях распределение степеней подчиняется степенному закону, а также показано, что доля ребер в графе, которые соединяют вершины с разницей “возрастов”, большей чем T , то есть вершины i и j с $|i - j| > T$, убывает экспоненциально с ростом T [5].

Результаты диссертации обоснованы в виде строгих математических доказательств. Точные формулировки установленных автором утверждений приведены ниже.

Методы исследования

При доказательстве результатов использовалась разнообразная техника. Существенное место в данной работе занимают различные методы доказательства концентрации случайных величин около своего математического ожидания. Например, в задачах, которые возникают в данной работе, можно использовать неравенство Азумы³⁰. Однако, в некоторых случаях оказывается полезнее использовать неравенство Талагранна³¹. Применение неравенства Талагранна неочевидно в случае моделей предпочтительного присоединения, в частности, нужно переопределить вероятностное пространство модели в терминах независимых случайных величин. Далее, проверка условий применения неравенства Талагранна является сложной комбинаторной задачей.

Теоретическая и практическая ценность

Разработана комбинаторная техника, позволяющая эффективно оценивать математические ожидания случайных величин. Переопределено вероятностное пространство моделей предпочтительного присоединения в терминах независимых случайных величин, что позволяет, в частности, применять неравенство Талагранна для этих моделей. Несмотря на то, что диссертация носит теоретический характер, некоторые результаты имеют и практическую ценность. Например, предложенная новая модель предпочтительного присоединения с устареванием может быть использована для разработки и сравнения различных методов информационного поиска.

Апробация работы

Результаты диссертации докладывались на следующих научно-исследовательских семинарах:

- Неоднократные выступления на семинаре “Вероятностные и алгебраические методы в комбинаторике” под руководством профессора А.М. Райгородского (механико-математический факультет МГУ, 2011-2013 гг.).
- Кафедральный семинар кафедры математической статистики под руководством профессора А.М. Зубкова (механико-математический факультет МГУ, 2014 г.).

³⁰К. Azuma. Weighted sums of certain dependent variables. *Tôhoku Math. J.*, 19:357–367, 1967.

³¹М. Talagrand. Concentration of measures and isoperimetric inequalities in product spaces. *Publications Mathématiques de l’I.H.E.S.*, 81:73–205, 1996.

- Семинар “Динамические системы и статистическая физика” под руководством профессора Б.М. Гуревича и профессора В.И. Оселедца (механико-математический факультет МГУ, 2013 г.).
- Семинар “Дискретный анализ” под руководством профессора А.А. Сапоженко (факультет ВМК МГУ, 2013 г.)
- Неоднократные выступления на семинаре под руководством профессора А.М. Райгородского (ФИВТ МФТИ, 2011-2014 гг.).

Результаты диссертации докладывались на следующих международных научных конференциях:

- Международная конференция “Instabilities and Control of Excitable Networks” (Долгопрудный, Россия, май 2012 г.).
- Международная конференция “4th Polish Combinatorial Conference” (Бедлево, Польша, сентябрь 2012 г.).
- Международная конференция “Workshop on Internet Topology and Economics” (Атланта, США, ноябрь 2012 г.).
- Международная конференция “Franco-Russian workshop on Algorithms, complexity and applications” (Москва, Россия, июнь 2013 г.).
- Международная конференция “International Conference on Random Structures and Algorithms” (Познань, Польша, август 2013 г.).
- Международная конференция “Palanga Conference in Combinatorics and Number Theory” (Паланга, Литва, сентябрь 2013 г.).
- Международная конференция “Workshop on Random Graphs and their Applications” (Москва, Россия, октябрь 2013 г.).
- Международная конференция “Workshop on Algorithms and Models for the Web Graph” (Кембридж, США, декабрь 2013 г.).
- Международная конференция “Moscow Workshop on Combinatorics and Number Theory” (Москва, Россия, январь 2014 г.).

- Международная конференция “11th International Vilnius Conference on Probability Theory and Mathematical Statistics” (Вильнюс, Литва, июль 2014 г.).
- Международная конференция “Sum(m)it240” (Будапешт, Венгрия, июль 2014 г.).

Работа автора поддержана грантом РФФИ N 12-01-00683 и грантом Президента МД-6277.2013.1.

Публикации

Результаты диссертации опубликованы в 5 работах автора (4 из них входят в перечень ВАК), список которых приведен в конце автореферата.

Структура диссертации

Диссертация состоит из введения, трех глав и списка литературы, насчитывающего 56 наименований. Общий объем диссертации составляет 119 страниц.

КРАТКОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Введение к диссертации состоит из трех частей. В них обсуждаются характеристики сложных сетей, различные подходы к моделированию сетей и основная идея, которая лежит в основе многих моделей, — идея предпочтительного присоединения.

Содержание главы 1

Первая глава посвящена анализу свойств классической модели Боллобаша–Риордана. Данная глава основана на статьях автора [1, 2, 4].

В **разделе 1.1** приводится определение модели и известные результаты. Обычно для графов в модели Боллобаша–Риордана используется обозначение G_m^n , где n — число вершин, m — фиксированное натуральное число. Граф G_m^n строится следующим образом. Сначала индуктивно строится G_1^n . Граф G_1^1 состоит из одной вершины и одной петли. Граф G_1^t получается из G_1^{t-1} посредством добавления вершины t и одного ребра между вершинами t и i , где i выбирается случайно

следующим образом:

$$P(i = s) = \begin{cases} d_{G_1^{t-1}}(s)/(2t-1) & \text{если } 1 \leq s \leq t-1, \\ 1/(2t-1) & \text{если } s = t. \end{cases}$$

Здесь $d_{G_1^t}(s)$ — степень вершины s в графе G_1^t . Иными словами, чем больше степень вершины в графе G_1^{t-1} , тем больше вероятность того, что следующая вершина будет соединена с ней. Граф G_m^n ($m > 1$) получается из графа G_1^{mn} следующим образом. Первые m вершин G_1^{mn} объединяются в первую вершину нового графа, следующие m — во вторую, и так далее. Ребра в некотором смысле сохраняются, а именно: если в G_1^{mn} ребро соединяло вершины i и j , то в полученном графе G_m^n это ребро соединяет те группы вершин, в которые попали i и j . Тем самым в графе могут возникнуть кратные ребра и кратные петли. Построенные таким образом графы G_m^n образуют вероятностное пространство \mathfrak{G}_m^n .

Известно, что граф G_m^n обладает степенным распределением степеней вершин³².

ТЕОРЕМА 1. Пусть $m \geq 1$. Тогда существует такая функция φ аргумента n , что $\varphi = o(n)$ при $n \rightarrow \infty$, и для любой такой степени d , что $m \leq d \leq n^{1/15}$, мы имеем

$$\lim_{n \rightarrow \infty} P \left(\left| \#_m^n(d) - \frac{2nm(m+1)}{d(d+1)(d+2)} \right| > \frac{\varphi(n)}{d(d+1)(d+2)} \right) = 0.$$

Здесь $\#_m^n(d)$ — количество вершин степени d в графе G_m^n .

В разделе 1.2 рассматриваются вторые степени вершин в графе G_1^n . Сначала оценивается математическое ожидание количества вершин со второй степенью d , затем доказывается концентрация. Случай $m \geq 2$ обсуждается в главе 2 данной работы.

Второй степенью вершины $t \in G_1^n$ называем величину

$$d_2(t) = |\{ij : i \neq t, j \neq t, it \in G_1^n, ij \in G_1^n\}|.$$

Другими словами, вторая степень вершины t — это количество ребер, соединенных с соседями вершины t кроме тех, которые ведут в саму вершину t . Через $X_n(d)$ обозначаем количество вершин второй степени d в G_1^n .

В разделе 1.2 доказаны следующие новые результаты.

³²В. Bollobás, O. M. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18(3):279–290, 2001.

ТЕОРЕМА 2. Для любого $k > 1$ имеем

$$\mathbb{E}X_n(k) = \frac{4n}{k^2} \left(1 + O\left(\frac{\ln^2 k}{k}\right) + O\left(\frac{k^2}{n}\right) \right).$$

ТЕОРЕМА 3. Для любого $\varepsilon > 0$ существует такая функция φ аргумента n , что $\varphi = o(n)$ при $n \rightarrow \infty$, и

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(|X_n(k) - \mathbb{E}X_n(k)| \geq \frac{\varphi(n)}{k^2} \right) = 0$$

для любого такого k , что $1 \leq k \leq n^{1/6-\varepsilon}$.

Эта теорема показывает, что для любого k , $1 \leq k \leq n^{1/6-\varepsilon}$, с вероятностью, стремящейся к 1, количество вершин второй степени k отличается от своего математического ожидания на что-то малое в сравнении с математическим ожиданием. Таким образом, доказан закон больших чисел или, как говорят в этой области науки, доказана концентрация. Итак, распределение вторых степеней тоже подчиняется (асимптотически) степенному закону.

При доказательстве теоремы 3 о концентрации используется мартингалное неравенство Азумы³³.

В разделе 1.3 изучаются k -е степени вершин. К определению k -й степени вершины можно подходить по-разному. Первое возможное определение k -й степени вершины t в некотором графе — количество ребер, которые выходят из вершин, находящихся на расстоянии $k - 1$ от t , кроме тех, которые ведут в вершины, находящиеся на расстоянии $k - 2$ от t . Кроме того, можно определять k -ю степень вершины t как количество вершин, находящихся на расстоянии k от нее.

В графе предпочтительного присоединения вводится естественная ориентация ребер — от вершин с большими номерами к вершинам с меньшими. Таким образом, можно рассматривать “входящую” степень вершин. Определим ее сперва по аналогии с первым определением. Будем пока рассматривать только граф G_1^n . Пусть множества $D_k(t)$ определяются индуктивно следующим образом: $D_0(t) = \{t\}$, для $k \geq 1$

$$D_k(t) = \{i : i \notin D_0(t) \cup D_1(t) \cup \dots \cup D_{k-1}(t), \exists j \in D_{k-1}(t), ij \in G_1^n\}.$$

Когда будем писать $ij \in G_m^n$, будем иметь в виду, что направленное ребро из вершины i в вершину j проведено в графе G_m^n . Таким образом, $D_k(t)$ — множество

³³K. Azuma. Weighted sums of certain dependent variables. *Tôhoku Math. J.*, 19:357–367, 1967.

вершин i , от которых можно добраться до вершины t самое меньшее за k шагов, двигаясь все время по направленным ребрам. Далее, пусть

$$S_k(t) = \{ji : i \in D_{k-1}(t), ji \in G_1^n\}, \quad k \geq 1.$$

Величина $|S_k(t)| = d_k(t)$ и есть k -я входящая степень вершины. Считаем $d_0(t) = 1$. Кроме того, можно было бы назвать k -й входящей степенью вершины t количество вершин, находящихся от t на “входящем” расстоянии k , то есть $|D_k(t)|$. Заметим, что в графе G_1^n при $k > 1$ оба определения совпадают.

В данном разделе для графа G_1^n вычисляется величина $\mathbb{E}g_k(t)$, где $g_k(t) = d_k(t) + d_{k-1}(t)$, $k \geq 1$.

Для графа G_m^n вычислена немного другая величина. Пусть $D'_k(t)$ — множество таких вершин графа G_m^n , из которых можно добраться до вершины t ровно за k шагов, двигаясь все время по направленным ребрам (но не по петлям). При этом возможность добраться за меньшее количество шагов допускается. Считаем $D'_0(t) = \{t\}$. Положим $d'_k(t) = |D'_k(t)|$. Вычисляем величину $\mathbb{E}g'_k(t)$, где $g'_k(t) = d'_k(t) + md'_{k-1}(t)$. Для $m = 1$ при $k > 2$ эта величина не отличается от определенной ранее, поскольку в графе G_1^n между двумя вершинами не может быть больше одного направленного пути.

Сформулируем основные результаты раздела 1.3. Сначала рассматривается граф G_1^n . Определим функцию $f_k(t, n)$. Пусть $f_0(t, n) = 1$, а для $k > 0$

$$f_k(t, n) = \frac{1}{(k-1)!} \sqrt{\frac{n}{t}} \left(\ln \sqrt{\frac{n}{t}} \right)^{k-1}.$$

ТЕОРЕМА 4. Для любого натурального $k \geq 2$ и любой вершины $t < n$ в графе G_1^n

$$\mathbb{E}g_k(t) = f_k(t, n) (1 + \theta(k, t, n)) (1 + \theta_1(t))^k,$$

где $|\theta(k, t, n)| \leq C \frac{k^2}{t \ln \sqrt{n/t}}$, $|\theta_1(t)| \leq C_1 \frac{1}{t}$, C и C_1 — некоторые константы.

Посмотрим, при каких условиях на k и t мы получаем асимптотику. Чтобы получить $|\theta(k, t, n)| = o(1)$, надо взять $k = o(\sqrt{t \ln \frac{n}{t}})$. Тогда при $t \geq \ln n$ имеем $(1 + \theta_1(t))^k = 1 + o(1)$. Чтобы иметь возможность взять $k \geq 2$, надо наложить ограничение $t \leq n - w(n)$, где $w(n)$ стремится к бесконечности с ростом n .

СЛЕДСТВИЕ 1. Пусть $w(n) \rightarrow \infty$ при $n \rightarrow \infty$. Пусть, далее, $\ln n \leq t \leq n - w(n)$ и $k = o(\sqrt{t \ln \frac{n}{t}})$, $k \geq 2$. Тогда

$$\mathbb{E}g_k(t) = f_k(t, n) (1 + o(1))$$

при $n \rightarrow \infty$.

Далее от графа G_1^n переходим к графу G_m^n с $m \geq 2$. Доказана следующая теорема.

ТЕОРЕМА 5. Пусть натуральные $m \geq 2$ и $k \geq 1$ фиксированы. Тогда для $t < n$

$$\mathbb{E}g'_k(t) = m^k f_k(t, n) \left(1 + O\left(\frac{1}{t \min\{1, \ln \frac{n}{t}\}}\right) \right).$$

В разделе 1.4 изучается некоторое обобщение понятия диаметра — r -диаметр:

$$\text{diam}_r(G) = \max_{\substack{A \subset V \\ |A|=r}} \min_{i, j \in A} \rho(i, j).$$

Иными словами, здесь рассматриваются не пары вершин, как в обычном диаметре, а подмножества мощности r . В каждом таком подмножестве мы находим две вершины на наименьшем расстоянии друг от друга. И нас интересует максимум этого расстояния по всем подмножествам. Заметим, что при $r = 2$ имеем в точности обычный диаметр графа G .

В 2004 году Боллобаш и Риордан доказали следующую теорему³⁴.

ТЕОРЕМА 6. Для любого $\varepsilon > 0$ и любого натурального $m \geq 2$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left((1 - \varepsilon) \frac{\ln n}{\ln \ln n} \leq \text{diam}(G_m^n) \leq (1 + \varepsilon) \frac{\ln n}{\ln \ln n} \right) = 1.$$

В разделе 1.4 доказано подобное утверждение для r -диаметра. А именно

ТЕОРЕМА 7. Пусть $\varepsilon > 0$ и натуральное $m \geq 2$ фиксированы. Пусть, далее, $r(n)$ — произвольная последовательность целых чисел, лежащих в пределах от 2 до n . Тогда

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{(1 - \varepsilon) \ln n - \ln r}{\ln \ln n} \leq \text{diam}_r(G_m^n) \leq \frac{(1 + \varepsilon) \ln n - \ln r}{\ln \ln n} \right) = 1.$$

Как и ожидалось, чем больше вершин в рассматриваемых множествах, тем меньше r -диаметр. Отметим, что если для любого $\alpha > 0$ выполнено $r(n) = o(n^\alpha)$, т.е. $r(n) = n^{o(1)}$, то результаты теорем 6 и 7, по существу, идентичны. Если же,

³⁴B. Bollobás and O. M. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24(1):5–34, 2004.

например, $r(n) = n^\alpha$, $\alpha > 0$, то константа в асимптотике из теоремы 7 принципиально другая: $1 \pm \varepsilon$ превращается в $1 - \alpha \pm \varepsilon$. Если, наконец, $r(n) = n^{1-o(1)}$, то асимптотики в теореме 7 нет.

Содержание главы 2

В данной главе рассмотрена модель, которая является обобщением модели Боллобаша и Риордана. С целью сделать модель более гибкой, две группы авторов^{35,36} предложили добавить в модель новый параметр — “привлекательность” вершины (константа, не зависящая от степени). Бакли и Остхус дали строгое определение этой модели³⁷. Модель позволяет получить степенное распределение степеней вершин с любым параметром из $(2; \infty)$. В главе 2 проанализировано распределение вторых степеней вершин в модели Бакли–Остхуса. Глава основана на работе автора [3].

Дадим строгое определение рассматриваемой модели. Пусть n — количество вершин в графе, а $m \in \mathbb{N}$ и $a \in \mathbb{R}_+$ — фиксированные параметры. Сначала рассмотрим случай $m = 1$ и индуктивно построим граф $H_{a,1}^n$. Граф $H_{a,1}^1$ состоит из одной вершины и одной петли. Предположим, что граф $H_{a,1}^{t-1}$ уже построен. На шаге t в граф добавляется одна вершина t и одно ребро между вершинами t и i , где i выбрана случайно следующим образом:

$$P(i = s) = \begin{cases} \frac{d_{H_{a,1}^{t-1}}(s) - 1 + a}{(a+1)t-1} & \text{если } 1 \leq s \leq t-1, \\ \frac{a}{(a+1)t-1} & \text{если } s = t. \end{cases}$$

Здесь $d_{H_{a,1}^t}(s)$ — степень вершины s в графе $H_{a,1}^t$.

Граф $H_{a,m}^n$ с $m > 1$ получается из графа $H_{a,1}^{mn}$. Вершины $1, \dots, m$ образуют первую вершину нового графа; вершины $m+1, \dots, 2m$ — вторую; и так далее. А ребра сохраняются: если ребро e соединяло вершины $im+k$ и $jm+l$, $1 \leq k, l \leq m$ в графе $H_{a,1}^{mn}$, то в графе $H_{a,m}^n$ будет проведено ребро e' между вершинами $i+1$ и $j+1$. Как и в случае графа Боллобаша–Риордана, в данном случае могут возникать петли и кратные ребра. Через $\mathfrak{H}_{a,m}^n$ обозначим полученное вероятностное пространство.

³⁵S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Phys. Rev. Lett.*, 85, 4633, 2000.

³⁶E. Drinea, M. Enachescu, and M. Mitzenmacher. Variations on random graph models for the web. *technical report, Harvard University, Department of Computer Science*, 2001.

³⁷P. G. Buckley and D. Osthus. Popularity based random graph models leading to a scale-free degree sequence. *Discrete Mathematics*, 282:53–63, 2004.

Бакли и Остхус доказали, что распределение степеней в графе $H_{a,m}^n$ подчиняется степенному закону с параметром $-2 - a$, если a — натуральное число.

Напомним, что *второй степенью* вершины $t \in H_{a,1}^n$ мы называем

$$d_2(t) = |\{ij : i \neq t, j \neq t, it \in H_{a,1}^n, ij \in H_{a,1}^n\}|,$$

то есть количество ребер, смежных с соседями вершины t , кроме тех, которые смежны с самой вершиной t . Вершину t назовем *k-вершиной*, если $d_2(t) \geq k$. Через $Y_n(k)$ обозначим количество k -вершин в $H_{a,1}^n$. Кроме того, рассматривается величина $X_n(k)$, равная количеству вершин второй степени k в $H_{a,1}^n$. Очевидно, что $Y_n(k) = \sum_{i \geq k} X_n(i)$.

В главе 2 доказано, что $Y_n(k)$ убывает как k^{-a} . Это значит, что распределение вторых степеней также подчиняется степенному закону. Чтобы доказать это, сначала оценивается математическое ожидание величины $Y_n(k)$, а затем показывается концентрация этой величины около математического ожидания за счет неравенства Талагранна. Применение неравенства Талагранна в данном случае неочевидно, в частности, нужно переопределить вероятностное пространство модели Бакли–Остхуса в терминах независимых случайных величин. После этого, неравенство Талагранна применяется в его общей несимметричной форме. Далее, проверка условий применения неравенства Талагранна является сложной комбинаторной задачей.

Математическое ожидание $Y_n(k)$ удастся оценить следующим образом.

ТЕОРЕМА 8. *Для любого $k > 1$ имеем*

$$\mathbb{E}Y_n(k) = \frac{(a+1)\Gamma(2a+1)}{\Gamma(a+1)k^a} n \left(1 + O\left(\frac{(\ln k)^{[a+1]}}{k}\right) + O\left(\frac{k^{1+a}}{n}\right) \right).$$

Из теоремы 8 несложно получить следующее следствие.

СЛЕДСТВИЕ 2. $\mathbb{E}Y_n(k) = \Theta\left(\frac{n}{k^a}\right)$ для $k = O\left(n^{\frac{1}{1+a}}\right)$.

Используя ту же технику, что и в доказательстве теоремы 8, можно доказать следующий результат.

ТЕОРЕМА 9. *Для любого $k \geq 1$ имеем*

$$\mathbb{E}X_n(k) = \frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)k^{a+1}} \left(1 + O\left(\frac{(\ln k)^{[a+1]}}{k}\right) + O\left(\frac{k^{1+a}}{n}\right) \right).$$

Из этой теоремы не сложно получить следующее следствие.

СЛЕДСТВИЕ 3. $\mathbb{E}X_n(k) = \Theta\left(\frac{n}{k^{1+a}}\right)$ для $k = O\left(n^{\frac{1}{1+a}}\right)$.

Далее, доказана теорема о концентрации.

ТЕОРЕМА 10. Пусть $\delta > 0$ и $k = O\left(n^{\frac{1}{2+a+\delta}}\right)$. Тогда для некоторого $\epsilon > 0$ имеем

$$\mathbb{P}\left(|Y_n(k) - \mathbb{E}(Y_n(k))| > (\mathbb{E}(Y_n(k)))^{1-\epsilon}\right) = o(1).$$

Можно также доказать подобный результат для величины $X_n(k)$.

ТЕОРЕМА 11. Пусть $\delta > 0$ и $k = O\left(n^{\frac{1}{4+a+\delta}}\right)$. Тогда для некоторого $\epsilon > 0$ имеем

$$\mathbb{P}\left(|X_n(k) - \mathbb{E}(X_n(k))| > (\mathbb{E}(X_n(k)))^{1-\epsilon}\right) = o(1).$$

Содержание главы 3

Если предыдущие две главы посвящены анализу классических моделей, то в главе 3 вводится новый класс моделей, основанных на идее предпочтительного присоединения, а именно — модели с устареванием. Глава основана на статье автора [5].

Несмотря на то, что модели предпочтительного присоединения отражают такие важные свойства реальных сетей, как маленький диаметр и степенное распределение степеней вершин, они не соответствуют реальности в других важных аспектах. В главе 3 обсуждается так называемое *свойство устаревания*. Свойство устаревания отражает тот факт, что в реальных сетях вершины чаще соединены с другими вершинами, близкими к ним по возрасту (времени появления). В случае моделей предпочтительного присоединения под временем появления понимается шаг n , на котором вершина добавилась в граф. Нетрудно понять, что свойство устаревания не выполняется для моделей предпочтительного присоединения: новые вершины стремятся присоединиться к уже популярным вершинам, а наиболее популярные вершины — “старые” вершины.

В главе 3 рассматривается величина $e(T)$ — доля ребер в графе, которые соединяют вершины с разницей “возрастов”, большей чем T , то есть вершины i и j с $|i - j| > T$. Известно³⁸, что в некоторых реальных сетях $e(T)$ убывает экспоненциально с ростом T .

³⁸D. Lefortier, L. Ostroumova, and E. Samosvat. Evolution of the media web. In *Algorithms and Models for the Web Graph*, pages 80–92. Springer International Publishing, 2013.

Класс моделей, который обсуждается в данной главе, является обобщением модели Бианкони и Барабаши³⁹, которая, в свою очередь, обобщает модели предпочтительного присоединения добавлением качества (fitness) вершин. В модели Бианкони и Барабаши, когда новая вершина добавляется в граф, она выбирает вершину, в которую проведет ребро, с вероятностью, пропорциональной произведению качества этой вершины и степени. В данной работе в функцию привлекательности вершины добавляется множитель, соответствующий устареванию вершины. Таким образом, вероятность того, что в данную вершину будет проведено новое ребро, определяется *привлекательностью* вершины, которая, в свою очередь, зависит от входящей степени вершины (текущая популярность), ее качества (некоторая константа) и возраста (новые страницы обладают большей привлекательностью).

В **разделе 3.2** дано формальное определение модели с устареванием. А именно, рассматривается последовательность случайных графов $\{G_n\}$. У этой последовательности есть следующие параметры: натуральное число m (количество ребер, добавляемых вместе с новой вершиной) и функция $N(n)$, принимающая целочисленные значения. Нам также потребуется последовательность независимых положительных случайных величин ζ_1, ζ_2, \dots с некоторым распределением. Каждый граф G_n определяется в соответствии со своей собственной процедурой построения, которая основана на идее предпочтительного присоединения.

Случайный граф G_n определяется следующим образом. В начале процедуры построения имеем две вершины и одно ребро между ними (граф \tilde{G}_2^m). Первые две вершины имеют качества $q(1) := \zeta_1$ и $q(2) := \zeta_2$. На шаге $t + 1$ ($2 \leq t \leq n - 1$) к графу \tilde{G}_t^m добавляется одна вершина и m ребер. Новая вершина $t + 1$ имеет качество $q(t + 1) := \zeta_{t+1}$. Новые ребра проводятся независимо друг от друга, и они соединяют новую вершину с одной из предыдущих вершин. Для каждого ребра вероятность того, что оно будет проведено в вершину i ($1 \leq i \leq t$), равна

$$\frac{\text{attr}_t(i)}{\sum_{j=1}^t \text{attr}_t(j)},$$

где

$$\text{attr}_t(i) = (1 \text{ or } q(i)) \cdot (1 \text{ or } d_t(i)) \cdot \left(1 \text{ or } I[i > t - N(n)] \text{ or } e^{-\frac{t-i}{N(n)}}\right)$$

³⁹G. Bianconi and A.-L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86(24):5632–5635, 2001.

и $d_t(i)$ — степени вершины i в \tilde{G}_t^n . Далее мы опускаем n в обозначении $N(n)$. Согласно определению, в графе нет петель, однако могут возникать кратные ребра.

Важно заметить, что, в отличие от стандартного определения предпочтительного присоединения, в нашем случае граф G_n не может быть получен из графа G_{n-1} . Каждый граф строится в соответствии со своей собственной процедурой, которая, в свою очередь, основана на предпочтительном присоединении. Такое необычное определение позволяет теоретически исследовать распределение степеней и поведение функции $e(T)$ в графе.

Ранее было экспериментально показано, что функция привлекательности, которая содержит качество и фактор устаревания, лучше отражает поведение некоторых частей интернета, чем функция привлекательности, которая содержит степень, качество и фактор устаревания. Поэтому далее рассматриваются только две модели, содержащие качество и один из двух факторов устаревания.

В **разделе 3.3** рассматривается модель с функцией привлекательности $q(i)I[i > t - N]$. Фактор устаревания в данном случае означает то, что вершина i может получить входящие ребра только в течение следующих N шагов после появления, и этот период называется *жизнью* вершины.

Мы также предполагаем, что случайные величины ζ_1, ζ_2, \dots имеют распределение Парето с плотностью $f(x) = \frac{\gamma a^\gamma I[x > a]}{x^{\gamma+1}}$, где $\gamma > 1$, $a > 0$.

В итоге у случайного графа есть несколько параметров: 1) количество вершин n , 2) исходящая степень вершин m , 3) продолжительность жизни N , 4) параметр распределения качества γ , 5) минимальное качество a .

В **параграфе 3.3.1** показано, что степенное распределение качества ведет к степенному распределению степеней вершин в графе.

Через $\#(d)$ обозначим количество вершин степени d в графе G_n . Доказана следующая теорема.

ТЕОРЕМА 12. Пусть $d = d(n)$ растет с ростом n и $d = o\left(\left(\frac{n}{N}\right)^{\frac{1}{\gamma+1}}\right)$. Если $\gamma > 2$ и $d = o\left(N^{\frac{1}{\gamma+3}}\right)$ или $1 < \gamma \leq 2$ и $d = o\left(N^{\frac{\alpha-1}{\gamma+\alpha+1}}\right)$ для некоторого α , $1 < \alpha < \gamma$, то

$$\frac{\mathbb{E}\#(d)}{n} = \frac{\gamma}{d^{\gamma+1}} \left(\frac{(\gamma-1)m}{\gamma} \right)^\gamma (1 + o(1)).$$

Теорема 12 говорит о том, что математическое ожидание количества вершин степени d убывает как $d^{-\gamma-1}$. Однако, чтобы показать степенное распределение

степеней вершин, необходимо также доказать концентрацию количества вершин степени d около своего математического ожидания.

ТЕОРЕМА 13. *Для любого d выполнено:*

$$\mathbb{P} \left(|\#(d) - \mathbb{E}\#(d)| \geq \sqrt{Nn \ln n} \right) \leq \frac{2}{\ln n}.$$

Заметим, что для $d = o \left(\left(\frac{n}{N \ln n} \right)^{1/2(\gamma+1)} \right)$ имеем $\sqrt{Nn \ln n} = o(n/d^{\gamma+1})$, поэтому теорема 13 действительно дает концентрацию.

В **параграфе 3.3.2** изучается свойство устаревания. Напомним, что $e(T)$ — доля ребер в графе, которые соединяют вершины с разницей возрастов, большей чем T , то есть такие вершины i и j , что $|i - j| > T$. В этом параграфе показано, что в рассматриваемой модели $e(T)$ убывает линейно.

ТЕОРЕМА 14. *Для любого натурального T*

$$\mathbb{E}e(T) = \begin{cases} 1 - \frac{T}{N} + O\left(\frac{N}{n}\right), & \text{если } T \leq N; \\ 0, & \text{если } T > N. \end{cases}$$

ТЕОРЕМА 15.

$$\mathbb{P} \left(|e(T) - \mathbb{E}e(T)| \geq \sqrt{\frac{\ln n}{n}} \right) \leq \frac{1}{m \ln n}.$$

В **разделе 3.4** изучается модель с экспоненциальным фактором устаревания. То есть функция привлекательности имеет вид $q(i)e^{-\frac{t-i}{N}}$. Опять же, предполагаем, что случайные величины ζ_1, ζ_2, \dots имеют распределение Парето с плотностью $f(x) = \frac{\gamma a^\gamma I[x>a]}{x^{\gamma+1}}$, где $\gamma > 1$, $a > 0$.

В **параграфе 3.4.1** проанализировано распределение степеней в этой модели. Полученные результаты похожи на результаты для модели с индикатором.

ТЕОРЕМА 16. *Пусть $d = d(n)$ растет с ростом n и $d = o \left(\left(\frac{n}{N \ln N} \right)^{\frac{1}{\gamma+1}} \right)$. Если $\gamma > 2$ и $d = o \left(N^{\frac{1}{3\gamma+5}} \right)$ или $1 < \gamma \leq 2$ и $d = o \left(N^{\frac{\alpha-1}{\alpha+(\gamma+1)(\alpha+1)}} \right)$ для некоторого α , $1 < \alpha < \gamma$, то*

$$\frac{\mathbb{E}\#(d)}{n} = \frac{\gamma}{d^{\gamma+1}} \left(\frac{(\gamma-1)m}{\gamma} \right)^\gamma (1 + o(1)).$$

Опять, математическое ожидание количества вершин степени d убывает как $d^{-\gamma-1}$. Следующая теорема показывает, что количество вершин степени d сконцентрировано около своего математического ожидания.

ТЕОРЕМА 17. Для любого d выполнено следующее неравенство:

$$\mathbb{P} \left(|\#(d) - \mathbb{E}\#(d)| > \sqrt{Nn \ln n} \right) = O \left(\frac{1}{\ln n} \right).$$

Как и ранее, для $d = o \left(\left(\frac{n}{N \ln n} \right)^{\frac{1}{2(\gamma+1)}} \right)$ выполнено $\sqrt{Nn \ln n} = o(\mathbb{E}\#(d))$ и теорема 17 дает концентрацию.

В параграфе 3.4.2 показано, что поведение $e(T)$ для модели с экспоненциальным убыванием популярности совпадает с реальностью, а именно, $e(T)$ убывает экспоненциально с ростом T .

Вычислено математическое ожидание $e(T)$.

ТЕОРЕМА 18. Для любого натурального T

$$\mathbb{E}e(T) = e^{-\frac{T}{N}} + O \left(\frac{N}{n} \right).$$

И показана концентрация.

ТЕОРЕМА 19. Для любого натурального T

$$\mathbb{P} \left(|e(T) - \mathbb{E}e(T)| \geq \sqrt{\frac{\ln n}{n}} \right) \leq \frac{1}{m \ln n}.$$

Эти теоремы показывают, что $e(T)$ убывает экспоненциально, как и в наблюдаемых сетях.

Благодарности. Автор признателен научному руководителю, доктору физико-математических наук, профессору Андрею Михайловичу Райгородскому за неоценимую помощь в работе.

Работы автора по теме диссертации

- [1] Л.А. Остроумова, *Математические ожидания k -х входящих степеней вершин в случайных графах в модели Боллобаша–Риордана*, Труды МФТИ, 13:29–40, 2012.
- [2] Л.А. Остроумова, *Об r -диаметрах случайных графов в модели Боллобаша–Риордана*, Труды МФТИ, 13:41–59, 2012.
- [3] A. Kupavskii, L. Ostroumova, D. Shabanov, and P. Tetali, *The distribution of second degrees in the Buckley–Osthus random graph model*, Internet Mathematics, 9(4):297–335, 2013. [А. Купавский доказал лемму 5 в теореме о концентрации. Д. Шабанов и П. Тетали помогли в редактировании текста. Все основные результаты получены Л. Остроумовой.]
- [4] L. Ostroumova and E. Grechnikov, *The distribution of second degrees in the Bollobás–Riordan random graph model*, Moscow Journal of Combinatorics and Number Theory, 2(2):85–110, 2012. [Е. Гречников помог свернуть сумму в доказательстве теоремы 2. Л. Остроумовой были сформулированы и доказаны все основные результаты работы.]
- [5] Л. А. Прохоренкова и Е. А. Самосват, *Модели предпочтительного присоединения с устареванием*, Деп. в ВИНТИ 20.11.14, № 319-В2014, 2014. Расширенная версия статьи находится в печати в журнале Internet Mathematics. [Е. Самосват предложил идею устаревания. Л. Прохоренковой была формализована модель, а также сформулированы и доказаны все основные результаты работы.]