УДК 519.175.4

Л. А. Остроумова

Кафедра дискретной математики факультета инноваций и высоких технологий МФТИ;
Отдел теоретических и прикладных исследований, ООО «Яндекс»

Об r-диаметрах случайных графов в модели Боллобаша—Риордана

Работа посвящена модели Боллобаша–Риордана случайного веб-графа. Эта модель адекватно описывает поведение реального веба. Рассмотрено обобщение понятия диаметра графа — так назывемый r-диаметр, который определяется как максимум по всем множествам вершин мощности r от минимума расстояний между парами вершин в данном множестве. Доказана теорема о том, что почти наверное веб-граф на n вершинах имеет r-диаметр, сколь угодно близкий к величине $\frac{\ln n - \ln r}{\ln \ln n}$.

Ключевые слова: случайный граф, веб-граф, диаметр.

1. Введение

Настоящая работа посвящена исследованию некоторых свойств случайных графов. Существует множество различных моделей таких графов, наиболее распространенная из них — классическая модель Эрдеша–Реньи, в которой граф $G(n,p)$ на n вершинах строится следующим образом: все ребра проводятся независимо друг от друга с вероятностью $p \in (0,1)$. Изучению свойств таких графов посвящено большое количество работ (см., например, [1, 2, 3]).

Однако в последнее время появилась потребность моделировать различные реальные структуры, свойства которых существенно отличаются от свойств случайных графов в модели Эрдеша–Реньи. Особенно важно создать модель, наиболее точно представляющую Интернет. Пусть вершинами графа будут сайты или страницы Интернета, а ребрами — гиперссылки между ними. Получим некоторый граф. Хочется иметь модель, отражающую основные его свойства. Например, в отличие от $G(n,p)$, граф, представляющий Интернет, обладает степенным распределением степеней вершин. Именно это свойство положено в основу различных моделей случайных графов. Некоторые из них изначально обладают нужным распределением, для других это свойство доказывается. Обзор различных моделей и основных результатов можно найти в работе [4].

Наиболее распространенная модель Интернета была впервые описана Барабаши и Альберт (см. [5]), а позже уточнена Боллобашем и Риорданом. Определенную ими модель мы и будем рассматривать в настоящей работе. Опишем, как строится граф в этой модели. Обычно для него используется обозначение G_m^n, где n — число вершин, m — фиксированный параметр. Сперва по индукции строится G_1^n. Граф G_1^1 состоит из одной вершины и одной петли. Граф G_1^t получается из G_1^{t-1} посредством добавления вершины t и одного ребра между вершинами t и i, где i выбирается случайно следующим образом:

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

Настоящая работа выполнена при финансовой поддержке гранта РФФИ № 12-01-00683 и гранта Президента РФ № MD-666.2012.1.

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

Итак, в основу построения графа G_m^n положен принцип предпочтительного присоединения. Это вполне соответствует цели получить модель Интернета, ведь вновь появившийся сайт скорее предпочтет сослаться на ту страницу, которая уже «популярна». Тот факт, что граф G_m^n обладает степенным распределением степеней вершин, был доказан Боллобашем и Риорданом в работе [6]. Кроме того, в статье [7] они показали, что диаметр G_m^n при $m \geqslant 2$ асимптотически равен $\frac{\ln n}{\ln \ln n}$. Точная формулировка этого результата будет дана в следующем разделе. В настоящей работе будет дано обобщение понятия диаметра графа и доказан соответствующий результат.

В разделе 2 будут введены основные понятия и сформулированы теоремы, которые будут доказаны в разделах 3 и 4.

2. Введение обозначений и формулировка результата

Пусть имеется некоторый граф $G = (V, E)$. Его *диаметром* называется следующая величина:

$$\operatorname{diam}(G) = \max_{\substack{i,j \in V \\ i \neq j}} \rho(i, j),$$

где $\rho(i, j)$ — длина кратчайшего пути между вершинами i и j в графе G.

Введем некоторое обобщение понятия диаметра — *r-диаметр*:

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

Иными словами, теперь мы рассматриваем не пары вершин, а подмножества мощности r. В каждом таком подмножестве мы находим две вершины на наименьшем расстоянии друг от друга. И нас интересует максимум этого расстояния по всем подмножествам. Заметим, что при $r = 2$ имеем в точности обычный диаметр графа G. В этой работе всегда будем полагать, что $r \geqslant 2$, чтобы понятие r-диаметра имело смысл.

В работе [7] Боллобаш и Риордан доказали следующую теорему.

Теорема 1. *Для любого $\varepsilon > 0$ и любого натурального $m \geqslant 2$*

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

В настоящей работе будет доказано подобное утверждение для r-диаметра. А именно

Теорема 2. *Пусть $\varepsilon > 0$ и натуральное $m \geqslant 2$ фиксированы. Далее, пусть $r(n)$ — произвольная последовательность целых чисел, лежащих в пределах от 2 до n. Тогда*

$$\lim_{n \to \infty} \mathrm{P}\left(\frac{(1-\varepsilon)\ln n - \ln r}{\ln \ln n} \leqslant \operatorname{diam}_r(G_m^n) \leqslant \frac{(1+\varepsilon)\ln n - \ln r}{\ln \ln n} \right) = 1.$$

Как и ожидалось, чем больше вершин в рассматриваемых множествах, тем меньше r-диаметр. Отметим, что если для любого $\alpha > 0$ выполнено $r(n) = o\,(n^{\alpha})$, т.е. $r(n) = n^{o(1)}$, то результаты теорем 1 и 2, по существу, идентичны. Если же, например, $r(n) = n^{\alpha}$, $\alpha > 0$, то константа в асимптотике из теоремы 2 принципиально другая: $1 \pm \varepsilon$ превращается в $1 - \alpha \pm \varepsilon$. Если, наконец, $r(n) = n^{1-o(1)}$, то асимптотики в теореме 2 нет.

Доказательство теоремы 2 естественно разбивается на два шага — оценка снизу и оценка сверху. Нижняя оценка будет доказана в разделе 3, верхняя — в разделе 4. При этом мы существенно будем опираться на работу [7], особенно в разделе 4.

3. Оценка снизу

3.1. Формулировка результата и вспомогательных лемм

Будем говорить, что событие происходит *с асимптотической вероятностью* 1, если его вероятность стремится к 1 при $n \to \infty$.

При доказательстве оценки снизу мы докажем немного больше, а именно

Теорема 3. *Пусть $m \geqslant 1$ фиксировано, тогда* $\mathrm{diam}_r(G_m^n) \geqslant \dfrac{\ln n - \ln r}{\ln(17m^2 \ln n)} - 1$ *с асимптотической вероятностью* 1.

Ясно, что требуемая оценка сразу следует из этой теоремы.

Разобьем $r(n)$ на две подпоследовательности. В одной из них $r(n) \geqslant \dfrac{n}{\ln n}$, в другой $r(n) < \dfrac{n}{\ln n}$. Заметим, что для первой подпоследовательности оценка снизу тривиальна — для достаточно больших n там просто стоит отрицательное число. Поэтому далее считаем, что $r < \dfrac{n}{\ln n}$.

Для доказательства нам потребуется несколько вспомогательных лемм. Во-первых, нас будет интересовать вероятность того, что граф G_1^n содержит некоторый подграф S. Под событием $\{S \subset G_1^n\}$ понимаем, что граф G_1^n содержит в точности граф S, а не изоморфный ему подграф. Иными словами, если в S проведено ребро ij, то в G_1^n должно быть ребро между вершинами с теми же номерами. В работе [7] доказано следующее утверждение.

Лемма 1. *Пусть $S = (V, E)$ — некоторый граф с множеством вершин $V = \{1, \ldots, n\}$, у которого нет петель и каждая вершина соединена не более чем с одной из вершин с меньшими номерами. Пусть далее $e(S) = |E|$, а $\Delta(S)$ — максимальная степень вершины в графе S. Тогда*

$$\mathrm{P}(S \subset G_1^n) \leqslant 2^{e(S)(\Delta(S)+2)} \prod_{ij \in E(S)} \frac{1}{\sqrt{ij}}.$$

Далее нам понадобится следующая конструкция. Пусть ξ_{ij} $(i < j,\ 1 \leqslant i, j \leqslant n)$ — индикаторные случайные величины на каком-то вероятностном пространстве, то есть с некоторой вероятностью $p_{ij} \in [0, 1]$ величина ξ_{ij} принимает значение 1, а с вероятностью $1 - p_{ij}$ принимает значение 0. Про зависимость между ξ_{ij} ничего не говорится. Пусть, кроме того, Ω_n — множество всех графов на вершинах $1, \ldots, n$ без петель и кратных ребер, $\mathscr{F}_n = 2^{\Omega_n}$ — стандартная сигма-алгебра, состоящая из всех подмножеств Ω_n, а вероятностная мера задается следующим образом: для любого $G = (V, E) \in \Omega_n$

$$\mathrm{P}_{n,\xi_{ij}}(G) = \mathrm{P}(\{\forall\ ij \in E(G) : \xi_{ij} = 1\} \cap \{\forall\ ij \notin E(G) : \xi_{ij} = 0\}).$$

Рассмотрим теперь вероятностное пространство $\mathfrak{G}(n, \xi_{ij}) = (\Omega_n, \mathscr{F}_n, \mathrm{P}_{n,\xi_{ij}})$. Для таким образом определенного случайного графа будет доказана следующая лемма.

Лемма 2. *Пусть* $r < \frac{n}{\ln n}$. *Пусть, кроме того, для индикаторных случайных величин* ξ_{ij} $(i < j,\ 1 \leqslant i, j \leqslant n)$ *выполнено:* $\mathrm{P}(\xi_{ij} = 1) = O\left(\frac{1}{r \ln n}\right)$. *Тогда с асимптотической вероятностью* 1 *в графе* $G \in \mathfrak{G}([n/2], \xi_{ij})$ *есть независимое множество размера хотя бы* r.

В последующих двух параграфах этого раздела будет сначала доказана лемма 2, а затем непосредственно оценка снизу.

3.2. Доказательство леммы 2

Заметим, что количество ребер в графе G с асимптотической вероятностью 1 не превосходит $\frac{n^2}{10\,r}$:

$$\mathrm{P}\left(|E(G)| > \frac{n^2}{10\,r}\right) \leqslant \frac{10\,r\,\mathsf{M}|E(G)|}{n^2} = \frac{10\,r\,C^2_{[n/2]}\,o(1)}{n^2\,r} = o(1).$$

Доказательство леммы проведем методом от противного. Допустим, что в любом подграфе G на r вершинах проведено хотя бы одно ребро. Пусть A_1 — независимое множество в графе G максимального размера. Тогда $|A_1| < r$ и, кроме того, из любой вершины $V(G) \setminus A_1$ ведет ребро в хотя бы одну вершину A_1 (иначе A_1 не максимальное). Далее, пусть A_2 — максимальное независимое множество в подграфе на вершинах $V(G) \setminus A_1$. Тогда $|A_2| < r$ и из любой вершины $V(G) \setminus A_1 \setminus A_2$ ведет ребро хотя бы в одну вершину A_2. Аналогичную операцию будем проводить до тех пор, пока не кончатся вершины графа G. В итоге придем к последовательности множеств $A_1, A_2, \ldots, A_l$. Теперь можем оценить количество ребер снизу:

$$|E(G)| \geqslant ([n/2] - |A_1|) + ([n/2] - |A_1| - |A_2|) + \ldots + ([n/2] - |A_1| - \ldots - |A_l|) = \\ = l\,[n/2] - l\,|A_1| - (l-1)|A_2| - \ldots - |A_l|.$$

Легко заметить, что если при фиксированной сумме $|A_1| + \ldots + |A_l| = [n/2]$ мы хотим минимизировать полученное выражение, мы должны взять максимально возможным $|A_1|$, затем $|A_2|$, и так далее, сколько получится. В итоге получим

$$|E(G)| \geqslant ([n/2] - r) + ([n/2] - 2r) + \ldots + ([n/2] - \left[\frac{n}{2r}\right] r) = [n/2]\left[\frac{n}{2r}\right] - \frac{r}{2}\left[\frac{n}{2r}\right]\left(\left[\frac{n}{2r}\right] + 1\right) = \\ = \frac{n^2}{4\,r}(1 + o(1)) - \frac{n^2}{8\,r}(1 + o(1)) = \frac{n^2}{8\,r}(1 + o(1)).$$

Получили, что с одной стороны, $|E(G)| \leqslant \frac{n^2}{10\,r}$ с асимптотической вероятностью 1, а с другой стороны, $|E(G)| \geqslant \frac{n^2}{8\,r}(1 + o(1))$. Таким образом, предположение неверно. Лемма доказана.

Заметим, что проведенное выше рассуждение — это один из способов доказательства известной теоремы Турана в экстремальной теории графов (см. [8]). Мы воспроизвели его прежде всего ради большей замкнутости работы.

3.3. Доказательство теоремы 3

Положим $L = \left[\frac{\ln n - \ln r}{\ln(17 m^2 \ln n)} - 1\right]$. Рассмотрим множество вершин $A = \{\lceil n/2 \rceil + 1, \ldots, n\}$ графа G^n_m. Заметим, что $|A| = [n/2]$. Докажем, что для любых вершин $i, j \in A$

$$\mathrm{P}(\rho(i,j) \leqslant L) = O\left(\frac{1}{r \ln n}\right).$$

Рассмотрим две вершины $i, j \in A$. Оценим вероятность того, что между этими двумя вершинами найдется путь длины $1 \leqslant l \leqslant L$. Пусть путь между этими вершинами в G_m^n следующий: $v_0 = i, v_1, v_2, \ldots, v_{l-1}, v_l = j$. Рассмотрим какой-нибудь граф G_1^{mn}, из которого получен граф G_m^n с нужным свойством. Тот факт, что в графе G_m^n соединены вершины v_t и v_{t+1}, обозначает то, что в G_1^{mn} есть ребро $u_t w_{t+1}$, причем $\left\lceil \frac{u_t}{m} \right\rceil = v_t$ и $\left\lceil \frac{w_{t+1}}{m} \right\rceil = v_{t+1}$. При данных $u_0, \ldots, u_l$ и $w_1, \ldots, w_{l+1}$ вероятность того, что в графе G_m^n проведены ребра $u_t w_{t+1}$ $(0 \leqslant t \leqslant l-1)$, по лемме 1 не превосходит

$$2^{4l} \prod_{t=0}^{l-1} \frac{1}{\sqrt{u_t w_{t+1}}} \leqslant 2^{4l} \prod_{t=0}^{l-1} \frac{1}{\sqrt{v_t v_{t+1}}} = 2^{4l} \frac{1}{\sqrt{ij}} \prod_{t=1}^{l-1} \frac{1}{v_t}.$$

Посмотрим, сколько всего есть графов, из которых получается G_m^n с данным конкретным путем длины l. Каждое ребро между v_t и v_{t+1} может быть получено m^2 способами: m возможностей выбрать вершину u_t и m возможностей выбрать вершину w_{t+1}. Ребер всего l, в итоге получаем m^{2l} графов. Итак, вероятность содержать конкретный путь длины l не превосходит

$$(4m)^{2l} \frac{1}{\sqrt{ij}} \prod_{t=1}^{l-1} \frac{1}{v_t}.$$

Суммируем по всем возможным $v_1, \ldots, v_{l-1}$, получаем, что вероятность того, что вершины i и j соединены путем длины l, не превосходит (для достаточно больших n)

$$\frac{(4m)^{2l}}{\sqrt{ij}} \sum_{1 \leqslant v_1, \ldots, v_{l-1} \leqslant n} \prod_{t=1}^{l-1} \frac{1}{v_t} \leqslant \frac{2(4m)^{2l}}{n} \left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right)^{l-1} \leqslant \frac{2(17m^2)^l (\ln n)^{l-1}}{n}.$$

В первом неравенстве воспользовались тем, что мы рассматриваем вершины с номерами, большими $n/2$. Просуммируем по всем $l \leqslant L$:

$$\frac{2}{n \ln n} \sum_{l=1}^{L} (17m^2 \ln n)^l \leqslant \frac{2(17m^2 \ln n)^{L+1}}{n \ln n} \leqslant \frac{2(17m^2 \ln n)^{\frac{\ln n - \ln r}{\ln(17m^2 \ln n)}}}{n \ln n} = O\left(\frac{1}{r \ln n}\right).$$

Что и требовалось.

Итак, для любых двух вершин из A вероятность того, что они соединены путем длины не более L, равна $O\left(\frac{1}{r \ln n}\right)$. Обратимся теперь к лемме 2. Рассмотрим граф G на множестве вершин A. Проведем ребро между двумя вершинами в графе G, если эти вершины соединены путем длины не больше L в G_m^n. Из леммы 2 следует, что с асимптотической вероятностью 1 в множестве A найдется независимое подмножество размера r. Иными словами, найдутся такие r вершин в графе G_m^n, что расстояние между любыми двумя из них больше L. Теорема доказана.

4. Оценка сверху

4.1. Формулировка результата и описание модели

Справедлива

Теорема 4. *Пусть* $\varepsilon > 0$, *тогда* $\operatorname{diam}_r(G_m^n) \leqslant \frac{(1+\varepsilon)\ln n - \ln r}{\ln \ln n}$ *с асимптотической вероятностью 1.*

При доказательстве теоремы 4 мы будем использовать другое, эквивалентное определение модели G_m^n — *хордовую диаграмму*, которая была описана в работе [7]. Для большей замкнутости данной работы мы приводим здесь полное описание этой модели.

Положим $N = mn$. Рассмотрим числа от 1 до $2N$ и разобьем их на пары. Как нетрудно видеть, всего таких разбиений $(2N)!/(N!2^N)$. Теперь выберем одно из разбиений случайным образом (вероятность выбрать каждое равна $N!2^N/(2N)!$). Представим, что точки от 1 до $2N$ расположены на прямой, и соединим хордами пары точек, на которые мы разбили наше множество. Получили N хорд. Сформируем граф. Из всех правых концов хорд выберем один с наименьшим номером. Пусть это число r_1. Тогда объединим точки $1, \ldots, r_1$ в первую вершину. Далее, точки $r_1 + 1, \ldots, r_2$, где r_2 — следующий правый конец, образуют вторую вершину. И так далее. Получим N вершин. А каждой хорде соответствует ребро. Ребро соединяет вершины i и j $(i \leqslant j)$, если имеется хорда между r_j и одной из точек $r_{i-1} + 1, \ldots, r_i - 1$ (считаем $r_0 = 0$).

Покажем, что такое определение графа эквивалентно обычному определению G_1^N. Будем рассматривать правые концы хорд слева направо. Рассмотрим первый правый конец. Из этой точки идет одна хорда, и она соответствует петле в первой вершине. Далее, пусть каким-либо образом проведены хорды из правых концов $r_1, \ldots, r_i$. Рассмотрим $i + 1$-й правый конец. Вероятность того, что из вершины $i + 1$ ведет ребро в вершину t, пропорциональна количеству отрезков, на которые в данный момент разбит отрезок $[r_{t-1}, r_t]$, ведь ровно столько есть соответствующих разбиений на пары. А количество отрезков, как нетрудно видеть, это степень вершины t в данный момент. Итак, получили ровно то же построение модели.

Преобразуем полученное определение в еще одно эквивалентное. Возьмем отрезок $[0, 1]$ и случайные величины $x_1, x_2, \ldots, x_{2N}$ — независимые и равномерно распределенные на этом отрезке. Пары теперь — это вершины x_{2i-1} и x_{2i}. Каждый порядок расположения точек равновероятен. При этом одному и тому же паросочетанию соответствует ровно 2^N различных взаимных расположений точек. Таким образом, все паросочетания равновероятны. Положим $L_i = \min\{x_{2i-1}, x_{2i}\}$, $R_i = \max\{x_{2i-1}, x_{2i}\}$. Тот же результат получим, если рассмотрим случайные величины R_i — независимые с плотностью $2x$ на отрезке $[0, 1]$ каждая. И при фиксированных R_i возьмем L_i равномерно распределенными на $[0, R_i]$. Аналогично тому, как делали это ранее, соединяем хордами L_i с соответствующими R_i. Считаем, что в графе проведено ребро ij $(i \geqslant j)$, если в полученной диаграмме проведена хорда, соединяющая R_i и L_i, где L_i лежит между R_{j-1} и R_j (полагаем $R_0 = 0$). С этим определением и будем работать далее.

Граф G_m^n получаем из G_1^N обычным образом. Положим $W_i = R_{mi}$ — нас интересуют именно эти правые концы, когда мы рассматриваем G_m^n. И пусть $w_i = W_i - W_{i-1}$ (считаем $W_0 = 0$).

Пусть a — наименьшее целое число, для которого $s = 2^a > \ln^7 n$, b — наибольшее целое число, для которого $2^b < 2n/3$. Положим $I_t = [2^t + 1, 2^{t+1}]$ для $a \leqslant t < b$. Нам потребуется следующая лемма (см. [7]):

Лемма 3. *Фиксируем $m \geqslant 2$. Введем несколько событий:*

$$E_1 = \left\{ \left| W_i - \sqrt{\frac{i}{n}} \right| \leqslant \frac{1}{10}\sqrt{\frac{i}{n}}, \text{ для всех } s \leqslant i \leqslant n \right\},$$

$$E_2 = \left\{ I_t \text{ содержит хотя бы } 2^{t-1} \text{ вершин } i \text{ с } w_i \geqslant \frac{1}{10\sqrt{in}}, \text{ для всех } a \leqslant t < b \right\},$$

$$E_3 = \left\{ w_i \geqslant \frac{\ln^2 n}{n}, \text{ для всех } i < n^{1/5} \right\},$$

$$E_4 = \left\{ w_i \leqslant n^{-4/5},\ \text{для всех } i > \frac{n}{\ln^2 n} \right\}.$$

Тогда вероятность каждого из этих событий стремится к 1 *при* $n \to \infty$.

Далее будем считать, что правые концы $R_1, \ldots, R_N$ фиксированы. Величины L_i независимы и равномерно распределены на $[0, R_i]$. Но мы для простоты будем полагать, что $L_{m(i-1)+j}$ $(1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m)$ равномерно распределены на $[0, W_i]$. Такое допущение просто увеличит вероятность петли в каждой из вершин и, следовательно, не уменьшит r-диаметр. Теперь заметим, что можно рассматривать только $m = 2$. Случай $m > 2$ сводится к этому удалением лишних ребер, что только увеличивает r-диаметр.

Итак, определили случайный граф G, с которым будем работать. Везде далее $W_1, \ldots, W_n$ фиксированы и выполняются события $E_1, \ldots, E_4$. Из каждой вершины i в вершины с меньшими номерами идет два независимых ребра — в вершины $l_{i,1}$ и $l_{i,2}$ (поскольку $m = 2$), причем $\mathrm{P}(l_{i,j} = k) = \dfrac{w_k}{W_i}$ для $k \leqslant i$.

В последующих двух параграфах мы сперва докажем некоторую техническую лемму, а затем и саму теорему 4.

4.2. Техническая лемма

Лемма 4. *Пусть* $\varepsilon > 0$, $r \geqslant \ln^{10} n$, $K = \dfrac{1}{2}\dfrac{(1+\varepsilon)\ln n - \ln r}{\ln \ln n} - 1$. *Далее, пусть* $f_0, f_1, \ldots$ — *последовательность действительных чисел с* $f_0 \geqslant \left(\ln^2 n\right)/n$ *и*

$$f_{k+1} \geqslant \frac{\min\{2\log_2(f_k n/\ln n) - 31, b - a - 1\}}{1000} f_k$$

для всех $k \geqslant 0$. *Тогда при достаточно большом* n *величина* $l = \min\{k : f_k \geqslant 4\ln n/\sqrt{r\,n}\}$ *существует и не превосходит* K.

Доказательство. Заметим, что $b - a - 1$ и $2\log_2(f_0 n/\ln n) - 31$ — растущие с ростом n функции. Значит, при достаточно большом n выполняется неравенство

$$\frac{\min\{2\log_2(f_0 n/\ln n) - 31, b - a - 1\}}{1000} \geqslant 2.$$

В таком случае $f_1 \geqslant 2f_0$. Далее, очевидно, $f_{k+1} \geqslant 2f_k$ (ведь $f_k \geqslant f_0$). В итоге $f_k \geqslant 2^k f_0$ и l существует.

Поскольку $b - a - 1 \geqslant \log_2 n - 8\log_2 \ln n$ для достаточно больших n, то неравенство $2\log_2(f_k n/\ln n) - 31 > b - a - 1$ выполняется лишь при $f_k > \dfrac{1}{(\ln^3 n)\sqrt{n}}$. Последнее выражение больше, чем требуемое нам $4\ln n/\sqrt{r\,n}$, при достаточно больших n. Из этого следует, что для $k < l$ минимум равен первому члену. Таким образом, при достаточно больших n

$$f_{k+1} \geqslant \frac{\log_2(f_k n/\ln n) - 16}{500} f_k \geqslant \frac{\log_2(2^k f_0 n/\ln n) - 16}{500} f_k \geqslant \frac{\log_2(2^k \ln^2 n/\ln n) - 16}{500} f_k = \\ = \frac{k + \log_2 \ln n - 16}{500} f_k \geqslant f_k(k+1)/500.$$

Отсюда

$$f_{l-1} \geqslant \frac{(l-1)!}{500^{l-1}} f_0 \geqslant \left(\frac{l-1}{500e}\right)^{l-1} f_0.$$

Поскольку $f_{l-1} \leqslant 4\ln n/\sqrt{r\,n}$, то

$$\left(\frac{l-1}{500e}\right)^{l-1} \leqslant \frac{4\sqrt{n}}{\sqrt{r}\ln n}.$$

Отсюда при больших n имеем

$$(l-1)\ln(l-1)-(l-1)\ln 500e \leqslant \ln 4+\ln\sqrt{n}-\ln\sqrt{r}-\ln\ln n.$$

Докажем, что с некоторого n имеем $l-1 \leqslant \dfrac{1+\varepsilon/2-\log_n r}{2}\dfrac{\ln n}{\ln\ln n}$. Если величина l ограничена с ростом n, то все очевидно. Иначе предположим противное: для бесконечной последовательности значений n выполнено $l-1 > \dfrac{1+\varepsilon/2-\log_n r}{2}\dfrac{\ln n}{\ln\ln n}$. Тогда (по той же последовательности значений n, начиная с некоторого момента)

$$(l-1)\ln(l-1)-(l-1)\ln 500e-\ln 4=(l-1)\ln(l-1)(1+o(1)) >$$
$$> \frac{1+\varepsilon/2-\log_n r}{2}\frac{\ln n}{\ln\ln n}\left(\ln\frac{1+\varepsilon/2-\log_n r}{2}+\ln\ln n-\ln\ln\ln n\right)(1+o(1)) =$$
$$=\frac{1+\varepsilon/2-\log_n r}{2}\ln n\,(1+o(1))=\big((1+\varepsilon/2)\ln\sqrt{n}-\ln\sqrt{r}\big)\,(1+o(1)) > \ln\sqrt{n}-\ln\sqrt{r}-\ln\ln n.$$

Воспользовались тем, что $\ln\dfrac{1+\varepsilon/2-\log_n r}{2}=o(\ln\ln n)$, поскольку $0\leqslant\log_n r\leqslant 1$. Итак, мы пришли к противоречию, а значит, при достаточно больших n имеем

$$l-1\leqslant\frac{1+\varepsilon/2-\log_n r}{2}\frac{\ln n}{\ln\ln n}=\frac{1}{2}\frac{(1+\varepsilon/2)\ln n-\ln r}{\ln\ln n}<K-1.$$

Лемма доказана. □

4.3. Доказательство теоремы 4

Наконец, можем приступить к доказательству оценки сверху. Опять разобьем $r(n)$ на две подпоследовательности. В одной из них $r(n)<\ln^{10} n$, в другой $r(n)\geqslant\ln^{10} n$. Заметим, что в первом случае $\ln r=o(\ln n)$ и нам фактически нужно доказать, что для любого $\varepsilon>0$ выполнено $\mathrm{diam}_r(G_m^n)\leqslant\dfrac{(1+\varepsilon)\ln n}{\ln\ln n}$ с асимптотической вероятностью 1. Но в работе [7] показано, что путь требуемой длины существует между любыми двумя вершинами графа. Поэтому далее можем считать, что $r(n)\geqslant\ln^{10} n$.

Назовем вершину i *полезной*, если $i\leqslant n/\ln^5 n$ и $w_i\geqslant\left(\ln^2 n\right)/n$. Воспользуемся леммой из работы [7].

Лемма 5. *Вероятность того, что каждая вершина v графа G соединена путем длины не более $8\ln\ln n$ с какой-нибудь полезной вершиной, равна $1-o(1)$ при $n\to\infty$.*

Сформулируем теперь основную лемму, которая фактически даст нам оценку сверху.

Лемма 6. *Рассматриваем граф G, определенный выше. Число $\varepsilon>0$ — фиксировано. Возьмем некоторую полезную вершину v. Тогда с вероятностью $1-o(n^{-1})$ существует путь длины не более $\dfrac{1}{2}\dfrac{(1+\varepsilon)\ln n-\ln r}{\ln\ln n}$ между этой вершиной и одной из вершин $1,\ldots,r-1$.*

Доказательство. Нас будет интересовать то, как растет количество вершин, до которых можно добраться от данной вершины не более чем за k шагов с ростом k. Из каждой вершины i в вершины с меньшими номерами идет два ребра — в вершины $l_{i,1}$ и $l_{i,2}$. В процессе доказательства эти ребра будем рассматривать отдельно. Назовем вершину i *хорошей*, если $w_i\geqslant\dfrac{1}{10\sqrt{in}}$. Определим множества $\Gamma_0,\Gamma_1,\ldots$ следующим образом. Положим $\Gamma_0=\{v\}$. Далее, пусть для $k\geqslant 1$ множество Γ_k состоит из таких вершин $j\in[s+1,2^b]\setminus(\Gamma_0\cup\ldots\cup\Gamma_{k-1})$, что j — хорошая и либо $l_{j,2} \in \Gamma_{k-1}$, либо найдется такая вершина $i \in \Gamma_{k-1}$, что $l_{i,1} = j$. Положим также $N_k = \Gamma_0 \cup \ldots \cup \Gamma_k$.

Будем рассматривать поведение следующей величины:

$$f_k = \sum_{i \in \Gamma_k} \frac{1}{\sqrt{in}}.$$

Поскольку для всех $k \geqslant 1$ множество Γ_k состоит только из хороших вершин, вес Γ_k (то есть $\sum_{i \in \Gamma_k} w_i$) не меньше $f_k/10$. Чтобы это выполнялось и для Γ_0, положим $f_0 = \left(\ln^2 n\right)/n$.

Назовем интервал I_t *полным* на шаге $k+1$, если $|N_{k+1} \cap I_t| \geqslant 2^{t-2}$.

Воспользуемся результатом Боллобаша и Риордана (см. [7]).

Лемма 7. *Пусть* $\Gamma_0, \ldots, \Gamma_k$ *фиксированы, тогда с вероятностью* $1 - O(n^{-6/5})$ *либо какой-то из* I_t *полный на шаге* $k+1$, *либо*

$$f_{k+1} \geqslant \frac{\min\{2\log_2(f_k n/\ln n) - 31, b - a - 1\}}{1000} f_k.$$

Итак, $\Gamma_0 = \{v\}$. Строим множества $\Gamma_1, \Gamma_2, \ldots, \Gamma_K$, где K — наименьший номер k, для которого либо $f_k \geqslant 4\ln n/\sqrt{r\,n}$, либо

$$f_{k+1} < \frac{\min\{2\log_2(f_k n/\ln n) - 31, b - a - 1\}}{1000} f_k.$$

Лемма 4 говорит нам о том, что $K \leqslant \dfrac{1}{2}\dfrac{(1+\varepsilon)\ln n - \ln r}{\ln\ln n} - 1$.

Если ни на каком шаге (до K) никакой интервал не становится полным, тогда (из леммы 7) при каждом фиксированном k с вероятностью $1 - O(n^{-6/5})$ выполняется

$$f_{k+1} \geqslant \frac{\min\{2\log_2(f_k n/\ln n) - 31, b - a - 1\}}{1000} f_k.$$

Вероятность того, что это неравенство выполняется на каждом шаге до K, будет равна

$$1 - O(n^{-6/5}\ln n) = 1 - o(n^{-1})$$

(поскольку $K \leqslant \ln n$). Следовательно, $f_K \geqslant 4\ln n/\sqrt{r\,n}$ с вероятностью $1 - o(n^{-1})$.

Пусть на некотором шаге $1 \leqslant k \leqslant K$ какой-то из интервалов I_t стал полным. Тогда в множестве $N_k \backslash \{v\}$ имеем хотя бы $2^{t-2} - 1$ хорошую вершину, лежащую в I_t. Поэтому

$$f_1 + \ldots + f_K \geqslant \frac{2^{t-2} - 1}{\sqrt{2^{t+1} n}} \geqslant \frac{\sqrt{2^t}}{12\sqrt{n}} \geqslant \frac{\sqrt{s}}{12\sqrt{n}} \geqslant \frac{\ln^3 n}{12\sqrt{n}}.$$

Поскольку для всех $k < K$ выполнено $f_k < 4\ln n/\sqrt{r\,n}$, отсюда следует, что для достаточно больших n неравенство

$$f_K \geqslant \frac{\ln^3 n}{12\sqrt{n}} - \frac{4\ln^2 n}{\sqrt{r\,n}} \geqslant \frac{4\ln n}{\sqrt{r\,n}}$$

справедливо и в этом случае тоже.

Итак, для любой вершины $i \in \Gamma_K$ имеем

$$\mathrm{P}(l_{i,1} \leqslant r - 1 | \Gamma_1, \ldots, \Gamma_K) = \frac{W_{r-1}}{W_i} \geqslant \frac{9}{10}\sqrt{\frac{r-1}{n}} \cdot \frac{10}{11}\sqrt{\frac{n}{i}} \geqslant \sqrt{\frac{r}{4i}}.$$

В первом неравенстве воспользовались тем, что выполняется событие E_1. В итоге вероятность того, что нужного ребра ни из одной из вершин не идет, оценивается сверху цепочкой неравенств

$$\prod_{i \in \Gamma_K}\left(1 - \sqrt{\frac{r}{4i}}\right) \leqslant \exp\left(-\sum_{i \in \Gamma_K}\sqrt{\frac{r}{4i}}\right) = \exp\left(-\frac{f_K\sqrt{rn}}{2}\right) \leqslant \exp\left(-2\ln n\right) = n^{-2}.$$

Отсюда следует, что с вероятностью $1 - o(n^{-1})$ вершина v соединена путем длины не более $K + 1 \leqslant \frac{1}{2}\frac{(1+\varepsilon)\ln n - \ln r}{\ln\ln n}$ с одной из вершин $1, \ldots, r-1$. Лемма доказана. $\square$

Итак, у нас есть r вершин. По лемме 5 с вероятностью $1 - o(1)$ каждая вершина G соединена путем длины не более $8\ln\ln n$ с полезной вершиной. Если вдруг для каких-нибудь из наших вершин эти полезные вершины совпали, то доказывать нечего. Иначе есть r полезных вершин. По лемме 6 (в качестве ε берем $\varepsilon/2$) с вероятностью $1 - o(1)$ каждая полезная вершина соединена путем длины не более $\frac{1}{2}\frac{(1+\varepsilon/2)\ln n - \ln r}{\ln\ln n}$ с одной из вершин $1, \ldots, r-1$. Значит, для каких-то двух из наших r полезных вершин искомая вершина из $1, \ldots, r-1$ совпадет. Следовательно, с вероятностью $1 - o(1)$ среди любых r вершин найдутся две, расстояние между которыми не превосходит (для достаточно больших n)

$$16\ln\ln n + \frac{(1+\varepsilon/2)\ln n - \ln r}{\ln\ln n} \leqslant \frac{(1+\varepsilon)\ln n - \ln r}{\ln\ln n}.$$

В определении графа G мы предполагали, что события $E_1, \ldots, E_4$ выполнены. А в нашем графе G_m^n они происходят с асимптотической вероятностью 1. Таким образом, теорема 4 доказана.

Литература

1. *B. Bollobás B.* Random Graphs.— Cambridge Univ. Press, 2001.
2. *Janson S.*, *Łuczak T.*, *Ruciński A.* Random graphs.— New York: Wiley, 2000.
3. *Колчин В. Ф.* Случайные графы.— М.: Физматлит, 2002.
4. *Bollobás B.*, *Riordan O. M.* Mathematical results on scale-free random graphs // Handbook of graphs and networks. — Wiley-VCH, Weinheim, 2003.
5. *Barabási A.-L.*, *Albert R.* Emergence of scaling in random networks // Science. — 1999. — V. 286. — P. 509—512.
6. *Bollobás B.*, *Riordan O. M.* The degree sequence of a scale-free random graph process // Random Structures and Algorithms. — 2001. — V. 18, N 3. — P. 279—290.
7. *Bollobás B.*, *Riordan O. M.* The diameter of a scale-free random graph // Combinatorica. — 2004. — V. 24, N 1. — P. 5—34.
8. *Харари Ф.* Теория графов.— М.: Мир, 1973.

Поступила в редакцию 27.06.2011