Издательский Дом
ИНТЕЛЛЕКТ

А.М. РАЙГОРОДСКИЙ

# МОДЕЛИ ИНТЕРНЕТА

А.М. РАЙГОРОДСКИЙ

# МОДЕЛИ ИНТЕРНЕТА


ДОЛГОПРУДНЫЙ
2013

**А.М. Райгородский**

Модели Интернета: Учебное пособие / А.М. Райгородский – Долгопрудный: Издательский Дом «Интеллект», 2013. – 64 с.
ISBN 978-5-91559-143-0

Учебное пособие посвящено моделированию Интернета, который был диковинкой для большинства из нас еще каких-то 15 лет назад. Сейчас мы ежедневно пользуемся ресурсами Интернета - поиском, электронной почтой, блогами и др. Сеть динамично развивается, растет и усложняется, а потому рядовому пользователю может казаться, что в Интернете царит полный хаос. Однако в реальности все устроено намного интереснее.

Многочисленные статистические исследования показывают, что есть ряд законов, которым подчиняется «всемирная паутина». В частности, эти законы связаны с интерпретацией Интернета как графа, вершины которого - сайты, а ребра - гиперссылки. В книге описаны основные законы такого типа и рассказано, как современная математика помогает их моделировать.

Для понимания книги читателю понадобится знание основ комбинаторики, теории графов и теории вероятностей. Книга будет полезна студентам, аспирантам и преподавателям технических ВУЗов, а также всем, кто интересуется приложениями математики к моделированию «сложных сетей» - Интернета, социальных, биологических, транспортных и других сетей.

ISBN 978-5-91559-143-0

# ОГЛАВЛЕНИЕ

# ВВЕДЕНИЕ

Еще каких-то 15 лет назад про существование сети Интернет мало кто знал. Сейчас же Интернет прочно вошел в нашу жизнь, став для нас незаменимым инструментом получения и распространения информации, а также, конечно, общения. Поиск, электронная почта, блоги, социальные сети, скайп и пр., — без этого многие уже не могут обойтись. Всемирная паутина невероятно разрослась за эти годы. Может показаться даже, что в ней царит полный хаос: так велика она и, на первый взгляд, неконтролируема. И тем не менее, существуют законы, по которым развивается Интернет. Изучение этих законов так же увлекательно, как и изучение законов природы. Начато оно было практически одновременно с появлением сети. И к настоящему времени возникла целая научная дисциплина, лежащая на стыке физики, математики и социологии.

Этой небольшой книгой мы открываем серию брошюр, посвященных исследованиям Интернета. Здесь мы сделаем акцент на интерпретации Интернета как графа, вершины которого, в зависимости от контекста, суть страницы или сайты (хосты), а ребра — (гипер)ссылки между ними. Об удивительных свойствах этого графа — графа, растущего с течением времени, — и о неожиданно простых моделях, в которых эти свойства реализуются, мы и поговорим в книге.

Для понимания книги читателю потребуются начальные знания в области комбинаторики, теории графов и теории вероятностей. Структурно книга устроена следующим образом. В первой главе мы расскажем о многочисленных характеристиках Интернета, которые изучались и продолжают изучаться исследователями. Эти-то характеристики и покажут, насколько ошибочно первоначальное впечатление о полном хаосе, якобы царящем во всемирной паутине. Во второй главе мы опишем

несколько вероятностных моделей веба и сформулируем теоретические результаты, которые выявят как сильные, так и слабые стороны моделей. При этом мы пройдем основные этапы того пути, который прошли исследователи Интернета: от простейших моделей конца 90-х годов XX века до значительно более точных моделей, возникших буквально в последние несколько лет. Правда, мы будем следовать лишь одному из важнейших направлений в идеологии построения моделей. А именно, мы обсудим только идею так называемого предпочтительного присоединения. Это уже огромный пласт современной Интернет-математики. Его мы и постараемся вскрыть в нашей первой книге. Наконец, в третьей главе мы приведем схемы доказательств некоторых теорем из второй главы.

ГЛАВА 1

# СВОЙСТВА ИНТЕРНЕТА

## 1.1. ОСНОВНЫЕ ОБЪЕКТЫ И ОБЩАЯ ИДЕОЛОГИЯ ИХ ИЗУЧЕНИЯ

Существуют два типа графов, возникающих при изучении всемирной паутины. С одной стороны, можно считать вершинами графа компьютеры, подключенные к Интернету, или серверы, через которые идет Интернет-коммуникация. В этом случае ребрами являются линии связи. Это такой граф «из железа». Он называется *Интернет-графом*, и он нас не будет интересовать. Нашим объектом послужит более «виртуальный» граф. Его вершины — это страницы или сайты в Интернете, а ребра — (гипер)ссылки между ними. Если вершины — страницы, граф называется *веб-графом*; если вершины — сайты (хосты), граф называется *хост-графом*. Хост-граф проще для статистического анализа, так как он гораздо меньше веб-графа. В хост-графе есть и кратные ребра (несколько различных ссылок с одного сайта на другой), и (кратные) петли (ссылки между страницами внутри сайта), и, конечно, хост-граф (как и веб-граф) ориентирован. Мы будем, как правило, говорить о хост-графах.

Предположим, нас интересует какая-то статистика хост-графа. Например, число ребер в нем. Разумеется, желая узнать ее приблизительное значение, мы действуем так же, как и всегда в статистическом анализе. Мы берем достаточно репрезентативную (большую) выборку, состоящую из хост-графов, которые получены в результате взятия надлежащего количества «временны́х срезов». И для каждого из этих графов мы считаем число ребер $x_i$. Получается выборка $x_1, \ldots, x_n$. Пользуясь теми или иными статистическими инструментами, мы делаем те

или иные заключения о природе числа ребер. Иными словами, если мы произносим фразу «хост-граф имеет такое-то количество ребер» или «степени вершин хост-графа распределены так-то», то мы подразумеваем не один хост-граф, но последовательность хост-графов — случайный процесс, статистики которого мы изучаем.

В настоящей книге нас не будут интересовать конкретные инструменты статистического анализа, с помощью которых получены различные наблюдения об Интернете. С ними можно будет познакомиться при чтении исходных публикаций, ссылки на которые мы будем давать. Для нас важнее будет сама картина Интернета, и именно на ее основе мы будем строить модели веба.

Итак, в следующих разделах мы опишем несколько весьма удивительных статистических закономерностей, которым подчиняются хост-графы.

## 1.2. КОЛИЧЕСТВО РЕБЕР

Хост-граф, да и веб-граф тоже, — графы разреженные. Принято считать, что если у такого графа $n$ вершин, то ребер у него $mn$, где $m$ — некоторая константа, не меньшая единицы. Разумеется, это лишь допущение. Скорее, статистика такова, что число ребер — это $\Theta(n)$, т. е. оно зажато в пределах от $m_1 n$ до $m_2 n$ с постоянными $m_1$, $m_2$.

Почему при таких характеристиках граф разрежен? Очень просто: даже если забыть о наличии в графе кратных ребер, за счет которых общее количество ребер могло бы быть сколь угодно большим, все равно на $n$ вершинах бывает вплоть до $\Theta(n^2)$ ребер, и эта функция растет существенно быстрее линейной функции $mn$.

Наблюдение о разреженности графов страниц и хостов — одно из самых первых и простых (см., например, [1–3]).

## 1.3. ГИГАНТСКАЯ КОМПОНЕНТА

Понятие *гигантской компоненты* является одним из центральных в теории графов. Как всегда, нелепо пытаться определить его для одного конкретного графа. Оно имеет смысл лишь для бесконечных последовательностей графов. Итак, пусть дана последовательность графов $\{G_n = (V_n, E_n)\}$, в которой $\lim\limits_{n \to +\infty} |V_n| = +\infty$. Говорят, что графы $G_n$ *содержат гигантскую компоненту связности*, если существует такая константа $\gamma > 0$, что для каждого $n$ размер наибольшей связной компоненты в $G_n$ не меньше $\gamma |V_n|$.

Одно из важнейших наблюдений в науке о «реальных сетях» (т. е., в частности, о веб- и хост-графах) состоит в том, что гигантская компонента в этих сетях всегда есть. Исследователи, которые не очень склонны к черезмерной математической аккуратности (а таких в этой области большинство), подходят к понятию гигантской компоненты даже менее формально, чем это сделали мы. Обычно они просто говорят, что у графа (отдельного графа!) есть гигантская связная компонента, коль скоро «существенная доля его вершин» образует такую компоненту. Разумеется, совершенно не ясно, что такое «существенная» доля. Грубо говоря, одна сотая — это не существено, а девять десятых точно годится. Факт тот, что даже в этом нестрогом смысле принято считать, что и веб-, и хост-граф содержат гигантскую компоненту. Тем более они ее содержат в смысле данного выше аккуратного определения.

И в строгом, и в нестрогом определении гигантская компонента, как правило, ровно одна. Интуиция за этим следующая. Трудно поверить, что, например, в хост-графе, у которого 1 000 000 000 (миллиард) вершин, может быть две связных компоненты размера 400 000 000 (четыреста миллионов) и ни одной ссылки между ними: слишком велика вероятность того, что хотя бы один владелец сайта из первой компоненты процитирует хотя бы один сайт из второй компоненты!

Повторим, что все сказанное только что — это лишь способ осознать суть понятия гигантской компоненты. Строгое определение дано, и в дальнейшем мы предпочтем апеллировать только к нему.

## 1.4. УСТОЙЧИВОСТЬ И УЯЗВИМОСТЬ

Исключительно важное для практики наблюдение о веб- и хост-графах состоит в том, что они устойчивы к случайным разрушениям (нарушениям функциональности) вершин, но уязвимы, коль скоро атакам целенаправленно подвергаются вершины максимальной степени — так называемые «хабы».

Строго указанные статистические результаты можно сформулировать в следующих двух теоремах (см. [4]).

**Теорема 1.** *Пусть $G_n$ — последовательность хост-графов, растущих с течением времени n. Пусть $p \in (0,1)$ и каждая вершина графа $G_n$ уничтожается с вероятностью p независимо от всех остальных вершин. Тогда с вероятностью, стремящейся к единице при $n \to \infty$, в последовательности случайных графов $G_{n,p}$ есть гигантская компонента.*

Разумеется, размер гигантской компоненты (константа $\gamma$) тем меньше, чем ближе $p$ к единице.

**Теорема 2.** *Пусть $G_n = (V_n, E_n)$ — последовательность хост-графов, растущих с течением времени n. Пусть $c \in (0,1)$. Упорядочим вершины $G_n$ по невозрастанию величины степени (имеется в виду сумма входящей и исходящей степеней). Удалим из $G_n$ первые $[c|V_n|]$ вершин (здесь $[x]$ — целая часть числа x). Тогда существует такая константа $c^*$, что при $c \leqslant c^*$ в графе $G_{n,c}$ есть гигантская компонента, а при $c > c^*$ в графе $G_{n,c}$ гигантской компоненты нет.*

В последней теореме мы наблюдаем исключительно важное для физики явление — так называемый *фазовый переход*: уязвимость к атакам на хабы возникает скачкообразно; до $c^*$ ее нет, но, едва мы преодолеваем порог $\sim c^*|V_n|$, и граф разваливается на мелкие компоненты — компоненты размера $o(|V_n|)$ каждая.

## 1.5. ДИАМЕТР

Как известно, *диаметр* графа — это максимальное расстояние между его вершинами, а *расстояние* в графе — это число ребер в кратчайшем реберном пути. Разумеется, у несвязного графа, каковым является веб-граф, диаметр не определен (или равен бесконечности). Однако вполне можно искать диаметр гигантской компоненты. И вот он все долгие годы исследований остается практически постоянным — не превосходящим 10–20, — и едва ли не уменьшается (см. [3]). Это при условии, что переходы осуществляются с учетом ориентации ребер. Если ориентацию убрать, то диаметр уменьшается и вовсе до 5–6. Примерно то же верно и для хост-графа. Это, по сути, знаменитый закон шести рукопожатий — статистическое наблюдение, состоящее в том, что любые два человека в мире знакомы друг с другом через не более чем пять посредников.

Отметим, что в свете разреженности хост-графа столь малый диаметр весьма неожидан и замечателен. По-английски это свойство принято называть «small-world phenomenon», т. е. «мир тесен» (ср. [5]).

## 1.6. СТЕПЕНИ ВЕРШИН

Обозначим $\operatorname{indeg} v$, $\operatorname{outdeg} v$, $\deg v$ — входящую, исходящую и полную степень вершины $v$ соответственно (петля дает вклад 1 и в $\operatorname{indeg} v$, и в $\operatorname{outdeg} v$, т. е. вклад 2 в $\deg v$). Замечательный статистический закон состоит в том, что каждая из этих степеней и в веб-графе, и в хост-графе, и во многих других реальных сетях (социальных,

биологических и пр.) подчиняется степенному закону распределения (см. [1,2]) с тем или иным показателем.

**Теорема 3.** *Пусть $G_n = (V_n, E_n)$ — последовательность реальных сетей, растущих с течением времени n. Пусть $d_n \in \mathbb{N}$. Пусть* adeg — *одна из трех видов степеней. Тогда существуют константы $\gamma$ и c, с которыми*

$$\frac{|\{v \in V_n\colon \operatorname{adeg} v = d_n\}|}{|V_n|} \sim \frac{c}{d_n^{\gamma}}.$$

Константа $\gamma$ полностью задает распределение: она является степенью, определяющей скорость убывания доли вершин данной степени $d_n$. Константу $c$ можно найти из тех соображений, что сумма всех долей (вероятностей того, что степень вершины равна $d_n$) есть 1.

Степенные законы в сетях стали изучать задолго до появления Интернета (см., например, [6]).

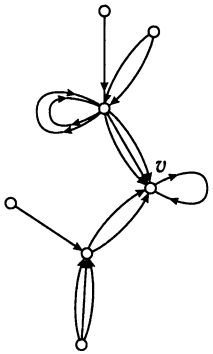
Как правило, $\gamma \in (2,3)$. Например, для adeg = deg и хост-графа $\gamma \approx 2{,}3$ (см. [7]).

## 1.7. ВТОРЫЕ СТЕПЕНИ ВЕРШИН

Степени вершин хост-графа, о которых мы писали в предыдущем разделе, свидетельствуют, в частности, об их популярности: чем больше ссылок на данный сайт, тем он, конечно, популярнее. Однако такую характеристику достаточно легко увеличить искусственно, создав множество фиктивных сайтов, которые будут цитировать сайт, подлежащий раскрутке. И, разумеется, спамеры и оптимизаторы этим активно пользуются. Поэтому важно не только количество ссылок на сайт, но и авторитетность тех сайтов, владельцы которых эти ссылки проставляют. В простейшем случае достаточно посмотреть на степени тех вершин, которые соединены ребром с данной вершиной $v$ хостграфа. В частности, можно изучать *вторую степень вершины v*. Ее, в свою очередь, можно определять по-разному. Например, можно считать ее равной числу вершин, отстоящих на расстояние 2 от $v$, или числу вершин, отстоящих на расстояние $\leqslant 2$ от $v$, или сумме степеней вершин, отстоящих на расстояние 1 от $v$, и т. д. На рис. 1 показано, насколько разнятся данные определения.

Можно показать, что вторые степени вершин подчиняются степенным законам, как и обычные степени.

Разумеется, давались определения и $k$-х степеней с $k > 2$. Но нас такие степени в дальнейшем интересовать не будут. В следующем разделе




$\operatorname{indeg} v = 6$

Число вершин на расстоянии 2 от $v$ равно 4.
Сумма степеней вершин на расстоянии 1 от $v$ равна 9

**Рис. 1**

мы сразу дадим определение *пейджранка*, который в некотором роде содержит в себе информацию обо всех степенях всех вершин графа.

## 1.8. ПЕЙДЖРАНК

Пейджранк — это, при всей своей простоте, одна из самых сильных характеристик вершины веб-графа, уточняющая в некотором смысле понятие степени, второй степени и т. д. Пусть $G_n = (V_n, E_n)$ — веб-граф (пока именно веб-граф). Обозначим $PR(v)$ пейджранк его вершины $v$. Тогда $PR(i)$, где $i \in \{1, \ldots, |V_n|\}$, определяется как решение системы линейных уравнений

$$PR(i) = c \sum_{j \to i} \frac{PR(j)}{\operatorname{outdeg} j} + \frac{c}{|V_n|} \sum_{j \in \mathcal{D}} PR(j) + \frac{1-c}{|V_n|}, \quad i = 1, \ldots, |V_n|,$$

где $c \in (0, 1)$ — константа, а $\mathcal{D}$ — множество вершин, исходящие степени которых равны нулю.

Смысл пейджранка очень простой. Мы хотим учесть не только количество ссылок на данную страницу веб-графа, но и качество ссылок. Пусть качество — это пейджранк. Тогда качество (пейджранк) страницы $i$ и выражается, грубо говоря, как сумма качеств (пейджранков) страниц $j$, $j \to i$, нормированных на величины $\operatorname{outdeg} j$. При этом надо учитывать, что есть страницы нулевой исходящей степени. Величина $1 - c$ называется *вероятностью телепортации*. Идея в том, что мы можем представлять себе человека, который блуждает по сети: с вероятностью $c$ он на очередном шаге переходит по ссылке, а с вероятностью $1 - c$ совершает «скачок» на случайную страницу.

Пейджранк придумали в конце 90-х годов XX века Л. Пейдж и С. Брин, основатели компании Гугл (см. [8]). Считается, что слово «пейдж» в названии пейджранка происходит не от английского «страница», а именно от фамилии первого автора. Тем не менее, часто, когда аналогичное определение дается для хостграфа, соответствующую характеристику называют *хостранком*. Просто так удобнее различать два объекта.

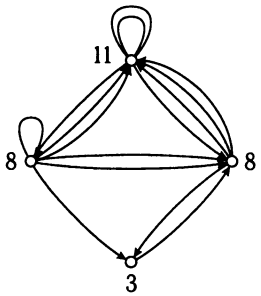
Пейджранк и хостранк подчиняются степенным законам, параметры которых очень близки к параметрам, задающим распределения степеней вершин веб- и хост-графа. Однако корреляция между ранком и степенью вершины низкая (см. [9]). Иными словами, бывают вершины большой степени и с малым пейджранком, а бывают вершины относительно малой степени с большим пейджранком. Именно в этом смысле пейджранк учитывает всю локальную структуру — не только степени, но и вторые степени, и т. д.

И пейджранк, и хостранк сыграли значительную роль в задачах информационного поиска. В частности, они служат сильными факторами, которые применяются для ранжирования документов по поисковому запросу.

В последнее время классические пейджранки практически не употребляются. Однако наука об их различных аналогах разрослась непомерно, и современные пейджранки по-прежнему активно используются поисковыми системами.

## 1.9. КОЛИЧЕСТВО РЕБЕР МЕЖДУ ВЕРШИНАМИ ЗАДАННЫХ СТЕПЕНЕЙ

Следующая характеристика веб- и хост-графа возникла совсем недавно и была мотивирована совершенно прикладными вопросами. Пусть $G_n = (V_n, E_n)$ — наш граф. Пусть $d_1, d_2 \in \mathbb{N}$. Определим $X_n(d_1, d_2)$ как общее число ребер между вершинами (суммарных) степеней $d_1$ и $d_2$ в $G_n$. Например, на рис. 2 изображен граф с четырьмя вершинами, степени которых суть 11, 8, 8, 3. Тогда для него $X_4(3,8) = 3$, $X_4(8,8) = 6$, $X_4(8,11) = 7$, $X_4(3,11) = 0$, $X_4(11,11) = 4$. Тут главное понимать, что при подсчете ребер между вершинами одинаковой степени мы каждое ребро учитываем дважды. Иными словами, мы суммируем по всем упорядоченным парам вершин количества ребер между ними, и в это суммирование мы включаем даже пары совпадающих вершин, дабы не упустить петли (петля — это, конечно, одно ребро, которое, однако, дает вклад 2 в суммарную степень).




$X(3,8) = 3$
$X(8,8) = 6$
$X(11,8) = 7$
$X(11,3) = 0$
$X(11,11) = 4$

**Рис. 2**

Разумеется, $X_n(d_1, d_2)$ — достаточно нетривиальная величина. Из нее легко получить, скажем, количество вершин степени $d \in \mathbb{N}$:

$$|\{v \in V_n\colon \deg v = d\}| = \frac{1}{d}\sum_{d_1} X_n(d_1, d).$$

А стало быть, зная ее распределение, можно узнать и распределение степеней вершин. Естественно, обратное сделать нельзя.

О распределении величины $X_n(d_1, d_2)$ мы поговорим подробнее в п. 2.6.9 (см. также [7]). А о том, откуда она возникла, мы скажем в последнем — 12-м — разделе этой главы.

## 1.10. КОРРЕЛЯЦИИ СТЕПЕНЕЙ ВЕРШИН

Еще одна важная характеристика сети — это, грубо говоря, ожидаемая средняя степень соседей случайной вершины степени $d$. Ее обозначают $d_{nn}(d)$ ($nn$ — это не число, но сокращение от «nearest neighbor», т. е. «ближайший сосед»), и ее определение будет гораздо понятнее, если дать его сперва для графа $G = (V, E)$ без петель, кратных ребер и ориентации:

$$d_{nn}(d) = \frac{1}{d \cdot |\{i\colon \deg i = d\}|} \sum_{i\colon \deg i = d} \sum_{j\colon \{i,j\} \in E} \deg j.$$

Если в графе есть петли, кратные ребра и ориентация, то нормировка и внешнее суммирование остаются прежними (берется именно суммарная степень), а внутреннее суммирование усложняется. Пусть фиксирована вершина $i$ степени $d$. Если раньше мы брали ее соседей и у каждого смотрели степень, то теперь нужно брать отдельно каждое ребро, смежное с $i$, и смотреть степень второго его конца. Иными словами,

величина $\deg j$ входит во внутреннее суммирование столько раз, сколько есть ребер с концами $i$ и $j$ в $E$. При этом, если есть петли $(i,i) \in E$, то каждая из них учитывается дважды. Например, для графа с рис. 2 и $d = 8$

$$d \cdot |\{i\colon \deg i = d\}| = 16,$$
$$16 d_{nn}(8) = (4 \cdot 11 + 2 \cdot 3 + 2 \cdot 8) + (3 \cdot 11 + 2 \cdot 8 + 1 \cdot 3 + 2 \cdot 8) = 134,$$

а для того же графа и $d = 11$

$$d \cdot |\{i\colon \deg i = d\}| = 11, \quad 11 d_{nn}(11) = 3 \cdot 8 + 4 \cdot 8 + 4 \cdot 11 = 100.$$

Очень удобно выразить величину $d_{nn}$ в общем случае через введенную в предыдущем разделе величину $X_n(d_1, d_2)$, которая оказывается, таким образом, не только важной надстройкой над числом вершин данной степени, но еще и не менее важной надстройкой над классической величиной $d_{nn}$:

$$d_{nn}(d) = \frac{\sum\limits_{d_1} d_1 X_n(d_1, d)}{\sum\limits_{d_1} X_n(d_1, d)}.$$

Величина $d_{nn}$ — это своего рода корреляция степеней. Часто ее называют *ассортативностью* вслед за М. Ньюманом, который ввел ее в [10]. Обычно в реальных сетях $d_{nn}(d) \sim d^{\delta}$. Если $\delta < 0$, то сеть называется *дисассортативной*; если $\delta > 0$, то сеть называется *ассортативной*. Веб-граф — это дисассортативная сеть, в то время как социальные сети, как правило, ассортативны (см. [11]).

## 1.11. КЛАСТЕРНЫЕ КОЭФФИЦИЕНТЫ

С корреляциями степеней вершин тесно связано понятие *кластерного коэффициента*. Идеологически кластерный коэффициент — это вероятность того, что соседи случайной вершины графа сами соединены ребром («знакомые случайного человека сами между собой знакомы»).

Существует множество различных кластерных коэффициентов. Из них наиболее употребительны два, и именно ими мы ограничимся. Как и в предыдущем разделе, нам проще будет давать определения сперва для графов без петель, кратных ребер и ориентации. Позже мы поговорим о возможных обобщениях, более близких к практическим нуждам.

Итак, пусть дан простой граф $G = (V, E)$, $|V| = n$. Пусть $v \in V$. Обозначим $N_v$ множество соседей вершины $v$ в графе $G$. Положим $n_v = |N_v|$,

т. е. $n_v = \deg v$. Если $n_v \geqslant 2$, то назовем *кластерным коэффициентом вершины* $v$ величину

$$C_v = \frac{|\{\{x,y\} \in E\colon x,y \in N_v\}|}{C_{n_v}^2}.$$

Иными словами, $C_v$ — это отношение реального числа ребер среди соседей $v$ и их максимально возможного количества.

Теперь можно двумя принципиально различными способами усреднить величину $C_v$ по всем вершинам графа. С одной стороны, можно взять

$$T(G) = \frac{\sum\limits_{v \in V} C_{n_v}^2 C_v}{\sum\limits_{v \in V} C_{n_v}^2}.$$

Нетрудно видеть, что если $\sharp(K_3, G)$ — это число треугольников (полных графов $K_3$) в графе $G$, а $\sharp(P_2, G)$ — число цепей длины 2, то

$$T(G) = \frac{3\sharp(K_3, G)}{\sharp(P_2, G)}. \tag{1}$$

С другой стороны, можно положить

$$C(G) = \frac{1}{n} \sum_{v \in V} C_v.$$

Величина $T(G)$ — это *глобальный кластерный коэффициент*, или *транзитивность*, а величина $C(G)$ — это *средний локальный кластерный коэффициент*, предложенный в [5].

Вообще говоря, величины $T$ и $C$ могут сильно разниться. В статье [12] приведен соответствующий пример. А именно, пусть $G$ — это полный двудольный граф $K_{2,n-2}$, к которому добавлено ребро между вершинами из доли размера 2. У него $n$ вершин и $2n-3$ ребра. Легко посчитать в нем и треугольники, и цепи длины 2, и кластерные коэффициенты всех вершин. Получится, что $C(G) \sim 1$ при $n \to \infty$, а $T(G) = \Theta(1/n)$.

Если $G$ — ориентированный граф без петель и кратных ребер, то легко модифицировать определения $T$ и $C$: достаточно лишь заменить нормировку $C_{n_v}^2$ нормировкой $n_v(n_v - 1)$. Если в $G$ есть кратные ребра и/или петли, то нет проблем с отысканием $T(G)$ по формуле (1). Однако величину $C(G)$ в этом случае определить затруднительно.

На практике имеет место довольно серьезная путаница. Про кластерные коэффициенты пишутся тысячи статей, но крайне редки случаи, когда авторы четко говорят, какой именно кластерный коэффициент они считают и как они поступают с кратными ребрами, петлями

и пр. Тем не менее, как правило, считают $C(G)$ (это намного проще с вычислительной точки зрения) и при этом, по-видимому, пренебрегают петлями и кратными ребрами. Большинство специалистов согласны с тем, что в веб- и хост-графах $C(G) = \Theta(1)$: по разным данным $C(G)$ варьируется от 0,1 до 0,3 (см., например, [11,13,14]). Это означает, что вероятность наличия ссылки между сайтами (страницами), на которые ссылается случайный сайт (случайная страница), крайне велика!

С транзитивностью все несколько хуже. Конечно, она меньше, чем средний локальный кластерный коэффициент. Однако при ее отыскании совсем нелепо пренебрегать кратностями ребер. И в то же время подсчет точного количества треугольников в графах с миллиардами вершин при нынешних вычислительных мощностях невозможен. Огромное число работ посвящено построению приближенных алгоритмов (см., например, [15]), и почти в каждой из них алгоритм применяется к *куску* веб- или хост-графа. В обзоре [13] в таблице Table II указано, что $T(WWW) = 0{,}11$, но никаких подробностей нет. Иначе говоря, есть некоторые основания предполагать, что транзитивность для веба так же асимптотически постоянна, как и величина $C(G)$. Но совсем серьезных исследований на эту тему нет.

Заметим еще, что в реальных сетях (например, в различных *кусках* веба) могут быть различные транзитивности при данном среднем локальном коэффициенте. Интересное исследование на эту тему можно найти в [16].

## 1.12. ЧИСЛО КОПИЙ ФИКСИРОВАННОГО ГРАФА

В предыдущем разделе мы уже видели, как важен подсчет числа треугольников и двухзвенных цепей в веб- и хост-графах. Сколь бы приблизительны ни были известные данные, все они свидетельствуют о том, что треугольников в наших графах на порядок больше, чем вершин. По-видимому, это соотношение только растет с ростом веба. Есть основания полагать, что если $n$ — число вершин, то $\sharp(K_3, G) \approx n^{\alpha}$, где $\alpha \approx 2$.

Разумеется, треугольниками дело не ограничивается. В принципе, интерес представляет любой фиксированный граф $H$ и число $\sharp(H, G)$ его вхождений в $G$. Проблема лишь в недостаточных вычислительных мощностях.

С точки зрения кластерных коэффициентов и их обобщений изучают циклы и цепи произвольной длины. Еще один важный класс графов, которые часто встречаются в Интернете, образуют двудольные графы. Во-первых, это связано с тем, что в Интернете постоянно возникают

группы владельцев сайтов, которых объединяет какой-то общий интерес: например, автомобили, фильмы, спиртное и т. д. Не мудрено, что те владельцы, которые любят, скажем, советское кино, ставят на своих сайтах более или менее одинаковые ссылки — ссылки на наиболее популярные сайты о кино, на базы данных, откуда это кино можно скачать, и пр. Получается двудольный граф, в одной доле которого сайты любителей фильмов, а в другой — их любимые сайты. Во-вторых, двудольные графы в Интернете служат отличными средствами раскрутки сайтов. В этом случае одна доля состоит из сайтов, продающих свои ссылки, а другая — из сайтов, владельцы которых хотят таким способом искусственно увеличить популярность своих детищ.

Проблема поиска двудольных структур в вебе очень серьезная. Ей посвящена масса работ (см., например, [17]). Однако не менее важна и проблема автоматического различения естественных сообществ и спамерских конструкций. Разумеется, в полной мере ее решить невозможно, но ею все равно занимаются, и, в частности, величина $X_n(d_1, d_2)$ из разд. 1.9 появилась именно в результате попытки осуществить классификацию двудольных графов в Интернете. Более подробно идею той классификации мы опишем в п. 2.6.10.

ГЛАВА 2

# МОДЕЛИ ХОСТ-ГРАФОВ

## 2.1. ОБЩАЯ КОНЦЕПЦИЯ

В первой главе мы продемонстрировали, насколько нетривиально устроена сеть Интернет. Оказывается, это разреженный граф, у которого, тем не менее, маленький диаметр; степени его вершин подчиняются весьма специфическому закону распределения; у него весьма большие кластерные коэффициенты, он дисассортативен, и т. д., и т. п. Разумеется, хороший исследователь не может просто накапливать подобные знания: настоящая наука начинается тогда, когда предпринимается попытка *объяснить* природу тех или иных явлений, понять, какие механизмы ими управляют. Для этого и нужны модели.

Что такое модель хост-графа? Более или менее простая и естественная модель хост-графа предполагает, что в каждый момент времени $t \in \mathbb{N}$ хост-граф — это случайный элемент из некоторого множества $\Omega_t = \{G = (V, E)\}$, состоящего из графов на $n = n(t)$ вершинах. Любая модель характеризуется зависимостью $n$ от $t$, содержимым множества $\Omega_t$, а также распределением нашего случайного элемента, т. е. вероятностями, с которыми этот элемент оказывается равным тем или иным конкретным графам $G$ из $\Omega_t$. В науке наш случайный элемент принято называть *случайным графом*.

Модель тем адекватнее реальности, чем больше свойств из гл. 1 выполняется в ней с вероятностью, стремящейся к 1 при $t \to \infty$.

Такое понимание модели отлично согласуется с общим статистическим подходом, изложенным в разд. 1.1. Проблема, однако, в том, что от модели еще хочется максимальной *простоты*. Хочется верить в то, что все многообразие закономерностей, которым подчиняется Интернет, удастся с достаточной точностью описать за счет нескольких сравни-

тельно простых механизмов, и эта вера, как мы скоро увидим, небеспочвенна. Тем удивительнее, конечно, будет то, о чем мы напишем ниже.

Еще раз подчеркнем следующий важный момент. Конечно, желая получить хорошую модель хост-графа, мы можем взять $\Omega_t$ состоящим из графов, которые обладают всеми известными свойствами реальной сети, и извлекать элементы из $\Omega_t$ согласно равномерному распределению (с одной и той же вероятностью $1/|\Omega_t|$). Такой подход, однако, ничего не объясняет и крайне тяжел для реализации: пойди перебери все графы с кучей нетривиальных свойств! Частично этот подход реализован в моделях, опирающихся только на свойство степенного распределения степеней вершин (см., например, [18]). Но ведь и степенной закон взялся не с потолка. Поэтому в дальнейшем мы поговорим о принципиально других моделях — моделях, в рамках которых свойства графов не задаются заранее, но возникают из некоторых естественных предположений о природе становления веба.

В следующем разделе мы приведем пример модели, которая является классической моделью случайного графа, но нужными свойствами не обладает.

## 2.2. МОДЕЛЬ ЭРДЕША–РЕНЬИ

Модель, о которой мы поговорим в этом разделе, была предложена в конце 50-х годов XX века. Ее принято называть моделью Эрдеша–Реньи, поскольку именно П. Эрдеш и А. Реньи подвергли ее по-настоящему глубокому анализу (см. [19–21]), в результате чего возникла вся современная теория случайных графов. До того были лишь разрозненные наблюдения и результаты.

Подробнейшую историю науки о случайных графах с массой результатов, проблем и методов можно найти, например, в книгах [22–25]. Здесь мы лишь поймем, как устроена модель Эрдеша–Реньи и как далека она, к сожалению, от того, что нам нужно.

Итак, в модели Эрдеша–Реньи случайный граф — это случайный элемент со значениями во множестве

$$\Omega_n = \{G = (V_n, E)\colon V_n = \{1, 2, \ldots, n\}\},$$

которое состоит из всех $n$-вершинных графов без петель, кратных ребер и ориентации, так что $|\Omega_n| = 2^{C_n^2}$. Распределение здесь биномиальное, т. е.

$$\mathbb{P}(G) = p^{|E|}(1-p)^{C_n^2 - |E|}, \quad p = p(n) \in [0, 1].$$

Иными словами, можно считать, что каждое ребро случайного графа проводится независимо от всех остальных ребер с одной и той же вероятностью $p$, величина которой может меняться с течением времени $n$.

Конечно, читатель скажет: «Очевидно, описанная модель не имеет никакого отношения к веб-графам. Нет ни кратных ребер, ни петель, ни ориентации. Да и связи в Интернете вряд ли совсем не зависят друг от друга.» Это верно, но, во-первых, кратные ребра, петли и ориентация — всем этим и на практике часто пренебрегают (ср. разд. 1.11), а во-вторых, есть у модели Эрдеша–Реньи и определенная гибкость: вероятность ребра может зависеть от времени. Поэтому стоит все же повнимательнее отнестись к модели.

Начнем с числа ребер. Его математическое ожидание за счет линейности считается мгновенно:

$$\mathbb{E}|E| = C_n^2 p \sim \frac{n^2 p}{2}.$$

Нам нужно, чтобы число ребер было примерно $\mathrm{const} \cdot n$. Ясно, что тогда нам следует брать $p \sim c/n$, где $c$ — константа.

Теперь обсудим распределение степеней вершин. Очевидно, для любого $i \in V_n$

$$\deg i \sim \mathrm{Binom}(n-1, p),$$

т. е. степень вершины имеет биномиальное распределение с «числом испытаний» $n-1$ и «вероятностью успеха» $p$: просто каждый из $n-1$ потенциальных соседей $i$ может стать таковым независимо от остальных с вероятностью $p$.

Хорошо известно, что при $p \sim c/n$ биномиальное распределение аппроксимируется пуассоновским. А именно, верна

**Теорема 4.** *Пусть $p \sim c/n$ и $\xi_n \sim \mathrm{Binom}(n-1, p)$. Тогда для любого $k$ при $n \to \infty$ имеет место асимптотика*

$$\mathbb{P}(\xi_n = k) \sim \frac{c^k e^{-c}}{k!}.$$

Доказательство. В самом деле,

$$\mathbb{P}(\xi_n = k) = C_{n-1}^k p^k (1-p)^{n-1-k} \sim \frac{n^k}{k!}\left(\frac{c}{n}\right)^k e^{-\frac{c}{n}(n-1-k)} \sim \frac{c^k e^{-c}}{k!}.$$

И вот это уже ставит крест на возможности использования модели Эрдеша–Реньи для Интернета: ввиду линейности математического ожидания

$$\mathbb{E}\left(\frac{|\{i\colon \deg i = k\}|}{n}\right) = \mathbb{P}(\deg i = k) \sim \frac{c^k e^{-c}}{k!},$$

а функция $c^k e^{-c}/k!$ совсем далека от степенной функции $\mathrm{const}/k^\alpha$.

У модели Эрдеша–Реньи есть естественное обобщение. Можно проводить ребра по-прежнему взаимно независимо, но вероятности ребер можно сделать зависящими от их концов. Понятно, что тут возникает масса параметров, и модель становится непомерно сложной. Впрочем, и тут степенного закона при $n \to \infty$ не будет.

## 2.3. МОДЕЛИ БАРАБАШИ–АЛЬБЕРТ

В 1999 г. Л.-А. Барабаши и Р. Альберт опубликовали работу [1], в которой предложили перенести идею *предпочтительного присоединения* (и до того возникавшую в разных работах в связи с исследованиями других сетей) на случай веб-графа. Интуитивно идея крайне простая, красивая и естественная. Она состоит в том, что новый сайт, появляясь на свет, *предпочтет* сослаться на сайт, у которого и так высокий индекс цитирования: так сказать, «к деньгам деньги» (вернее, к ссылкам ссылки).

Барабаши и Альберт писали примерно так: «...стартуя с малого числа ($m_0$) вершин, мы в каждый момент времени добавляем новую вершину с $m$ ($\leqslant m_0$) выходящими из нее ребрами, которые соединяют эту новую вершину с $m$ различными вершинами, уже присутствующими в системе. Дабы включить в модель предпочтительное присоединение, мы предполагаем, что вероятность $\Pi$ того, что новая вершина присоединится к некоторой вершине $i$, зависит от степени $k_i$ этой вершины, причем $\Pi(k_i) = k_i / \sum_j k_j$. После $t$ шагов модель порождает случайную сеть с $t + m_0$ вершинами и $mt$ ребрами».

Понятно, что отношение числа ребер к числу вершин в конструкции Барабаши–Альберт заведомо правильное: оно равно примерно $m$, где $m \in \mathbb{N}$, как и хотелось. Однако дальнейшей ясности нет. В статье [12] Б. Боллобаш подверг серьезной критике приведенную нами цитату. И был, несомненно, прав, так как уже к тому времени было написано множество статей об эмпирических свойствах модели, и никто не удосужился пояснить, как именно он эту модель конкретизировал.

Приведем основные возражения Боллобаша.

1. Похоже, что начальный граф $G_0$ на $m_0$ вершинах авторам модели безразличен. Вернее, кажется, что в нем нет ребер. Но в таком случае как мы будем делить что-либо на сумму степеней вершин, которая здесь, на старте, равна нулю? Если же в начале процесса каждая из $m_0$ вершин имеет степень не меньше единицы, то снова зависимость от $G_0$ весьма существенная. Например, если $m = 1$, то процесс Барабаши–Альберт задает дерево (каждая новая вершина отправляет ребро в одну из своих предшественниц, и циклы не

возникают)... — а дерево ли? Не совсем! Дерево будет лишь при условии, что $G_0$ — дерево. И таких проблем масса.

2. С самим принципом предпочтительного присоединения не все ясно. Тут проблемы лишь при $m \geqslant 2$. А именно, когда мы добавляем очередную вершину (скажем, с номером $t+1$), то мы должны, следуя Барабаши и Альберт, соединить ее со случайным множеством $N_{t+1}$, состоящим из $m$ прежде имевшихся вершин (как мы помним, эти $m$ вершин должны быть различными). Если считать, как обычно, что вершины графа нумеруются натуральными числами и что граф, построенный до появления вершины $t+1$, — это некий граф $G_t$, то описание Барабаши–Альберт говорит лишь о том, что для любого $i \in \{1, \ldots, t\}$

$$\mathbb{P}(i \in N_{t+1}) = \frac{m \deg_{G_t} i}{\sum\limits_{j=1}^{t} \deg_{G_t} j}. \tag{2}$$

Конечно, в формуле Барабаши–Альберт не было сомножителя $m$. Если допустить, что ребра можно добавлять последовательно и независимо друг от друга, то получится в точности формула (2). Но и этого нельзя, ведь нам точно сказано, что $m$ вершин должны быть различными. В итоге ясно, что ничего лучшего, нежели формула (2), мы из описания Барабаши–Альберт не извлечем. А тогда возникает вопрос: как именно мы конкретизируем распределение самих $N_{t+1}$ среди всех $m$-элементных подмножеств множества вершин графа $G_t$? Это можно сделать изрядным числом способов!

Вообще-то, могло статься, что ничто особенно не зависит от конкретизации распределения $N_{t+1}$. Тогда возражения Боллобаша были бы напрасными. Но не тут-то было. Боллобаш в той же статье [12] доказывает совершенно замечательную теорему.

**Теорема 5.** *Пусть $f(n)$, $n \geqslant 2$, — произвольная функция, принимающая натуральные значения с условиями $f(2) = 0$, $f(n) \leqslant f(n+1) \leqslant f(n)+1$ для любого $n \geqslant 2$, и $f(n) \to \infty$ при $n \to \infty$. Тогда существует процесс, подчиняющийся правилу* (2), *в котором с вероятностью* 1 *при всех достаточно больших $n$ ровно $f(n)$ треугольников.*

Изумительная теорема, особенно если учесть, как важны треугольники для исследований веб-графа. Захотим — смоделируем процесс с $\sim \ln n$ треугольниками, захотим — смоделируем процесс с $\sim \sqrt{n}$ треугольниками. И т. д. Совершенно не ясно, как после этого читать, например, статью [26] тех же Барабаши и Альберт, в которой они, кроме всего прочего, предсказывают, что кластерный коэффициент в их модели поведет себя как $\Theta(1/n^{0,75})$.

Короче, идея предпочтительного присоединения отличная — и естественная, и простая, — но четкой модели пока нет. В следующем разделе мы опишем модель Боллобаша–Риордана, которая максимально близка идеям Барабаши–Альберт и в то же время абсолютно аккуратна.

## 2.4. МОДЕЛЬ БОЛЛОБАША–РИОРДАНА: ОПРЕДЕЛЕНИЯ

### 2.4.1. Динамическое определение модели

Пусть $m \in \mathbb{N}$ — параметр, отвечающий за отношение числа ребер к числу вершин в будущем случайном графе. Сам случайный граф в момент времени $n \in \mathbb{N}$ мы обозначим $G_m^n$. Определим сперва $G_1^n$.

Положим $G_1^1 = (\{1\}, \{(1,1)\})$, т. е. у самого первого графа одна вершина 1 и петля в ней. Допустим, граф $G_1^n$ с некоторым $n \geqslant 1$ построен. Тогда его вершины суть $1, \ldots, n$, и ребер у него тоже $n$ штук. Добавим вершину $n+1$ и ровно одно ребро, выходящее из нее:

$$\mathbb{P}(n+1 \to n+1) = \frac{1}{2n+1}, \quad \mathbb{P}(n+1 \to i) = \frac{\deg_{G_1^n} i}{2n+1}, \quad i = 1, \ldots, n.$$

Иными словами, $n+1$ может сослаться на себя (образовать петлю) с вероятностью $1/(2n+1)$, а может процитировать свою предшественницу $i$, и вероятность последнего исхода пропорциональна степени вершины $i$ в графе $G_1^n$. Таким образом, мы реализуем идею предпочтительного присоединения. Знаменатель имеет величину $2n+1$, так как сумма вероятностей должна равняться единице, а сумма степеней вершин графа $G_1^n$, у которого $n$ ребер, равна, конечно, $2n$.

Для пущей наглядности мы на рис. 3 изобразим одну из траекторий процесса вплоть до $n = 4$. Видно, что получается ориентированный лес с петлями: звучит довольно зловеще, но такова жизнь. Правда, ориентация и здесь весьма условная. Исходящая-то степень у каждой вершины равна 1, и лишь входящая степень варьируется. Ну, другого мы и не ждали в нынешнем простейшем случае.

Пусть теперь $m \geqslant 2$. Рассмотрим граф $G_1^{mn}$. У него вершины $1, 2, \ldots, mn$ и $mn$ ребер. Разобьем вершины на $n$ групп $\{1, \ldots, m\}$, $\{m+1, \ldots, 2m\}$, $\ldots$, $\{(m-1)n+1, \ldots, mn\}$. Назовем эти группы $v_1, \ldots, v_n$ соответственно. Будем считать $v_1, \ldots, v_n$ вершинами нового графа. Проведем между $v_i$ и $v_j$ $(i \leqslant j)$ столько ребер, направленных от $v_j$ к $v_i$, сколько в графе $G_1^{mn}$ было ребер с одним концом в группе $v_j$, а другим в группе $v_i$. Образуется граф $G_m^n$, у которого $n$ вершин, $mn$ ребер, есть кратные ребра и даже кратные петли. Пример изображен на рис. 4.

Опять ориентация графа $G_m^n$ достаточно условная: исходящая степень каждой вершины равна $m$. Но именно это и предлагали Барабаши

и Альберт. Не правда ли, удивительно будет, если столь простая модель окажется в чем-то похожей на реальность?

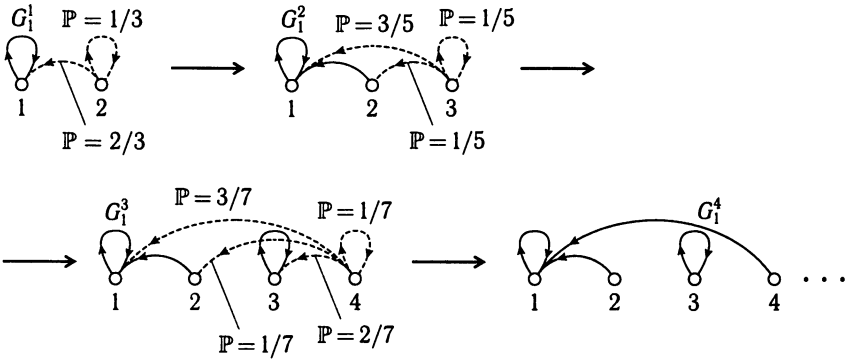



**Рис. 3**

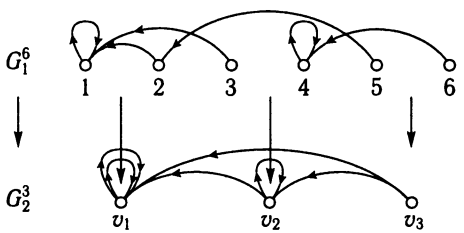



**Рис. 4**

Часто спрашивают, почему Боллобаш и Риордан «схлопывают» в одну именно последовательные вершины графа $G_m^n$. Это обусловлено желанием реализовать идею предпочтительного присоединения не только при $m = 1$, но и при бóльших $m$. Просто граф $G_m^n$ строится так, что с высокой вероятностью его хабы (вершины максимальной степени) — это его первые вершины. И наоборот, чем дальше от начала, тем вероятнее, что степени будут маленькими. Поэтому странно было бы схлопывать как-то по-другому: явно повысилась бы вероятность уйти от предпочтительного присоединения и его ожидаемых последствий, в которые скептически настроенному читателю верится пока с трудом, но которые вскоре появятся.

Имеется еще один естественный вопрос (помимо разных иронических выпадов в стиле «да тут столько натяжек, что это ни за что не сработает!», на которые мы сейчас не реагируем, зная, что при всей

своей искусственности модель Боллобаша–Риордана уже многое объясняет, о чем ниже). Вопрос такой: «А почему бы не добавлять на каждом шаге сразу $m$ ребер последовательно и взаимно независимо? Кратные ребра в модели Боллобаша–Риордана ведь все равно есть.» Ответ очень простой. Дело в том, что у модели Боллобаша–Риордана, которую мы только что определили динамически, есть красивая статическая интерпретация, называемая LCD-моделью. Название происходит от выражения «Linearized Chord Diagram», означающего «Линейная Хордовая Диаграмма». Мы опишем эту модель в следующем пункте.

Напоследок мы все же определим вариант модели Барабаши–Альберт, в котором на каждом шаге добавляется сразу $m$ ребер. По своим свойствам он крайне близок к модели Боллобаша–Риордана, но он нам еще пригодится.

Итак, строим последовательность случайных графов $G_m^{(n)}$ (добавили скобки вокруг $n$ в обозначении, дабы не возникла путаница). Граф $G_m^{(1)}$ — это граф с вершиной 1 и $m$ петлями. Если уже построен граф $G_m^{(n)}$ с вершинами $1, \dots, n$ и $mn$ ребрами, то для построения графа $G_m^{(n+1)}$ добавляем вершину $n+1$ и выпускаем из нее $m$ независимых в совокупности ребер с распределением

$$\mathbb{P}(n+1 \to n+1) = \frac{1}{2n+1}, \quad \mathbb{P}(n+1 \to i) = \frac{\deg_{G_m^{(n)}} i}{m(2n+1)}, \quad i = 1, \dots, n.$$

Все корректно, так как

$$\sum_{i=1}^{n+1} \mathbb{P}(n+1 \to i) = \frac{1}{2n+1} + \frac{2mn}{m(2n+1)} = \frac{m+2mn}{m(2n+1)} = 1.$$

### 2.4.2. Статическое определение модели

Новому определению подлежат только графы $G_1^n$. Графы $G_m^n$ получаются из них обычной склейкой вершин.

Прежде всего назовем *линейной хордовой диаграммой* произвольный объект типа того, что изображен на рис. 5. А именно, мы ставим числа $1, 2, \dots, 2n$ на прямую и в верхней полуплоскости соединяем пары этих чисел дугами так, чтобы у дуг были разные концы и свободных концов не оставалось. Иначе говоря, мы берем произвольное неупорядоченное разбиение множества наших чисел на $n$ непересекающихся пар и каждой паре в разбиении сопоставляем дугу. Всего таких разбиений $t_n = \dfrac{(2n)!}{n!\,2^n}$.

По данной линейной хордовой диаграмме легко построить граф на $n$ вершинах. Двигаемся вдоль диаграммы слева направо, пока не встретим первый правый конец $i_1$ какой-то дуги. Называем вершиной $v_1$ графа группу чисел $\{1, \ldots, i_1\}$. Аналогично находим второй правый конец $i_2$ некоторой дуги и полагаем $v_2 = \{i_1 + 1, \ldots, i_2\}$. И т. д. В итоге у нас возникают $n$ вершин $v_1, \ldots, v_n$. Ребро соединяет $v_j$ с $v_i$ тогда и только тогда, когда между соответствующими группами есть дуга. Например, на рис. 6 изображен граф, построенный по диаграмме с рис. 5.

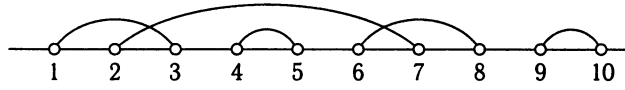



**Рис. 5**

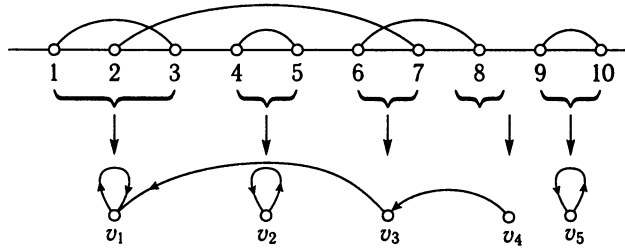



**Рис. 6**

В итоге случайный граф $G_1^n$ получается из случайной хордовой диаграммы, т. е. из диаграммы, вероятность которой полагается равной $1/t_n$. Почему мы получили в точности тот же объект, что и прежде? Очень просто. Снова последовательно будем рассматривать правые концы дуг при движении слева направо вдоль диаграммы. Рассмотрим первый правый конец $i_1$. Из этой точки идет одна дуга, и она соответствует петле в первой вершине. Далее, пусть каким-то образом проведены дуги из правых концов $i_1, \ldots, i_k$. Рассмотрим $i_{k+1}$. Вероятность того, что из вершины $v_{k+1}$ идет ребро в вершину $v_t$, пропорциональна количеству отрезков, на которые в данный момент разбит отрезок $[i_{t-1}, i_t]$, ведь ровно столько есть соответствующих разбиений на пары. А количество отрезков — это и есть степень вершины $v_t$ в данный момент. Таким образом, имеем действительно $G_1^n$.

Отметим, что линейные хордовые диаграммы весьма важны для топологии, благодаря чему они глубоко изучены (см. [27]). Впрочем, для наших целей это не существенно.

## 2.5. МОДЕЛЬ БОЛЛОБАША–РИОРДАНА: РЕЗУЛЬТАТЫ

### 2.5.1. Гигантская компонента, устойчивость и уязвимость

Модель Боллобаша–Риордана исключительно проста, и потому ее активно и всесторонне изучали. Более того, если бы с ходу не оказалось, что она очень многое в Интернете объясняет, то, наверное, ее бы забросили. Однако психологический, по сути, закон предпочтительного присоединения оказался на удивление значимым с точки зрения понимания механизма становления веба.

В этом пункте мы сформулируем результаты, которые свидетельствуют о том, что в случайном графе Боллобаша–Риордана есть гигантская компонента и, что еще замечательнее, что этот граф, подобно реальному хост-графу, устойчив к случайным ошибкам, но уязвим по отношению к атакам на хабы (ср. теоремы 1 и 2).

Следующая теорема доказана в статье [28].

**Теорема 6.** *Пусть $m \geqslant 2$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, а $p \in [0,1)$ — фиксированное число. Пусть $G_{m,p}^n$ — случайный граф, полученный из $G_m^n$ удалением каждой вершины с вероятностью $p$ независимо от остальных вершин. Тогда существует такая константа $c = c(m,p)$, что с вероятностью, стремящейся к единице при $n \to \infty$, в $G_{m,p}^n$ есть ровно одна компонента размера $(c + o(1))n$.*

Теорема утверждает, что в случайном графе в модели Боллобаша–Риордана гигантская компонента есть и что случайный граф устойчив к случайному уничтожению вершин. Разумеется, как и в реальности, $c$ тем меньше, чем ближе $p$ к единице. Авторы теоремы — Боллобаш и Риордан — оценили скорость стремления к нулю величины $c(m,p)$ при каждом данном $m$ функцией вида $e^{-\mathrm{const}(m)/(1-p)}$. Для «практической проверки» этого результата данных заведомо не хватает. Тем не менее, как минимум, качественная близость модели к этому аспекту реальности очевидна.

В той же статье [28] доказана и теорема об уязвимости.

**Теорема 7.** *Пусть $m \geqslant 2$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, а $c \in (0,1)$ — фиксированное число. Пусть $G_{m,c}^n$ — случайный граф, полученный из $G_m^n$ удалением первых $[cn]$ вершин. Положим $c_m^* = \dfrac{m-1}{m+1}$. Тогда при $c < c_m^*$ с вероятностью, стремящейся к единице при $n \to \infty$, в $G_{m,c}^n$ есть гигантская компонента. Если же $c > c_m^*$, то с вероятностью, стремящейся к единице при $n \to \infty$, в $G_{m,c}^n$ гигантской компоненты нет.*

Опять же, трудно сравнить с эмпирическими данными величину $c_m^*$, но качественно все идеально соответствует наблюдениям: даже фазовый переход есть.

### 2.5.2. Диаметр

И здесь все замечательно. Снова сами авторы модели доказали следующий результат (см. [29]).

**Теорема 8.** *Пусть $m \geqslant 2$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, а $\varepsilon > 0$ — сколь угодно малое фиксированное число. Тогда*

$$\mathbb{P}\left((1-\varepsilon)\frac{\ln n}{\ln\ln n} \leqslant \operatorname{diam} G_m^n \leqslant (1+\varepsilon)\frac{\ln n}{\ln\ln n}\right) \to 1, \quad n \to \infty.$$

Пафос результата в том, что даже при нынешних размерах хост-графа величина $\dfrac{\ln n}{\ln\ln n}$ совсем невелика: она не превосходит семи, и это крайне похоже на эмпирические 5–6. Правда, есть гипотеза, что в реальности диаметр вовсе не растет, но проверить эту гипотезу в сочетании с альтернативой $\operatorname{diam} G \approx \dfrac{\ln n}{\ln\ln n}$ не представляется возможным. Пока что совпадение просто фантастическое!

### 2.5.3. Степени вершин

Тут история чуть более сложная. Сперва классики — Боллобаш и Риордан — совместно со Спенсером и Тушнади опубликовали статью [30], в которой доказали следующую теорему.

**Теорема 9.** *Пусть $m \geqslant 1$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, а $\varepsilon > 0$ — сколь угодно малое фиксированное число. Тогда для любого $d \leqslant n^{1/15}$*

$$\mathbb{E}\left|\left\{i = 1, \ldots, n\colon \deg_{G_m^n} i = d\right\}\right| \sim \frac{2mn(m+1)}{d(d+1)(d+2)}, \quad n \to \infty,$$

$$\mathbb{P}\left((1-\varepsilon)\frac{2mn(m+1)}{d(d+1)(d+2)} \leqslant \left|\left\{i = 1, \ldots, n\colon \deg_{G_m^n} i = d\right\}\right| \leqslant (1+\varepsilon)\frac{2mn(m+1)}{d(d+1)(d+2)}\right) \to 1, \quad n \to \infty.$$

С одной стороны, теорема замечательная. Из нее следует, что случайный граф Боллобаша–Риордана в асимптотике подчиняется степенному закону, ведь с вероятностью, стремящейся к единице при $n \to \infty$, доля вершин степени $d$ в этом графе примерно равна

$$\frac{2m(m+1)}{d(d+1)(d+2)} \approx \frac{2m(m+1)}{d^3}.$$

С другой стороны, есть две проблемы. Во-первых, показатель степенного закона равен трем, а не ожидаемой величине $\gamma \in (2,3)$: вроде и попадание опять почти «в яблочко», но слишком уж важен показатель $\gamma$ для реальных сетей, и ясно, что варьировать его модель не позволяет. Во-вторых, в теореме присутствует ограничение $d \leqslant n^{1/15}$, которое выглядит ужасно как с теоретической, так и — тем более — с практической точки зрения: даже для графа с миллиардами вершин корень пятнадцатой степени сохраняет лишь величины $d < 5$, что нелепо.

Об устранении первой проблемы мы поговорим в разделах 2.6 и 2.7. А по второй проблеме было некоторое количество работ, приведших в итоге к ее полному разрешению. Видимо, самый точный результат доказан в статье [31], но он слегка громоздкий, и мы приведем здесь почти столь же сильный результат, полученный недавно Е. А. Гречниковым в качестве простого частного случая более общей и существенно более сложной теоремы о распределении величины $X_n(d_1, d_2)$ (ср. разд. 1.9 и п. 2.5.6, а также [32]).

**Теорема 10.** *Пусть $m \geqslant 1$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, а $\varepsilon > 0$ — сколь угодно малое фиксированное число. Пусть $\mathbb{I}(A)$ — индикатор свойства $A$, т. е. $\mathbb{I}(A) = 1$, если $A$ верно, и $\mathbb{I}(A) = 0$ иначе. Тогда для любого $d$*

$$\mathbb{E}\left|\left\{i = 1, \ldots, n\colon \deg_{G_m^n} i = d\right\}\right| = \\ = \mathbb{I}(d \geqslant m)\frac{(2mn+1)(m+1)}{d(d+1)(d+2)} - \frac{\mathbb{I}(d=m)}{m} + O_m\left(\frac{d}{n}\right).$$

В теореме Гречникова получена *трехчленная* асимптотика математического ожидания числа вершин степени $d$ без каких-либо ограничений на $d$. В гл. 3 мы расскажем об идеях доказательства Боллобаша с соавторами и об идеях доказательства Гречникова.

### 2.5.4. Вторые степени вершин

Если с первыми степенями вершин случайного графа в модели Боллобаша–Риордана еще сравнительно легко работать, то со вторыми степенями все намного сложнее. Поэтому как следует они изучены лишь при $m = 1$. В этом случае определение второй степени вершины $t$ становится довольно однозначным:

$$d_2(t) = \left|\left\{\{i,j\}\colon i \neq t, j \neq t, \{i,t\} \in E(G_1^n), \{i,j\} \in E(G_1^n)\right\}\right|.$$

Иными словами, вторая степень вершины $t$ — это количество ребер, которые инцидентны соседям $t$, кроме тех, концом которых является $t$. Следует уточнить, что если в вершине $i$, соединенной с $t$, есть петля,

то она дает вклад 2 во вторую степень (в определении пара $\{i, i\}$ учитывается дважды). На рис. 7 приведен пример графа $G_1^6$, и у каждой его вершины написана ее вторая степень. Подчеркнем, что при $m = 1$ мы всегда получаем лес с петлями.

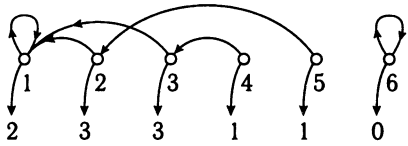



**Рис. 7**

Данное определение с подачи автора этой книги изучала Л. А. Остроумова, которая при участии Гречникова доказала в [33] следующую теорему.

**Теорема 11.** *Для любого* $d > 0$

$$\mathbb{E}|\{i = 1, \ldots, n\colon d_2(i) = d\}| = \frac{4n}{d^2}\left(1 + O\left(\frac{\ln^2 d}{d}\right) + O\left(\frac{d^2}{n}\right)\right).$$

Видно, что теорема дает асимптотику математического ожидания при всех растущих $d$, квадрат которых мал по сравнению с $n$. Сама эта асимптотика является степенной с показателем 2. Разумеется, этот факт еще не означает наличия асимптотического степенного закона, которому бы подчинялись вторые степени вершин. Но степенной закон есть, и его можно сформулировать следующим образом (см. [33]).

**Теорема 12.** *Для любого* $\delta > 0$, *любого* $\varepsilon > 0$ *и* $d \in \{1, \ldots, n^{1/6-\delta}\}$

$$\mathbb{P}\left((1-\varepsilon)\frac{4n}{d^2} \leqslant |\{i = 1, \ldots, n\colon d_2(i) = d\}| \leqslant (1+\varepsilon)\frac{4n}{d^2}\right) \to 1, \quad n \to \infty.$$

Как видно, здесь есть ограничение $d \leqslant n^{1/6-\delta}$. Оно не столь ужасно, как ограничение в теореме 9, но это и не случай теоремы 10, где ограничений вовсе нет. Граница для $d$ позднее была уточнена в более общей модели. Мы об этом еще поговорим в п. 2.6.5.

Существенно более общие, но и гораздо менее точные результаты о структуре окрестностей данной вершины случайного дерева можно найти в статье [34].

### 2.5.5. Пейджранк

Непосредственно в модели Боллобаша–Риордана $G_m^n$ (или в ее модификации $G_m^{(n)}$) пейджранк не считали, так как работать с петлями в этом контексте крайне неприятно. С целью некоторого упро-

щения ситуации в работе [35] был предложен еще один вариант конкретизации модели Барабаши–Альберт, который очень похож на известные нам модели, но все же отличается от них.

Итак, как обычно, добавляем в каждый момент времени одну вершину и $m$ ребер. Стартуем с вершины 0. Никаких петель в ней нет. Однако по определению полагаем вес вершины 0 равным $m$. Затем берем вершину 1. Ей ничего не остается, как поставить $m$ ссылок на 0. Полагаем вес вершины 0 равным $2m$, а вес вершины 1 равным $m$. Дальнейшие вершины, появляясь на свет, выставляют свои $m$ ссылок независимо, каждую с вероятностью, пропорциональной весам существующих вершин, которые равны сумме их входящих степеней и $m$:

$$\mathbb{P}(n+1 \to i) = \frac{\operatorname{indeg} i + m}{\sum\limits_{k=0}^{n}(\operatorname{indeg} k + m)} = \frac{\operatorname{indeg} i + m}{2mn + m}.$$

При $m = 1$ это уже не леса с петлями, но просто деревья. Впрочем, авторы статьи [35] доказывают, что если $\pi_v^m(n)$ — это пейджранк вершины $v$ в случайном графе, построенном к моменту времени $n$, то $\pi_v^m(n) = \pi_v^1(n)$, так что можно обозначить пейджранк просто $\pi_v(n)$. Правда, не стоит забывать о параметре $c$, отвечающем за телепортацию: от него-то зависимость никуда не денется.

Основной результат статьи [35] — это

**Теорема 13.** *При $v > 0$*

$$\mathbb{E}\pi_v(n) = \frac{1-c}{1+n}\left(\frac{1}{1+c} + \frac{c\Gamma\left(v+\frac{1}{2}\right)\Gamma\left(n+\frac{c}{2}+1\right)}{(1+c)\Gamma\left(v+\frac{c}{2}+1\right)\Gamma\left(n+\frac{1}{2}\right)}\right) \approx$$

$$\approx \frac{1-c}{1+n}\left(\frac{1}{1+c} + \frac{c}{1+c}\left(v+\frac{1}{2}\right)^{-\frac{1+c}{2}}\left(n+\frac{1}{2}\right)^{\frac{1+c}{2}}\right),$$

*где $\Gamma$ — гамма-функция Эйлера. В то же время*

$$\mathbb{E}\pi_0(n) = \frac{1}{1+n}\left(\frac{1}{1+c} + \frac{2\sqrt{\pi}\Gamma\left(n+\frac{c}{2}+1\right)}{(1+c)\Gamma\left(\frac{c}{2}\right)\Gamma\left(n+\frac{1}{2}\right)}\right) \approx \frac{2\sqrt{\pi}}{(1+c)\Gamma\left(\frac{c}{2}\right)} n^{-\frac{1-c}{2}}.$$

Все случайные величины, которые мы изучали до сих пор (степень вершины, вторая степень вершины и пр.), были дискретными. Однако величина $\pi_v(n)$ абсолютно непрерывна. Из теоремы 13 авторы статьи [35] легко вывели следствие о величине плотности $p_n(x)$ пейджранка.

**Теорема 14.** *Выполнено*

$$p_n(x) = \frac{2}{1-c}(n+1)\Big(1+\frac{1}{2n}\Big)c^{\frac{2}{1+c}}\Big(\frac{1+c}{1-c}(n+1)x-1\Big)^{-\frac{3+c}{1+c}} \approx \Theta\left(\frac{1}{x^{\frac{3+c}{1+c}}}\right).$$

Да ведь это снова степенной закон! Закон с параметром $\dfrac{3+c}{1+c}$. Это отлично коррелирует с наблюдениями, о которых мы писали в разд. 1.8. Правда, параметр степенного закона несколько отличается от того, который мы имеем в случае степеней вершин, но эту проблему мы знаем, и ее нам еще предстоит решать. Заметим, что при $c = 0{,}85$ (именно такой принято брать величину $c$ на практике) $\dfrac{3+c}{1+c} \approx 2{,}1$, и это весьма близко к эмпирическим данным из статьи [9].

### 2.5.6. Количество ребер между вершинами заданных степеней

Впервые величина $X_n(d_1, d_2)$ для модели Боллобаша–Риордана была изучена в работе [32]. Надо только иметь в виду, что в той работе определение слегка отличалось от данного нами в разд. 1.9. А именно, в нынешнем определении мы не станем учитывать петли. В остальном изменений не будет.

В следующей теореме найдена трехчленная асимптотика для математического ожидания величины $X_n(d_1, d_2)$.

**Теорема 15.** *Пусть $m \geqslant 1$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, $X = X_n(d_1, d_2)$. Если $d_1 < m$, $d_2 < m$ или $d_1 = d_2 = m$, то $X = 0$. Если же $d_1 \geqslant m$, $d_2 \geqslant m$ и $d_1 + d_2 \geqslant 2m+1$, то математическое ожидание величины $X$ выражается формулой*

$$\begin{aligned}\mathbb{E}X = {} & \frac{m(m+1)}{d_1(d_1+1)d_2(d_2+1)}\left(1-\frac{C_{2m+2}^{m+1}C_{d_1+d_2-2m}^{d_1-m}}{C_{d_1+d_2+2}^{d_1+1}}\right)(2mn+1) - \\ & - \sum_{i=1}^{m}\frac{C_{d_1+d_2-2i}^{d_1-i}}{d_1d_2C_{d_1+d_2}^{d_1}}\left(\frac{(2i)!}{i!(i+1)!}\frac{m+1}{2m} + \mathbb{I}(i=m)\frac{(2m)!}{2(m-1)!^2}\right) - \\ & - \mathbb{I}(d_1=m)\frac{(m-1)(m+1)}{2md_2(d_2+1)} - \mathbb{I}(d_2=m)\frac{(m-1)(m+1)}{2md_1(d_1+1)} + O_{m,d_1,d_2}\Big(\frac{1}{n}\Big).\end{aligned}$$

А ниже дается асимптотика распределения величины $X$.

**Теорема 16.** *Пусть $d_1$, $d_2$ таковы, что $(d_1+d_2)d_1^2d_2^2 = o(\sqrt{n})$. Тогда для любого $\varepsilon > 0$ выполнено*

$$\mathbb{P}((1-\varepsilon)\mathbb{E}X \leqslant X \leqslant (1+\varepsilon)\mathbb{E}X) \to 1, \quad n \to \infty.$$

Насколько точно и удивительно результаты этого пункта согласуются с реальностью, мы узнаем в п. 2.6.9.

### 2.5.7. Кластерные коэффициенты

Тут ситуация прямо противоположна той, что описана в разд. 1.11. А именно, здесь нормально посчитана только транзитивность. В статье [12] установлена

**Теорема 17.** *Пусть $m \geqslant 1$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана, $T$ — кластерный коэффициент. Тогда*

$$\mathbb{E}T \sim \frac{m-1}{8}\frac{\ln^2 n}{n}, \quad n \to \infty.$$

Пожалуй, это первый результат, который совсем не соответствует наблюдениям, в свете которых должна получаться константа, а вовсе не $\Theta(\ln^2 n/n)$. Это отдельная проблема, о которой мы поговорим в п. 2.7.5.

Кстати, в конце разд. 2.3 мы уже упоминали статью [26], в которой авторы предсказывают величину $\Theta(1/n^{0,75})$ транзитивности в модели Барабаши–Альберт. Боллобаш в [12] также отмечает разницу между этим предсказанием и утверждением теоремы 17, не без сарказма добавляя, что из работы [36] следует, в свою очередь, «предсказание» $T = \Theta(1/n^{0,25})$. Все-таки как ни важна идея, а математическая строгость необходима.

Отметим еще, что, по-видимому, в модели Боллобаша–Риордана средний локальный кластерный коэффициент лишь в константу раз больше транзитивности. Однако строгого доказательства мы в литературе не встречали.

### 2.5.8. Число копий фиксированного графа

Пусть $S$ — конкретный граф на вершинах $1, 2, \ldots$, в котором разрешены петли. Ориентируем каждое его ребро $\{i, j\}$ с $i \leqslant j$ от $j$ к $i$. Сперва нас будет интересовать вероятность $p_S$ его реализации в качестве подграфа графа Боллобаша–Риордана $G_1^n$. Подчеркнем, что нам важно вхождение самого $S$ в $G_1^n$ — не какой-то его изоморфной копии.

Следуя Боллобашу (см. [12]), введем обозначения $V^+(S)$ и $V^-(S)$ для множеств вершин графа $S$, из которых выходят и в которые входят ребра соответственно. Эти множества, конечно, могут пересекаться. Для $i \in V(S)$ определим $d_S^{in}(i)$ как входящую степень $i$ в $S$. Аналогично определим $d_S^{out}(i)$. Заметим, что петля в $i$ дает вклад 1 и в $d_S^{in}(i)$, и в $d_S^{out}(i)$. Пусть, наконец, $C_S(t)$ — число ребер в $S$, которые пересекают $t$, т.е. ребер $\{i, j\}$ с $i \leqslant t$ и $j \geqslant t$. Назовем $S$ *допустимым для* $G_1^n$, если $d_S^{out}(i) \leqslant 1$ для каждого $i$. В [12] Боллобаш доказывает следующую теорему.

**Теорема 18.** *Пусть $S$ допустим для $G_1^n$. Тогда*

$$p_S = \prod_{i\in V^-(S)} d_S^{in}(i)! \prod_{i\in V^+(S)} \frac{1}{2i-1} \prod_{t\notin V^+(S)} \left(1+\frac{C_S(t)}{2t-1}\right) =$$

$$= \prod_{i\in V^-(S)} d_S^{in}(i)! \prod_{\{i,j\}\in E(S)} \frac{1}{2\sqrt{ij}} \exp\Bigl(O\Bigl(\sum_{i\in V(S)} C_S(i)^2/i\Bigr)\Bigr).$$

Еще более громоздкую формулу можно было бы написать для роизвольного $m$. Однако этого Боллобаш уже не делает, но лишь приводит несколько частных результатов, которые мы и перечислим.

**Теорема 19.** *Пусть $m \geqslant 1$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана. Тогда число треугольников подчиняется соотношению*

$$\mathbb{E}\sharp(K_3, G_m^n) \sim \frac{m(m-1)(m+1)}{48} \ln^3 n, \quad n\to\infty.$$

**Теорема 20.** *Пусть $m \geqslant 1$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана. Тогда для любого $\varepsilon > 0$ число двухзвенных цепей подчиняется соотношению*

$$\mathbb{P}\Bigl((1-\varepsilon)\frac{m(m+1)}{2} n\ln n \leqslant \sharp(P_2, G_m^n) \leqslant (1+\varepsilon)\frac{m(m+1)}{2} n\ln n\Bigr) \to 1, \quad n\to\infty.$$

Из этих двух теорем как раз и следует теорема 17 о величине кластерного коэффициента. Более того, из первой теоремы видно, что несоответствие величины транзитивности ее же оценкам в реальных сетях в первую очередь связано с тем, что в модели Боллобаша–Риордана крайне мало треугольников: если в реальности их $n^\alpha$ (см. разд. 1.12), то здесь их только $\Theta(\ln^3 n)$. О способах борьбы с этой проблемой мы поговорим в п. 2.7.5.

Еще одна специальная теорема Боллобаша говорит о числе циклов.

**Теорема 21.** *Пусть $m \geqslant 2$, $G_m^n$ — случайный граф в модели Боллобаша–Риордана. Тогда число циклов длины $l$ подчиняется соотношению*

$$\mathbb{E}\sharp(C_l, G_m^n) \sim C_{m,l} \ln^l n, \quad n\to\infty.$$

*Более того, при увеличении $m$ множитель $C_{m,l}$ растет как $\Theta(m^l)$.*

На самом деле все эти теоремы допускают обобщение. В статье [37] оно доказано в случае модели $G_m^{(n)}$ (см. п. 2.4.1). Однако разница между моделями крайне невелика, и потому следующая теорема верна, повидимому, и для $G_m^n$.

**Теорема 22.** *Пусть $m \geqslant 1$, $G_m^{(n)}$ — случайный граф в модели Барабаши–Альберт, определенной в п. 2.4.1. Пусть $S$ — фиксированный граф. Обозначим $d_i$ число его вершин степени $i$. Тогда*

$$\mathbb{E}\sharp(S, G_m^{(n)}) = \Theta(1)\left(n^{d_0}\left(\sqrt{n}\right)^{d_1}(\ln n)^{d_2}\right) m^{|E(S)|}.$$

Иными словами, получается, что математическое ожидание числа копий графа $S$ в случайном графе $G_m^{(n)}$ зависит только от количества вершин графа $S$, имеющих степени 0, 1 и 2. Например, число полных графов на четырех вершинах уже не превосходит константы, и это нелепо. Вообще, теорема 22 дает очень приятную формулу. Из нее легко следуют чуть ослабленные утверждения теорем 19–21 и многие другие подобные факты (если, конечно, пренебречь разницей в определениях моделей).

## 2.6. УТОЧНЕНИЯ МОДЕЛИ БОЛЛОБАША–РИОРДАНА: НАЧАЛЬНАЯ ПРИТЯГАТЕЛЬНОСТЬ ВЕРШИНЫ

### 2.6.1. Несколько вводных замечаний

В предыдущем разделе мы обсудили соответствие свойств модели Боллобаша–Риордана свойствам реального веба и пришли к выводу, что во многих отношениях модель удивительно похожа на хост-граф. Вместе с тем мы заметили и ряд досадных несоответствий. Во-первых, степенной закон распределения степеней вершин получился не с тем параметром. Во-вторых, совсем не так, как хотелось бы, распределены треугольники, а стало быть, и кластерные коэффициенты.

Ниже мы рассмотрим две очень похожие модели, обобщающие модель Боллобаша–Риордана. Некоторые из свойств Интернета мы сравним со свойствами первой из них, а некоторые — со свойствами второй. Это связано с тем, что модели близки друг к другу, и часть результатов в разное время была получена для одной модели, часть — для другой. Но принцип общий. Мы не будем писать о гигантской компоненте, устойчивости, уязвимости и диаметре, поскольку здесь существенных изменений не будет. Не станем мы писать и о пейджранке, так как он и без того посчитан не в модели Боллобаша–Риордана, а в некоторой ее модификации, которая, как мы увидим скоро, чем-то напоминает обе модели, о которых пойдет речь.

### 2.6.2. Модель Бакли–Остгуса

Идея этой модели независимо пришла в голову исследователям, работавшим сразу в нескольких исследовательских центрах

(см. [38,39]). Однако именно Бакли и Остгус аккуратно формализовали ее в [40].

Пусть $a > 0$ — параметр модели. Случайный граф $H^n_{a,m}$ строится по сути так же, как строился случайный граф $G^n_m$: сперва определяется $H^n_{a,1}$, затем производится склейка наборов из $m$ последовательных вершин в новые вершины. При этом $H^n_{a,1}$ тоже образуется по принципу образования графа $G^n_1$, только вероятности слегка другие:

$$\mathbb{P}(n+1 \to n+1) = \frac{a}{(a+1)n+1}, \quad \mathbb{P}(n+1 \to i) = \frac{\deg_{H^n_{a,1}} i - 1 + a}{(a+1)n+1}, \quad i = 1, \ldots, n.$$

Тем самым, при $a = 1$ мы имеем в точности модель Боллобаша–Риордана. Параметр $a$ называется *начальной притягательностью вершины*. Оказывается, именно он в ответе за правильные показатели степенных распределений.

### 2.6.3. Модель Мори

Модель Мори почти такая же, как модель Бакли–Остгуса. В ней за притягательность вершины отвечает параметр $\beta > 0$. Случайный граф мы обозначим соответственно $H^{(n)}_{\beta,m}$. Его построение начинается с графа $H^{(2)}_{\beta,1}$, у которого две вершины и ребро между ними. А дальше все, как обычно, но петель нет:

$$\mathbb{P}(n+1 \to i) = \frac{\deg_{H^{(n)}_{\beta,1}} i + \beta}{(\beta+2)n - 2}, \quad i = 1, \ldots, n.$$

Снова распределение задано корректно, и граф $H^{(n)}_{\beta,m}$ получается из $H^{(mn)}_{\beta,1}$ стандартной склейкой. В нем уже петли бывают.

Модель описана Мори в работе [41]. Если положить в модели $\beta = 0$, то получится практически в точности модель Боллобаша–Риордана.

### 2.6.4. Степени вершин

Здесь ситуация очень похожа на ситуацию из п. 2.5.3. Сперва авторы модели — Бакли и Остгус — доказали довольно слабую теорему, в которой, однако, уже заключалась вся суть (см. [40]).

**Теорема 23.** *Пусть $m \geqslant 1$, $a \geqslant 1$ — фиксированные целые числа. Положим*

$$\alpha_{a,m,d} = (a+1)(am+a)!\, C^{am-1}_{d+am-1} \frac{d!}{(d+am+a+1)!}.$$

*Пусть $H^n_{a,m}$ — случайный граф в модели Бакли–Остгуса. Пусть $\varepsilon > 0$ — сколь угодно малое фиксированное число. Тогда для любого*

$d \leqslant n^{1/100(a+1)}$

$$\mathbb{E}\left|\left\{i = 1, \ldots, n\colon \deg_{H^n_{a,m}} i = d\right\}\right| \sim n\alpha_{a,m,d}, \quad n \to \infty,$$

$$\mathbb{P}\left((1-\varepsilon)n\alpha_{a,m,d} \leqslant \left|\left\{i = 1, \ldots, n\colon \deg_{H^n_{a,m}} i = d\right\}\right| \leqslant (1+\varepsilon)n\alpha_{a,m,d}\right) \to 1,$$
$$n \to \infty.$$

С одной стороны, теорема замечательная, ведь, если расписать по формуле Стирлинга, то получится

$$\alpha_{a,m,d} = \Theta(d^{-2-a}).$$

Иными словами, можно варьировать параметр степенного закона.

С другой стороны, $a$ — *целое*, и мы снова не можем получить $\gamma \in (2,3)$, как хотелось бы. Кроме того, условие $d \leqslant n^{1/100(a+1)}$ еще ужаснее ограничения $d \leqslant n^{1/15}$ из теоремы 9.

Тут снова есть улучшения из работы [31] и более компактное улучшение Гречникова (см. [42]). Его мы и приведем.

**Теорема 24.** *Пусть $m \geqslant 1$, $a > 0$. Пусть $H^n_{a,m}$ — случайный граф в модели Бакли–Остгуса. Пусть $\varepsilon > 0$ — сколь угодно малое фиксированное число. Положим*

$$\beta_{a,m,d} = \frac{\mathrm{B}(d - m + ma, a + 2)}{\mathrm{B}(ma, a + 1)},$$

*где* B *— бета-функция. Тогда для любого $d \geqslant m$*

$$\mathbb{E}\left|\left\{i = 1, \ldots, n\colon \deg_{H^n_{a,m}} i = d\right\}\right| = n\beta_{a,m,d} + O_{a,m}\left(\frac{1}{d}\right),$$

*а при условии $d = o\left(n^{1/(a+2)}\right)$*

$$\mathbb{P}\left((1-\varepsilon)n\beta_{a,m,d} \leqslant \left|\left\{i = 1, \ldots, n\colon \deg_{H^n_{a,m}} i = d\right\}\right| \leqslant (1+\varepsilon)n\beta_{a,m,d}\right) \to 1,$$
$$n \to \infty.$$

Ограничение стало осязаемым (а для математического ожидания его и вовсе нет). А главное, $a$ любое, причем

$$\beta_{a,m,d} = \Theta(d^{-2-a}),$$

и мы, наконец, можем задавать распределение по своему усмотрению.

### 2.6.5. Вторые степени вершин

В модели Бакли–Остгуса вторые степени вершин, как и в случае модели Боллобаша–Риордана, были изучены Остроумовой — в этот раз при участии А. Б. Купавского, Д. А. Шабанова и П. Тетали

(см. [43]). Разумеется, и здесь рассмотрено только $m = 1$. Полным аналогом теоремы 11 служит

**Теорема 25.** *Пусть $a > 0$. Пусть $H^n_{a,1}$ — случайный граф в модели Бакли–Остгуса с $m = 1$. Тогда для любого $d > 0$*

$$\mathbb{E}|\{i = 1, \dots, n\colon d_2(i) = d\}| = \\ = \frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)d^{a+1}}\left(1 + O\left(\frac{(\ln d)^{\lceil a+1\rceil}}{d}\right) + O\left(\frac{d^{a+1}}{n}\right)\right).$$

При $a = 1$ это в точности результат теоремы 11. Зато следующая теорема, говорящая о наличии степенного закона, даже слегка точнее аналогичной теоремы 12.

**Теорема 26.** *Для любого $\delta > 0$, любого $\varepsilon > 0$ и $d \in \{1, \dots, n^{1/(4+a)-\delta}\}$*

$$\mathbb{P}\left((1-\varepsilon)\frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)d^{a+1}} \leqslant |\{i = 1, \dots, n\colon d_2(i) = d\}| \leqslant \right. \\ \left. \leqslant (1+\varepsilon)\frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)d^{a+1}}\right) \to 1, \quad n \to \infty.$$

Бóльшая точность в ограничении на $d$: если в теореме 12 оно имело вид $d \leqslant n^{1/6-\delta}$, то в теореме 26 при $a = 1$ оно превращается в неравенство $d \leqslant n^{1/5-\delta}$. В пунктах 3.4.3 и 3.4.4 мы объясним, как получено улучшение.

### 2.6.6. Количество ребер между вершинами заданных степеней

Как и в п. 2.5.6, основной результат принадлежит Гречникову (см. [42]).

**Теорема 27.** *Пусть $d_1 \geqslant m$ и $d_2 \geqslant m$. Пусть $X = X_n(d_1, d_2)$. Существует такая функция $c_X(d_1, d_2)$, что*

$$\mathbb{E}X = c_X(d_1, d_2)n + O_{a,m}(1)$$

*и*

$$c_X(d_1, d_2) = \frac{\Gamma(d_1 - m + ma)\Gamma(d_2 - m + ma)\Gamma(d_1 + d_2 - 2m + 2ma + 3)}{\Gamma(d_1 - m + ma + 2)\Gamma(d_2 - m + ma + 2)\Gamma(d_1 + d_2 - 2m + 2ma + a + 2)} \times \\ \times ma(a+1)\frac{\Gamma(ma + a + 1)}{\Gamma(ma)} \times \\ \times \left(1 + \theta(d_1, d_2)\frac{(d_1 - m + ma + 1)(d_2 - m + ma + 1)}{(d_1 + d_2 - 2m + 2ma + 1)(d_1 + d_2 - 2m + 2ma + 2)}\right),$$

*где*

$$-4 + \frac{2}{1 + ma} \leqslant \theta(d_1, d_2) \leqslant a\frac{\Gamma(ma+1)\Gamma(2ma+a+3)}{\Gamma(2ma+2)\Gamma(ma+a+2)}.$$

*Когда и $d_1$, и $d_2$ растут, величина $c_X$ асимптотически ведет себя следующим образом:*

$$c_X(d_1, d_2) =$$
$$= ma(a+1)\frac{\Gamma(ma+a+1)}{\Gamma(ma)}\frac{(d_1+d_2)^{1-a}}{d_1^2 d_2^2}\left(1 + O_{a,m}\left(\frac{1}{d_1} + \frac{1}{d_2} + \frac{d_1 d_2}{(d_1+d_2)^2}\right)\right).$$

Таким образом, если в модели Боллобаша–Риордана

$$\mathbb{E}X = \Theta\left(\frac{1}{d_1^2 d_2^2}\right),$$

то в модели Бакли–Остгуса

$$\mathbb{E}X = \Theta\left(\frac{(d_1+d_2)^{1-a}}{d_1^2 d_2^2}\right),$$

и эти факты отлично согласуются.

Стоит отметить, что если говорить не о порядке роста, но об асимптотике математического ожидания, то ее теорема 27 дает лишь в случаях, когда $d_1 = d_1(n) \to \infty$, $d_2 = d_2(n) \to \infty$ и $d_1/d_2 \not\to \gamma \in (0, \infty)$ при $n \to \infty$. Последнее ограничение недавно устранил сам Гречников. Однако результат выглядит еще более устрашающе, да и статья пока не опубликована.

Полным аналогом теоремы 16 служит

**Теорема 28.** *Пусть $d_1$, $d_2$ таковы, что $(d_1+d_2)^a d_1^2 d_2^2 = o(\sqrt{n})$. Тогда для любого $\varepsilon > 0$ выполнено*

$$\mathbb{P}((1-\varepsilon)\mathbb{E}X \leqslant X \leqslant (1+\varepsilon)\mathbb{E}X) \to 1, \quad n \to \infty.$$

### 2.6.7. Кластерные коэффициенты

Как и в случае модели Боллобаша–Риордана, здесь посчитана только транзитивность $T$, причем в модели Мори (см. [44]).

**Теорема 29.** *Пусть $m \geqslant 1$, $\beta > 0$, $H^{(n)}_{\beta,m}$ — случайный граф в модели Мори, $T$ — кластерный коэффициент. Тогда*

$$\mathbb{E}T \sim A_{\beta,m}\frac{\ln n}{n}, \quad n \to \infty,$$

*где $A_{\beta,m}$ — некоторая величина, зависящая только от констант $\beta$, $m$.*

Как отмечают сами авторы статьи [44], где доказана теорема 29, любопытно то, что при всех $\beta > 0$

$$\mathbb{E}T = \Theta\left(\frac{\ln n}{n}\right),$$

тогда как у Боллобаша–Риордана (см. п. 2.5.7), т. е. по сути при $\beta = 0$

$$\mathbb{E}T = \Theta\Big(\frac{\ln^2 n}{n}\Big).$$

Как видно, никакого улучшения с точки зрения величины кластерного коэффициента модель Мори не дает. Если со степенями вершин мы разобрались, то с транзитивностью или средним локальным кластерным коэффициентом нам еще предстоит разбираться.

### 2.6.8. Число копий фиксированного графа

Здесь вся теория построена для модели Мори в статье [44], и результаты весьма похожи на результаты из п. 2.5.8. Как и в том пункте, все начинается с подсчета вероятности $p_S$ вхождения допустимого графа $S$ в $H^{(n)}_{\beta,1}$.

**Теорема 30.** *Пусть $S$ допустим для $H^{(n)}_{\beta,1}$. Тогда*

$$p_S = \frac{\beta}{\beta + d^{in}_S(1)} \prod_{i \in V^-(S)} \frac{\Gamma\big(1+\beta+d^{in}_S(i)\big)}{\Gamma(1+\beta)} \prod_{i \in V^+(S)} \frac{1}{(2+\beta)(i-1)-2} \times$$

$$\times \prod_{t \notin V^+(S)} \Big(1 + \frac{C_S(t)}{(2+\beta)(t-1)-2}\Big) = \frac{\beta}{\beta + d^{in}_S(1)} \prod_{i \in V^-(S)} \frac{\Gamma\big(1+\beta+d^{in}_S(i)\big)}{\Gamma(1+\beta)} \times$$

$$\times \prod_{\{i,j\} \in E(S)} \frac{1}{(2+\beta)\big(i^{1+\beta} j\big)^{1/(2+\beta)}} \cdot \exp\Big(O\Big(\sum_{i \in V(S)} C_S(i)^2/i\Big)\Big).$$

Далее следуют теоремы о числе треугольников и числе цепей длины 2.

**Теорема 31.** *Пусть $m \geqslant 1$, $\beta > 0$, $H^{(n)}_{\beta,m}$ — случайный граф в модели Мори. Тогда число треугольников подчиняется соотношению*

$$\mathbb{E}\sharp\big(K_3, H^{(n)}_{\beta,m}\big) = \Big(m(m-1)\frac{(1+\beta)^2}{\beta^2} + (m-1)^2\frac{(1+\beta)^3}{\beta^2(2+\beta)}\Big)\ln n + O(1),$$
$$n \to \infty.$$

**Теорема 32.** *Пусть $m \geqslant 1$, $\beta > 0$, $H^{(n)}_{\beta,m}$ — случайный граф в модели Мори. Положим*

$$\alpha_{\beta,m} = \frac{2+5\beta}{2\beta} m^2 + \frac{2-\beta}{2\beta} m.$$

*Тогда для любого $\varepsilon > 0$ число двухзвенных цепей подчиняется соотношению*

$$\mathbb{P}\Big((1-\varepsilon)\alpha_{\beta,m} n \leqslant \sharp\big(P_2, H^{(n)}_{\beta,m}\big) \leqslant (1+\varepsilon)\alpha_{\beta,m} n\Big) \to 1, \quad n \to \infty.$$

Как видно, и число треугольников, и число цепей в модели Мори ведут себя не так, как в модели Боллобаша–Риордана. Однако по-прежнему теорема 29 следует как раз из приведенных результатов. В частности,

$$A_{\beta,m} = \frac{3}{\alpha_{\beta,m}}\left(m(m-1)\frac{(1+\beta)^2}{\beta^2} + (m-1)^2\frac{(1+\beta)^3}{\beta^2(2+\beta)}\right).$$

Аналога теоремы 22 здесь пока нет.

### 2.6.9. Удивительное соответствие модели Бакли–Остгуса реальному хост-графу

В предыдущих двух пунктах мы отметили досадные различия между моделью Бакли–Остгуса и реальным хост-графом, связанные с неадекватностью кластерного коэффициента. Однако до того мы наблюдали куда более радужную картину: например, с точки зрения реберных характеристик — число вершин данной степени и пр. — модель неизменно оказывалась на высоте. Ниже мы обсудим недавние исследованияя, проведенные в компании Яндекс и засвидетельствовашие, что с указанной точки зрения все даже более удивительно, чем пока представляется (см. также [7]).

Идея следующая. Давайте предположим, что реальный хост-граф «строился» в модели Бакли–Остгуса, и зададимся вопросом: с каким параметром $a$ шло это «строительство»? Можно пытаться оптимизировать этот параметр разными способами. Например, можно посмотреть, когда распределение степеней вершин случайного графа в модели Бакли–Остгуса ближе всего к аналогичной характеристике реального веба. А можно то же самое сделать в отношении вероятности ребра между вершинами данных степеней. Априори характеристики сильно разные, корреляция между ними мала, и совершенно не очевидно, что в обоих случаях получится, скажем, одно и то же оптимальное значение начальной притягательности. Но оно получается, и это кажется почти невероятным!

Строго говоря, в работе [7] сделано следующее. Прежде всего, как водится, и хост-граф, и случайный граф Бакли–Остгуса очищаются от кратных ребер, петель и ориентации. Как показано в [7], это почти не отражается на формулировках теорем про свойства модели. Далее, вводятся величины $\sharp_{\mathrm{Host}}(j)$ — число вершин степени $j$ в хост-графе, —

$$\tilde{\sharp}_{\mathrm{Host}}(d) = \sum_{j>d} \sharp_{\mathrm{Host}}(j),$$

$X_{\mathrm{Host}}(j_1, j_2)$ — число ребер между вершинами со степенями $j_1, j_2$ в хост-графе, —

$$\tilde{X}_{\mathrm{Host}}(d_1, d_2) = \sum_{j_1 \geqslant j_2, j_1 > d_{\max}, j_2 > d_{\min}} X_{\mathrm{Host}}(j_1, j_2),$$

где

$$d_{\max} = \max\{d_1, d_2\}, \quad d_{\min} = \min\{d_1, d_2\},$$

и, наконец,

$$\widetilde{\rho}_{\mathrm{Host}}(d_1, d_2) = \frac{\widetilde{X}_{\mathrm{Host}}(d_1, d_2)}{\widetilde{\sharp}_{\mathrm{Host}}(d_1)\,\widetilde{\sharp}_{\mathrm{Host}}(d_2)}.$$

Величины $\widetilde{\sharp}_{\mathrm{Host}}(d)$, $\widetilde{X}_{\mathrm{Host}}(d_1, d_2)$, в которых ведется суммирование, устойчивее по отношению к возможным флуктуациям данных и играют роль своего рода функций распределения. Последняя же величина — это, по сути, распределение числа ребер между случайными вершинами степеней $d_1$, $d_2$, откуда и знаменатель.

Естественной аппроксимацией для величины $\widetilde{\sharp}_{\mathrm{Host}}(d)$ служит функция

$$f_{a_1,b_1}(d) = b_1 d^{-1-a_1},$$

ведь мы предположили, что хост-граф сформировался в модели Бакли–Остгуса, а для этой модели есть теорема 24: в ней говорится о числе вершин степени $d$, которое оказывается по порядку равным $d^{-2-a}$; у нас же сейчас речь идет о числе вершин со степенями, большими $d$, т. е. о сумме величин типа $j^{-2-a}$ по $j = d+1, d+2, \ldots$, которая как раз и есть $b_1 d^{-1-a_1}$.

Аналогично, ввиду теорем 24 и 27, аппроксимируем величину $\widetilde{\rho}_{\mathrm{Host}}(d_1, d_2)$ функцией

$$g_{a_2,b_2}(d_1, d_2) = b_2 (d_1 + d_2)^{1-a_2} d_1^{a_2} d_2^{a_2}.$$

Именно поэтому в статье искались $a_1$, $b_1$ и $a_2$, $b_2$, минимизирующие выражения

$$\frac{1}{|\mathcal{D}_1|} \sum_{d \in \mathcal{D}_1} \left( \sqrt{\widetilde{\sharp}_{\mathrm{Host}}(d)} - \sqrt{f_{a_1,b_1}(d)} \right)^2,$$

$$\frac{1}{|\mathcal{D}_2|} \sum_{d_1, d_2 \in \mathcal{D}_2} \left( \sqrt{\widetilde{\rho}_{\mathrm{Host}}(d_1, d_2)} - \sqrt{g_{a_2,b_2}(d_1, d_2)} \right)^2,$$

где

$$\mathcal{D}_1 = [10^{2,9}, 10^{5,9}], \quad \mathcal{D}_2 = \{(d_1, d_2) \in D_1^2 \colon d_1/d_2 > 10\}.$$

Здесь фактически применен нелинейный метод наименьших квадратов. Есть несколько причин, по которым был задействован именно этот метод. Например, обычная линейная регрессия, как показано в статье [7], дает очень плохое приближение. Кроме того, в модели Бакли–Остгуса порядок роста вариации и среднего значения величины $\widetilde{\sharp}_{\mathrm{Host}}(d)$ один и тот же (это можно доказать), а стало быть, величина $\sqrt{\widetilde{\sharp}_{\mathrm{Host}}(d)}$ ограничена константой, что приводит к наилучшей сбалансированности слагаемых в минимизируемых выражениях.

Области $\mathcal{D}_1$, $\mathcal{D}_2$, по которым ведутся суммирования, выбраны исключительно для удобства.

Совершенно замечательно то, что в результате расчетов и статистических проверок, описанных подробно в [7], получилось, что $a_1 \approx 0{,}2762$, $a_2 \approx 0{,}2774$. Иными словами, близость к *одной и той же* модели подтверждается сразу двумя — по существу независимыми — измерениями. Вот если бы еще с кластерным коэффициентом разобраться, то и вовсе чудесно будет. Ясно, однако, что тут нужна дополнительная идея: если предпочтительное присоединение столь точно описывает реберные характеристики, то, по-видимому, некий иной закон правит треугольниками, т. е. отвечает за высокую распространенность в Интернете явления «мои друзья дружат между собой».

### 2.6.10. Классификация ссылочного спама

В разд. 1.12 мы говорили о необходимости классификации двудольных подграфов хост-графа с целью автоматического выявления линковой (ссылочной) накрутки. Одна из идей, как можно эту классификацию проводить, основана на фактическом использовании величины $\widetilde{\rho}_{\mathrm{Host}}(d_1, d_2)$ и ее модельного аналога

$$\rho_n(d_1, d_2) = \frac{X_n(d_1, d_2)}{\sharp_n(d_1)\,\sharp_n(d_2)}.$$

Выбор модели произволен (чем адекватнее модель, тем лучше), однако видно, что модель Бакли–Остгуса вполне годится.

Итак, пусть дан некий двудольный подграф $G = (V_1 \sqcup V_2, E)$ хост-графа. Вычислим степени $d_1(v_i)$ и $d_2(v_j)$ для каждой вершины $v_i \in V_1$, $v_j \in V_2$ (полные степени на всем хост-графе) и посчитаем величину

$$S = \sum_{v_i \in V_1, v_j \in V_2} \rho_n(d_1(v_i), d_2(v_j)) \approx \sum_{v_i \in V_1, v_j \in V_2} b_2 (d_1(v_i) + d_2(v_j))^{1-a_2} d_1^{a_2}(v_i) d_2^{a_2}(v_j).$$

Здесь мы пользуемся оптимальной аппроксимацией из предыдущего пункта. Понятно, что $S$ играет роль ожидаемого числа ребер в графе $G$. И если $|E| \gg S$, то что-то не так: либо это сообщество (ср. разд. 1.12), либо накрутка.

## 2.7. ДАЛЬНЕЙШИЕ УТОЧНЕНИЯ МОДЕЛИ БОЛЛОБАША–РИОРДАНА

### 2.7.1. Несколько вводных замечаний

В настоящем разделе мы скажем несколько слов о различных уточнениях модели Боллобаша–Риордана и Бакли–Остгуса. Проблемы, с которыми мы будем бороться, — это неадекватность кластер-

ного коэффициента, а также отсутствие реальной ориентации в моделях, которые мы до сих пор рассматривали (число исходящих ребер у вершин было фиксировано). Поэтому ниже мы опишем несколько моделей, но для каждой из них скажем лишь, в чем ее преимущество по сравнению с Боллобашем–Риорданом или Бакли–Остгусом. С точки зрения математического анализа своих свойств все эти модели гораздо сложнее, и не удивительно, что про них почти ничего не доказано.

### 2.7.2. Модель Боллобаша–Боргса–Риордана–Чайес

Здесь мы опишем одну из немногих существующих моделей предпочтительного присоединения, в которой ориентация ребер играет гораздо более важную роль, чем то было во всех предыдущих моделях. Ее придумали Боллобаш и Риордан в соавторстве с Боргсом и Чайес в работе [45].

У модели пять параметров: $\alpha$, $\beta$, $\gamma$, $\delta_{\text{in}}, \delta_{\text{out}}$. Эти числа неотрицательны, причем первые три из них в сумме дают единицу. Последнее число — это начальная притягательность каждой вершины, а предпоследнее — ее начальная «тяга к простановке ссылок». Процесс построения случайного графа устроен более или менее стандартно. Вначале есть граф $G_0 = G(t_0)$ с $t_0$ ребрами. Для определенности считаем, что это одна вершина с $t_0$ петлями. Далее на каждом шаге добавляется ровно одно ребро, так что в момент времени $t$ мы имеем граф $G(t)$ с $t$ ребрами. Однако вершины могут не добавляться, поскольку иногда новое ребро проводится между уже существующими вершинами. В результате число вершин в момент времени $t$ равно некоторому $n(t)$, которое по факту окажется случайной величиной.

Для более детального описания процесса введем полезную терминологию. Скажем, что вершина $v$ графа $G(t)$ выбирается согласно $\operatorname{indeg} + \delta_{\text{in}}$, если

$$\mathbb{P}(v = v_i) = \frac{\operatorname{indeg}(v_i) + \delta_{\text{in}}}{t + \delta_{\text{in}} n(t)}.$$

Аналогично можно говорить, что вершина $v$ графа $G(t)$ выбирается согласно $\operatorname{outdeg} + \delta_{\text{out}}$.

При $t \geqslant t_0$ граф $G(t+1)$ получается из графа $G(t)$ по следующим правилам:

1) с вероятностью $\alpha$ добавляются новая вершина $v$ и ребро, выходящее из $v$ в какую-то из существующих вершин $w$, которая выбирается согласно $\operatorname{indeg} + \delta_{\text{in}}$;
2) с вероятностью $\beta$ добавляется ребро из существующей вершины $v$ в существующую вершину $w$, где $v$ и $w$ выбираются независимо: $v$ — согласно $\operatorname{outdeg} + \delta_{\text{out}}$, а $w$ — согласно $\operatorname{indeg} + \delta_{\text{in}}$;

3) с вероятностью $\gamma$ добавляются новая вершина $w$ и ребро, выходящее из существующей вершины $v$ в $w$, где $v$ выбирается согласно $\mathrm{outdeg} + \delta_{\mathrm{out}}$.

Ясно, что в процессе могут возникать и кратные ребра, и кратные петли. Разумеется, следует считать, что $\alpha + \gamma > 0$, иначе Интернет не растет. Кроме того, интуитивно понятно, что в рамках варианта 1 появляется обычный сайт, а в рамках варианта 3 — сайт, который создан исключительно для демонстрации своего «контента» (содержимого). В последнем случае сайт вряд ли когда-нибудь поставит исходящие ссылки. Поэтому естественно полагать в модели $\delta_{\mathrm{out}} = 0$. Авторы модели подчеркивают, что они ввели этот параметр для симметрии, а также в надежде, что для других реальных сетей он окажется более полезным.

Модель сложна для математического анализа. Ее преимущество исключительно в наличии более реалистичного распределения исходящих степеней. О нем (а равно и о распределении входящих степеней) говорит следующая теорема, доказанная в [45].

**Теорема 33.** *Пусть фиксировано натуральное $i \geqslant 1$. Пусть $x_i(t)$ — число вершин графа $G(t)$, имеющих входящую степень $i$. Аналогично определим $y_i(t)$ в случае исходящих степеней. Положим*

$$c_1 = \frac{\alpha + \beta}{1 + \delta_{in}(\alpha + \gamma)}, \quad c_2 = \frac{\beta + \gamma}{1 + \delta_{out}(\alpha + \gamma)}.$$

*Тогда существуют такие $p_i$, $q_i$, являющиеся константами для данного $i$, и такие $\varphi_i$, $\psi_i$, являющиеся функциями аргумента $t$, которые при данном $i$ бесконечно малы в сравнении с $t$, что с вероятностью 1 для каждого $t$ выполнено*

$$x_i(t) = p_i t + \varphi_i(t), \quad y_i(t) = q_i t + \psi_i(t).$$

*Более того, если $\alpha\delta_{in} + \gamma > 0$ и $\gamma < 1$, то существует такая константа $C_{in} > 0$, что при $i \to \infty$*

$$p_i \sim C_{in} i^{-1-\frac{1}{c_1}}.$$

*Если же $\gamma\delta_{out} + \alpha > 0$ и $\alpha < 1$, то существует такая константа $C_{out} > 0$, что при $i \to \infty$*

$$q_i \sim C_{out} i^{-1-\frac{1}{c_2}}.$$

Прежде всего теорема показывает, что и входящие, и исходящие степени вершин подчиняются в модели асимптотически степенным законам распределения. Ничего подобного в моделях типа Бакли–Остгуса не было, хотя реальность именно такова. И параметры модели можно подобрать так, чтобы показатели степенных законов были реалистичны-

ми. Правда, зона действия теоремы крайне узкая. Здесь $i$ — величина степени — и $t$ — момент времени — не могут одновременно стремиться к бесконечности. Сейчас Гречников работает над устранением этой проблемы. Когда она будет устранена, а также будет посчитана асимптотика числа ребер между вершинами заданных степеней в модели, очень уместно будет сделать для модели такую же статистическую работу, какую мы описали в п. 2.6.9. Есть все основания полагать, что здесь ситуация будет еще лучше.

### 2.7.3. Модель копирования

Эта модель имеет великое множество модификаций (см. [46]), но мы опишем здесь ее лишь в простейшей форме. Придумана она была специально для того, чтобы объяснить феномен возникновения сообществ — двудольных графов — в Интернете. Этот феномен ни одна из ранее описанных в этой книге моделей не улавливала. Отметим сразу, что эта модель слегка выбивается из всей линейки, поскольку основывается не совсем на принципе предпочтительного присоединения. В следующем пункте мы ее в некотором смысле сочетаем с предпочтительным присоединением, и тогда все становится на свои места.

Пусть дано натуральное число $d \geqslant 1$ и число $\alpha \in (0, 1)$. Нам хочется построить процесс, в котором на каждом шаге добавляются одна вершина и $d$ исходящих ребер (как всегда!), и при этом мы желаем с некоторой вероятностью строить ребра, как бы «копируя ссылки», а с дополнительной вероятностью (это и есть $\alpha$) запускать ребра в случайные вершины. Интуиция за этим такая: создавая новый сайт, владелец либо уже принадлежит определенному сообществу — например, сообществу любителей советской песни 30-х–50-х годов XX века, — либо не принадлежит; в первом случае он выберет сайт с интересной ему тематикой (скажем, http://sovmusic.ru/), который с точки зрения стороннего наблюдателя «случаен», и скопирует некоторые его ссылки к себе; во втором случае он просто выберет «случайный» сайт и на него сошлется (даже — для простоты — не учитывая его популярность, т. е. не привязываясь к степеням).

Строго процесс описывается так. В нулевой момент времени есть некоторый граф $G_0$, у которого $t_0$ вершин и каждая вершина имеет исходящую степень не меньше $d$. Понятно, что если на очередных шагах процесса добавлять ровно одну вершину и ровно $d$ ребер, выходящих из нее, то и у любого графа $G_n$ будет исходящая степень каждой вершины не меньше $d$. Пусть при $n \geqslant 0$ граф $G_n$ уже построен. Добавим к нему вершину $v$ и станем последовательно и взаимно независимо выпускать ребра $e_i$, $i = 1, \ldots, d$, из $v$ в вершины графа $G_n$. Сперва выберем случайную вершину $p \in V(G_n)$ согласно равномерному

распределению, т. е. с вероятностью $\frac{1}{|V(G_n)|} = \frac{1}{t_0+n}$. Это будет аналог сайта http://sovmusic.ru/, т. е. сайта, с которого владелец нового сайта (вершины) $v$ будет копировать ссылки. Далее, пусть дано $i \in \{1, \ldots, d\}$, и нам надо найти второй конец для ребра $e_i$. С вероятностью $\alpha$ выбираем этот конец случайно согласно равномерному распределению (он даже может совпасть с $p$). А с вероятностью $1-\alpha$ выбираем $i$-ю по счету ссылку с сайта $p$ и проводим ребро из $v$ в цитируемую вершину.

Основной результат о модели (см. [46]) состоит в том, что и копирование приводит к степенному закону распределения степеней вершин. Правда, поскольку исходящие степени здесь снова фиксированы, то речь идет лишь о входящих степенях. Обозначим $N_{t,r}$ число вершин входящей степени $r$ в графе $G_n$ с $t$ вершинами. Справедлива

**Теорема 34.** *Пусть* $d = 1$, $r > 0$. *Положим* $P_r = \lim\limits_{t\to\infty} \frac{\mathbb{E}N_{t,r}}{t}$. *Эта величина определена корректно и задается формулой*

$$P_r = P_0 \prod_{i=1}^{r} \frac{1+\alpha/(i(1-\alpha))}{1+2/(i(1-\alpha))}.$$

*При этом*

$$P_r = \Theta\left(r^{-\frac{2-\alpha}{1-\alpha}}\right).$$

*Более того, для любого* $\varepsilon > 0$ *и для любого* $r < \ln t$ *выполнено*

$$\mathbb{P}\left((1-\varepsilon)P_r \leqslant \frac{N_{t,r}}{t} \leqslant (1+\varepsilon)P_r\right) \to 1, \quad t \to \infty.$$

Получается, что подбором параметра $\alpha$ мы можем добиться того степенного закона, какого пожелаем, т. е. в этом смысле все не хуже, чем в модели Бакли–Остгуса. Однако полных двудольных графов в модели Бакли–Остгуса почти не возникало (ср. теорему 22). Что же с моделью копирования?

Обозначим $K_{i,j}$ полный двудольный граф с долями размера $i$ и $j$, в котором все ребра направлены от первой доли ко второй. Авторы статьи [46] доказали следующую теорему.

**Теорема 35.** *Пусть* $d$ — *фиксированное натуральное число,* $\alpha \in (0,1)$, $t \to \infty$, *а* $i \leqslant \ln t$. *Тогда существует такая константа* $c = c(d)$, *что*

$$\mathbb{P}\left(\sharp(K_{i,d}, G_t) \geqslant cte^{-i}\right) \to 1, \quad t \to \infty.$$

Иными словами, большие двудольные графы весьма часто встречаются в графах $G_t$. Например, при $d = 10$, $i = \ln t/2$ с высокой вероятностью число сообществ, каждое из которых имеет размер $\ln t/2$ и цитирует

всего 10 сайтов, представляющих для него интерес, по порядку не меньше величины $\sqrt{t}$. И $\ln t/2$, и $\sqrt{t}$ — довольно значительные величины, ведь если, скажем, $t = 10^{10}$ (примерно, как в Интернете), то $\ln t/2 \approx 11$, а $\sqrt{t} = 100\,000$, т. е. в Интернете, построенном согласно модели, ожидается 100 000 двудольных подграфов с долями размера 11 и 10. Первое число совсем близко к реальному, но 11 чуть-чуть поменьше наблюдаемого в природе. Тем не менее, модель удивительно реалистична!

Отметим, что в теореме 35 от $\alpha$ зависит лишь скорость стремления вероятности к единице.

### 2.7.4. Модель Купера–Фриза

Эта модель исключительно громоздка, и потому она почти на грани осмысленности. Ее компютерная симуляция крайне затруднена из-за огромного числа параметров, а ее строгий анализ дает лишь частные результаты, точность которых весьма низкая. Тем не менее, мы опишем ее здесь по двум причинам. Во-первых, есть все основания ожидать, что при правильном подборе параметров модель даст очень адекватную картину Интернета. Во-вторых, задача аккуратного доказательства весьма точных результатов о свойствах случайного графа в модели все же не кажется совсем неподъемной.

В модели есть и элементы предпочтительного присоединения, и некоторые элементы копирования, и элементы, предвосхищающие модель Боллобаша–Боргса–Риордана–Чайес.

Стартует процесс с одной вершины и нуля ребер — с графа $G(0)$. На каждом шаге построения случайного графа $G(t)$ в модели выбирается одна из двух процедур — OLD или NEW. В первом случае ребра проводятся между уже существующими вершинами. Во втором случае добавляется новая вершина и ребра с концом в ней. Авторы модели подчеркивают, что не следят за ориентацией, т. е. считают ребра ненаправленными.

В рамках обеих процедур осуществляется, в свою очередь, выбор между двумя способами определения того, какие вершины будут соединяться новыми ребрами. С одной стороны, можно выбирать вершины согласно равномерному распределению (как в модели копирования). С другой стороны, их можно выбирать согласно степеням (повторим, что степень здесь всегда обычная, так как ни входящих, ни исходящих степеней просто нет).

Итак, есть куча параметров: натуральные числа $j_0 \geqslant 1$, $j_1 \geqslant 1$ и вещественные числа $\alpha$, $\beta$, $\gamma$, $\delta$, $p_1, \ldots, p_{j_0}$, $q_1, \ldots, q_{j_1}$. Здесь

$$p_i \in (0,1), \quad q_i \in (0,1), \quad p_1 + \ldots + p_{j_0} = 1, \quad q_1 + \ldots + q_{j_1} = 1,$$
$$\alpha, \beta, \gamma, \delta \in (0,1).$$

Пусть граф $G(t-1)$ построен. Работаем с шагом $t$. С вероятностью $\alpha$ выбираем процедуру OLD, с вероятностью $1-\alpha$ — NEW.

**Процедура NEW.** Появляется новая вершина. Числа $p_i$ — это вероятности того, что она породит $i$ ребер, т. е. новая вершина порождает *случайное* количество $i$ ребер в пределах от одного до $j_0$. Это новшество.

Сами $i$ ребер проводятся последовательно и взаимно независимо. При этом число $\beta$ — это вероятность того, что конец очередного ребра выбирается согласно равномерному распределению. Соответственно, с вероятностью $1-\beta$ конец этого ребра выбирается согласно степеням вершин.

**Процедура OLD.** Новая вершина не появляется. Числа $q_i$ — это вероятности того, что между старыми вершинами возникнет $i$ ребер, т. е. теперь старые вершины порождают *случайное* количество $i$ ребер в пределах от одного до $j_1$. Это тоже новшество.

Пусть $i$ — число ребер, которые предстоит провести. Проводим их последовательно и взаимно независимо. На каждом шаге выбираем две концевых вершины независимо друг от друга. Число $\delta$ — это вероятность того, что на очередном шаге первая вершина ребра выбирается согласно равномерному распределению. Соответственно, с вероятностью $1-\delta$ первая вершина выбирается согласно степени. Такую же роль играет $\gamma$ для второй вершины ребра.

Процесс полностью описан. Попытаемся хоть как-то сформулировать, что же о нем известно. Известно, что во многих случаях степени вершин распределены по степенному закону. Для сравнительно точной формулировки придется ввести ряд громоздких обозначений:

$$\mu_p = \sum_{j=0}^{j_0} jp_j, \quad \mu_q = \sum_{j=0}^{j_1} jq_j, \quad \theta = 2((1-\alpha)\mu_p + \alpha\mu_q),$$

$$a = 1 + \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha} + \frac{\alpha\delta}{1-\alpha},$$

$$b = \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta} + \frac{\alpha(1-\delta)}{\theta}, \quad c = \beta\mu_p + \frac{\alpha\gamma\mu_q}{1-\alpha},$$

$$d = \frac{(1-\alpha)(1-\beta)\mu_p}{\theta} + \frac{\alpha(1-\gamma)\mu_q}{\theta}, \quad e = \frac{\alpha\delta}{1-\alpha}, \quad f = \frac{\alpha(1-\delta)}{\theta}.$$

Далее последовательность $d_k$, $k = 0, 1, 2, \ldots$, определяется рекурсией $d_0 = 0$,

$$d_k(a+bk) = (1-\alpha)p_k + (c + d(k-1))d_{k-1} + \sum_{j=1}^{k-1}(e + f(k-j))q_j d_{k-j}.$$

Нетрудно проверить, что последовательна задана однозначно.

**Теорема 36.** *Пусть $G(t)$ — случайный граф в модели Купера–Фриза. Тогда при фиксированных $k \in \mathbb{N}$ выполнено*

$$\mathbb{E}\,|\{v \in V(G(t))\colon \deg v = k\}| \sim td_k, \quad t \to \infty.$$

Степенной закон забрезжит лишь при условии, что $d_k$ убывает как степень $k$. Несколько подобных утверждений доказаны и у Купера–Фриза (см. [47]). Например,

- если $j_1 = 1$, то с некоторым постоянным $C$ выполнено

$$d_k \sim Ck^{-1-\frac{1}{d+l}}, \quad k \to \infty;$$

- если $f = 0$, то с некоторым постоянным $C$ выполнено

$$d_k \sim Ck^{-1-\frac{1}{d}}, \quad k \to \infty.$$

Есть еще разные случаи, но мы отошлем уставшего читателя к оригинальной статье [47].

### 2.7.5. Модель Холма–Кима

Здесь мы поговорим об одной из многочисленных попыток справиться с проблемой неадекватности кластерного коэффициента в моделях типа Барабаши–Альберт своему аналогу в реальной сети. В целом, все подобные попытки достаточно искусственны, и та, которую мы сейчас опишем, не является исключением. Просто она довольно типична.

Идея совсем простая. Предлагается при построении случайного графа перемежать шаги предпочтительного присоединения с шагами, на которых бы формировались дополнительные треугольники. В отличие от моделей чистого предпочтительного присоединения, моделей копирования и их смесей, когда в основу случайного процесса, формирующего граф, закладываются один или несколько интуитивно обоснованных принципов (типа «к ссылкам ссылки» или «цитата с интересного мне сайта — и моя цитата тоже») и эти принципы в итоге оказываются подтвержденными эмпирически (т. е. за счет них получаются нужные распределения степеней вершин и пр.), модель с упомянутыми дополнительными шагами не имеет интуитивной мотивировки: она заточена под увеличение кластерного коэффициента, и это все.

Конкретная модель Холма–Кима (см. [48]) устроена следующим образом (ср. модель $G_m^{(n)}$ из п. 2.4.1). Строится последовательность случайных графов $G_m^{(n)}$. Граф $G_m^{(1)}$ состоит из одной вершины и одной петли. Если уже построен граф $G_m^{(n)}$ с вершинами $1, \ldots, n$ и $mn$ ребрами, то для построения графа $G_m^{(n+1)}$ добавляется вершина $n+1$ и $m$ ребер из

нее. Ребра добавляются последовательно и взаимно независимо. Первое ребро добавляется согласно предпочтительному присоединению, т. е. согласно распределению

$$\mathbb{P}(n+1 \to n+1) = \frac{1}{2n+1}, \quad \mathbb{P}(n+1 \to i) = \frac{\deg_{G_m^{(n)}} i}{m(2n+1)}, \quad i = 1, \ldots, n.$$

Пусть уже добавлены $i$ ребер, $1 \leqslant i < m$. Надо добавить ребро с номером $i+1$. С некоторой заранее заданной вероятностью $p$ (возможно, зависящей от $n$: в этом случае заранее задается функция $p = p(n)$) новое ребро добавляется опять-таки согласно указанному выше распределению. С вероятностью $1-p$ делается следующее. Отыскивается последний момент $j \leqslant i$, когда ребро из $n+1$ проводилось согласно предпочтительному присоединению (такой момент заведомо найдется, так как $j = 1$ точно годится), и берется вторая вершина $v$ того ребра. Если вершина $n+1$ еще не соединена со всеми *соседями* вершины $v$, то ребро из $n+1$ проводится в случайного соседа вершины $v$ (согласно равномерному распределению). Иначе ребро из $n+1$ снова выбирается согласно предпочтительному присоединению.

Строго говоря, случайный граф в модели стоило бы обозначить $G_m^n(p)$, так как весь процесс существенно зависит от того, с какой вероятностью делается выбор между предпочтительным присоединением и искусственной попыткой добавить в сеть треугольник. Эксперименты, проведенные Холмом и Кимом (см. [48]), показали, что при многих $p$ степенной закон распределения степеней вершин случайного графа в модели (асимптотически) выполнен и его параметр равен, как и следовало ожидать, трем. Более того, при правильном подборе $p$ средний локальный кластерный коэффициент ограничен снизу константой. Про транзитивность, однако, Холм и Ким умалчивают, и это не случайно: конечно, ее тяжело считать, но главная беда в том, что транзитивность все же стремится к нулю! Иными словами, если в реальной сети скорее всего оба кластерных коэффициента постоянны (см., впрочем, дискуссию в разд. 1.11), то в модели Холма–Кима постоянен лишь больший из двух. Это отнюдь не победа, но заведомо лучше, чем Боллобаш–Риордан.

Почему-то начальную притягательность вершины Холм и Ким вводить не стали. Ясно, что с ее введением тройка в степенном законе превратится в $2 + a$. Таким образом, обобщенная модель Холма–Кима строго лучше модели Бакли–Остгуса. Правда, ни одной математической теоремы здесь нет, но очевидно, что при определенных усилиях (по-видимому, значительных) такие теоремы можно доказать.

ГЛАВА

# 3 СХЕМЫ И ИДЕИ НЕКОТОРЫХ ДОКАЗАТЕЛЬСТВ

## 3.1. НЕСКОЛЬКО ВВОДНЫХ СЛОВ

Мы не собираемся перегружать читателя массой технических деталей. И, конечно, у нас нет цели обсудить здесь доказательства всех трех с лишним десятков теорем, сформулированных в гл. 2. Нам хочется лишь дать представление о некоторых математических методах работы с моделями, которые мы описали ранее. Поэтому мы поступим следующим образом. В разд. 3.2 мы поговорим о доказательстве первой части теоремы 9. Иными словами, мы объясним, как Боллобаш с соавторами нашли асимптотику математического ожидания числа вершин полной степени $d$ в модели Боллобаша–Риордана и откуда у них взялось ограничение $d \leqslant n^{1/15}$. В разд. 3.3 мы приведем схему доказательства теоремы Гречникова, в которой устраняется ограничение. А разд. 3.4 мы посвятим обсуждению того, как из результатов о математическом ожидании получать утверждения об асимптотическом распределении (ср. теоремы 9, 12, 16, 20, 23, 24, 26, 28, 32).

## 3.2. СХЕМА ДОКАЗАТЕЛЬСТВА ТЕОРЕМЫ 9

Наша задача — убедиться в том, что для любого $d \leqslant n^{1/15}$

$$\mathbb{E}\left|\left\{i = 1, \ldots, n\colon \deg_{G_m^n} i = d\right\}\right| \sim \frac{2mn(m+1)}{d(d+1)(d+2)}, \quad n \to \infty.$$

Еще более упростим себе жизнь, сочтя, что $m = 1$ (на самом деле это и есть основной случай, так что $m > 1$ мы не станем рассматривать). Положим для краткости $d_i = \deg_{G_1^n} i$.

Сперва изучим случайную величину

$$D_k = \sum_{i=1}^{k} d_i,$$

равную суммарной степени первых $k$ вершин графа Боллобаша–Риордана. Посчитаем вероятность того, что $D_k = 2k + s$, $0 \leqslant s \leqslant n - k$. Воспользуемся линейными хордовыми диаграммами из п. 2.4.2. Очевидно, что интересующее нас событие состоит в том, что $k$-й по счету правый конец дуги в хордовой диаграмме попадает в точку $2k + s$. Искомая вероятность есть отношение числа таких хордовых диаграмм к их общему количеству $\frac{(2n)!}{2^n n!}$. Числитель искомой дроби равен произведению количества способов выбрать левый конец дуги с правым концом в $2k + s$ (это количество есть $2k + s - 1$), количества способов из еще не задействованных $2k + s - 2$ точек слева от точки $2k + s$ выбрать ровно $s$ точек, которые послужат левыми концами дуг, имеющими правые концы правее точки $2k + s$ (это количество есть $C^s_{2k+s-2}$), количества способов разбить на пары $2(k - 1)$ точек, оставшихся слева от точки $2k + s$ (это количество есть $\frac{(2k-2)!}{2^{k-1}(k-1)!}$), количества способов найти $s$ правых концов правее точки $2k + s$ для ранее выбранных $s$ дуг (это количество есть $C^s_{2n-2k-s}$), количества способов сочетать левые концы указанных дуг с правыми концами (это количество есть $s!$) и количества способов разбить на пары $2n - 2k - 2s$ точек, оставшихся не задействованными справа от точки $2k + s$ (это количество есть $\frac{(2n-2k-2s)!}{2^{n-k-s}(n-k-s)!}$). После несложных выкладок имеем

$$\mathbb{P}(D_k = 2k + s) = \frac{(2k + s - 1)!\,(2n - 2k - s)!\,n!\,2^{s+1}}{s!\,(k - 1)!\,(n - k - s)!\,(2n)!}.$$

Положим

$$p_s = \mathbb{P}(D_k = 2k + s), \quad r_s = \frac{p_{s+1}}{p_s} = 2\frac{(2k + s)(n - k - s)}{(s + 1)(2n - 2k - s)}.$$

Нетрудно проверить, что функция $r_s$ убывающая и что при больших $n$ выполнено

$$\frac{r_{s+1}}{r_s} \leqslant e^{-\frac{1}{2n}}.$$

Значит, единственное решение уравнения $r_s = 1$ — это

$$s = -2k + \sqrt{4kn - 2n + \frac{1}{4}} + \frac{1}{2},$$

откуда следует, что число

$$s_0 = \left\lceil -2k + \sqrt{4kn - 2n + \frac{1}{4}} + \frac{1}{2} \right\rceil$$

является одним из наиболее вероятных значений случайной величины $D_k - 2k$.

Для $x > 0$ из неравенств $r_{s_0} \leqslant 1$ и $r_{s+1}/r_s \leqslant e^{-\frac{1}{2n}}$ следуют оценки $r_{s_0+x} \leqslant e^{-\frac{x}{2n}}$ и $p_{s_0+x} \leqslant e^{-\frac{x(x-1)}{4n}}$, а также аналогичные оценки для $r_{s_0-x}$ и $p_{s_0-x}$. В итоге

$$\mathbb{P}\big(|D_k - (2k + s_0)| \geqslant 3\sqrt{n \ln n}\big) = o(n^{-1}).$$

Более того, поскольку

$$\big|s_0 - \big(2\sqrt{kn} - 2k\big)\big| \leqslant 2\sqrt{n}$$

при всех $k$, получаем

$$\mathbb{P}\big(|D_k - 2\sqrt{kn}| \geqslant 4\sqrt{n \ln n}\big) = o(n^{-1}).$$

Комбинаторные рассуждения с хордовыми диаграммами показывают (мы не станем вдаваться в детали), что

$$\mathbb{P}(d_{k+1} = d + 1 \,|\, D_k = 2k + s) = \\ = (s + d)2^d \frac{(n - k - s)(n - k - s - 1) \cdot \ldots \cdot (n - k - s - d + 1)}{(2n - 2k - s)(2n - 2k - s - 1) \cdot \ldots \cdot (2n - 2k - s - d)}.$$

Ниже мы используем обозначения $o(\cdot)$, $O(\cdot)$, $\sim$, подразумевая, что константы в них абсолютные и асимптотики зависят только от $n$ — не от $d \leqslant n^{1/15}$ или $k$.

Итак, пусть $M = [n^{4/5}/\ln n]$, а $k$ — любая функция со значениями в пределах $[M, n - M]$. Исходя из этих ограничений и пресловутого неравенства $d \leqslant n^{1/15}$ (именно здесь оно и нужно), можно показать, что равномерно по всем $D$, удовлетворяющим условию $|D - 2\sqrt{kn}| \leqslant 4\sqrt{n \ln n}$, имеет место асимптотика

$$\mathbb{P}(d_{k+1} = d + 1 \,|\, D_k = D) \sim \sqrt{\kappa}(1 - \sqrt{\kappa})^d,$$

где $\kappa = k/n$.

По формуле полной вероятности

$$\mathbb{P}(d_{k+1} = d + 1) = o(n^{-1}) + (1 + o(1))\sqrt{\kappa}(1 - \sqrt{\kappa})^d.$$

Значит, среднее число вершин полной степени $d + 1$ есть

$$O(M) + o(1) + \sum_{k=M}^{n-M} (1 + o(1))\sqrt{\frac{k}{n}}\left(1 - \sqrt{\frac{k}{n}}\right)^d.$$

Последняя сумма асимптотически равна интегралу

$$\int_0^1 n\sqrt{\kappa}(1-\sqrt{\kappa})^d d\kappa \sim \frac{4n}{(d+1)(d+2)(d+3)},$$

откуда и вытекает заявленный в теореме результат.

## 3.3. СХЕМА ДОКАЗАТЕЛЬСТВА ТЕОРЕМЫ 10

Здесь мы поговорим сразу о произвольном $m \geqslant 1$. Вспомним, что граф $G_m^n$ получается из графа $G_1^{mn}$ путем склейки последовательных блоков из $m$ вершин в новые вершины: $v_1 = \{1, \ldots, m\}$, $v_2 = \{m+1, \ldots, 2m\}$, ..., $v_n = \{m(n-1)+1, \ldots, mn\}$. В частности, оправдано обозначение $i \in v_\beta$ с тем или иным $\beta \in \{1, \ldots, n\}$.

Обозначим, далее, $\mathbb{P}_t$ вероятностную меру на пространстве графов $G_1^t$. Заметим, что если нас интересует изменение степени какой-то вершины $v_\beta$ при переходе от графа $G_m^n$ к графу $G_m^{n+1}$, то мы можем следить за ним, добавляя к уже сформированным вершинам $v_1, \ldots, v_n$ графа $G_m^n$ вершины $mn+1, mn+2, \ldots$ графов $G_1^{mn+1}, G_1^{mn+2}, \ldots$, покуда не сформируется вершина $v_{n+1} = \{mn+1, \ldots, m(n+1)\}$, а вместе с ней и граф $G_m^{n+1}$.

В текущих обозначениях искомая величина есть

$$r(d,n) = \mathbb{E}\,|\{\beta = 1, \ldots, n\colon \deg v_\beta = d\}| = \sum_{\beta=1}^{n} \mathbb{P}_{mn}\left(\deg v_\beta = d\right).$$

Положим

$$r(d,n,k) = \sum_{\beta=1}^{n} \mathbb{P}_{mn+k}\left(\deg v_\beta = d\right).$$

Тогда $r(d,n) = r(d,n,0)$. Напишем рекуррентное соотношение для $r(d,n,k)$. В самом деле, если $i$ — конец единственного ребра, выходящего из вершины $mn+k+1$ графа $G_1^{mn+k+1}$, то

$$\begin{aligned}
\mathbb{P}_{mn+k+1}(\deg v_\beta = d) &= \mathbb{P}_{mn+k+1}(\deg v_\beta = d,\ i \in v_\beta) + \\
&+ \mathbb{P}_{mn+k+1}(\deg v_\beta = d,\ i \notin v_\beta) = \mathbb{P}_{mn+k+1}\left(\deg_{mn+k} v_\beta = d-1,\ i \in v_\beta\right) + \\
&+ \mathbb{P}_{mn+k+1}\left(\deg_{mn+k} v_\beta = d,\ i \in v_\beta\right) = \\
&= \mathbb{P}_{mn+k}\left(\deg_{mn+k} v_\beta = d-1\right)\frac{d-1}{2mn+2k+1} + \\
&+ \mathbb{P}_{mn+k}\left(\deg_{mn+k} v_\beta = d\right)\left(1 - \frac{d}{2mn+2k+1}\right).
\end{aligned}$$

Суммируя по всем $\beta$, получаем

$$r(d,n,k+1) = r(d-1,n,k)\frac{d-1}{2mn+2k+1} + r(d,n,k)\Bigl(1-\frac{d}{2mn+2k+1}\Bigr).$$

Начальными условиями рекурсии служат

$$r(d,n+1,0) = r(d,n,m) + \mathbb{P}(\deg v_{n+1} = d), \quad r(d,1,0) = \mathbb{I}(d=2m).$$

Завершение доказательства теоремы состоит в рутинной проверке того факта, что функция вида

$$\mathbb{I}(d \geqslant m)\frac{(2mn+2k+1)(m+1)}{d(d+1)(d+2)} - \frac{\mathbb{I}(d=m)}{m}(k+1) + O_{m+k}\Bigl(\frac{d}{n}\Bigr)$$

удовлетворяет описанной рекурсии.

## 3.4. НЕРАВЕНСТВА ПЛОТНОЙ КОНЦЕНТРАЦИИ И ТЕОРЕМЫ ОБ АСИМПТОТИЧЕСКОМ РАСПРЕДЕЛЕНИИ

### 3.4.1. Несколько вступительных слов

Все теоремы, в которых доказывается утверждение типа «для любого $\varepsilon > 0$ выполнено $\mathbb{P}((1-\varepsilon)f \leqslant \xi \leqslant (1+\varepsilon)f) \to 1$», получены из теорем об асимптотиках математического ожидания соответствующих величин $\xi$ применением тех или иных неравенств концентрации меры около среднего. Среди таких неравенств самым простым и общеизвестным является неравенство Чебышёва, и о нем мы поговорим в п. 3.4.2. Однако наиболее употребляемым по факту оказывается неравенство Азумы–Хеффдинга, о котором мы расскажем в п. 3.4.3. Наконец, совсем недавно удалось использовать весьма продвинутую вероятностную технологию, придуманную Талаграном в 90-е годы XX века, и этому мы уделим внимание в п. 3.4.4.

### 3.4.2. Неравенство Чебышёва

Классическое неравенство Чебышёва, доказательство которого можно найти в любой стандартной книге по теории вероятностей, выглядит так:

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geqslant \delta) \leqslant \frac{\mathbb{D}\xi}{\delta^2}.$$

Здесь $\mathbb{D}\xi$ — дисперсия случайной величины $\xi$. По-другому можно переписать следующим образом:

$$\mathbb{P}(\mathbb{E}\xi - \delta \leqslant \xi \leqslant \mathbb{E}\xi + \delta) \leqslant \frac{\mathbb{D}\xi}{\delta^2}.$$

Следовательно, для получения результатов нужного нам вида достаточно брать $\delta = \varepsilon\mathbb{E}\xi$. Как правило, при наличии асимптотики математического ожидания удается даже заменить произвольную константу $\varepsilon > 0$ на функцию $\varepsilon = o(1)$.

В то же время у неравенства Чебышёва есть и большой минус: вычислить асимптотику дисперсии на несколько порядков сложнее, чем асимптотику среднего, которая и сама-то, как мы знаем, крайне тяжело поддается отысканию. В итоге среди всех теорем этой книги есть лишь одна, доказанная с помощью оценок дисперсии. Это теорема 24. А именно, пусть $R(d,n)$ — число вершин степени $d$ в графе $G_m^n$, а $r(d,n)$, как и в разд. 3.3, — ее математическое ожидание (только сейчас речь идет о модели Бакли–Остгуса). Справедлива

**Теорема 37.** *Пусть $d_1 \geqslant m$, $d_2 \geqslant m$. Тогда в случайном графе $H_{a,m}^n$ в модели Бакли–Остгуса выполнено*

$$\operatorname{cov}(R(d_1,n),R(d_2,n)) = O_{a,m}\left(\left(d_1^{-2-a} + d_2^{-2-a}\right)n + d_1^{-1}d_2^{-1}\right).$$

В теореме 24 асимптотика среднего есть $n\beta_{a,m,d} = \Theta\left(nd^{-2-a}\right)$. Возьмем в неравенстве Чебышёва

$$\delta = \left(\sqrt{d^{-2-a}n} + d^{-1}\right)\psi(n),$$

где $\psi(n)$ — произвольная функция, стремящаяся к бесконечности (сколь угодно медленно). Тогда по теореме 37 имеем $\mathbb{D}\xi/\delta^2 \to 0$ при $n \to \infty$, откуда и получается теорема 24 с $\varepsilon = \delta/(\mathbb{E}\xi)$. При этом $\varepsilon$ стремится к нулю, коль скоро $\sqrt{nd^{-2-a}} \to \infty$, и этим обусловлено ограничение $d = o\left(n^{-1/(a+2)}\right)$ в теореме 24.

Заметим, что если, наоборот, $n^{1/(a+2)} = o(d)$, то неравенство Чебышёва говорит нам, что с асимптотической вероятностью 1 выполнено $R(d,n) < 1$, т. е. $R(d,n) = 0$. Это означает, что в детализированной только что теореме 24 не только математическое ожидание числа вершин данной степени, но и асимптотика распределения этого числа найдена фактически при совершенно произвольных $d$. В этом преимущество неравенства Чебышёва перед другими неравенствами, о которых мы поговорим ниже. Беда лишь в том, что не всегда хватает сил найти асимптотику дисперсии.

### 3.4.3. Неравенство Азумы–Хёффдинга

Здесь читателю потребуется знание о том, что такое условное математическое ожидание и как определяется мартингал в теории вероятностей. В максимально элементарном и полностью пригодном для наших нынешних нужд виде эти сведения даны в книге [49], поэтому сейчас мы лишь напомним формулировку неравенства Азумы–Хёффдинга.

**Теорема 38.** *Пусть $\xi_0, \xi_1, \ldots, \xi_s$ — мартингал на некотором вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, причем существует такая величина $c > 0$, что $|\xi_i - \xi_{i-1}| \leqslant c$ для всех допустимых $i$ и всех $\omega \in \Omega$. Тогда для любого $a > 0$ выполнено*

$$\mathbb{P}(|\xi_s - \xi_0| \geqslant a) \leqslant 2e^{-\frac{a^2}{2sc^2}}.$$

Продемонстрируем, как работает неравенство на примере второй части теоремы 9. Нетрудно проверить, что последовательность из величин $X_t = \mathbb{E}(R(d,n) \mid G_m^t)$, $t = 0, \ldots, n$, образует мартингал. Более того, $X_n = R(d,n)$, $X_0 = \mathbb{E}R(d,n) = r(d,n)$. Наконец, $|X_i - X_{i-1}| \leqslant 2$. Значит, по теореме 38

$$\mathbb{P}\bigl(|R(d,n) - r(d,n)| \geqslant \sqrt{n \ln n}\bigr) \leqslant 2e^{-(\ln n)/8} \to 0, \quad n \to \infty.$$

Однако $\sqrt{n \ln n} = o(r(d,n))$ (по крайней мере в условиях теоремы, т. е. при $d \leqslant n^{1/15}$), откуда и результат.

Аналогично доказываются теоремы 12, 16, 20, 23, 28 и 32. В них наиболее муторное место — оценка разности соседних величин в мартингале. Кроме того, хотя они дают заведомо более точную оценку плотности концентрации, чем неравенство Чебышёва, они, в отличие от этого неравенства, работают не на всей области изменения параметров. Поэтому и возникают ситуации, в которых возможны уточнения. Например, именно так случилось с улучшением теоремы 12, найденном в теореме 26. О нем мы и поговорим в следующем пункте.

### 3.4.4. Неравенство Талаграна

Неравенство Талаграна — это один из мощных и притом совсем недавних инструментов доказательства плотной концентрации случайной величины около своего среднего значения. Любопытно то, что здесь концентрация ищется не около математического ожидания, но около медианы. Впрочем, если концентрации около медианы есть, то и математическое ожидание, конечно, находится неподалеку. Подробно и популярно неравенство Талаграна изложено в книге [50].

Напомним вкратце неравенство Талаграна. Оно работает на вероятностных пространствах $(\Omega, \mathcal{F}, \mathbb{P})$, в которых $\Omega = \Omega_1 \times \ldots \times \Omega_n$ и вероятность — это тоже произведение вероятностных мер, заданных на $\Omega_i$, $i = 1, \ldots, n$. Вводится расстояние от события $A \in \mathcal{F}$ до элементарного исхода $\mathbf{x} = (x_1, \ldots, x_n) \in \Omega$:

$$\operatorname{dist}(A, \mathbf{x}) = \max_{\alpha \in S^{n-1}} \min_{\mathbf{y} \in A} \sum_{i:\, x_i \neq y_i} \alpha_i,$$

где $S^{n-1} \subset \mathbb{R}^n$ — единичная сфера. Кроме того, $A_t = \{\mathbf{x}\colon \operatorname{dist}(A, \mathbf{x}) \leqslant t\}$. Имеет место

**Теорема 39.** *Для любого $t > 0$ и $A \in \mathcal{F}$ выполнено*

$$\mathbb{P}(A)(1 - \mathbb{P}(A_t)) \leqslant e^{-t^2/4}.$$

Поскольку мы собираемся говорить о теореме 26, то у читателя в первую очередь должен возникнуть вопрос: «Какое отношение модель Бакли–Остгуса имеет к вероятностным пространствам, задаваемым произведениями?» В самом деле, если бы речь шла о модели Эрдеша–Реньи, то все было бы ясно. Но тут случайный граф $H^n_{a,m}$ находится в прямой зависимости от графа $H^{n-1}_{a,m}$. Оказывается, можно описать модель Бакли–Остгуса и в терминах независимых случайных величин. Положим $m = 1$, как в теореме 26.

Рассмотрим последовательность

$$1, \xi_1, 2, \xi_2, \ldots, n, \xi_n,$$

где величины $\xi_1, \ldots, \xi_n$ взаимно независимы, определены на некоторых пространствах $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $i = 1, \ldots, n$, и принимают значения $\{1, 2, \ldots, 2i-1\}$ с вероятностями

$$\mathbb{P}_i(\xi_i = 2j-1) = \frac{a}{(a+1)i - 1}, \quad j = 1, \ldots, i,$$

$$\mathbb{P}_i(\xi_i = 2j) = \frac{1}{(a+1)i - 1}, \quad j = 1, \ldots, i-1.$$

Последовательность интерпретируется следующим образом. Каждое число $i$ — это вершина графа. Значение соответствующей величины $\xi_i$ определяет второй конец ребра, выходящего из вершины $i$. Если $\xi_i = 2j - 1$, то ребро идет просто в вершину $j$. Если же $\xi_i = 2j$, то мы говорим, что ребро из вершины $i$ идет в ту же вершину, что и ребро из вершины $j$. Значение величины $\xi_j$ может само быть четным (скажем, $2j'$), и тогда мы снова перенаправляем конец ребра, выходящего из $i$, согласно значению $\xi_{j'}$. В конце концов процесс останавливается на некотором нечетном значении $2v - 1$, и мы говорим, что $\xi_i$ (а равно $\xi_j$ и $\xi_{j'}$) ведет в вершину $v$. Нетрудно проверить, что возникает в точности $H^n_{a,1}$.

За счет такой интерпретации модели Бакли–Остгуса удается применить теорему 39, и в результате довольно сложных комбинаторных рассуждений возникает

**Теорема 40.** *Пусть $Y_n(d)$ — число вершин второй степени не ниже $d$ в случайном графе $H^n_{a,1}$. Тогда для любого $t > 0$, любых $d, s \in \mathbb{N}$ и любой функции $f(s)$, удовлетворяющей условию $f^2(s) > (2d+1)(4d+5)s$, выполнено*

$$\mathbb{P}(Y_n(d) \leqslant s - tf(s))\,\mathbb{P}(Y_n(d) \geqslant s) \leqslant e^{-t^2/4}.$$

Если положить в этой теореме $t = 2\ln n$, med — медиана распределения величины $Y_n(d)$, $s = \mathrm{med} + t(\mathbb{E}Y_n(d))^{1-\varepsilon}$, $f(s) = (\mathbb{E}Y_n(d))^{1-\varepsilon}$, то можно показать, что, как следствие, существует $\varepsilon' > 0$, с которым

$$\mathbb{P}\big(|Y_n(d) - \mathbb{E}Y_n(d)| > (\mathbb{E}Y_n(d))^{1-\varepsilon'}\big) \to 0, \quad n \to \infty,$$

коль скоро $\delta > 0$ и $d = O(n^{1/(2+a)-\delta})$.

Наконец, несложными манипуляциями из неравенства для $Y_n(d)$ выводится неравенство для

$$X_n(d) = |\{i = 1, \ldots, n\colon d_2(i) = d\}|,$$

имеющее вид

$$\mathbb{P}\big(|X_n(d) - \mathbb{E}X_n(d)| > (\mathbb{E}X_n(d))^{1-\varepsilon'}\big) \to 0, \quad n \to \infty,$$

коль скоро $\delta > 0$ и $d = O(n^{1/(4+a)-\delta})$. В последнем неравенстве и состоит утверждение теоремы 26.

# СПИСОК ЛИТЕРАТУРЫ


1. *Barabási L.-A., Albert R.* Emergence of scaling in random networks // Science. — 1999. — V. 286. — P. 509–512.
2. *Barabási L.-A., Albert R., Jeong H.* Scale-free characteristics of random networks: the topology of the world-wide web // Physica. — 2000. — V. A281. — P. 69–77.
3. *Albert R., Jeong H., Barabási L.-A.* Diameter of the world-wide web // Nature. — 1999. — V. 401. — P. 130–131.
4. *Albert R., Jeong H., Barabási A.-L.* Attack and error tolerance of complex networks // Nature. — 2000. — V. 406. — P. 378.
5. *Watts D. J., Strogatz S. H.* Collective dynamics of «small-world» networks // Nature. — 1998. — V. 393. — P. 440–442.
6. *Newman M. E. J.* Power laws, Pareto distributions and Zipf's law // Contemporary Physics. — 2005. — V. 46, No. 5. — P. 323–351.
7. *Grechnikov E. A., Gusev G. G., Ostroumova L. A., Pritykin Yu. L., Raigorodskii A. M., Serdyukov P., Vinogradov D. V., Zhukovskiy M. E.* Empirical Validation of the Buckley–Osthus Model for the Web Host Graph // The proceedings of The 21st ACM Conference on Information and Knowledge Management, 2012. — P. 1577–1581.
8. *Brin S., Page L.* The anatomy of a large-scale hypertextual web search engines // In: Proceedings of the 7th WWW Conference, 1998.
9. *Pandurangan G., Raghavan P., Upfal E.* Using PageRank to Characterize Web Structure // Internet Math. — 2006. — V. 3, No. 1. — P. 1–20.
10. *Newman M. E. J.* Assortative mixing in networks // Phys. Rev. Letter. — 2002. — V. 89. — P. 208701.
11. *Pastor-Satorras R., Vázquez A., Vespignani A.* Dynamical and Correlation Properties of the Internet // Phys. Rev. Lett. — 2001. — V. 87, No. 25. — P. 258701.
12. *Bollobás B.* Mathematical results on scale-free random graphs // Handbook of graphs and networks, 1–34. — Weinheim: Wiley-VCH, 2003.

13. *Newman M. E. J.* The structure and function of complex networks // SIAM Review. — 2003. — V. 45. — P. 167–256.
14. *Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D. -U.* Complex networks: Structure and dynamics // Physics Reports. — 2006. — V. 424. — P. 175–308.
15. *Kolountzakis M. N., Miller G. L., Peng R., Tsourakakis Ch. E.* Efficient Triangle Counting in Large Graphs via Degree-Based Vertex Partitioning // Internet Math. — 2011. — V. 8, No. 1. — P. 161–185.
16. *Durak N., Pinar A., Kolda T. G., Seshadhri C.* Degree Relations of Triangles in Real-world Networks and Graph Models // In: Proceedings of CIKM, 2012.
17. *Gibson D., Kumar R., Tomkins A.* Discovering Large Dense Subgraphs in Massive Graphs // Proceedings of the 31st VLDB Conference, 2005.
18. *Aiello W., Chung F., Lu L.* A random graph model for power law graphs // Experiment. Math. — 2001. — V. 10. — P. 53–66.
19. *Erdős P., Rényi A.* On random graphs I // Publ. Math. Debrecen. — 1959. — V. 6. — P. 290–297.
20. *Erdős P., Rényi A.* On the evolution of random graphs // Publ. Math. Inst. Hungar. Acad. Sci. — 1960. — V. 5. — P. 17–61.
21. *Erdős P., Rényi A.* On the evolution of random graphs // Bull. Inst. Int. Statist. Tokyo. — 1961. — V. 38. — P. 343–347.
22. *Bollobás B.* Random Graphs. — 2nd. ed. — Cambridge Univ. Press, 2001.
23. *Janson S., Łuczak T., Ruciński A.* Random graphs. — NY: Wiley, 2000.
24. *Колчин В. Ф.* Случайные графы. — М.: ФИЗМАТЛИТ, 2002.
25. *Райгородский А. М.* Модели случайных графов. — М.: МЦНМО, 2011.
26. *Barabási L. -A., Albert R.* Statistical mechanics of complex networks // Reviews of Modern Physics. — 2002. — V. 74, No. 1. — P. 47–97.
27. *Stoimenow A.* Enumeration of chord diagrams and an upper bound for Vassiliev invariants // J. Knot Theory Ramifications. — 1998. — V. 7, No. 1. — P. 93–114.
28. *Bollobás B., Riordan O.* Robustness and vulnerability of scale-free random graphs // Internet Math. — 2003. — V. 1, No. 1. — P. 1–35.
29. *Bollobás B., Riordan O.* The diameter of a scale-free random graph // Combinatorica. — 2004. — V. 24, No. 1. — P. 5–34.
30. *Bollobás B., Riordan O., Spencer J., Tusnády G.* The degree sequence of a scale-free random graph process // Random Structures Algorithms. — 2001. — V. 18, No. 3. — P. 279–290.
31. *Deijfen M., H. van den Esker, R. van der Hofstad, Hooghiemstra G.* A preferential attachment model with random initial degrees // Ark. Mat. — 2009. — V. 47, No. 1. — P. 41–72.
32. *Grechnikov E. A.* An estimate for the number of edges between vertices of given degrees in random graphs in the Bollobás–Riordan model // Moscow Journal of Combinatorics and Number Theory. — 2011. — V. 1, No. 2. — P. 40–73.

33. *Ostroumova L. A.*, *Grechnikov E. A.* The distribution of second degrees in the Bollobás–Riordan random graph model // Moscow Journal of Combinatorics and Number Theory. — 2012. — V. 2, No. 2. — P. 85–110.
34. *Rudas A.*, *Tóth B.*, *Valko B.* Random trees and general branching processes // Random Structures Algorithms. — 2007. — V. 31. — P. 186–202.
35. *Avrachenkov K.*, *Lebedev D.* PageRank of Scale-Free Growing Networks // Internet Math. — 2006. — V. 3, No. 2. — P. 207–232.
36. *Farkas I. J.*, *Derényi I.*, *Barabási A.-L.*, *Vicsek T.* Spectra of «real-world» graphs: Beyond the semicircle law // Phys. Rev. E. — 2001. — V. 64. — P. 026704.
37. *Рябченко А. А.*, *Самосват Е. А.* О числе подграфов в случайном графе Барабаши–Альберт // Изв. РАН. Сер. матем. — 2012. — Т. 6, №3. — С. 183–202.
38. *Drinea E.*, *Enachescu M.*, *Mitzenmacher M.* Variations on Random Graph Models for the Web // Harvard Technical Report TR-06-01 (2001).
39. *Dorogovtsev S. N.*, *Mendes J. F. F.*, *Samukhin A. N.* Structure of growing networks with preferential linking // Phys. Rev. Lett. — 2000. — V. 85. — P. 4633.
40. *Buckley P. G.*, *Osthus D.* Popularity based random graph models leading to a scale-free degree sequence // Discrete Math. — 2004. — V. 282. — P. 53–68.
41. *Móri T.F.* The maximum degree of the Barabási–Albert random tree // Combinatorics, Probability and Computing. — 2005. — V. 14. — P. 339–348.
42. *Grechnikov E. A.* The degree distribution and the number of edges between nodes of given degrees in the Buckley–Osthus model of a random web graph // Internet Math. — 2012. — V. 8, No. 3. — P. 257–287.
43. *Kupavskiy A. B.*, *Ostroumova L. A.*, *Shabanov D. A.*, *Tetali P.* The distribution of second degrees in the Buckley–Osthus random graph model // Internet Math., 2013.
44. *Eggemann N.*, *Noble S. D.* The clustering coefficient of a scale-free random graph // Discrete Applied Mathematics. — 2011. — V. 159, No. 10. — P. 953–965.
45. *Bollobás B.*, *Borgs Ch.*, *Chayes J.*, *Riordan O.* Directed scale-free graphs // ACM-SIAM Symposium on Discrete Algorithms. — 2003. — P. 132–139.
46. *Kumar R.*, *Raghavan P.*, *Rajagopalan S.*, *Sivakumar D.*, *Tomkins A.*, *Upfal E.* Stochastic models for the web graph // Proc. 41st Symposium on Foundations of Computer Science, 2000.
47. *Cooper C.*, *Frieze A.* On a general model of web graphs // Random Structures and Algorithms. — 2003. — V. 22. — P. 311–335.
48. *Holme P.*, *Kim B. J.* Growing scale-free networks with tunable clustering // Phys. Rev. E. — 2002. — V. 65. — P. 026107.
49. *Райгородский А. М.* Комбинаторика и теория вероятностей. — Долгопрудный: Издательский Дом «Интеллект», 2013.
50. *Alon N. and Spencer J.* The probabilistic method // Wiley-Interscience Series in Discrete Math. and Optimization. — 2nd ed., 2000. Русский перевод: *Алон Н.*, *Спенсер Дж.* Вероятностный метод. — М.: Бином. Лаборатория знаний, 2007.

# А.М. Райгородский

# Экстремальные задачи теории графов и Интернет

**Учебное издание**

Заявки на книги присылайте по адресам:
**zakaz@id-intellect.ru**
**solo@id-intellect.ru**
**id-intellect@mail.ru**

тел. (495) 579-96-45
факс (495) 617-41-88

В заявке обязательно указывайте
свои реквизиты (для организаций) и почтовый адрес!

Подробная информация о книгах на сайте
**http://www.id-intellect.ru**

Книжный магазин «Интеллект» в МФТИ
тел. (495) 408-73-55

**Андрей Михайлович Райгородский**

**МОДЕЛИ ИНТЕРНЕТА**

Компьютерная верстка – А.А. Пярнпуу
Корректура – автора
Ответственный за выпуск – Л.Ф. Соловейчик

Формат 60x90/16. Печать офсетная.
Гарнитура Ньютон.
Печ. л. 4. Тираж 500 экз. Зак. № 850.
Бумага офсетная № 1, плотность 80 г/м$^2$

Издательский Дом «Интеллект»
141700, Московская обл., г. Долгопрудный,
Промышленный пр-д, д. 14,
тел. (495) 617-41-85

Отпечатано в ГУП МО «Коломенская типография»
140400, г. Коломна, ул. III Интернационала, д. 2а.
Тел. 8(4966) 18-69-33, 18-60-16. ИНН 5022013940.
E-mail: bab40@yandex.ru, www.kolomna-print.ru

УЧЕБНОЕ ПОСОБИЕ ПО
МАТЕМАТИЧЕСКИМ МОДЕЛЯМ
СЛОЖНЫХ СЕТЕЙ

РАЙГОРОДСКИЙ АНДРЕЙ МИХАЙЛОВИЧ
Зав. кафедрой дискретной математики МФТИ, профессор кафедры математической статистики и случайных процессов мехмата МГУ, руководитель отдела теоретических и прикладных исследований компании Яндекс. В область научных интересов входят комбинаторика, теория графов и гиперграфов, теория вероятностей и математическая статистика, теория алгоритмов, большие сети типа Интернета. Лауреат премии Президента РФ для молодых ученых в области науки и инноваций 2011 года.

ISBN 978-5-91559-143-0

www.id-intellect.ru