# 2. Модель Эрдёша — Реньи случайного графа

В этой главе мы расскажем о самой глубоко изученной модели случайного графа — модели Эрдёша — Реньи.

## 2.1. Введение

Каждый из нас хорошо знает, что такое граф. Нет, разумеется, речь идет не о титуле, но о математическом объекте, который сплошь и рядом встречается как в «чистой» комбинаторике, так и в ее приложениях. Именно этот объект должен быть знаком всякому, кто хотя бы немного занимался или просто интересовался математикой — в особенности ее «дискретным» аспектом. Давайте только, на всякий случай, договоримся о терминологии и обозначениях.

Произнося слово «граф», мы подразумеваем, что в нашем графе нет петель, кратных ребер и ориентации. Говоря формально, граф у нас — это любая пара $G=(V, E)$, в которой $V$ — это (как правило, конечное) множество вершин, а $E$ — совокупность таких пар $(x, y)$ с $x, y \in V$, что

1) $x \neq y$ (отсутствие петель),
2) $(x, y) = (y, x)$ (отсутствие ориентации),
3) в $E$ нет совпадающих элементов (отсутствие кратных ребер).

Коль скоро мы пожелаем избавиться от ограничения 1), слово «граф» нам придется заменить на «псевдограф»; коль скоро мы захотим обойтись без условия 2), вместо слова «граф» мы скажем «орграф» (т. е. ориентированный граф); коль скоро мы пренебрежем свойством 3), «граф» превратится в «мультиграф»; возможны «мультипсевдографы», «псевдоорграфы» и т. д. Тем не менее до тех пор, пока не оговорено противное, граф — это именно граф в указанном, так сказать, узком смысле.

Про графы есть много книг: [13–19] и др. Имеется масса задач о графах. Некоторые из этих задач мы изучим в дальнейшем с вероятностной точки зрения.

В следующем параграфе мы дадим строгое определение модели Эрдёша — Реньи случайного графа.

## 2.2. Определение модели Эрдёша — Реньи

Зафиксируем произвольное натуральное число $n$ и рассмотрим множество $V = \{1, \ldots, n\}$. Это будет множество вершин случайного графа. Собственно случайными будут только ребра. Пусть $N = C_n^2$ и $e_1, \ldots, e_N$ суть все возможные ребра, которые можно провести на парах элементов из $V$, коль скоро строим мы именно *граф*. Иными словами, $e_1, \ldots, e_N$ — ребра *полного* графа $K_n$. Зададимся некоторым $p \in [0, 1]$ и станем выбирать ребра из множества $\{e_1, \ldots, e_N\}$ согласно схеме Бернулли с вероятностью успеха $p$ (см. §1.2), т. е. в случае успеха кладем очередное ребро в строящееся множество ребер $E$, а в случае неудачи — не кладем. Возникает случайный граф $G = (V, E)$.

Фактически, как и в схеме Бернулли, граф — это последовательность $\omega = (x_1, \ldots, x_N)$ из нулей и единиц. Просто теперь мы последовательность в известном смысле представляем графом: $x_i = 1$ — значит, $e_i \in E$; $x_i \neq 1$ — значит, $e_i \notin E$.

В итоге имеем вероятностное пространство $G(n, p) = (\Omega_n, \mathscr{F}_n, P_{n,p})$, в котором, как и положено,

$$|\Omega_n| = 2^N = 2^{C_n^2}, \quad P_{n,p}(G) = p^{|E|} q^{C_n^2 - |E|}.$$

По-другому можно еще сказать так: в модели Эрдёша — Реньи каждое ребро независимо от всех остальных ребер входит в случайный граф с вероятностью $p$. Хотя через $G(n, p)$ мы обозначили вероятностное пространство, мы будем говорить зачастую о «модели $G(n, p)$». Эта небольшая вольность речи нужна нам будет исключительно для краткости и удобства.

Каждое событие из $\mathscr{F}$ есть, разумеется, множество графов. Однако естественно интерпретировать события как свойства графов. В самом деле: что значит посчитать вероятность того, что граф обладает данным свойством $A$? Это и значит описать множество $A$ тех графов, которые данным свойством обладают, а затем найти вероятность $A$. Например, если $A$ — множество всех связных графов, то вероятность события «случайный граф связен» и есть $P_{n,p}(A) = \sum\limits_{G \in A} P_{n,p}(G)$.

Наука о случайном графе изучает вероятности тех или иных свойств графов. Наибольший интерес представляет «динамика» этих вероятностей, т. е. их изменение с ростом $n$. А числу вершин и впрямь ничто не мешает расти. При этом вероятность возникновения ребра графа тоже вполне может эволюционировать. Иными словами, мы интерпретируем $G(n, p)$ даже не как схему Бернулли, но как схему серий (ср. §1.3). В обозначениях параграфа 1.3 $\{n_1, n_2, n_3, \ldots\} =$

$=\{C_2^2, C_3^2, C_4^2, \ldots\}$ и $p = p(n)$. Может статься, скажем, что $p = \frac{1}{2}$ для всех $n$, а может статься, $p = \frac{1}{n}$ или $p = 1 - \frac{1}{\ln n}$, — как угодно.

Пусть $\{A_n\}$ — некоторая последовательность свойств-событий в пространствах $G(n, p)$. Например, $A_n$ состоит в связности графа, и тогда зависимости от $n$, по сути, нет, или $A_n$ сводится к наличию в графе не менее $n$ треугольников (троек вершин, попарно соединенных ребрами), и тогда индекс $n$ важен, и т. д. Мы говорим, что $A_n$ *выполнено почти наверное* (*п. н.*), если $P_{n,p}(A_n) \to 1$ при $n \to \infty$. Иными словами, вероятностная доля тех графов, для которых свойство выполняется, с ростом числа вершин все ближе к единице. Мы увидим в дальнейшем массу ситуаций, когда выполнение того или иного свойства почти наверное влечет поистине удивительные следствия.

Модель $G(n, p)$ была предложена П. Эрдёшем и А. Реньи в конце 50-х годов XX века. С тех пор наука шагнула далеко вперед, и о многих аспектах проблематики полувековой давности известно «почти все». Тем не менее даже здесь остается великое множество нерешенных задач.

Несколько весьма любопытных вопросов такого типа мы и изучим в настоящей главе.

## 2.3. Одна естественная модификация модели

Зачастую схему Бернулли, лежащую в основе модели $G(n, p)$, заменяют классической вероятностью. Делают это так. Рассматривают $M \leqslant N$ и выбирают $M$ случайных элементов из множества $\{e_1, \ldots, e_N\}$, получая, таким образом, случайный граф $G = (V, E)$. Если соответствующее вероятностное пространство обозначить $G(n, M) = (\Omega_n, \mathscr{F}_n, P_{n,M})$, то здесь, очевидно, $P_{n,M}(G) = \frac{1}{C_N^M}$. Опыт показывает, что работать с базовой (бернуллиевской) моделью значительно проще, нежели с только что описанной. В то же время, повозившись, можно показать, что во многих случаях есть прямая связь между вероятностями свойств в двух моделях. Заинтересованный читатель найдет необходимые сведения в книге [1]. Мы же для простоты будем работать только с моделью $G(n, p)$.

## 2.4. Треугольники в случайных графах

Этот параграф мы посвятим одной несложной задаче о случайном графе. Она не столь интересная, зато весьма иллюстративная. Речь пойдет о распределении треугольников в случайных графах.

### 2.4.1. Постановка задачи и формулировки результатов

Легче всего почувствовать вкус задач о случайных графах на примере задачи о распределении треугольников в них. Обозначим через $T_{3,n}$ случайную величину на пространстве $G(n, p)$, равную количеству треугольников в случайном графе. Например, $T_{3,n}(K_n) = C_n^3$, а для любого графа $G$ с двумя ребрами $T_{3,n}(G) = 0$. Сформулируем основные утверждения и прокомментируем их.

**Теорема 10.** *Пусть $\alpha$ — любая функция натурального аргумента $n$, стремящаяся к нулю при $n \to \infty$. Предположим, $p(n) = \dfrac{\alpha(n)}{n}$ для каждого $n \in \mathbb{N}$. Тогда почти наверное $T_{3,n} = 0$ (т. е. граф не содержит треугольников).*

**Теорема 11.** *Пусть $p(n) \sim \dfrac{c}{n}$, где $c > 0$ — константа. Тогда $T_{3,n}$ имеет асимптотически пуассоновское распределение с параметром $\lambda = \dfrac{c^3}{6}$.*

**Теорема 12.** *Пусть $\omega$ — любая функция натурального аргумента $n$, стремящаяся к бесконечности при $n \to \infty$. Предположим, $p(n) = \dfrac{\omega(n)}{n}$ для каждого $n \in \mathbb{N}$. Тогда почти наверное $T_{3,n} \geqslant 1$ (т. е. граф содержит треугольники).*

Поясним, почему в совокупности теоремы 10—12 дают практически исчерпывающую картину в задаче о треугольниках. Действительно, в теореме 10 говорится о функциях $p(n)$, которые «бесконечно малы» по сравнению с $\dfrac{1}{n}$ (в математическом анализе принято обозначение $p = o\left(\dfrac{1}{n}\right)$, см. приложение); в теореме 11 имеем $p(n) \sim \dfrac{c}{n}$; наконец, в теореме 12 имеем $\dfrac{1}{n} = o(p)$. Понятно, что разумные функции как-то так и устроены: либо малы в сравнении с дробью $\dfrac{1}{n}$, либо в константу раз от нее отличаются, либо велики в сравнении с ней. Остальные изыски нам вряд ли интересны в данном контексте.

Получается, что в первой ситуации (когда вероятность ребра совсем маленькая) треугольников п.н. нет; в третьей ситуации (когда вероятность ребра совсем большая) треугольники п.н. есть; и лишь в пограничной — второй — ситуации вероятность отсутствия треугольника (равно как и вероятность его наличия) не стремится ни к нулю, ни к единице. Точнее, во второй ситуации

$$P_{n,p}(T_{3,n} = 0) \sim \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\frac{c^3}{6}} \in (0, 1).$$

В науке о случайных графах, коль скоро возникла подобная картина, принято говорить о *фазовом переходе*, а функцию типа нынешней $\frac{1}{n}$ принято называть *пороговой*. Смысл в том, что, едва мы преодолеваем «порог», как от «почти наверное выполнено данное свойство» мы переходим к «почти наверное не выполнено данное свойство», и только, «стоя прямо на пороге», мы, возможно, не сваливаемся ни в одну из «воронок».

В последующих пунктах мы приведем доказательства теорем 10, 11 и 12.

### 2.4.2. Доказательство теоремы 10

Нам нужно убедиться в том, что $P_{n,p}(T_{3,n}=0)\to 1$ при $n\to\infty$. Это равносильно соотношению $P_{n,p}(T_{3,n}\geqslant 1)\to 0$. Воспользуемся неравенством Маркова (см. § 1.11): $P_{n,p}(T_{3,n}\geqslant 1)\leqslant MT_{3,n}$. Найдем, стало быть, математическое ожидание числа треугольников в случайном графе и проверим, что оно стремится к нулю с ростом числа вершин графа. Применим линейность (см. § 1.8), а именно, занумеруем все тройки вершин из множества $V=\{1,\dots,n\}$, обозначая их, скажем, $\tau_1,\dots,\tau_{C_n^3}$, и для каждого $i\in\{1,\dots,C_n^3\}$ положим $T_{3,n,i}(G)=1$, если вершины из $\tau_i$ попарно соединены ребрами в $G$ (образуют треугольник), и $T_{3,n,i}(G)=0$ в противном случае. Очевидно,

$$T_{3,n}=T_{3,n,1}+\ldots+T_{3,n,C_n^3},$$

так что

$$MT_{3,n}=MT_{3,n,1}+\ldots+MT_{3,n,C_n^3}.$$

Однако $MT_{3,n,i}=p^3$ для любого $i$, и, значит, $MT_{3,n}=C_n^3p^3$.

Вспоминаем комбинаторику и тот факт, что у нас $p=\frac{\alpha}{n}$:

$$MT_{3,n}=C_n^3p^3=\frac{n!}{6(n-3)!}\cdot\frac{\alpha^3}{n^3}=\frac{n(n-1)(n-2)\alpha^3}{6n^3}\sim\frac{\alpha^3}{6}\to 0$$

ввиду условий, наложенных нами на функцию $\alpha$ в формулировке теоремы. Теорема доказана.

### 2.4.3. Доказательство теоремы 12

Здесь нам нужно установить неравенство, противоположное тому, которое мы обосновали в предыдущем пункте. Прежде всего обратимся к помощи неравенства Чебышёва (см. § 1.11):

$$P_{n,p}(T_{3,n}=0)=P_{n,p}(T_{3,n}\leqslant 0)=P_{n,p}(-T_{3,n}\geqslant 0)=$$
$$=P_{n,p}(MT_{3,n}-T_{3,n}\geqslant MT_{3,n})\leqslant\frac{DT_{3,n}}{(MT_{3,n})^2}.$$

Если теперь нам удастся показать, что $\frac{DT_{3,n}}{(MT_{3,n})^2}\to 0$, то утверждение теоремы будет у нас в кармане. Как видно, отныне нам не хватит знания об асимптотическом поведении среднего; нужно еще и со вторым моментом разобраться.

Итак, $DT_{3,n}=MT_{3,n}^2-(MT_{3,n})^2$, причем вычитаемое мы уже нашли ранее. В обозначениях п. 2.4.2

$$MT_{3,n}^2=M(T_{3,n,1}+\ldots+T_{3,n,C_n^3})^2.$$

Ни о какой линейности речь тут, разумеется, уже не идет, тем более что индикаторы $T_{3,n,i}$ отнюдь не являются независимыми. После возведения в квадрат получаем (за счет обычной линейности математического ожидания)

$$MT_{3,n}^2=MT_{3,n,1}^2+\ldots+MT_{3,n,C_n^3}^2+\sum_{i\neq j}M(T_{3,n,i}T_{3,n,j}).$$

Здесь последнее суммирование идет по всем *упорядоченным* парам несовпадающих индексов $i$, $j$; оттого и двойки нет в качестве сомножителя.

Заметим, что $T_{3,n,i}^2=T_{3,n,i}$, так что

$$MT_{3,n,1}^2+\ldots+MT_{3,n,C_n^3}^2=MT_{3,n}.$$

В то же время полезно понимать, что

$$\sum_{i\neq j}M(T_{3,n,i}T_{3,n,j})=M_f^2T_{3,n}.$$

Эта величина есть среднее количество упорядоченных пар $i$, $j$, для которых обе тройки вершин $\tau_i$, $\tau_j$ случайного графа образуют треугольник в нем.

Промежуточный итог следующий:

$$DT_{3,n}=MT_{3,n}^2-(MT_{3,n})^2=MT_{3,n}+M_f^2T_{3,n}-(MT_{3,n})^2.$$

Допустим, мы показали, что одновременно выполнены два свойства: 1) $MT_{3,n}\to\infty$; 2) $M_f^2T_{3,n}\sim(MT_{3,n})^2$ или, что то же самое, $M_f^2T_{3,n}=$

$=(1+o(1))(MT_{3,n})^2$. Тогда

$$P_{n,p}(T_{3,n}=0)\leqslant\frac{DT_{3,n}}{(MT_{3,n})^2}=\frac{MT_{3,n}+M_f^2T_{3,n}-(MT_{3,n})^2}{(MT_{3,n})^2}=$$
$$=\frac{1}{MT_{3,n}}+\frac{(1+o(1))(MT_{3,n})^2-(MT_{3,n})^2}{(MT_{3,n})^2}=o(1)+o(1)\to 0,$$

и дело в шляпе. Проверим утверждения 1) и 2).

С утверждением 1) все совсем просто:

$$MT_{3,n}=C_n^3p^3=\frac{n!}{6(n-3)!}\cdot\frac{\omega^3}{n^3}=\frac{n(n-1)(n-2)\omega^3}{6n^3}\sim\frac{\omega^3}{6}\to\infty,$$

и вся недолга.

Что же до утверждения 2), то тут придется чуть-чуть попотеть. В самом деле, нам нужно найти асимптотику для

$$M_f^2T_{3,n}=\sum_{i\neq j}M(T_{3,n,i}T_{3,n,j}).$$

Фактически каждое слагаемое $M(T_{3,n,i}T_{3,n,j})$ есть вероятность того, что треугольники с вершинами в $\tau_i\neq\tau_j$ принадлежат случайному графу. Если $\tau_i\cap\tau_j=\varnothing$, то такая вероятность равна $p^6$; если $|\tau_i\cap\tau_j|=1$, то ответ тот же; если же $|\tau_i\cap\tau_j|=2$, то вероятность равна $p^5$. Разобьем искомую сумму на две части: в первой будут слагаемые величины $p^6$, во второй — $p^5$. Простая комбинаторика показывает, что в первой части слагаемых $C_n^3C_{n-3}^3+3C_n^3C_{n-3}^2$, а во второй — $3C_n^3(n-3)$. Имеем

$$M_f^2T_{3,n}=\sum_{i\neq j}M(T_{3,n,i}T_{3,n,j})=(C_n^3C_{n-3}^3+3C_n^3C_{n-3}^2)p^6+3C_n^3(n-3)p^5.$$

Остается доказать, что

$$\frac{M_f^2T_{3,n}}{(MT_{3,n})^2}\to 1.$$

Действительно,

$$C_n^3\sim C_{n-3}^3\sim\frac{n^3}{6},\quad C_{n-3}^2\sim\frac{n^2}{2},\quad n-3\sim n,$$

поэтому

$$M_f^2T_{3,n}\sim\left(\frac{n^6}{36}+\frac{n^5}{4}\right)p^6+\frac{n^4}{2}p^5\sim\frac{n^6}{36}p^6+\frac{n^4}{2}p^5.$$

Более того,

$$\frac{n^4p^5}{n^6p^6}=\frac{1}{n^2p}=\frac{1}{n\omega(n)}\to 0,$$

так что
$$M_f^2 T_{3,n} \sim \frac{n^6}{36} p^6.$$
В то же время
$$(MT_{3,n})^2 = (C_n^3 p^3)^2 \sim \frac{n^6}{36} p^6,$$
ввиду чего
$$\frac{M_f^2 T_{3,n}}{(MT_{3,n})^2} \sim 1.$$

Утверждение 2) доказано, и теорема доказана вместе с ним.

### 2.4.4. Доказательство теоремы 11

Всем, кто читал вводную главу 1 этой книги, а также всем, кто знаком с *методом моментов*, идея доказательства понятна с ходу: раз нам нужно установить асимптотическую пуассоновость некоторой случайной величины, значит, наверное, нам следует обратиться к технологии из параграфа 1.9.

Разумеется, наша задача состоит, тем самым, в отыскании асимптотик для факториальных моментов величины $T_{3,n}$. Начнем с первого из них, т. е. просто с математического ожидания. Ну, его-то мы уже не раз находили, и равно оно $C_n^3 p^3$. В текущей ситуации $p \sim \frac{c}{n}$ и, стало быть,
$$MT_{3,n} = \frac{n(n-1)(n-2)}{6} \cdot p^3 \sim \frac{n^3}{6} \cdot \frac{c^3}{n^3} = \frac{c^3}{6}.$$
Все, дальнейшее уже ясно: если мы верим в справедливость теоремы 11, то происхождение величины $\lambda$ из ее формулировки очевидно и нам остается только проверить, что для каждого $r$ величина $M_f^r T_{3,n}$ надлежащим образом близка к $\lambda^r$ (см. теорему 3); благо при $r=1$ близость стопроцентная.

Итак, нам нужно найти асимптотику для $M_f^r T_{3,n}$ при любом фиксированном $r$ и $n \to \infty$. Для $r=1$ и даже для $r=2$ (см. п. 2.4.3) мы ее знаем: в первом случае это $\lambda$, во втором — $\lambda^2$. Попробуем посмотреть на общий случай.

В пункте 2.4.3 мы специально обратили внимание читателя на то, что второй факториальный момент величины $T_{3,n}$ есть сумма по всем упорядоченным парам несовпадающих чисел $i$, $j$ математических ожиданий произведений индикаторов $T_{3,n,i}$, $T_{3,n,j}$. Аналогично устроен и $r$-й факториальный момент:
$$M_f^r T_{3,n} = \sum_{i_1,\ldots,i_r} M(T_{3,n,i_1} \cdot \ldots \cdot T_{3,n,i_r}),$$

где суммирование ведется по всем упорядоченным кортежам длины $r$ попарно различных индексов, каждый из которых лежит в пределах от 1 до $C_n^3$. Рассмотрим отдельно слагаемые, отвечающие тем кортежам $(i_1, \ldots, i_r)$, для которых тройки $\tau_{i_1}, \ldots, \tau_{i_r}$ не только попарно различны, но еще и попарно не пересекаются. Сумму этих слагаемых назовем $\Sigma_1 = \Sigma_1(n, r)$; сумму всех остальных математических ожиданий обозначим $\Sigma_2 = \Sigma_2(n, r)$.

Справедливы следующие два факта: 1) $\Sigma_1 \sim \lambda^r$; 2) $\Sigma_2 = o(\Sigma_1)$. Ниже мы последовательно эти факты докажем и тем завершим доказательство теоремы 11.

Начнем с факта 1). Каждое слагаемое в сумме $\Sigma_1$ есть вероятность того, что некоторые $r$ попарно не пересекающихся троек вершин из множества $\{1, \ldots, n\}$ образуют $r$ треугольников в случайном графе. Очевидно, такая вероятность равна $p^{3r}$. Количество слагаемых в сумме (т. е. количество способов выбрать $r$ троек из $n$-элементного множества) есть 0, коль скоро $3r > n$, и

$$C_n^3 \cdot C_{n-3}^3 \cdot \ldots \cdot C_{n-3(r-1)}^3$$

в противном случае. Здесь мы пользуемся тем, что кортежи из наших троек упорядочены. Таким образом, при условии $3r \leqslant n$

$$\Sigma_1 = C_n^3 \cdot C_{n-3}^3 \cdot \ldots \cdot C_{n-3(r-1)}^3 \cdot p^{3r}.$$

При фиксированном $r$

$$\lambda^r \sim (C_n^3 p^3)^r, \quad n \to \infty.$$

Оценим

$$\frac{\Sigma_1}{\lambda^r} \sim \frac{C_n^3}{C_n^3} \cdot \frac{C_{n-3}^3}{C_n^3} \cdot \ldots \cdot \frac{C_{n-3(r-1)}^3}{C_n^3}.$$

Для каждого $i \in \{1, \ldots, r\}$ имеем $\frac{C_{n-3(i-1)}^3}{C_n^3} \leqslant 1$ и

$$\begin{aligned}\frac{C_{n-3(i-1)}^3}{C_n^3} &= \frac{(n-3i+3)(n-3i+2)(n-3i+1)}{n(n-1)(n-2)} = \\ &= \left(1 - \frac{3i-3}{n}\right)\left(1 - \frac{3i-3}{n-1}\right)\left(1 - \frac{3i-3}{n-2}\right) \geqslant \\ &\geqslant \left(1 - \frac{3r}{n}\right)\left(1 - \frac{3r}{n-1}\right)\left(1 - \frac{3r}{n-2}\right) \geqslant \left(1 - \frac{3r}{n-2}\right)^3 \sim 1.\end{aligned}$$

Таким образом, при любом фиксированном $r$ справедлива асимптотика $\Sigma_1 \sim \lambda^r$, $n \to \infty$, и факт 1) обоснован.

Перейдем к факту 2). Нам нужно убедиться в том, что $\Sigma_2 = o(\Sigma_1)$, $n \to \infty$. Однако мы только что установили асимптотику $\Sigma_1 \sim \lambda^r = \mathrm{const}$. Таким образом, наше утверждение эквивалентно равенству $\Sigma_2 = o(1)$. В сумме $\Sigma_2$ суммирование идет по всем кортежам $(i_1, \ldots, i_r)$, для которых хотя бы две тройки в множестве $\{\tau_{i_1}, \ldots, \tau_{i_r}\}$ пересекаются (но никакие две тройки не совпадают). Поэтому можно представить $\Sigma_2$ в виде

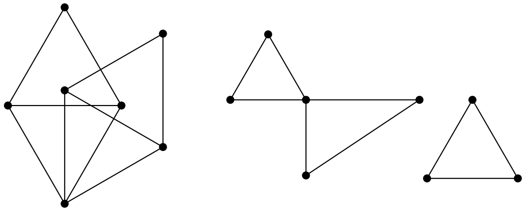
$$\Sigma_2 = \sum_{s=4}^{3r-1} \Sigma^s,$$

где в $\Sigma^s$ суммируются прежние $M(T_{3,n,i_1} \cdot \ldots \cdot T_{3,n,i_r})$ по всем кортежам, удовлетворяющим условию $|\tau_{i_1} \cup \ldots \cup \tau_{i_r}| = s$.

Практически очевидно, что если фиксированы $s$ и $i_1, \ldots, i_r$ со свойством $|\tau_{i_1} \cup \ldots \cup \tau_{i_r}| = s$, то совокупное число ребер $t$ у треугольников, порождаемых тройками $\tau_{i_1}, \ldots, \tau_{i_r}$, строго больше $s$ (см. примеры на рис. 1). Значит,

$$M(T_{3,n,i_1} \cdot \ldots \cdot T_{3,n,i_r}) = p^t \sim \frac{c^t}{n^t} = \frac{1}{n} \cdot O\left(\frac{1}{n^s}\right),$$

где значок $O$ имеет стандартный аналитический смысл, объясняемый в приложении.



*а)* $r=4$, $s=7$, $t=10$ *б)* $r=3$, $s=8$, $t=9$

Рис. 1

С другой стороны, при каждом $s$ количество слагаемых в $\Sigma^s$ есть $C_n^s \theta(s,r)$, где $\theta(s,r)$ является константой по $n$. При этом $C_n^s = O(n^s)$, и мы имеем в итоге

$$\Sigma_2 = \sum_{s=4}^{3r-1} \Sigma^s = \sum_{s=4}^{3r-1} C_n^s \cdot \theta(s,r) \cdot \frac{1}{n} \cdot O\left(\frac{1}{n^s}\right) =$$

$$= \sum_{s=4}^{3r-1} O(n^s) \cdot \frac{1}{n} \cdot O\left(\frac{1}{n^s}\right) = O\left(\frac{1}{n}\right) = o(1).$$

Факт 2) доказан, и теорема 11 доказана вместе с ним.

## 2.5. Связность случайного графа

Задача о треугольниках, которую мы изучали до сих пор, отлично иллюстрирует базовые методы теории случайных графов. Однако она слегка ученическая. Здесь мы перейдем к рассмотрению куда более глубокой и богатой приложениями проблематики. А именно, нас будет интересовать свойство связности случайного графа. Разумеется, мы не дадим уже столь исчерпывающей картины, какую мы имели в случае треугольников, но все же мы «снимем» самые «сливки» науки.

### 2.5.1. Формулировки результатов и комментарии к ним

Начнем с простейшего утверждения.

**Теорема 13.** *Пусть* $p = \frac{c\ln n}{n}$. *Если* $c > 1$, *то почти наверное случайный граф связен. Если* $c < 1$, *то почти наверное случайный граф связным не является.*

Опять фазовый переход! Но это и все, пожалуй, что понятно с ходу. Мы поступим следующим образом. Сперва сформулируем упрощенный (хотя и уточненный в то же время) результат. Затем прокомментируем его. В пункте 2.5.2 мы его докажем.

**Теорема 14.** *Пусть* $p = \frac{c\ln n}{n}$. *Если* $c \geqslant 3$ *и* $n \geqslant 100$, *то*

$$P_{n,p}(G\ \textit{связен}) > 1 - \frac{1}{n}.$$

Теорема 14 дает упрощенный вариант первого утверждения теоремы 13. Зато в ней конкретизируется выражение «почти наверное»: не просто вероятность стремится к единице, но делает она это с известной скоростью. Так практичнее.

Константа 100 в формулировке взята с большим запасом. Да не в ней суть.

Есть очень наглядная интерпретация теоремы 14. Представим себе, что в некоторой стране есть 2000 стратегически важных пунктов, причем страна столь богата, что каждые два из упомянутых пунктов связаны в ней *отдельным* железнодорожным путем. Из пункта М можно проехать в пункт В, минуя И, из М можно попасть в Ма через В, и т. д. Короче говоря, 2000-вершинный граф железных дорог полон (см. рис. 2).

Но есть у страны и «потенциальный противник». Допустим, в случае конфликта он уничтожает отдельно взятый железнодорожный путь с вероятностью $q$ независимо от всех остальных путей. Спрашивается, насколько большим может быть $q$, если страна желает

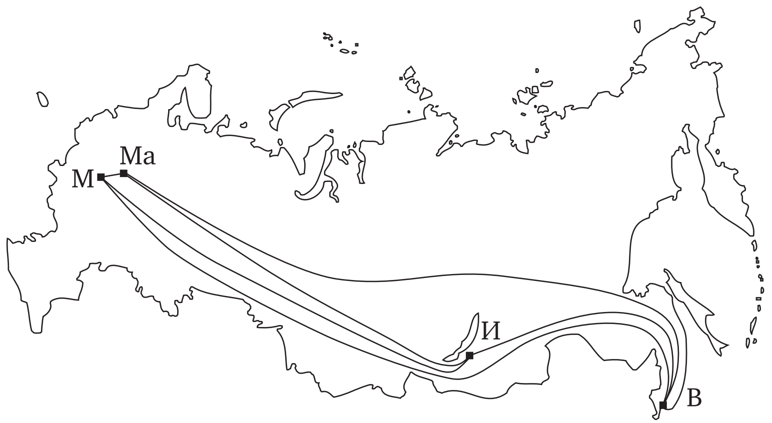



Рис. 2

сохранить свою «инфраструктуру» с достаточно большой вероятностью? Что такое инфраструктура? А это и есть связность! Если хоть каким-нибудь окольным путем мы можем добраться из любого пункта в любой другой, это уже кое-что. Иначе все: развал и дробление.

Заметим, что уничтожение путей с вероятностью $q$ — это ровно то же, что их возникновение с вероятностью $p$. Так ведь на выходе у нас будет как раз случайный граф железных дорог. По теореме 14, если $p = \frac{3\ln 2000}{2000} \approx 0{,}01$ (а стало быть, $q \approx 0{,}99$), то этот граф связен с вероятностью не меньше $1 - \frac{1}{2000} > 0{,}999$. Не правда ли, впечатляющий результат? Мы «разрешаем» противнику уничтожать каждый путь с вероятностью 0,99 и, тем не менее, сохраняем целостность дорожной сети с вероятностью 0,999!

Именно на этом примере отчетливо видно, почему, изучая случайный граф, мы не ограничиваемся схемой Бернулли с фиксированной вероятностью успеха, но рассматриваем схему серий, в рамках которой вероятность успеха может меняться с ростом числа испытаний. Если в стране окажется не 2000 стратегически важных пунктов, а, скажем, 100 000, то результат применения теоремы 14 будет еще более удивительным. Правда, стране тоже придется поднапрячься: шутка ли — в деле $C_{100000}^2$ железных дорог...

На самом деле, пример с железными дорогами взят лишь для наглядности. Любая сеть, в которой связи между узлами нарушаются независимо друг от друга с равными вероятностями, служит аналогичным примером. Например, речь может идти о телефонных станциях и линиях связи, об интернет-серверах, о различных социальных и биологических сетях и пр., и пр. Таким образом, практическая значимость теоремы 14 очевидна, и в свете сказанного она обычно интерпретируется прикладниками как теорема о *надежности сети*.

Разумеется, тут есть два крайне существенных ограничения: независимость, с которой уничтожаются связи в сети, и полнота исходного графа связей (т. е., собственно, сети). От этих ограничений необходимо избавляться, и кое-что мы скажем об этом в следующей главе. Пока же разберемся и с более простой ситуацией, благо и она совсем не так уж проста.

### 2.5.2. Доказательство теоремы 14

Мы слегка упростим себе жизнь и докажем лишь, что в условиях теоремы вероятность связности стремится к единице. Оценку скорости стремления мы оставим читателю в качестве упражнения.

Введем случайную величину на пространстве $G(n, p)$:

$$X_n = X_n(G) = \begin{cases} 0, & \text{если } G \text{ связен,} \\ k, & \text{если у } G\ k \text{ компонент.} \end{cases}$$

Таким образом, $X_n$ принимает неотрицательные целые значения, причем $X_n \neq 1$. Нам нужно показать, что $P_{n,p}(X_n = 0) \to 1$ при $n \to \infty$. Это равносильно асимптотике $P_{n,p}(X_n \geqslant 1) \to 0$. По неравенству Маркова (см. § 1.11) $P_{n,p}(X_n \geqslant 1) \leqslant MX_n$, и нам остается обосновать стремление к нулю математического ожидания.

Представим $X_n$ в виде суммы

$$X_n = X_{n,1} + \ldots + X_{n,n-1},$$

где $X_{n,k} = X_{n,k}(G)$ — число $k$-вершинных компонент графа $G$. Занумеруем все $k$-элементные подмножества множества вершин $V$ случайного графа в некотором (произвольном) порядке: $K_1, \ldots, K_{C_n^k}$. Тогда, в свою очередь,

$$X_{n,k} = X_{n,k,1} + \ldots + X_{n,k,C_n^k},$$

коль скоро

$$X_{n,k,i} = X_{n,k,i}(G) = \begin{cases} 1, & \text{если } K_i \text{ образует компоненту в } G, \\ 0 & \text{иначе.} \end{cases}$$

В итоге

$$MX_n = \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} MX_{n,k,i}.$$

Очевидно,

$$MX_{n,k,i} = P_{n,p}(K_i \text{ образует компоненту в } G) \leqslant$$
$$\leqslant P_{n,p}(\text{из } K_i \text{ в } V \setminus K_i \text{ нет ребер в } G).$$

Получая последнее неравенство, мы просто пренебрегли условием связности той части графа $G$, которая «сидит» на множестве вершин $K_i$ (такую часть принято называть *индуцированным подграфом* и обозначать $G|_{K_i}$). Далее (см. рис. 3),

$$P_{n,p}(\text{из } K_i \text{ в } V \setminus K_i \text{ нет ребер в } G) = q^{k(n-k)},$$

и, значит,

$$MX_n \leqslant \sum_{k=1}^{n-1} \sum_{i=1}^{C_n^k} q^{k(n-k)} = \sum_{k=1}^{n-1} C_n^k q^{k(n-k)}.$$

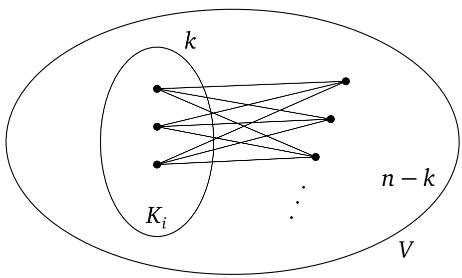



Рис. 3

Последняя сумма симметрична в том смысле, что ее слагаемые при $k$ и $n-k$ равны. Рассмотрим $k=1$:

$$nq^{n-1} = n(1-p)^{n-1} \leqslant ne^{-p(n-1)} \leqslant e^{-\frac{3(\ln n)(n-1)}{n}}.$$

При достаточно больших $n$ имеем $\dfrac{n-1}{n} \geqslant 0{,}9$. Следовательно,

$$nq^{n-1} \leqslant e^{-2{,}7\ln n} = \frac{1}{n^{2{,}7}}.$$

Возьмем теперь отношение соседних слагаемых в исследуемой сумме:

$$\frac{C_n^{k+1} q^{(k+1)(n-k-1)}}{C_n^k q^{k(n-k)}} = \frac{n-k}{k+1} \cdot q^{-k+n-k-1}.$$

Ввиду упомянутой выше симметрии достаточно смотреть лишь на $k \leqslant \frac{n}{2}$. Если, например, $k \leqslant \frac{n}{8}$, то

$$\frac{n-k}{k+1} \cdot q^{-k+n-k-1} \leqslant (n-1)q^{\frac{3n}{4}-1} \leqslant (n-1)e^{-\frac{9\ln n}{4}+p} \leqslant ne^{-\frac{9\ln n}{4}+1}.$$

При достаточно больших $n$ имеем

$$ne^{-\frac{9\ln n}{4}+1} \leqslant ne^{-2\ln n} = \frac{1}{n}.$$

Все это означает, что при $k \leqslant \frac{n}{8}$ каждое слагаемое в сумме заведомо меньше первого. При $k > \frac{n}{8}$ выполнены другие неравенства:

$$C_n^k < 2^n, \quad q^{k(n-k)} \leqslant q^{\frac{n}{8}\cdot\frac{n}{2}} = q^{\frac{n^2}{16}} \leqslant e^{-\frac{pn^2}{16}} \leqslant e^{-\frac{3(\ln n)n}{16}}.$$

В итоге

$$C_n^k q^{k(n-k)} \leqslant 2^n \cdot n^{-\frac{3n}{16}}.$$

При достаточно больших $n$ последнее выражение снова ничтожно мало по сравнению с $\frac{1}{n^{2,7}}$.

Итак, слагаемое с $k=1$ в искомой сумме наибольшее (с огромным запасом). Значит,

$$MX_n \leqslant \sum_{k=1}^{n-1} \frac{1}{n^{2,7}} < \frac{n}{n^{2,7}} \to 0, \quad n \to \infty,$$

и теорема доказана.

### 2.5.3. Вокруг теоремы 13

В этом пункте мы скажем еще несколько слов о теореме 13, которая все-таки точнее теоремы 14 с теоретической точки зрения.

Во-первых, понятно, что доказательство той части теоремы 13, которая касается $c > 1$, практически дословно повторяет доказательство теоремы 14. Просто надо еще аккуратнее провести выкладки.

Во-вторых, что такое случай $c < 1$? По сути нам нужно доказать в его рамках, что $P_{n,p}(X_n > 1) \to 1$, $n \to \infty$. Это очень похоже на теорему 12. Так и есть: доказательство следует осуществлять с помощью неравенства Чебышёва. Мы эти (довольно громоздкие) выкладки проводить не станем. Но заинтересованный читатель справится с ними и сам. Благо все инструменты у него уже есть.

Итак, мы снова имеем дело с фазовым переходом. Однако в прошлый раз (при изучении задачи о треугольниках) мы разобрались даже с ситуацией «балансирования на пороге». Сейчас ее аналогом служит та, в которой $p \sim \frac{\ln n}{n}$. Функций указанного вида великое множество. И вот для одного их класса удается получить аналог теоремы 11.

**Теорема 15.** *Пусть* $p = \frac{\ln n + c + o(1)}{n}$. *Тогда*

$$P_{n,p}(G \text{ связен}) \to e^{-e^{-c}}.$$

*В частности, при* $p = \frac{\ln n}{n}$ *вероятность стремится к* $e^{-1}$.

Доказательство теоремы 15 можно найти в книге [1], и мы его не приводим. Оно довольно сложное, но сам факт выглядит весьма привлекательно.

### 2.5.4. Гигантская компонента

В этом пункте мы расскажем еще об одном удивительном фазовом переходе, относящемся к свойству связности.

**Теорема 16.** *Пусть* $p = \frac{c}{n}$. *Тогда при любом* $c < 1$ *существует такая константа* $\beta = \beta(c) > 0$, *что почти наверное каждая компонента случайного графа имеет не более* $\beta \ln n$ *вершин. При любом* $c > 1$ *существует такая константа* $\gamma = \gamma(c) \in (0, 1)$, *что почти наверное среди компонент случайного графа есть одна, число вершин которой не меньше* $\gamma n$.

В действительности, это один из результатов, которые Эрдёш и Реньи получили в своих основополагающих работах [20–22]. Именно с него, по большому счету, пошла вся наука о случайных графах. Вокруг этого результата есть масса литературы, и в ней можно найти его разнообразные доказательства. Сошлемся, например, на книги [1] и [23]. Здесь мы теорему 16 доказывать не станем.

Сразу отметим, что и о «внутренности фазового перехода», т. е. о случае $p \sim \frac{1}{n}$ известно очень много (см. [1, 23]). Однако тут все гораздо сложнее, нежели то было в предшествующих ситуациях, и мы не будем вдаваться в соответствующие детали.

Если вернуться к «стратегической» интерпретации задачи о связности, можно сказать так. «Разрешив» противнику уничтожать каждый железнодорожный путь с вероятностью $1 - \frac{1}{n}$, мы с вероятностью, близкой к единице, сохраним если не всю страну, то хотя бы ее огромный кусок — кусок размера $\gamma n$. Этот кусок принято называть *гигантской компонентой*. Если же мы дадим противнику еще больше воли, то развал почти неизбежен, и это развал катастрофический:

островки связности, скорее всего, имеют ничтожный — логарифмический — размер. Кстати, даже при наличии гигантской компоненты все остальные компоненты крошечные — логарифмические. Это тоже можно строго обосновать.

В принципе так устроена любая крупная сеть. Например, так устроен Интернет. В нем узлы — это сайты, а связи — ссылки между сайтами. Есть гигантская связная компонента и масса мелких, разрозненных «окраин». Ну, об Интернете мы поговорим отдельно.

Завершим пункт еще одной стратегической интерпретацией. Итак, можно представлять себе, что $p$ — это своего рода время, меняющееся на отрезке $[0, 1]$. Если задано достаточно большое $n$, то при $p \ll \frac{1}{n}$ мы имеем феодальную раздробленность; время идет, и возникает империя; наконец, при $p \gg \frac{\ln n}{n}$ империя получает мировое господство. Такая вот история.

## 2.6. Хроматическое число случайного графа

Этот параграф мы посвятим одному из самых изящных и нетривиальных сюжетов теории случайных графов: речь пойдет о хроматическом числе.

### 2.6.1. Определения, формулировки и комментарии

Напомним, что *хроматическим числом* графа $G = (V, E)$ называется величина $\chi(G)$, равная минимальному количеству цветов, в которые можно так покрасить все вершины графа, чтобы концы любого ребра были разноцветными. Иными словами,

$$\chi(G) = \min\{\chi : V = V_1 \sqcup \ldots \sqcup V_\chi,\ \forall i \ \ \forall x, y \in V_i \ \ (x, y) \notin E\}.$$

Здесь посредством значка «$\sqcup$» мы подчеркиваем тот факт, что множества $V_i$ попарно не пересекаются.

Хроматическое число — это одна из важнейших «экстремальных» характеристик графа. Можно представлять себе, например, такую модельную задачу. В тюрьме сидит тысяча заключенных. Некоторые из них настолько ненавидят друг друга, что их опасно сажать в одну камеру. Спрашивается, каково минимальное число камер, при котором возможна безопасная рассадка заключенных? Понятно, что в данном случае вершины графа — это заключенные, и мы соединяем пару вершин ребром, если соответствующие заключенные несовместимы.

В каждой из книг [13–19] можно найти разделы, посвященные хроматическим числам. Разумеется, нас будут интересовать вероятностные аспекты проблематики.

Прежде чем перейти к формулировкам наиболее ярких результатов, напомним еще пару экстремальных характеристик графа, которые тесно связаны с хроматическим числом. А именно, положим

$$\alpha(G) = \max\bigl\{|W|\colon W \subseteq V,\ \forall x, y \in W\ (x, y) \notin E\bigr\},$$
$$\omega(G) = \max\bigl\{|W|\colon W \subseteq V,\ \forall x, y \in W\ (x, y) \in E\bigr\}.$$

Величина $\alpha(G)$ называется *числом независимости* графа $G$, а величина $\omega(G)$ носит название *кликового числа*. Вообще, любой полный подграф данного графа называется *кликой* в нем, а любой «пустой» подграф (т. е. индуцированный подграф без ребер) называется *независимым множеством*.

Практически очевидно, что $\chi(G) \geqslant \omega(G)$ и $\chi(G) \geqslant \frac{|V|}{\alpha(G)}$. Смысл последнего неравенства в том, что в любой раскраске каждый цветовой класс образует независимое множество, и лучшее, на что мы можем рассчитывать, строя оптимальную раскраску, — это придумать разбиение на цвета, в каждом из которых наибольшее возможное количество вершин. Это неравенство нам будет очень полезно в дальнейшем, и позже мы поймем, насколько оно сильнее первой оценки.

Теперь сформулируем и прокомментируем две теоремы.

**Теорема 17.** *Пусть* $p = \frac{1}{2}$. *Тогда существует такая функция* $\varphi = \varphi(n) = o\left(\frac{n}{\log_2 n}\right)$, *что*

$$P_{n,p}\left(\left|\chi(G) - \frac{n}{2\log_2 n}\right| > \varphi(n)\right) \to 0, \quad n \to \infty.$$

**Теорема 18.** *Пусть* $p = n^{-\alpha}$, *где* $\alpha > \frac{5}{6}$. *Тогда существует такая функция* $u = u(n, \alpha)$, *что*

$$P_{n,p}(u \leqslant \chi(G) \leqslant u + 3) \to 1, \quad n \to \infty.$$

Теорема 17 говорит о том, что хроматическое число «почти всякого» графа на $n$ вершинах асимптотически ведет себя как $\frac{n}{2\log_2 n}$. Это значит, грубо говоря, что при случайном заполнении тюрьмы преступниками камер потребуется лишь в малое (хотя и медленно растущее) число раз меньше, нежели количество самих преступников. Мы употребили выражение «почти всякого», поскольку при $p = \frac{1}{2}$ вероятность графа подчиняется фактически классическому определению вероятности (см. § 1.1), т. е. вероятность — это доля графов, обладающих данным свойством. Мы написали «грубо говоря» не только по причине некоторой искусственности задачи о тюрьме, но еще и потому, что на самом деле «социальные» связи не вполне правильно моделировать в том виде, в каком мы это делаем сейчас. О социальных сетях и об их адекватных моделях речь пойдет в последней главе. Тем не менее смысл остается прежним: с ростом числа вершин хроматическое число случайного графа с «подавляющей» вероятностью оказывается заключенным между $(1-\varepsilon)\frac{n}{2\log_2 n}$ и $(1+\varepsilon)\frac{n}{2\log_2 n}$, где $\varepsilon > 0$ — сколь угодно малое наперед заданное число; хроматическое число *плотно сконцентрировано* около величины $\frac{n}{2\log_2 n}$ (ср. конец параграфа 1.15).

Теорема 18 на порядок более удивительна и неожиданна, нежели теорема 17. Действительно, мы ведь утверждаем, что с вероятностью, близкой к единице, хроматическое число принимает всего четыре из возможных $n$ значений! Вот уж концентрация, так концентрация, — плотнее некуда. Раньше-то функция $\varphi$ вполне могла расти вместе с $n$ (оставаясь, конечно, бесконечно малой в сравнении с $\frac{n}{\log_2 n}$); теперь никакой функции нет и в помине. Да, вероятность ребра случайного графа в теореме 18 быстро убывает, но с ходу это говорит лишь о том, что, по-видимому, у типичного графа хроматическое число не будет большим. А вот предугадать то, что это число вряд ли примет какое-либо значение вне множества $\{u, \ldots, u+4\}$, куда труднее. При этом мы ничего не говорим о поведении самой функции $u$. Но это и не самое интересное. Впрочем, мы и об этом скажем в надлежащий момент.

Теоремы 17 и 18 как бы выдернуты из общего многообразия потенциальных утверждений подобного типа. Мы же ничего не говорим о $p = \mathrm{const} \neq \frac{1}{2}$; ничего не говорим мы и о функциях $p$, отличных от функций вида $n^{-\alpha}$. Однако самая суть науки о хроматическом числе случайного графа улавливается именно тут, а к исчерпывающему изложению теории мы и не стремимся. Более широкий обзор достижений в рассматриваемой области читатель сможет найти в книгах [1, 3, 23].

Заметим, кстати, что условие теоремы 18 несколько избыточно: можно ослабить ограничение $\alpha > \frac{5}{6}$, можно даже $u+3$ заменить на $u+1$. Просто мы докажем относительно несложный и потому максимально проясняющий существо дела вариант утверждения.

Теоремы 17 и 18 принадлежат Б. Боллобашу — одному из крупнейших современных специалистов по случайным графам. В пункте 2.6.2 мы разовьем технику, необходимую для доказательства теорем, в пунктах 2.6.3 и 2.6.4 мы докажем теорему 18, в пунктах 2.6.5—2.6.11 мы опишем общую схему доказательства теоремы 17 (опустив лишь одну черезмерно громоздкую выкладку), в пункте 2.6.12 мы обсудим поведение функции $u$ из формулировки теоремы 18.

### 2.6.2. Мартингалы реберного и вершинного типов

Мы недаром сделали отсылку к параграфу 1.15, говоря о плотной концентрации. Именно техника мартингальных неравенств поможет нам доказать теоремы 17 и 18. Здесь мы введем понятия мартингала *реберного типа* и мартингала *вершинного типа*.

Пусть $\xi$ — произвольная случайная величина на пространстве $G(n, p)$. Свяжем с ней некоторую последовательность случайных величин $Y_0, \ldots, Y_N$, где $N = C_n^2$. Действительно, пусть $e_1, \ldots, e_N$ — ребра полного графа на $n$ вершинах. Для каждого $G \in \Omega_n$ и для всякого $i \in \{0, \ldots, N\}$ положим

$$Y_i(G) = M(\xi | \{H \in \Omega_n \colon \forall j \leqslant i \ e_j \in E(H) \iff e_j \in E(G)\}).$$

Здесь $E(G)$ — множество ребер графа, а условное математическое ожидание берется относительно *одного* события, т. е.

$$M(\xi | A) = \sum_{i=1}^{k} y_i P_{n,p}(\xi = y_i | A).$$

Мы как бы добавляем потенциальные ребра случайного графа одно за другим и следим за тем, как меняется среднее значение $\xi$ при таком добавлении. Читателю предлагается самому убедиться в том, что $Y_0, \ldots, Y_N$ — это мартингал (см. § 1.14), тем более, что интуитивно это понятно. Именно этот мартингал и называется мартингалом реберного типа или мартингалом *проявления ребер* (ср. [23]).

Легко сообразить, что $Y_0 = M\xi$, тогда как $Y_N = \xi$. Последнее обстоятельство особенно важно.

Мартингал вершинного типа или *мартингал проявления вершин* строится аналогично:

$$Y_i(G) = M(\xi | \{H \in \Omega_n \colon \forall x, y \leqslant i \ (x, y) \in E(H) \Leftrightarrow (x, y) \in E(G)\}), \quad i = 1, \ldots, n.$$

Мартингал $Y_1, \ldots, Y_n$ — это своего рода подмартингал в мартингале реберного типа. Как и в случае с проявлением ребер, $Y_1 = M\xi$, $Y_n = \xi$.

Назовем случайную величину $\xi$ *липшицевой по ребрам* (*по вершинам*), если $|\xi(G)-\xi(H)|\leqslant 1$, коль скоро графы $G$ и $H$ отличаются на не более чем одно ребро (одну вершину). Несложное упражнение состоит в том, чтобы обосновать такой факт: если $\xi$ липшицева по ребрам (вершинам), то соответствующий мартингал реберного (вершинного) типа удовлетворяет неравенству Азумы, т. е. $|Y_i-Y_{i-1}|\leqslant 1$ для всех $i$. Вот это и есть основа дальнейшего метода. Берем случайную величину $\xi$, которая липшицева либо по ребрам, либо по вершинам. Сопоставляем ей мартингал подходящего типа. В нем последняя величина и есть $\xi$, причем для нее (т. е., собственно, для $\xi$) выполнено неравенство Азумы. Главное — не забыть заменить обозначение $m$ из параграфа 1.15 на $N$ в случае проявления ребер и на $n-1$ в случае проявления вершин.

Рассмотрим естественный пример: в качестве $\xi$ возьмем $\chi$, т. е. хроматическое число. Случайная величина $\chi$ липшицева и по ребрам, и по вершинам. Стало быть, можно написать неравенства

$$P_{n,p}(|\chi(G)-M\chi|\geqslant a)\leqslant 2e^{-\frac{a^2}{2(n-1)}},$$

$$P_{n,p}(|\chi(G)-M\chi|\geqslant a)\leqslant 2e^{-\frac{a^2}{2N}}.$$

Очевидно, первое неравенство гораздо сильнее. Оно носит название теоремы Шамира — Спенсера.

### 2.6.3. Доказательство теоремы 18: формулировка леммы 1 и вывод из нее

Нам потребуется следующая лемма.

**Лемма 1.** *Пусть $p=n^{-\alpha}$, $\alpha>\frac{5}{6}$. Пусть также $V$ — множество вершин случайного графа в модели $G(n,p)$. Тогда при достаточно больших $n$ имеет место соотношение*

$$P_{n,p}\big(\forall S\subset V,\ |S|\leqslant\sqrt{n}\ln n,\ \textit{выполнено}\ \chi(G|_S)\leqslant 3\big)\geqslant 1-\frac{1}{\ln n}.$$

Лемма говорит о том, что почти наверное хроматическое число каждого не слишком большого индуцированного подграфа случайного графа не превосходит тройки, коль скоро вероятность ребра случайного графа надлежащим образом мала. Это интуитивно понятное, хотя и нетривиальное утверждение. Мы докажем его в пункте 2.6.4. А пока воспользуемся им.

Итак, давайте считать, что $n$ достаточно велико. Насколько именно — будет ясно из контекста. Зафиксируем это $n$, а также $\alpha>\frac{5}{6}$. По таким $n$ и $\alpha$ выберем наименьшее $u = u(n,\alpha)$, для которого

$$P_{n,p}(\chi(G) \leqslant u) > \frac{1}{\ln n}.$$

Такое $u$ корректно определено при $n \geqslant 3$, поскольку

$$P_{n,p}(\chi(G) \leqslant n) = 1$$

и, чем меньше $u \leqslant n$, тем меньше вероятность. Ввиду минимальности $u$ имеем

$$P_{n,p}(\chi(G) \leqslant u-1) \leqslant \frac{1}{\ln n} \ \Rightarrow\ P_{n,p}(\chi(G) \geqslant u) \geqslant 1 - \frac{1}{\ln n}.$$

Иными словами, $u$ — это уже такая функция, что первое неравенство из теоремы 18 для нее выполнено с вероятностью, стремящейся к единице.

Надо бы как-то еще показать, что с большой вероятностью $\chi(G) \leqslant u+3$. Дальнейшее рассуждение напоминает магию: ловкость рук — и никакого обмана.

Рассмотрим случайную величину $Y$, равную наименьшей мощности множества $S \subseteq V$, удаление которого из $V$ порождает граф с хроматическим числом не больше $u$:

$$Y(G) = \min\{|S| : S \subseteq V,\ \chi(G|_{V\setminus S}) \leqslant u\}.$$

Ясно, что $Y$ липшицева по вершинам. Применим технологию из предыдущего пункта:

$$P_{n,p}(Y(G) - MY \geqslant a) \leqslant e^{-\frac{a^2}{2(n-1)}},$$

$$P_{n,p}(Y(G) - MY \leqslant -a) \leqslant e^{-\frac{a^2}{2(n-1)}}.$$

Положим $a = \sqrt{2(n-1)\ln\ln n}$. Тогда

$$P_{n,p}(Y(G) - MY \geqslant a) \leqslant \frac{1}{\ln n},$$

$$P_{n,p}(Y(G) - MY \leqslant -a) \leqslant \frac{1}{\ln n}.$$

Предположим, $MY \geqslant a$. Если так, то вторая оценка влечет цепочку неравенств

$$P_{n,p}(Y(G) \leqslant 0) \leqslant P_{n,p}(Y(G) \leqslant MY - a) \leqslant \frac{1}{\ln n}.$$

Однако

$$P_{n,p}(Y(G) \leqslant 0) = P_{n,p}(\chi(G) \leqslant u) > \frac{1}{\ln n}.$$

Противоречие.

Значит, $MY < a$. Отсюда за счет первого мартингального неравенства получаем

$$P_{n,p}(Y(G) \geqslant 2a) \leqslant P_{n,p}(Y(G) \geqslant MY + a) \leqslant \frac{1}{\ln n},$$

т. е.

$$P_{n,p}(Y(G) \leqslant 2\sqrt{2(n-1)\ln\ln n}) \geqslant 1 - \frac{1}{\ln n}.$$

Обозначим через $A_1$ множество графов из леммы 1, т. е. множество, каждый элемент которого обладает свойством

$$\forall S \subset V,\ |S| \leqslant \sqrt{n}\ln n,\ \text{выполнено}\ \chi(G|_S) \leqslant 3.$$

По лемме 1 имеем $P_{n,p}(A_1) \geqslant 1 - \frac{1}{\ln n}$.

Обозначим через $A_2$ множество графов, для которых $\chi(G) \geqslant u$. Мы знаем, что $P_{n,p}(A_2) \geqslant 1 - \frac{1}{\ln n}$.

Пусть, наконец, $A_3$ — множество графов со свойством

$$Y(G) \leqslant 2\sqrt{2(n-1)\ln\ln n}.$$

Разумеется, $P_{n,p}(A_3) \geqslant 1 - \frac{1}{\ln n}$.

Рассмотрим событие $A_1 \cap A_2 \cap A_3$. Каждый граф $G$ в нем таков, что:

1) $\chi(G) \geqslant u$;
2) $Y(G) \leqslant 2\sqrt{2(n-1)\ln\ln n} < \sqrt{n}\ln n$;
3) $\chi(G|_S) \leqslant 3$, каким бы ни было $S$ с условием $|S| \leqslant \sqrt{n}\ln n$.

Для данного $G \in A_1 \cap A_2 \cap A_3$ берем то самое $S \subseteq V$, для которого $\chi(G|_{V\setminus S}) \leqslant u$. За счет 2) и 3) имеем $\chi(G|_S) \leqslant 3$. Следовательно, $\chi(G) \leqslant u + 3$. В итоге

$$P_{n,p}(u \leqslant \chi(G) \leqslant u+3) \geqslant P(A_1 \cap A_2 \cap A_3) \geqslant 1 - \frac{3}{\ln n} \to 1, \quad n \to \infty.$$

Теорема доказана.

### 2.6.4. Доказательство теоремы 18: доказательство леммы 1

Утверждение леммы равносильно неравенству

$$P_{n,p}\big(\exists S \subset V,\ |S| \leqslant \sqrt{n}\ln n,\ \text{выполнено}\ \chi(G|_S) > 3\big) \leqslant \frac{1}{\ln n},$$

коль скоро $n$ достаточно велико. Будем оценивать указанную вероятность сверху, предполагая, что $n$ большое.

Прежде всего заметим, что событие

$$\big\{\exists S \subset V,\ |S| \leqslant \sqrt{n}\ln n,\ \text{выполнено}\ \chi(G|_S) > 3\big\}$$

совпадает с событием

$$\{\exists S \subset V,\ 4 \leqslant |S| \leqslant \sqrt{n}\ln n,\ \text{выполнено}\ \chi(G|_S) > 3\},$$

поскольку у любого графа на не более трех вершинах хроматическое число не превосходит тройки. Более того, можно считать, что $S$ минимально по вложению, т. е., что $S$ обладает описанным свойством, но для любого $x \in S$ разность $S \setminus \{x\}$ этим свойством не обладает. Перепишем, стало быть, интересующее нас событие так:

$$A = \{\exists T \subset V,\ 4 \leqslant |T| \leqslant \sqrt{n}\ln n,$$
$$T\ \text{минимально, и выполнено}\ \chi(G|_T) > 3\}.$$

Ясно, что если ввести обозначение $\tau = [\sqrt{n}\ln n]$, то

$$A = \bigcup_{t=4}^{\tau} \bigcup_{\substack{T\subset V,\ |T|=t,\\ T\ \text{минимально}}} A_{t,T},$$

где

$$A_{t,T} = \{\text{для данного минимального}\ T,\ |T| = t,\ \text{выполнено}\ \chi(G|_T) > 3\}.$$

Если граф $G$ обладает свойством $A_{t,T}$ с конкретными $t$ и $T$, то

$$|E(G|_T)| \geqslant \frac{3t}{2}.$$

Действительно, возьмем любую вершину $x \in T$. Мы знаем, что

$$\chi(G|_{T\setminus\{x\}}) \leqslant 3,$$

хотя и верно, что $\chi(G|_T) \geqslant 4$. Следовательно, степень вершины $x$ в графе $G|_T$ (т. е. число ребер из $E(G|_T)$, которые имеют $x$ одним из своих концов) не меньше трех. Отсюда и вытекает неравенство $|E(G|_T)| \geqslant \frac{3t}{2}$.

В итоге мы получаем импликацию $A_{t,T} \Rightarrow B_{t,T}$, где

$$B_{t,T} =$$
$$= \left\{\text{для данного минимального}\ T,\ |T| = t,\ \text{выполнено}\ |E(G|_T)| \geqslant \frac{3t}{2}\right\}.$$

Итак,

$$P_{n,p}(A) \leqslant \sum_{t=4}^{\tau} \sum_{\substack{T\subset V,\ |T|=t,\\ T\ \text{минимально}}} P_{n,p}(A_{t,T}) \leqslant \sum_{t=4}^{\tau} \sum_{\substack{T\subset V,\ |T|=t,\\ T\ \text{минимально}}} P_{n,p}(B_{t,T}) \leqslant$$
$$\leqslant \sum_{t=4}^{\tau} \sum_{\substack{T\subset V,\ |T|=t,\\ T\ \text{минимально}}} C_r^s p^s,$$

где $r = C_t^2$, $s = \left\lceil \frac{3t}{2} \right\rceil$ (обрамление $\lceil \cdot \rceil$ обозначает верхнюю целую часть, т. е. минимальное целое число, не меньшее данного). Последняя двойная сумма не превышает величины

$$\sum_{t=4}^{\tau} C_n^t C_r^s p^s.$$

Применим неравенство $C_a^b \leqslant \left(\frac{ae}{b}\right)^b$, где $e = 2{,}71828\ldots$ — основание натурального логарифма. Оно доказывается по индукции, и мы на этом не останавливаемся. Имеем

$$\sum_{t=4}^{\tau} C_n^t C_r^s p^s \leqslant \sum_{t=4}^{\tau} \left(\frac{ne}{t}\right)^t \cdot \left(\frac{re}{s}\right)^s \cdot p^s \leqslant \sum_{t=4}^{\tau} \left(\frac{ne}{t}\right)^t \cdot \left(\frac{t^2 e}{2 \cdot 3t/2}\right)^{3t/2+1} \cdot n^{-\alpha \cdot 3t/2} =$$
$$= \sum_{t=4}^{\tau} \left(\frac{ne}{t} \cdot \left(\frac{te}{3}\right)^{3/2} \cdot n^{-3\alpha/2}\right)^t \cdot \frac{te}{3} \leqslant \sum_{t=4}^{\tau} t\left(3 \cdot n \cdot \sqrt{t} \cdot n^{-3\alpha/2}\right)^t \leqslant$$
$$\leqslant \sum_{t=4}^{\tau} t\left(3 \cdot n \cdot \sqrt{\tau} \cdot n^{-3\alpha/2}\right)^t \leqslant \sum_{t=4}^{\tau} t\left(3 \cdot n \cdot \sqrt{\sqrt{n}\ln n} \cdot n^{-3\alpha/2}\right)^t \leqslant$$
$$\leqslant \sum_{t=4}^{\tau} t\left(3n^{5/4-3\alpha/2}\sqrt{\ln n}\right)^t.$$

Вспоминаем, что $\alpha > \frac{5}{6}$. Значит, $\frac{5}{4} - \frac{3\alpha}{2} = -\beta < 0$. Иными словами,

$$\sum_{t=4}^{\tau} t\left(3n^{5/4-3\alpha/2}\sqrt{\ln n}\right)^t = \sum_{t=4}^{\tau} t\left(3n^{-\beta}\sqrt{\ln n}\right)^t =$$
$$= \sum_{t=4}^{[\ln n]} t\left(3n^{-\beta}\sqrt{\ln n}\right)^t + \sum_{t=[\ln n]+1}^{\tau} t\left(3n^{-\beta}\sqrt{\ln n}\right)^t \leqslant$$
$$\leqslant [\ln n] \cdot [\ln n] \cdot \left(3n^{-\beta}\sqrt{\ln n}\right)^4 + n \cdot n \cdot \left(3n^{-\beta}\sqrt{\ln n}\right)^{[\ln n]} < \frac{1}{\ln n}.$$

Последнее неравенство при больших $n$ выполнено с колоссальным запасом, и лемма доказана.

### 2.6.5. Нижняя оценка в теореме 17

Для доказательства теоремы 17 достаточно обосновать два следующих утверждения.

1) Существует такая функция $\varphi_1 = \varphi_1(n) = o\left(\frac{n}{\log_2 n}\right)$, что

$$P_{n,1/2}\left(\chi(G) \geqslant \frac{n}{2\log_2 n} - \varphi_1(n)\right) \to 1, \quad n \to \infty;$$

2) Существует такая функция $\varphi_2 = \varphi_2(n) = o\left(\frac{n}{\log_2 n}\right)$, что

$$P_{n,1/2}\left(\chi(G) \leqslant \frac{n}{2\log_2 n} + \varphi_2(n)\right) \to 1, \quad n \to \infty.$$

В этом пункте мы докажем утверждение 1. С этой целью мы установим асимптотику

$$P_{n,1/2}(\alpha(G) \leqslant [2\log_2 n] - 1) \to 1, \quad n \to \infty.$$

Почему? Да потому, что, как мы знаем (см. п. 2.6.1), $\chi(G) \geqslant \frac{|V|}{\alpha(G)}$, и, стало быть, обещанная асимптотика мгновенно влечет нужный результат.

Положим $k = [2\log_2 n]$ и рассмотрим случайную величину $X_k$, которая для данного графа $G$ равна числу независимых множеств мощности $k$ в нем. Тогда

$$P_{n,1/2}(\alpha(G) \leqslant [2\log_2 n] - 1) = P_{n,1/2}(X_k = 0).$$

Ввиду неравенства Маркова последняя вероятность не меньше, чем $1 - MX_k$, так что для завершения рассуждения остается лишь убедиться в том, что $MX_k \to 0$.

Воспользуемся линейностью математического ожидания:

$$MX_k = C_n^k \left(\frac{1}{2}\right)^{C_k^2}.$$

Имеем

$$C_n^k \left(\frac{1}{2}\right)^{C_k^2} \leqslant \frac{n^k}{k!} \cdot 2^{-\frac{k^2}{2} + \frac{k}{2}} \leqslant \frac{n^{2\log_2 n}}{k!} \cdot 2^{-(2\log_2 n - 1)^2/2 + \log_2 n} =$$

$$= \frac{1}{k!} \cdot 2^{2\log_2^2 n - 2\log_2^2 n + 3\log_2 n - 1/2} = \frac{1}{k!} \cdot 2^{3\log_2 n - 1/2} \leqslant \left(\frac{e}{k}\right)^k \cdot 2^{3\log_2 n},$$

поскольку $k! \geqslant \left(\frac{k}{e}\right)^k$ (доказывается по индукции). Продолжаем выкладку:

$$MX_k \leqslant 2^{3\log_2 n + k\log_2 e - k\log_2 k} \leqslant 2^{3\log_2 n + 2(\log_2 e)\log_2 n - [2\log_2 n]\log_2 [2\log_2 n]}.$$

Видно, что вычитаемое в показателе последней экспоненты растет значительно быстрее уменьшаемого. Следовательно, $MX_k \to 0$, и первое утверждение доказано.

### 2.6.6. Верхняя оценка в теореме 17: план действий

Ниже мы «почти» докажем второе утверждение. Не хватит лишь некоторых чисто технических деталей, связанных с чересчур громоздкими вычислениями.

Сперва в самых общих чертах поймем, чего нам будет достаточно для доказательства. Потом все это аккуратно конкретизируем.

Заметим, что для специалиста в области случайных графов нижняя оценка из предыдущего пункта — тривиальное упражнение. То, о чем мы будем говорить дальше, на несколько порядков сложнее. Именно в том и состояла заслуга Боллобаша, что он придумал всю кухню, о которой мы здесь расскажем.

### 2.6.7. Верхняя оценка в теореме 17: идея доказательства

Предположим, что существуют целочисленные функции $k=k(n)\sim\sim 2\log_2 n$ и $m=m(n)=o\left(\frac{n}{\log_2 n}\right)$, с которыми $P_{n,1/2}(A)\to 1$, $n\to\infty$, коль скоро

$$A=\{\forall S\subset V,\ |S|=m,\ \text{выполнено } \alpha(G|_S)\geqslant k\}.$$

Тогда возьмем любой граф $G$ из $A$. За счет свойства $A$ в нем есть независимое множество $K_1$ размера $k$ (ведь $n>m$). Покрасим все вершины из $K_1$ в цвет 1. Удалим $K_1$ из $V$. Если все еще $n-k\geqslant m$, то снова за счет свойства $A$ найдем независимое множество $K_2\subseteq V\setminus K_1$, элементы которого покрасим в цвет 2. Действуем так до тех пор, пока после удаления очередного независимого множества $K_l$ не останется меньше $m$ вершин. Очевидно,

$$l\leqslant\left\lceil\frac{n-m}{k}\right\rceil\sim\frac{n}{2\log_2 n}.$$

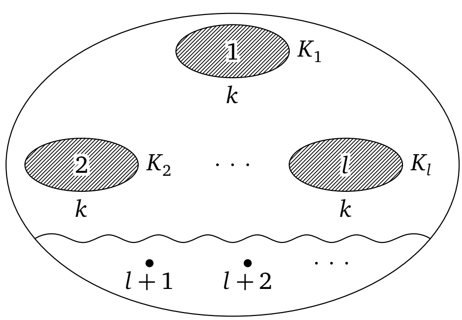



Рис. 4

По окончании процедуры каждую из недокрашенных вершин покрасим в отдельный цвет с номером $l+1, l+2, \ldots$ Всего получится не больше, нежели
$$l+m \sim \frac{n}{2\log_2 n}$$
цветов (см. рис. 4). Это значит, что
$$P_{n,1/2}(\chi(G) \leqslant l+m) \geqslant P_{n,1/2}(A) \to 1,$$
и верхняя оценка в теореме 17 доказана.

### 2.6.8. Верхняя оценка в теореме 17: выбор параметров и оценка вероятности

Во-первых, положим $m = \left[\frac{n}{\log_2^2 n}\right]$. Далее, можно показать, что существует $k \sim 2\log_2 m \sim 2\log_2 n$, с которым
$$C_m^k 2^{-C_k^2} > m^4.$$
Соответствующая выкладка крайне похожа на выкладку из пункта 2.6.5, и мы ее не проводим, предоставляя ее добросовестному читателю в качестве упражнения.

Хочется теперь убедиться в том, что с выбранными $m$ и $k$ выполнено соотношение $P_{n,1/2}(A) \to 1$. Мы проверим, что, напротив, $P_{n,1/2}(\overline{A}) \to 0$. Запишем
$$\overline{A} = \{\exists S \subset V,\ |S| = m,\ \text{выполнено } \alpha(G|_S) < k\} = \bigcup_{S\subset V,\ |S|=m} A_S,$$
где
$$A_S = \{\text{для данного множества } S,\ |S| = m,\ \text{выполнено } \alpha(G|_S) < k\}.$$

Очевидно, величина $P_{n,1/2}(A_S)$ не зависит от $S$. Таким образом,
$$P_{n,1/2}(\overline{A}) \leqslant \sum_{S\subset V, |S|=m} P_{n,1/2}(A_S) = C_n^m P_{m,1/2}(\alpha(G) < k).$$

Пусть $Y_k$ — это случайная величина на $G(m, 1/2)$, равная максимальному количеству независимых множеств мощности $k$ в случайном графе, любые два из которых имеют не более одной общей вершины:
$$Y_k(G) = \max\{|\mathcal{K}| : \mathcal{K} = \{K_1, \ldots\}, K_i \text{ — независимое множество в } G \text{ и } |K_i \cap K_j| \leqslant 1\}.$$

Тогда

$$P_{m,1/2}(\alpha(G) < k) = P_{m,1/2}(Y_k = 0).$$

Заметим, что $Y_k$ липшицева по ребрам (см. п. 2.6.2), чего нельзя сказать, например, про $X_k$ (см. п. 2.6.5). Именно последнее обстоятельство поспособствовало тому, что событие $\{\alpha(G) < k\}$ мы заменили «странным» событием $\{Y_k = 0\}$, а не вполне, казалось бы, естественным $\{X_k = 0\}$. Позже мы вернемся к этим тонкостям и прокомментируем их. А пока продолжим доказательство.

Раз $Y_k$ липшицева, применим неравенство Азумы. Имеем

$$P_{m,1/2}(Y_k = 0) = P_{m,1/2}(Y_k \leqslant 0) =$$
$$= P_{m,1/2}(-Y_k \geqslant 0) = P_{m,1/2}(MY_k - Y_k \geqslant MY_k) \leqslant e^{-\frac{(MY_k)^2}{2C_m^2}}.$$

Справедлива

**Лемма 2.** *Выполнено неравенство*

$$MY_k \geqslant \frac{m^2}{2k^4}\psi(m),$$

*где* $\psi(m) \to 1$.

Лемму мы докажем в следующем пункте, а пока применим ее:

$$C_n^m P_{m,1/2}(\alpha(G) < k) < 2^n P_{m,1/2}(Y_k = 0) \leqslant 2^n e^{-\frac{(MY_k)^2}{2C_m^2}} \leqslant 2^n e^{-\frac{m^4\psi^2}{4k^8m^2}} = 2^n e^{-\frac{m^2\psi^2}{4k^8}}.$$

Величина, стоящая после знака минус в показателе последней экспоненты, растет гораздо быстрее, нежели $n$. Значит, все произведение стремится к нулю, и «по модулю леммы 2» теорема 17 доказана.

Читатель может спросить: а где же мы использовали столь специфический выбор параметров, который мы осуществили в начале этого пункта? Ответ очевиден: он нужен для доказательства леммы 2, и как раз к нему мы теперь переходим.

### 2.6.9. Верхняя оценка в теореме 17: доказательство леммы 2

Нам нужно доказать, что в среднем случайный граф содержит достаточно много «почти непересекающихся» независимых множеств размера $k$. Представим себе на минуту, что граф не случаен, а фиксирован. Как бы в нем найти побольше нужных нам объектов? А что, если и их выбирать случайно? Как ни странно, это именно та идея, которая работает.

Пусть для данного графа $G \in \Omega_m$ совокупность $\mathscr{K} = \{K_1, \ldots, K_{x(G)}\}$ состоит из всех его независимых множеств мощности $k$. Если $G$

считать случайным, то и величина $x(G)$ случайная; собственно, это и есть $X_k$, просто мы ввели более краткое обозначение.

Для данного $G$ проведем $x(G)$ испытаний Бернулли с вероятностью успеха $p'$, которую мы оптимально выберем позднее. Если в испытании с номером $i$ произошел успех, то «вытаскиваем» из $\mathscr{K}$ множество $K_i$, иначе — пропускаем. На выходе возникает случайная совокупность $\mathscr{C} = \mathscr{C}(G)$ независимых $k$-подмножеств в $G$ (состоящая из вытащенных множеств).

Теперь вспомним, что сам граф $G$ случаен. Строго говоря, появившуюся «двойную случайность» — графа и совокупности подмножеств в нем — следует описывать путем введения нового вероятностного пространства. В нем элементарные события суть пары $(G, \mathscr{C}(G))$, а вероятность каждой пары есть произведение вероятности графа на вероятность совокупности, т. е.

$$P_{m,1/2,p'}((G, \mathscr{C}(G)) = P_{m,1/2}(G)(p')^{|\mathscr{C}(G)|}(1-p')^{x(G)-|\mathscr{C}(G)|}.$$

Нетрудно видеть, что

$$M|\mathscr{C}| = p'M|\mathscr{K}| = p'Mx = p'MX_k = p'C_m^k 2^{-C_k^2}.$$

Здесь важно понимать, что первое математическое ожидание в цепочке определено на новом вероятностном пространстве, тогда как все остальные — на старом, т. е. на $G(m, 1/2)$.

Рассмотрим еще два случайных множества. Одно из них будет определено на пространстве $G(m, 1/2)$, другое — на новом пространстве. Итак,

$$W = W(G) = \{(K_i, K_j) : K_i, K_j \in \mathscr{K},\ 2 \leqslant |K_i \cap K_j| \leqslant k-1\},$$
$$W' = W'((G, \mathscr{C}(G))) = \{(K_i, K_j) : K_i, K_j \in \mathscr{C},\ 2 \leqslant |K_i \cap K_j| \leqslant k-1\}.$$

Иными словами, в обоих случаях речь идет о совокупности (неупорядоченных) пар (различных) $k$-вершинных независимых множеств, которые цепляют друг друга слишком сильно — не так, как это происходит при определении величины $Y_k$.

Ясно, что $M|W'| = (p')^2 M|W|$. Однако поиск величины $M|W|$ — куда более противное занятие. Нет, в принципе и тут все понятно. Действительно,

$$|W| = \sum_{r=2}^{k-1} |W_r|,$$

где

$$W_r = \big\{(K_i, K_j) : K_i, K_j \in \mathscr{K},\ |K_i \cap K_j| = r\big\}.$$

За счет линейности математического ожидания (см. рис. 5) имеем

$$M|W| = \sum_{r=2}^{k-1} M|W_r| = \frac{1}{2}\sum_{r=2}^{k-1} C_m^k C_{m-k}^{k-r} C_k^r 2^{-2C_k^2+C_r^2}.$$

Вот это-то и ужасно. Пойди запиши такую сумму в сколь-нибудь удобоваримом виде! По счастью, это все-таки возможно. Мы не станем мучить читателя малоинтересными выкладками и лишь приведем ответ. Оказывается,

$$M|W| \sim \frac{(MX_k)^2 k^4}{2m^2}.$$

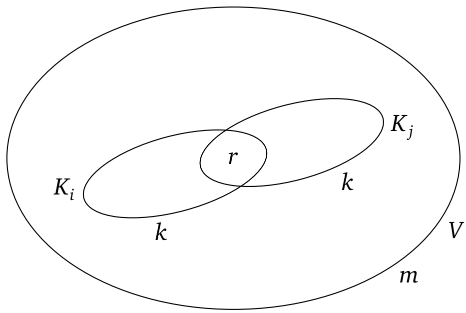



Рис. 5

Пришло время применить ловкость рук. Для данного $G$ из отвечающей ему совокупности $\mathscr{C}(G)$ удалим по одному элементу из каждой пары в множестве $W'$. Образуется совокупность $\mathscr{C}^* = \mathscr{C}^*(G)$, в которой *нет пар множеств, пересекающихся по двум и более элементам*. Значит,

$$MY_k \geqslant M|\mathscr{C}^*| \geqslant M|\mathscr{C}| - M|W'| = p'MX_k - \frac{(p')^2(MX_k)^2k^4\psi_1(m)}{2m^2},$$

где $\psi_1(m) \to 1$. Положим

$$p' = \frac{m^2}{(MX_k)k^4\psi_1(m)}.$$

Это, в самом деле, вероятность, так как у нас при больших $m$

$$MX_k = C_m^k 2^{-C_k^2} > m^4 \ \Rightarrow\ 0 < p' < \frac{1}{m^2k^4\psi_1(m)} < 1.$$

(Вот оно, где ружье-то выстрелило!) Получаем

$$MY_k \geqslant \frac{m^2}{k^4\psi_1(m)} - \frac{m^4k^4\psi_1(m)}{2m^2k^8\psi_1^2(m)} = \frac{m^2}{2k^4}\psi(m),$$

где $\psi = \dfrac{1}{\psi_1}$. Лемма доказана.

### 2.6.10. Комментарий к лемме 2

Доказательство теоремы 17 проведено практически полностью, дадим теперь различные комментарии к нему. Начнем с комментария к лемме 2.

Дело в том, что более или менее ясно, откуда появилось ее утверждение. А именно, понятно, что для любого графа $G$ на $m$ вершинах

$$Y_k(G) \leqslant c\frac{m^2}{k^2}, \quad c > 0.$$

Это связано с тем, что в каждом независимом множестве, которое фигурирует в определении величины $Y_k$, ровно $C_k^2$ пар элементов (отсутствующих ребер), причем любые два таких множества имеют непересекающиеся совокупности указанных пар (иначе бы они сами пересекались хотя бы по двум элементам). Всего же пар — $C_m^2$. Таким образом, $Y_k(G) \leqslant C_m^2/C_k^2$, откуда и вытекает обещанное неравенство.

Разумеется, приведенное рассуждение влечет оценку

$$MY_k \leqslant c\frac{m^2}{k^2}, \quad c > 0,$$

и, стало быть, лемма 2 говорит о том, что при известном соотношении между параметрами $m$ и $k$ эта оценка почти точна. Зазор лишь в $k^2 \sim 4\log_2^2 m$. С одной стороны, зазор мизерный, но, с другой стороны, это все-таки зазор. Так вот проблема в том, что никто не умеет его устранять. Цепочка неравенств

$$\frac{m^2}{2k^4}\psi(m) \leqslant MY_k \leqslant c\frac{m^2}{k^2}, \quad c > 0,$$

стоит незыблемо, и ни нижняя, ни верхняя оценка улучшению не поддаются. Хорошая задачка? Гипотеза такова, что верхняя оценка точна.

Заметим, что лемму 2 мы использовали для получения оценки

$$P_{m,1/2}(Y_k = 0) \leqslant e^{-\frac{m^2\psi^2}{4k^8}},$$

и если бы удалось доказать гипотезу, то возникла бы чуть лучшая оценка

$$P_{m,1/2}(Y_k = 0) \leqslant e^{-\frac{m^2}{ck^4}}.$$

Эта оценка уже установлена за счет других мощных инструментов теории вероятностей (неравенства Талаграна и Янсона, см. [3, 23, 24]); однако вопрос о математическом ожидании, как мы уже говорили, по-прежнему открыт.

### 2.6.11. Чем $Y_k$ лучше $X_k$, или почему не работает неравенство Чебышёва?

Здесь нам хочется объяснить читателю, почему вместо простой и естественной случайной величины $X_k$, которую мы к тому же с успехом использовали в пункте 2.6.5, нам приходится применять несколько замудренную, на первый взгляд, величину $Y_k$.

Случилось это все (возникла величина $Y_k$) в пункте 2.6.8. Нам нужно было доказать, что

$$C_n^m P_{m,1/2}(\alpha(G) < k) \to 0.$$

Мы переписали левую часть неравенства в виде $C_n^m P_{m,1/2}(Y_k = 0)$, хотя могли бы написать $C_n^m P_{m,1/2}(X_k = 0)$. Что ж, попробуем так и сделать. Как мы заметили еще в п. 2.6.5, $X_k$ не является липшицевой ни в каком из смыслов. Значит, заведомо неравенство Азумы не применимо. Тогда нам ничего не остается, как прибегнуть к помощи Чебышёва (других средств мы не знаем, не знали и те, кто работал в свое время над задачей). Вот выкладка (ср. п. 2.4.3 и п. 2.6.5):

$$P_{m,1/2}(X_k = 0) = P_{m,1/2}(X_k \leqslant 0) = P_{m,1/2}(MX_k - X_k \geqslant MX_k) \leqslant \frac{DX_k}{(MX_k)^2}.$$

Давайте поймем, как с ростом $m$ ведут себя математическое ожидание и дисперсия. Для пущей наглядности продолжим работать с теми $m$ и $k$, которые введены в начале пункта 2.6.8. Ну, $MX_k > m^4$, и это, пожалуй, все. А что с дисперсией? По сути, нас интересует второй момент $MX_k^2$ или даже его факториальный аналог:

$$M_f^2 X_k = \sum_{i \neq j} M(X_{k,i} X_{k,j}), \quad i, j \in \{1, \ldots, C_m^k\},$$

где $X_{k,i}$ — это индикатор свойства «$i$-е по счету $k$-элементное подмножество множества вершин графа независимо». Фактически мы считаем здесь среднее число (упорядоченных) пар различных независимых множеств вершин случайного графа. Ничего не напоминает? Да ведь это почти дословно $M|W|$ из пункта 2.6.9! Получается, что отношение $\frac{DX_k}{(MX_k)^2}$ не превосходит функции, обратной некоторому полиному от $m$.

В то же время

$$C_n^m = \frac{n(n-1)\ldots(n-m+1)}{m!} = \\ = \frac{n^m}{m!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{m-1}{n}\right) \geqslant \frac{n^m}{e(m/2)^m}\left(1 - \frac{m}{n}\right)^m$$

(неравенство $m! \leqslant e(m/2)^m$ легко доказать по индукции). Нетрудно убедиться в том, что в последнем выражении доминирующая по скорости роста функция — это

$$\left(\frac{n}{m}\right)^m \geqslant (\log_2^2 n)^m,$$

и растет она куда быстрее любого полинома. Значит, при нынешней технологии нам не удастся получить асимптотику

$$C_n^m P_{m,1/2}(\alpha(G) < k) \to 0,$$

и этим объясняется наше нежелание использовать $X_k$.

Конечно, наши рассуждения служат лишь намеком на суть проблемы: в конце концов мы можем менять $m$ и $k$. Однако это не помогает, и только мартингальная техника спасает ситуацию.

Подытожить следует так. Оказывается, существуют $k_1 \sim 2\log_2 n$ и $k_2 \sim 2\log_2 n$, которые, как видно, очень близки и с которыми, тем не менее, выполнены два совершенно противоположных результата:

1) в случайном графе в модели $G(n, 1/2)$ почти наверное нет независимых множеств мощности $k_1$;
2) в случайном графе в модели $G(n, 1/2)$ почти наверное каждое множество вершин размера $m$ содержит независимое подмножество мощности $k_2$.

Если бы мы хотели доказать, что почти наверное в случайном графе есть независимое подмножество мощности $k_2$, то нам бы хватило неравенства Чебышёва. Нам же потребовалось второе утверждение, и пришлось обратиться к неравенству Азумы и соответствующим изыскам.

### 2.6.12. О функции *u* в теореме 18

Теорема 18 — самая загадочная в серии теорем о хроматическом числе. Комментируя ее, мы заметили, что интуиция подсказывает малость функции $u$ из ее формулировки. Сейчас мы покажем, что во всяком случае функция $u$ может стремиться к бесконечности.

Пусть, например, $p = n^{-0,9}$. Покажем, что

$$P_{n,p}(\chi(G) \geqslant \ln n) \to 1.$$

Этого заведомо хватит. В свою очередь, для этого достаточно проверить, что (ср. п. 2.6.5)

$$P_{n,p}\left(\alpha(G) \leqslant \frac{n}{\ln n}\right) \to 1$$

или что
$$MX_k \to 0, \quad k = \left[\frac{n}{\ln n}\right].$$

Итак, при больших $n$ имеем
$$MX_k = C_n^k (1-p)^{C_k^2} \leqslant \left(\frac{ne}{k}\right)^k e^{-pC_k^2} \leqslant (3\ln n)^k e^{-\frac{pk^2}{4}} \leqslant$$
$$\leqslant e^{k\ln\ln n + k\ln 3 - n^{1,1}/16\ln^2 n} \to 0$$
с колоссальным запасом.

## 2.7. О числе независимости и кликовом числе случайного графа

Из пункта 2.6.5 мы знаем, что
$$P_{n,1/2}(\alpha(G) \leqslant [2\log_2 n] - 1) \to 1, \quad n \to \infty.$$
Говоря словами, число независимости случайного графа в модели $G(n, 1/2)$ почти наверное логарифмически мало. Практически очевидно, что в той же модели в точности такой же результат верен и относительно величины $\omega(G)$ — кликового числа графа. Значит, для почти всякого графа неравенство $\chi(G) \geqslant \omega(G)$ несопоставимо слабее неравенства $\chi(G) \geqslant \dfrac{n}{\alpha(G)}$. Это — по поводу комментария, который мы дали перед теоремой 17 в пункте 2.6.1.

И все-таки бывают графы, у которых $\chi(G) = \omega(G)$ (хоть их и крайне мало). Простейший пример такого графа — полный граф. Заметим, что полный граф обладает даже более сильным свойством: для любого его индуцированного подграфа $H$ выполнено $\chi(H) = \omega(H)$. Конечно, для полного графа указанный факт тривиален. Гораздо менее тривиален он для произвольного графа. В самом деле, как понять, обладает данный граф описанным свойством или нет?

Графы, у которых кликовое число каждого индуцированного подграфа совпадает с его хроматическим числом, называются *совершенными*. Несмотря на свою сравнительную редкость (с известной нам вероятностной точки зрения), совершенные графы «собираются» в целые классы. Например, любой *двудольный* граф (т. е. граф с хроматическим числом 2) является совершенным. Или, скажем, всякий *интервальный* граф обладает этим свойством (интервальным называется граф, у которого вершины суть отрезки на прямой, а ребра — пары пересекающихся отрезков).

Априори кажется очень трудным придумать какой-либо критерий «совершенства» графа. Правда: пойди пойми, какой граф совершенный, а какой нет. К. Берж, который ввел сам термин «совершенный граф», высказал в 1963 году сильную гипотезу: граф совершенный тогда и только тогда, когда он не содержит индуцированных циклов нечетной длины $\geqslant 5$ и их дополнений (до полного подграфа). На рис. 6 приведен пример графа с индуцированным циклом длины 7 и графа с индуцированным подграфом, который служит дополнением до цикла длины 7.

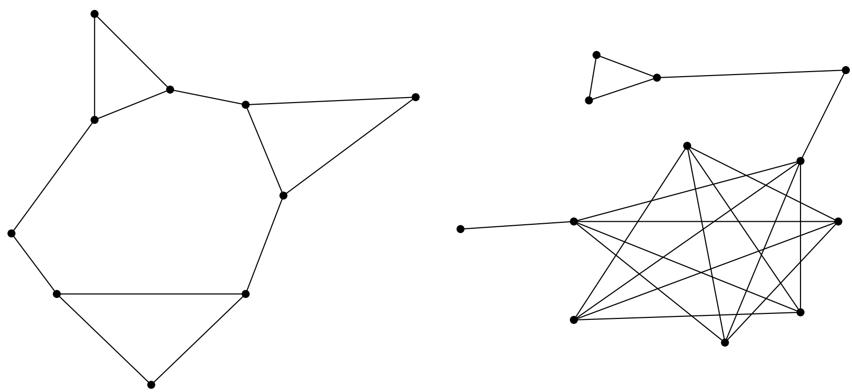


Рис. 6

По-видимому, ясно, откуда возникли циклы и «антициклы» (дополнения до циклов), причем ясно даже, зачем нужно условие «длина цикла больше либо равна пяти». Действительно, у цикла *нечетной длины* $\geqslant 5$ хроматическое число равно трем, а кликовое — двум. Поэтому такой цикл заведомо совершенству вредит. То же и с антициклами. Но вот почему отсутствия указанных вредных объектов *достаточно* для совершенства, отнюдь не понятно.

Тем не менее гипотеза Бержа верна. Ее доказали только в 2002 году (а опубликовали в 2006 году). Статья (см. [25]) занимает 179 страниц! Да вот, и такое бывает.

## 2.8. Числа Рамсея

Вероятно, читатель еще не догадывается, к чему идет дело. Да и немудрено: человеку, привыкшему связывать понятие числа Рамсея с утверждением о том, что среди любых шести людей либо трое знакомы друг с другом, либо трое друг друга не знают, невдомек, что предыдущий параграф был по сути о них — о числах Рамсея. Такие вот удивительные ходы бывают в математике.

Давайте вспомним стандартное определение числа Рамсея. Итак, пусть $s, t \in \mathbb{N}$. Назовем *числом Рамсея* величину $R(s, t)$, равную минимальному натуральному $n$, такому что при любой раскраске ребер графа $K_n$ в красный и синий цвета либо найдется $K_s \subseteq K_n$, у которого все ребра красные, либо найдется $K_t \subseteq K_n$, у которого все ребра синие.

Понятно, как числа Рамсея связаны с упомянутым выше утверждением о знакомствах. Это утверждение о том, что $R(3, 3) \leqslant 6$: мы просто соединяем пару людей красным ребром, если эти люди знакомы, и синим ребром, если они друг друга не знают; какова бы ни была раскраска, либо найдется красный, либо отыщется синий треугольник. Нетрудно, кстати, показать, что $R(3, 3) = 6$.

Чуть менее понятно, как связаны числа Рамсея с темой предыдущего параграфа. Сейчас мы дадим эквивалентное определение величины $R(s, t)$, и все станет ясно. Действительно, можно сказать так: $R(s, t)$ — это наименьшее натуральное $n$, при котором для любого графа $G$ на $n$ вершинах либо $\omega(G) \geqslant s$, либо $\alpha(G) \geqslant t$. Мы как бы красим ребра графа $G$ в красный цвет, а ребра, которых в $G$ нет, — в синий цвет. Тогда наличие большой красной клики ($K_s \subseteq K_n$) как раз отвечает неравенству $\omega(G) \geqslant s$, а наличие большой синей клики ($K_t \subseteq K_n$) в аккурат соответствует оценке $\alpha(G) \geqslant t$.

Прежде чем вывести из результатов предыдущего параграфа важное свойство чисел Рамсея, расскажем о них немного. Впервые они появились в статье [26] в 1930 году. Там они возникли в связи с одной задачей математической логики. В 1935 году вышла в свет одна из наиболее значимых статей по комбинаторике и комбинаторной геометрии XX века (см. [27]). Ее авторы, П. Эрдёш и Д. Секереш, переоткрыли числа Рамсея и доказали про них существенно более точные утверждения, нежели те, которые установил сам Ф. П. Рамсей в [26]. За прошедшие с тех пор 75 лет наука разрослась до такой степени, что теперь мы говорим о теории Рамсея, и в этой теории лишь небольшой (хотя и весьма почетный) уголок отведен задачам о величинах $R(s, t)$ и их обобщениях.

Имеется обширная литература по числам Рамсея. Мы процитируем лишь несколько значимых источников: [1, 12, 23, 28–31]. Ниже мы еще поговорим об этих замечательных числах.

Крайне мало известно о точных значениях для $R(s, t)$. Легко понять, что $R(s, t) = R(t, s)$, $R(1, t) = 1$ и $R(2, t) = t$. И это почти все. Список остальных найденных величин можно без труда взять, например, на странице [32], и мы его здесь не приводим.

Зато известно, что всегда $R(s,t)<\infty$. Более того, Эрдёш и Секереш (см. [12, 27]) установили рекуррентное неравенство

$$R(s,t) \leqslant R(s-1,t)+R(s,t-1),$$

которое мгновенно влечет оценку $R(s,t)\leqslant C_{s+t-2}^{t-1}$. Давайте обсудим эту оценку в случае $s=t$ (в таком случае $R(s,s)$ называется *диагональным числом Рамсея*).

Воспользуемся формулой Стирлинга для факториала (см., например, [33]):

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Имеем

$$C_{2s-2}^{s-1}=\frac{(2s-2)!}{((s-1)!)^2}\sim\frac{\sqrt{2\pi(2s-2)}(2s-2)^{2s-2}e^{2-2s}}{(2\pi(s-1))(s-1)^{2s-2}e^{2-2s}}=\frac{4^{s-1}}{\sqrt{\pi(s-1)}}.$$

Таким образом, диагональное число Рамсея растет не быстрее $4^s$. Фантастика, но, несмотря на колоссальные усилия, которые были брошены на уточнение указанной оценки, и на многие годы, прошедшие с момента ее появления в печати, воз и ныне практически там. Все, что удалось сделать, — увеличить «знаменатель». Сейчас наилучшая оценка для $R(s,s)$ выглядит так (см. [34]):

$$R(s,s) \leqslant \frac{4^s}{e^{\gamma \ln^2 s/\ln\ln s}}.$$

Здесь $\gamma$ — некоторая положительная константа, и смысл в том, что, какова бы она ни была,

$$e^{\gamma \ln^2 s/\ln\ln s}=o(4^s),$$

т. е. мы вольны написать

$$R(s,s) \leqslant (4-\delta(s))^s, \quad \delta(s)\to 0,$$

но не умеем заменить $\delta(s)$ какой-либо постоянной $\delta>0$. Р. Грэхем предложил деньги тому, кто первым обоснует оценку $R(s,s)\leqslant(4-\delta)^s$ (пускай даже $\delta=10^{-6}$ или $\delta=10^{-1000}$), но и материальное стимулирование пока не помогло. Соответственно, интересны нижние оценки чисел Рамсея.

Из параграфа 2.7 мы знаем, что у почти всякого графа $G$ на $n$ вершинах одновременно $\alpha(G)<[2\log_2 n]$ и $\omega(G)<[2\log_2 n]$. Введя обозначение $s=[2\log_2 n]$, немедленно получаем неравенство $R(s,s)>n$. Остается выразить $n$ через $s$. Но это совсем легко:

$$s\leqslant 2\log_2 n \;\Rightarrow\; n \geqslant 2^{s/2}.$$

Если в пункте 2.6.5 (на который мы и ссылаемся в параграфе 2.7) чуть аккуратнее провести выкладку, то в конечном счете получится неравенство

$$R(s,s) \geqslant (1+o(1))\frac{1}{e\sqrt{2}}s2^{s/2}.$$

Этот результат также принадлежит Эрдёшу и Секерешу (см. [27] и [1, 23, 30]), и он также почти не был улучшен за прошедшие годы. Сейчас рекорд лишь вдвое (!) точнее упомянутой оценки (см. [1, 23, 30]).

В итоге

$$(1+o(1))\frac{\sqrt{2}}{e}s2^{s/2} \leqslant R(s,s) \leqslant \frac{4^s}{e^{\gamma \ln^2 s/\ln\ln s}},$$

или, короче,

$$(\sqrt{2}+o(1))^s \leqslant R(s,s) \leqslant (4+o(1))^s.$$

Заметим, что графы, у которых число независимости и кликовое число одновременно не превосходят функции, асимптотически равной $2\log_2 n$, называются *рамсеевскими*. Они играют огромную роль в теории алгоритмов (в вопросах теоретической сложности вычислений).

Заметим также, что отдельный интерес представляет задача о явном конструировании рамсеевских графов (мы лишь доказали их существование с помощью случайного графа). Эта задача крайне далека от своего решения (см. [31])!

## 2.9. Хроматическое число и обхват графа

В предыдущем параграфе мы убедились в том, что случайные графы интересны не только как комбинаторно-вероятностные объекты, свойства которых мы изучаем, но и как инструменты для доказательства замечательных утверждений в тех областях, которые, на первый взгляд, к случайным графам вовсе отношения не имеют. Сейчас мы докажем еще один результат подобного типа.

Назовем *обхватом* графа $G$ длину самого короткого простого цикла в нем. Обозначим обхват графа через $g(G)$. Обозначение происходит от английского «girth» — «обхват».

Кажется естественным предположить, что чем больше обхват графа, тем меньше его хроматическое число. В самом деле, величина обхвата говорит о разреженности графа, о наличии в множестве его ребер больших дыр. Однако у нас есть уже некоторый опыт работы с аналогичными эффектами: мы ведь знаем, что у *почти всякого* графа весьма ощутимое хроматическое число ($\chi(G) \geqslant \frac{n}{2\log_2 n}$) при полном отсутствии сколь-нибудь значимых клик ($\omega(G) \leqslant 2\log_2 n$).

Ну, клики — кликами, их, может, и нет, а вот поверить в отсутствие коротких циклов у графа с большим хроматическим числом все равно тяжело. Одно дело, когда на данных $k$ вершинах нет каких-то из $C_k^2 \sim \frac{k^2}{2}$ ребер, и совсем иное дело, если отсутствуют какие-либо из $k$ ребер потенциального цикла: $k$ куда меньше, нежели $\frac{k^2}{2}$.

Теория случайных графов и тут преподносит нам сюрприз. Следующая теорема принадлежит Эрдёшу.

**Теорема 19.** *Для любых $k, l \in \mathbb{N}$ существует граф $G$, у которого $g(G) > l$ и $\chi(G) > k$.*

Не правда ли, удивительно?

**Доказательство теоремы 19.** Зафиксируем произвольное вещественное $\theta \in \left(0, \frac{1}{l}\right)$ и рассмотрим случайный граф в модели $G(n, p)$, где $p = n^{\theta-1}$. Обозначим через $Z_l$ случайную величину, равную числу циклов длины не больше $l$ в случайном графе. Линейность математического ожидания влечет равенство

$$MZ_l = \sum_{i=3}^{l} C_n^i \frac{(i-1)!}{2} p^i.$$

Здесь $C_n^i \frac{(i-1)!}{2}$ — количество $i$-вершинных циклов в полном графе $K_n$ (как известно, из $i$ драгоценных камней можно составить $\frac{(i-1)!}{2}$ различных ожерелий), и $i$ меняется не от единицы, а от тройки до $l$, поскольку циклов длины 1 и 2 не бывает.

Осуществим оценку:

$$MZ_l \leqslant \sum_{i=3}^{l} \frac{n^i}{i!} \cdot \frac{(i-1)!}{2} \cdot p^i < \sum_{i=3}^{l} (np)^i.$$

Мы знаем, что $i \leqslant l$, а $\theta < \frac{1}{l}$. Следовательно,

$$(np)^i \leqslant n^{l\theta} = o(n).$$

В итоге $MZ_l = o(n)$, ведь $l$ — константа.

За счет неравенства Маркова имеем

$$P_{n,p}\left(Z_l \geqslant \frac{n}{2}\right) \leqslant \frac{MZ_l}{n/2} = o(1).$$

Существует такое $n_1$, что для всех $n > n_1$ выполнено

$$P_{n,p}\left(Z_l \geqslant \frac{n}{2}\right) < \frac{1}{2} \Rightarrow P_{n,p}\left(Z_l < \frac{n}{2}\right) > \frac{1}{2}.$$

Положим

$$t = \left\lceil \frac{3\ln n}{p} \right\rceil.$$

Снова воспользуемся неравенством Маркова:

$$P_{n,p}(\alpha(G) \geqslant t) = P_{n,p}(X_t \geqslant 1) \leqslant MX_t = C_n^t(1-p)^{C_t^2} \leqslant n^t e^{-p\frac{t(t-1)}{2}} =$$
$$= \left(ne^{-\frac{p(t-1)}{2}}\right)^t \leqslant \left(ne^{-\frac{p((3\ln n/p)-1)}{2}}\right)^t \leqslant \left(ne^{-1{,}5\ln n+p/2}\right)^t =$$
$$= \left(\frac{1}{\sqrt{n}}(1+o(1))\right)^t = o(1).$$

Значит, снова есть такое $n_2$, что при всех $n > n_2$ выполнено

$$P_{n,p}(\alpha(G) \geqslant t) < \frac{1}{2} \Rightarrow P_{n,p}(\alpha(G) < t) > \frac{1}{2}.$$

Выходит, что при $n > \max\{n_1, n_2\}$

$$P_{n,p}\left(Z_l < \frac{n}{2},\ \alpha(G) < t\right) > 0,$$

т. е. существует граф $G$ на $n$ вершинах, у которого одновременно меньше чем $\frac{n}{2}$ циклов длины, не превосходящей $l$, и нет независимых множеств размера $t$.

Удалим из каждого $i$-вершинного цикла ($i \leqslant l$) в найденном графе $G$ одну вершину вместе со всеми смежными с ней ребрами. Получится граф $G'$, в котором не меньше $\frac{n}{2}$ вершин, вовсе нет циклов длины, не превосходящей $l$, и по-прежнему $\alpha(G') < t$. Последний факт говорит о том, что

$$\chi(G') \geqslant \frac{n/2}{t} \sim \frac{np}{6\ln n} = \frac{n^\theta}{6\ln n} \to \infty.$$

При всех $n > n_3$ имеем $\chi(G') > k$, т. е. при всех $n > \max\{n_1, n_2, n_3\}$ имеем $g(G') > l$ и $\chi(G') > k$. Теорема доказана. $\square$

## 2.10. Законы нуля или единицы

В классической теории вероятностей есть так называемые законы нуля или единицы. О них есть изрядная литература (см., например, [6]), но то, о чем речь пойдет ниже, практически никакого отношения к ним не имеет. Законы нуля или единицы для случайных графов — это особый раздел теории, который, на наш взгляд, едва ли не самый красивый и удивительный во всей этой науке. Для того чтобы их сформулировать, нам нужно будет понять, что такое логический язык первого порядка для описания свойств графов.

### 2.10.1. Язык первого порядка для графов

В этом пункте мы не собираемся делать экскурс в логику. Нам достаточно будет объяснить лишь тот аспект науки о «высказываниях», который касается графов и их свойств. Поэтому читателя, который пожелает глубже разобраться с логическими основами того, о чем мы будем говорить, мы отсылаем к книгам [35, 36].

Допустим, хочется нам как-нибудь формально записать свойство «граф содержит треугольник». Можно воспользоваться кванторами и символом $\sim$, который в данном контексте обозначает не асимптотическое равенство, но смежность двух вершин в графе (наличие ребра между ними):

$$\exists x\ \exists y\ \exists z\ (x \sim y) \cap (x \sim z) \cap (y \sim z).$$

Здесь символом «$\cap$» мы обозначили логическое «и», т. е. конъюнкцию. Аналогично, через $\cup$ можно обозначать дизъюнкцию («или»). В итоге с помощью кванторов существования и всеобщности, конъюнкций, дизъюнкций, импликаций ($\Longrightarrow$), отрицаний, значка «$\sim$» и просто «$=$», а также с помощью букв, обозначающих вершины графа, можно построить массу *конечных* «фраз», каждая из которых будет выражать некоторое свойство графа. Например, свойство «граф не имеет изолированных вершин» или свойство «в графе нет циклов длины не больше трех» выразимы на описанном языке. Именно этот язык и называется языком первого порядка для графов.

Разумеется, не составляет труда придумать свойства, которые на языке первого порядка не записать. Например, таково свойство связности или свойство «хроматическое число графа не превосходит половины от числа его вершин»: конечной фразой эти свойства не представить. И все-таки мир свойств, изложимых на языке первого порядка, весьма богат. Оказывается, он подчиняется исключительно красивым законам.

### 2.10.2. Формулировки результатов

**Теорема 20.** *Если $p = p(n)$ — такая функция, что $pn^{\alpha} \to \infty$ и $(1-p)n^{\alpha} \to \infty$ для любого $\alpha > 0$, а $A$ — произвольное свойство графа, которое можно выразить на языке первого порядка, то либо почти* *наверное свойство $A$ выполнено, либо почти наверное оно не имеет места. Иными словами, либо $P_{n,p}(A) \to 1$, либо $P_{n,p}(A) \to 0$.*

Это и есть закон нуля или единицы. Фактически, это закон, которому подчиняется вероятность ребра $p$. Оказывается, что если $p$ — это константа или функция, которая не слишком быстро стремится к нулю (подходит $\frac{1}{\ln n}$, но не подходит $\frac{1}{\sqrt[1/10000]{n}}$) либо к единице, то любое свойство первого порядка для случайного графа имеет асимптотическую вероятность 0 или 1. Промежуточные варианты невозможны!

Теорему 20 доказали Ю. В. Глебский, Д. И. Коган, М. И. Лиогонький и В. А. Таланов в 1969 году, а переоткрыл ее Р. Фагин в 1976 году. Еще более впечатляющий результат получили С. Шеллах и Дж. Спенсер в 1988 году.

**Теорема 21.** *Если $p = n^{-\alpha}$, где $\alpha > 0$ — иррациональное число, а $A$ — произвольное свойство графа, которое можно выразить на языке первого порядка, то либо почти наверное свойство $A$ выполнено, либо почти наверное оно не имеет места. Иными словами, либо $P_{n,p}(A) \to 1$, либо $P_{n,p}(A) \to 0$.*

На первый взгляд, теорема 21 кажется чем-то совсем запредельным. Но, может быть, иррациональность $\alpha$ не по делу? Тогда пафоса сразу становится куда как меньше. А вот и нет! Иррациональность нужна, и без нее теорема перестает работать. Действительно, рассмотрим свойство $A$ — «граф содержит треугольник» и возьмем $\alpha = 1$. По теореме 11 с $c = 1$ имеем

$$P_{n,p}(A) \sim 1 - e^{-1/6} \notin \{0, 1\}.$$

Ни нуля, ни единицы. Нет закона.

Теорему 21, ввиду ее сравнительной сложности, мы доказывать не будем. Что же до теоремы 20, то мы приведем схему ее доказательства. В основе этого доказательства лежит игра, предложенная А. Эренфойхтом в 1960 году. Об этой игре мы расскажем в следующем пункте.

### 2.10.3. Игра Эренфойхта

Даны два графа — $G = (V, E)$ и $H = (W, F)$. И есть два игрока — Новатор и Консерватор. В игре $k$ раундов. В каждом раунде Новатор и Консерватор делают по одному ходу. Сперва Новатор выбирает любой из двух графов и вершину в нем. Затем Консерватор обращается ко второму графу (тому, который «отверг» Новатор) и вытаскивает вершину из него. Например, пусть на шаге с номером $i$ Новатор предпочел граф $G$ и взял вершину $x_i \in V$. Тогда Консерватору ничего не остается делать, как вытащить некоторую вершину $y_i \in W$. После окончания игры образуются два множества вершин — $x_1, \ldots, x_k \in V$ и $y_1, \ldots, y_k \in W$. При этом часть вершин из первого множества и часть вершин из второго обязаны своим появлением Новатору; остальные возникли благодаря Консерватору. Консерватор выигрывает, если индуцированные подграфы

$$G|_{\{x_1,\ldots,x_k\}}, \quad H|_{\{y_1,\ldots,x_y\}}$$

*изоморфны*, т. е. для любых $i, j \in \{1, \ldots, k\}$ выполнено $(x_i, x_j) \in E$ тогда и только тогда, когда $(y_i, y_j) \in F$.

Смысл в том, что Новатор всякий раз смело переходит к новому графу, а Консерватор старается так скопировать его действия, чтобы на выходе получились одинаковые по сути структуры.

Для краткости обозначим описанную игру $\mathrm{EHR}(G, H, k)$ в честь Эренфойхта, чье имя в оригинале пишется *Ehrenfeucht* (см. [35]).

Сейчас мы без доказательства приведем теорему, которая позволяет перейти от логики и языков первого порядка к играм на случайных графах. Прежде аккуратно опишем объекты, с которыми будем иметь дело.

Скажем, что Консерватор побеждает в игре $\mathrm{EHR}(G, H, k)$, если существует выигрышная стратегия, применив которую, Консерватор добьется победы. Пусть даны два пространства случайных графов — $G(n, p(n))$ и $G(m, p(m))$. Здесь множества вершин случайных графов — $V$ (мощности $n$) и $W$ (мощности $m$) — не пересекаются, а функция $p$ в обоих случаях одна и та же, только аргументы у нее $n$ и $m$ соответственно. Для любых двух графов $G$ (из $G(n, p(n))$ и $H$ (из $G(m, p(m))$ определим вероятность пары $(G, H)$ как величину

$$P_{n,m,p}((G, H)) = P_{n,p(n)}(G) \cdot P_{m,p(m)}(H).$$

Вероятность произвольного события есть, как обычно, сумма вероятностей его элементов. Например, понятно, как устроена величина

$$P(n, m, k) = \\ = P_{n,m,p}\big(\{(G, H)\colon \text{Консерватор побеждает в игре } \mathrm{EHR}(G, H, k)\}\big).$$

Нужное нам утверждение — это

**Теорема 22.** *Функция $p$ подчиняется закону нуля или единицы тогда и только тогда, когда для любого $k$*

$$\lim_{n,m\to\infty} P(n, m, k) = 1.$$

Идею доказательства теоремы 22 можно найти в книге [23], а также в популярной статье [37].

В следующем пункте мы опишем выигрышную стратегию для Консерватора, оценим вероятность из теоремы 22 и завершим тем самым доказательство теоремы 20.

### 2.10.4. Выигрышная стратегия для Консерватора

Скажем, что граф $G=(V,E)$ обладает *свойством полного расширения уровня* $k$, если для любых $a, b$, удовлетворяющих условию $a+b\leqslant k$, и для любых различных вершин $u_1, \ldots, u_a, v_1, \ldots, v_b \in V$ найдется вершина $x\in V$, для которой $(x,u_i)\in E$ и $(x,v_i)\notin E$ при всех $i$.

Если $G$ и $H$ обладают свойством полного расширения уровня $k$, то выигрышная стратегия для Консерватора в игре $\mathrm{EHR}(G,H,k)$ почти тривиальна. В самом деле, пусть к моменту времени $i\leqslant k$ уже выбраны вершины $x_1, \ldots, x_{i-1}$ из $G$ и $y_1, \ldots, y_{i-1}$ из $H$. На $i$-м шаге Новатор берет (например) $x_i$ в $G$. Консерватор просто смотрит на связи между вершинами $x_1, \ldots, x_i$ и выбирает такую вершину $y_i$ из $H$, которая сохраняет те же связи на множестве вершин $y_1, \ldots, y_i$. Он может это сделать по причине свойства полного расширения.

Таким образом, если окажется, что почти наверное случайный граф обладает свойством полного расширения любого наперед заданного уровня $k$, то и асимптотика $P(n,m,k)\to 1$ будет у нас в кармане.

Что ж, давайте докажем нужное утверждение. Обозначим через

$$A_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}$$

событие, состоящее в том, что в случайном графе вершина $x$ смежна с каждой вершиной $u_i$ и не смежна ни с одной из вершин $v_i$. Вершины $u_1, \ldots, u_a, v_1, \ldots, v_b, x$ фиксированы. Очевидно,

$$P_{n,p}(A_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}) = p^a(1-p)^b.$$

Далее,

$$P_{n,p}\big(\cap_{x\notin\{u_1,\ldots,u_a,v_1,\ldots,v_b\}}\overline{A_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}\big) = \big(1-p^a(1-p)^b\big)^{n-a-b}.$$

Наконец,

$$P_{n,p}\big(\cup_{a,b\colon a+b\leqslant k}\cup_{u_1,\ldots,u_a,v_1,\ldots,v_b}\cap_{x\notin\{u_1,\ldots,u_a,v_1,\ldots,v_b\}}\overline{A_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}\big) \leqslant \\ \leqslant k^2n^k(1-\varepsilon^k)^{n-k}, \quad \varepsilon=\min\{p, 1-p\}.$$

Длиннющее событие, стоящее под знаком последней вероятности, представляет собой отрицание свойства полного расширения: существуют $a$ и $b$, удовлетворяющие неравенству $a+b\leqslant k$, и существуют такие вершины $u_1, \ldots, v_b$, что для любой другой вершины $x$ либо нарушена хотя бы одна из связей $x \sim u_i$, либо присутствует хотя бы одна из связей $x \sim v_i$. Все, что нам нужно теперь, это убедиться в стремлении к нулю вероятности данного события. Итак,

$$k^2 n^k (1-\varepsilon^k)^{n-k} \leqslant k^2 n^k e^{-\varepsilon^k (n-k)}.$$

Понятно, что сомножитель $k^2$ и вычитаемое $k$ в показателе экспоненты на асимптотику не влияют, поскольку $k$ — константа (хотя и сколь угодно большая). Кроме того,

$$\varepsilon n^{1/(2k)} \to \infty,$$

каково бы ни было $k$, за счет условия теоремы. Значит, при больших $n$

$$\varepsilon^k n = (\varepsilon n^{1/(2k)})^k \cdot \sqrt{n},$$

т. е.

$$e^{-\varepsilon^k n} < e^{-\sqrt{n}}.$$

Ясно, что

$$n^k e^{-\sqrt{n}} \to 0,$$

и теорема доказана.

## 2.11. Еще ряд сюжетов

Задач о случайных графах — великое множество. Как мы уже говорили, мы не стремимся здесь к сколь-нибудь исчерпывающему изложению теории, но лишь повествуем о наиболее ярких и любопытных, с нашей точки зрения, сюжетах. Скажем несколько слов еще о ряде исследовательских направлений. Доказательств мы приводить не будем, лишь дадим небольшие обзоры.

### 2.11.1. Деревья в случайных графах

Напомним, что *дерево* — это связный граф без циклов. По-другому, дерево — это связный граф, у которого при данном количестве вершин наименьшее число ребер: $n-1$ ребро на $n$ вершинах. Изучают, например, распределение случайной величины $T_k$, равной количеству компонент, каждая из которых образует $k$-вершинное дерево в случайном графе. Сформулируем результат, который верен для любого постоянного $k \geqslant 2$.

**Теорема 23.** *Имеют место следующие ситуации*:

1) *если* $p = o(n^{-k/(k-1)})$, *то почти наверное* $T_k = 0$;

2) *если* $p \sim cn^{-k/(k-1)}$, *то* $T_k$ *имеет асимптотически пуассоновское распределение с параметром* $\lambda = \frac{c^{k-1}k^{k-2}}{k!}$;
3) *если* $pn^{k/(k-1)} \to \infty$ *и* $pkn - \ln n - (k-1)\ln\ln n \to -\infty$, *то почти наверное* $T_k \geqslant 1$;
4) *если* $pkn - \ln n - (k-1)\ln\ln n \sim x$, *то* $T_k$ *имеет асимптотически пуассоновское распределение с параметром* $\lambda = \frac{e^{-x}}{k \cdot k!}$;
5) *если* $pkn - \ln n - (k-1)\ln\ln n \to \infty$, *то почти наверное* $T_k = 0$.

Происхождение теоремы интуитивно ясно. Когда вероятность $p$ совсем маленькая (первая ситуация), ребра просто «почти» не собираются в деревья: вероятность появления дерева слишком мала; когда величина $p$ становится побольше, деревья возникают, но не сверхчасто; еще чуть больше $p$, и деревьев почти наверное полно... Так что же случается, когда $p$ дальше возрастает? Почему происходит обратное затухание? А дело в том, что деревьев-то огромное количество, но начинает уменьшаться вероятность того, что они образуют компоненты: мы же знаем, что при $p \gg \frac{\ln n}{n}$ случайный граф и вовсе почти наверное связен; откуда же в нем взяться компонентам-деревьям?

Формальное доказательство теоремы весьма похоже на доказательство теорем 10—12. Его можно найти в книге [1], но можно и воспроизвести его своими силами. Для ситуации 1 достаточно убедиться в стремлении к нулю величины $MT_k$; для ситуации 2 потребуется отыскание асимптотик для факториальных моментов; для ситуации 3 хватит неравенства Чебышёва; для ситуации 4 снова придется использовать метод моментов; для ситуации 5 нужно сослаться на теорему 13, а для непокрытых этой теоремой случаев опять сапеллировать к тому факту, что $MT_k \to 0$. При этом полезно иметь в виду, что различных деревьев на $k$ вершинах $k^{k-2}$. Это классическая формула Кэли (см. [16]).

Существует не только пуассоновская, но и нормальная аппроксимация для $T_k$ (ср. § 1.10). Ее получил А. Д. Барбур в 1982 году (см. [38]).

**Теорема 24.** *Пусть* $k \geqslant 2$ *и* $c > 0$ — *фиксированные числа. Пусть также* $p \sim \frac{c}{n}$. *Положим*

$$\lambda_k = nk^{k-2}c^{k-1}\frac{e^{-ck}}{k!}, \quad \sigma_k^2 = \lambda_k\left(1 + \frac{(c-1)(ck)^{k-1}e^{-ck}}{k!}\right),$$

$$\Phi(x) = \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt.$$

*Тогда*

$$\sup_x \left| P_{n,p}\left(\frac{T_k - \lambda_k}{\sigma_k} \leqslant x\right) - \Phi(x)\right| = O\left(\frac{1}{\sqrt{n}}\right).$$

Следующий по сложности после дерева связный граф — это граф *унициклический*: у него на $k$ вершинах $k$ ребер (т. е. в нем ровно один цикл). Справедлива

**Теорема 25.** *Пусть* $p = \frac{c}{n}$, $c < 1$. *Тогда почти наверное все компоненты случайного графа суть либо деревья, либо унициклические графы.*

Теорема 25 доказывается с помощью неравенства Чебышёва (см. [1]). Она любопытна еще и тем, что проливает дополнительный свет на теорему 16. Оказывается, при $p = \frac{c}{n}$, $c < 1$, все компоненты не только «феодально-раздробленные» (логарифмически маленькие), но они к тому же крайне просто устроены — не сложнее графа с одним циклом. При фазовом переходе (т. е. при скачке функции $p$ через границу $\frac{1}{n}$) картина разительно меняется.

### 2.11.2. Еще несколько слов о хроматическом числе случайного графа

Сейчас мы готовы дать еще один симпатичный комментарий к теореме 18. Действительно, в конце предыдущего пункта мы узнали, что при $p = \frac{c}{n}$, $c < 1$, почти наверное все компоненты случайного графа суть деревья или графы с одним циклом. Очевидно, стало быть, что при таких $p$ почти наверное хроматическое число не превосходит тройки. Иными словами, в ситуациях, когда $p = n^{-\alpha}$, $\alpha \geqslant 1$, теорема 18 становится тривиальной: концентрация хроматического числа в трех значениях и так есть. Более того, при $\alpha > 1$ случайный граф — это почти наверное лес (достаточно убедиться в том, что математическое ожидание числа циклов стремится к нулю), т. е. его хроматическое число не превосходит двойки, а при $\alpha > 2$ в случайном графе и вовсе почти наверное нет ребер, так что он красится в один цвет. В итоге получается, что лишь при $\frac{5}{6} < \alpha < 1$ вся кухня из теоремы 18 становится существенной. Именно тогда хроматическое число может расти с ростом числа вершин графа (ср. п. 2.6.12), оставаясь при этом крайне плотно сконцентрированным. И именно тогда теорема 18 по-настоящему удивительна.

### 2.11.3. Планарность случайного графа

Напомним, что граф называется *планарным*, если его можно изобразить на плоскости без пересечения ребер не по вершинам. Например, граф $K_4$ планарен (см. рис. 7), а граф $K_5$ планарным не является. Вообще, хорошо известен критерий планарности графа, доказанный К. Куратовским в 1932 году (см. [13]): граф планарен тогда и только тогда, когда он не содержит в качестве подграфа гомеоморфной копии графа $K_5$ или графа $K_{3,3}$. Здесь $K_{3,3}$ — это двудольный граф с долями размера 3 (см. рис. 8), и два графа называются *гомеоморфными* друг другу, если один из другого может быть получен добавлением вершин на ребрах (см. пример на рис. 9).

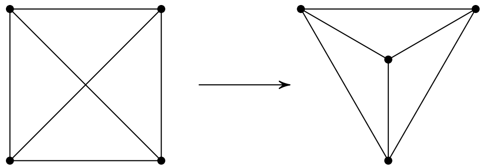


Рис. 7

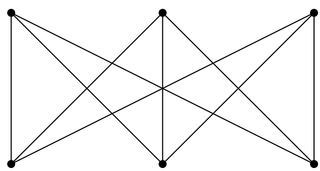


Рис. 8

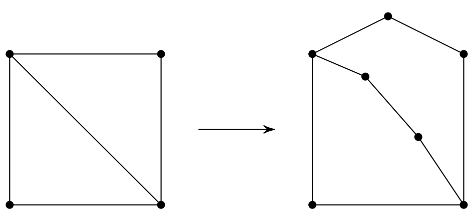


Рис. 9

Понятие планарности сыграло колоссальную роль в становлении теории графов и топологии. Фактически даже великая проблема четырех красок (см. [13]) — это утверждение о том, что хроматическое число любого планарного графа не превосходит четырех.

Если бы в критерии Куратовского не было гомеоморфизма и для планарности хватало бы просто отсутствия в графе подграфов $K_5$ и $K_{3,3}$, то задача о планарности случайного графа стала бы почти тривиальной. Нужное нам свойство можно было бы записать на языке первого порядка, да и без подобных изысков все было бы крайне просто. Однако гомеоморфизмы портят жизнь. Несмотря на это, имеется следующая теорема.

**Теорема 26.** *Пусть* $p=\frac{c}{n}$. *Тогда при* $c<1$ *почти наверное случайный граф планарен, а при* $c>1$ *почти наверное планарности нет.*

Случай $c<1$ мгновенно вытекает из теоремы 25, поскольку деревья и уницикличеcкие графы, конечно, планарны. Случай $c>1$ сложнее. Чуть более слабое утверждение нетрудно вывести из того факта, что у всякого планарного графа на $n$ вершинах не более $3n-6$ ребер. Подумайте над этим!

Интересно, кстати, каким значимым оказывается порог $\frac{1}{n}$. Вот и для планарности именно на нем происходит фазовый переход.

### 2.11.4. Степени вершин случайного графа

Имеется огромная и очень важная наука о распределении степеней вершин (*степень вершины* — это количество смежных с ней ребер) случайного графа. Значимость этой науки мы еще продемонстрируем как в следующем пункте, так и в четвертой главе этой книги. Сейчас скажем лишь несколько слов об известных результатах.

Прежде всего понятно, что для любой вершины $v$ случайного графа в модели $G(n, p)$ случайная величина $\deg v$, как раз обозначающая степень $v$, имеет биномиальное распределение $\mathrm{Binom}(n-1, p)$.

Более интересным является изучение последовательности степеней вершин случайного графа, упорядоченных по невозрастанию величины: $d_1 \geqslant d_2 \geqslant \ldots \geqslant d_n$. Массу результатов о величинах $d_i$ можно найти в книге [1]. Мы же упомянем лишь один из них, так как именно он пригодится нам в следующем пункте.

**Теорема 27.** *Пусть* $p=\frac{1}{2}$. *Тогда почти наверное для всех* $i \leqslant$ $\leqslant [3\log_2 n]$ *выполнено*

$$d_i - d_{i+1} \geqslant 3.$$

Теорема говорит о том, что у почти всякого графа зазоры между соседними величинами $d_i$ при довольно многих $i$ достаточно велики. Эта теорема допускает разнообразные обобщения, но мы о них писать не станем.

### 2.11.5. Изоморфизм случайных графов

Одной из наиболее глубоких проблем комбинаторики и теории алгоритмов является *проблема изоморфизма* графов. Пусть даны два графа с одинаковым числом вершин $n$. Спрашивается: существует ли алгоритм, работающий за полиномиальное по $n$ время и говорящий на выходе, изоморфны данные графы или нет? Эта проблема до сих пор крайне далека от своего решения. Известны лишь классы графов, внутри которых нужный алгоритм найден. Например (см. [39–43]), таковы все деревья, все планарные графы, все интервальные графы, все графы, у которых степени вершин ограничены общей константой...

В 1980 году Л. Бабаи, Эрдёш и С. Селков доказали следующую теорему.

**Теорема 28.** *Пусть $p = \frac{1}{2}$. Тогда для каждого $n$ существует такое множество графов $\mathcal{K}_n$, что $P_{n,1/2}(\mathcal{K}_n) \to 1$, и такой алгоритм $A$, что для любых двух графов $G, H \in \mathcal{K}_n$ за $O(n^2)$ операций $A$ дает ответ на вопрос об изоморфизме $G$ и $H$.*

Более выигрышно звучит такой укороченный вариант формулировки: существует квадратичный по сложности алгоритм, который решает проблему изоморфизма для почти всякой пары графов.

Опишем этот алгоритм. Для этого сперва построим $\mathcal{K}_n$. Возьмем произвольный граф $G \in \Omega_n$. Упорядочим вершины графа $G$ по величине их степени: вершину степени $d_i$ (см. предыдущий пункт) обозначим $x_i$. Посмотрим на первые $m = [3\log_2 n]$ вершин. Если хотя бы для одного $i$ в указанном диапазоне $d_i - d_{i+1} \leqslant 2$, то говорим, что $G \notin \mathcal{K}_n$ (ср. теорему 27). Иначе полагаем $a(i, j) = 1$, коль скоро $x_i \sim x_j$ или $i = j$, и $a(i, j) = 0$ в противном случае, после чего вычисляем

$$f(x_i) = \sum_{j=1}^{m} a(i, j) 2^j, \quad i = 1, \ldots, n.$$

Если найдутся $i, j$, для которых $f(x_i) = f(x_j)$, то говорим, что $G \notin \mathcal{K}_n$. Все, с $\mathcal{K}_n$ разобрались. Очевидно, для данных двух графов $G, H \in \Omega_n$ проверка их принадлежности множеству $\mathcal{K}_n$ требует $O(n^2)$ операций.

Если же мы убедились, что $G, H \in \mathcal{K}_n$, то $A$ тривиален. Переупорядочиваем вершины графов по *убыванию* их величины $f$ (это возможно, поскольку графы прошли второй этап проверки); графы $G$, $H$ изоморфны тогда и только тогда, когда новые графы совпадают. Опять имеем $O(n^2)$ операций.

Единственное, в чем еще осталось удостовериться, — это то, почему $P_{n,1/2}(\mathcal{K}_n)$ стремится к единице. Одним из оснований к этому служит теорема 27: ввиду нее мы отвергаем граф после первой проверки с ничтожно малой вероятностью. Гораздо труднее объяснить, почему в рамках второй проверки мы также отсеиваем пренебрежимо мало графов. Этот факт заведомо опирается на наше знание о «дырках» между соседними величинами $d_i$ (полученное в результате первой проверки), и его обоснование можно найти в книге [1].