Московский государственный университет имени М.В. Ломоносова
механико-математический факультет

На правах рукописи

Прохоренкова Людмила Александровна

СВОЙСТВА СЛУЧАЙНЫХ ВЕБ-ГРАФОВ, ОСНОВАННЫХ НА ПРЕДПОЧТИТЕЛЬНОМ ПРИСОЕДИНЕНИИ

01.01.05 — теория вероятностей и математическая статистика

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель
д.ф.-м.н. А.М. Райгородский

Москва, 2014

Оглавление

Список основных обозначений

$V(G)$ — множество вершин графа G;

$E(G)$ — множество ребер графа G;

$ij \in G$ — ребро ij принадлежит графу G;

$t \in G$ — вершина t принадлежит графу G;

$f(n) = o(g(n))$ — для любого числа $c > 0$ существует такое число n_0, что для любого $n > n_0$ выполнено неравенство $|f(n)| \leqslant c|g(n)|$;

$f(n) = O(g(n))$ — существуют такие числа $C > 0$ и n_0, что для любого $n > n_0$ выполнено неравенство $|f(n)| \leqslant C|g(n)|$;

$f(n) = \Theta(g(n))$ — существуют такие числа $C_1 > 0$, $C_2 > 0$ и n_0, что для любого $n > n_0$ выполнено неравенство $C_1 \cdot g(n) \leqslant f(n) \leqslant C_2 \cdot g(n)$;

$f(\cdot) = \theta(g(\cdot))$ — выполнено неравенство $|f(\cdot)| \leqslant |g(\cdot)|$;

$f(x) \sim g(x)$ — функции асимптотически равны при $x \to \infty$, то есть $f(x) = g(x) \cdot (1 + o(1))$;

$I(X)$ — индикатор события X. То есть $I(X) = 1$ в случае, если X выполняется, и 0 — иначе;

C_n^k — число сочетаний из n элементов по k;

$|A|$ — мощность множества A.

Введение

Данная работа состоит из трех глав. Первая и вторая главы посвящены, соответственно, анализу моделей Боллобаша–Риордана и Бакли–Остхуса. В третьей главе изучается новая модель случайного графа — модель предпочтительного присоединения с устареванием.

Актуальность темы

Диссертация посвящена изучению моделей сложных сетей и их свойств. Под сложными сетями понимают всевозможные сети, которые встречаются в природе: компьютерные, биологические, социальные, экономические, транспортные и так далее. Классический пример такой сети — граф сети Интернет. Вершины графа — это веб-страницы, а ребра — ссылки между ними. С появлением сети Интернет началось интенсивное изучение сложных сетей. Было замечено, что все эти сети обладают некоторыми общими свойствами: малый диаметр, степенной закон распределения степеней вершин, кластерная структура и другие [28, 32, 33, 42, 43, 44, 47, 50]. Возник естественный вопрос: почему сети столь различной природы обладают схожими свойствами? Возможно, в основе всех этих сетей лежит какой-то общий принцип формирования? Так стали появляться первые модели сложных сетей, в том числе вероятностные модели [8, 17, 18, 19, 25, 26, 30, 33, 36, 37, 38, 44, 48]. Вероятностная модель — это случайный элемент со значениями в множестве графов и некоторым вероятностным распределением на этом множестве.

В 1999 году А.-Л. Барабаши и Р. Альберт [9] предложили моделировать сложные сети с помощью так называемого принципа предпочтительного присоединения. Основная идея их подхода состоит в том, что новые страницы, появляющиеся в Интернете, скорее предпочтут сослаться на уже популярные страницы (то есть на те, на которые ведет много ссылок). Работа Барабаши и Альберт послужила толчком к развитию области вероятностного моделирования сложных сетей. Например, Р. Кумар, П. Рагхаван, С. Раджагопалан, Д. Сивакумар, А. Томкинс и Э. Упфал предложили так называемую модель копирования [39], в которой некоторые ребра, проведенные из новой вершины, "копируют" ребра одной из предшествующих страниц. Еще изучалась конфигурационная модель [1, 2, 41], которая позволяет построить граф с заданным распределением степеней вершин. Эта модель может применяться для моделирования графа Интернета, если взять подходящее распределение степеней. Обзор перечисленных и многих других моделей можно найти, например, в работах [3, 4, 5, 17, 18]. Модели сложных сетей применяются в различных областях: в математике, физике, биологии, области информационных технологий.

В данной работе проведен анализ классических моделей сложных сетей. Кроме того, предложен и проанализирован новый класс моделей, который, в отличие от других существующих моделей, отражает так называемое свойство устаревания в сложных сетях.

Характеристики сложных сетей

Наиболее распространенная характеристика сложных сетей — это распределение степеней вершин. А именно, если выбрать случайную вершину в сети и посмотреть на ее степень (количество примыкающих ребер), то получим случайную величину, распределение которой интересно изучать. Для большинства наблюдаемых сетей оказалось, что доля вершин степени d убывает как $d^{-\gamma}$ с некоторым параметром γ, который обычно лежит в интервале $(2, 3)$ [14, 17, 22, 24, 31]. Таким образом, принято говорить, что в сложных сетях распределение степеней вершин подчиняется (асимптотически) *степенному закону* [43].

В данной работе вводится обобщение понятия степени вершины — рассматривается вторая степень. Вторая степень вершины — это количество путей длины два, ведущих в данную вершину. Распределение вторых степеней представляет практический интерес, поскольку вторые степени являются хорошим приближением PageRank [12, 45, 46]. PageRank — это мера важности (популярности, качества) веб-страниц, он активно используется в информационном поиске. Основная идея PageRank заключается в том, что важность вершины определяется важностью (степенью, популярностью) ее соседей. Оказалось, что в классических моделях предпочтительного присоединения распределение вторых степеней тоже подчиняется степенному закону. Кроме того, в работе исследуются k-е входящие степени вершин.

Другой важной характеристикой сети является ее диаметр, то есть максимальная по всем парам вершин длина кратчайшего пути между двумя вершинами. Сложные сети, встречающиеся в природе, имеют малый диаметр (теория 6 рукопожатий) [10, 50]. В работе [20] Б. Боллобаш и О. Риордан доказали, что диаметр графов в классической модели предпочтительного присоединения ведет себя асимптотически как $\frac{\ln n}{\ln \ln n}$, где n — количество вершин в графе. В данной работе вводится обобщение понятия диаметра — r-*диаметр*. 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. При $r = 2$ имеем в точности обычный диаметр графа G. Для r-диаметра классических моделей предпочтительного присоединения тоже удается изучить асимптотическое поведение.

В данной работе обсуждается еще так называемое *свойство устаревания*. Свойство устаревания отражает тот факт, что в реальных сетях вершины чаще соединены с другими вершинами, близкими к ним по возрасту (времени появления). Классические модели предпочтительного присоединения не обладают свойством устаревания. В данной работе предложена и проанализирована модель, которая обладает этим свойством.

Предпочтительное присоединение

Самый распространенный способ построить модель сложных сетей — рассмотреть некоторый случайный процесс, в рамках которого в граф добавляются вершины и ребра. Обычно на шаге n добавляется новая вершина и несколько ребер, соединяющих ее с предыдущими вершинами.

Классические модели сложных сетей основаны на идее предпочтительного присоединения (preferential attachment), которая была предложена в 1999 году А.-Л. Барабаши и Р. Альберт [14, 15]. Идея состоит в том, что в Интернете новые страницы предпочитают цитировать те страницы, которые в настоящий момент более популярны. С помощью идеи предпочтительного присоединения удалось объяснить малый диаметр Интернета, а также степенной закон распределения степеней вершин в нем.

В 2001 году Б. Боллобаш и О. Риордан формализовали идею, предложенную Барабаши и Альберт [21]. На каждом шаге в граф добавляется вершина и m ребер. Ребра добавляются по очереди, и вероятность того, что ребро будет проведено в какую-то предыдущую вершину i, пропорциональна степени этой вершины. Именно анализу этой модели посвящена первая глава данной работы. Однако модель Боллобаша и Риордана оказалась недостаточно гибкой: она позволяет получить граф со степенным распределением степеней вершин с параметром 3, то есть количество вершин степени d в этой модели убывает как d^{-3}. Реальные сети, в свою очередь, могут обладать различными параметрами степенного распределения степеней вершин, которые обычно лежат в интервале $(2, 3)$. Эта проблема решается в модели, предложенной П. Бакли и Д. Остхусом [23]. Бакли и Остхус добавили в модель еще один параметр — начальную привлекательность вершины (константа, не зависящая от степени вершины). Этот параметр делает модель более гибкой и позволяет получить более широкий класс распределений степеней вершин. Анализу модели Бакли и Остхуса посвящена вторая глава данной работы.

Благодарности. Автор признателен профессору Андрею Михайловичу Райгородскому за неоценимую помощь в работе.

Глава 1

Анализ модели Боллобаша–Риордана

Наиболее известная модель сложных сетей — это модель Боллобаша и Риордана [21], которые формализовали идею, предложенную Барабаши и Альберт в работе [14]. Некоторые свойства модели Боллобаша–Риордана будут обсуждаться в этой главе. А именно, в разделе 1.1 дано определение модели и перечислены основные классические результаты; в разделе 1.2 обсуждается распределение вторых степеней в модели; k-е степени проанализированы в разделе 1.3; раздел 1.4 посвящен изучению r-диаметра.

Данная глава основана на работах автора [52, 53, 56].

1.1 Определение модели и известные результаты

Опишем, как строится граф в модели Боллобаша и Риордана. Обычно для графов этой модели используется обозначение G_m^n, где n — число вершин, m — фиксированное натуральное число. Сначала индуктивно строится G_1^n. Граф G_1^1 состоит из одной вершины и одной петли. Граф G_1^t получается из G_1^{t-1} посредством добавления вершины t и одного ребра между вершинами t и i, где i выбирается случайно следующим образом:

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

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

Тот факт, что граф G_m^n обладает степенным распределением степеней вершин, был доказан в работе [21].

Теорема 1. *Пусть $m \geqslant 1$. Тогда существует такая функция φ аргумента n, что $\varphi(n) = o(n)$ при $n \to \infty$, и для любой такой степени d, что $m \leqslant d \leqslant n^{1/15}$, мы имеем*

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

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

В 2011 году в работе [34] Е.А. Гречников значительно улучшил теорему 1.

1.2 Распределение вторых степеней вершин

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

1.2.1 Введение обозначений и формулировка результатов

Мы рассматриваем граф G_1^n. *Второй степенью* вершины $t \in G_1^n$ назовем

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

Другими словами, вторая степень вершины t — это количество ребер, соединенных с соседями вершины t, кроме тех, которые ведут в саму вершину t.

Через $N_n(d)$ обозначим количество вершин степени d в графе G_1^n, а через $X_n(d)$ — количество вершин второй степени d в G_1^n.

В данном разделе доказываются следующие новые результаты.

Теорема 2. *Для любого $k > 1$ имеем*

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

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

$$\lim_{n\to\infty} \mathsf{P}\left(|X_n(k) - \mathrm{E}X_n(k)| \geqslant \frac{\varphi(n)}{k^2}\right) = 0$$

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

Эта теорема показывает, что для любого k, $1 \leqslant k \leqslant n^{1/6-\varepsilon}$, с вероятностью, стремящейся к 1, количество вершин второй степени k отличается от своего математического ожидания на что-то малое в сравнении с математическим ожиданием. Таким образом, получена концентрация. Итак, распределение вторых степеней тоже подчиняется (асимптотически) степенному закону.

Для доказательства теоремы 2 нам потребуется следующее определение. Через $N_n(l,k)$ обозначим количество вершин в графе G_1^n без петель, со степенью l и со второй степенью k:

$$N_n(l,k) = \left|\{t \in G_1^n : d(t) = l, d_2(t) = k, tt \notin G_1^n\}\right|.$$

Докажем следующую вспомогательную теорему.

Теорема 4. *В графе* G_1^n

$$\mathrm{E}N_n(l,k) = n\,c(l,k)\,(1+\theta(n,l,k)),$$

где $|\theta(n,l,k)| < (2l+k-1)^2/n$. *Константы* $c(l,k)$ *определяются следующим образом:*

$$\begin{aligned} c(l,0) &= c(0,k) = 0, \quad l \geqslant 0, k \geqslant 0,\\ c(1,k) &= \frac{2k^2+14k}{(k+1)(k+2)(k+3)(k+4)}, \quad k \geqslant 1,\\ c(l,k) &= c(l,k-1)\frac{l+k-1}{2l+k+2} + c(l-1,k)\frac{l-1}{2l+k+2}, \quad l \geqslant 2, k \geqslant 1. \end{aligned}$$

Чтобы доказать теоремы, мы будем использовать следующие леммы.

Лемма 1. *Пусть* $d \geqslant 1$ *— натуральное число. Тогда*

$$\mathrm{E}N_n(d) = \frac{4n}{d(d+1)(d+2)}\left(1+\tilde{\theta}(n,d)\right),$$

где $|\tilde{\theta}(n,d)| < d^2/n$.

Через $P_n(l,k)$ обозначим количество вершин в графе G_1^n с петлей, со степенью l и со второй степенью k.

Лемма 2. *Для любого* n *имеем*

$$\mathrm{E}P_n(l,k) \leqslant p(l,k),$$

где

$$\begin{aligned} p(2,0) &= 1,\\ p(l,k) &= p(l,k-1)\frac{l+k-3}{2l+k-2} + p(l-1,k)\frac{l-1}{2l+k-2}, \quad l \geqslant 3, k \geqslant 0. \end{aligned}$$

Для всех других значений l *и* k *имеем* $p(l,k) = 0$.

В следующем параграфе приведены доказательства. Сначала будут доказаны теорема 4 и теорема 2; затем будут доказаны леммы. После этого будет доказана теорема 3 о концентрации.

1.2.2 Доказательства теорем

Доказательство теоремы 4

Из определения графа G_1^n следует, что $N_n(l,0) = N_n(0,k) = 0$. Действительно, в графе нет вершин степени 0, поэтому $N_n(0,k) = 0$. Поскольку вершины с петлями не входят в $N_n(l,k)$, то в $N_n(l,k)$ не учитываются вершины второй степени 0 и, следовательно, $N_n(l,0) = 0$. В итоге $\mathrm{E}N_n(l,0) = \mathrm{E}N_n(0,k) = 0$.

Докажем, что $\mathrm{E}N_n(1,k) = n\,c(1,k)\,(1+\theta(n,1,k))$. Докажем это утверждение индукцией по k. Для $k = 0$ утверждение верно. Предположим, что для всех $j < k$

$$\mathrm{E}N_n(1,j) = n\,c(1,j)\,(1+\theta(n,1,j)),$$

где

$$|\theta(n,1,j)| < (j+1)^2/n,$$

$$c(1,j) = \frac{2j^2+14j}{(j+1)(j+2)(j+3)(j+4)}.$$

Очевидно, $\mathrm{E}N_1(1,k) = 0$. Для $i \geqslant 1$ имеем

$$\mathrm{E}(N_{i+1}(1,k)|N_i(1,k), N_i(1,k-1), N_i(k)) =$$
$$= N_i(1,k)\left(1 - \frac{k+2}{2i+1}\right) + \frac{kN_i(1,k-1)}{2i+1} + \frac{k\,N_i(k)}{2i+1}. \quad (1.1)$$

Поясним это равенство. Предположим, что у нас имеется некоторый граф G_1^i. Мы добавляем к нему одну вершину и одно ребро. В графе G_1^i есть $N_i(1,k)$ вершин степени 1 и второй степени k. Вероятность "испортить" одну из таких вершин равна $(k+2)/(2i+1)$. Также есть $N_i(1,k-1)$ вершин степени 1 и второй степени $k-1$. Вероятность того, что одна из таких вершин будет иметь степень 1 и вторую степень k в G_1^{i+1}, равна $k/(2i+1)$. Наконец, с вероятностью $kN_i(k)/(2i+1)$ вершина $i+1$ сама будет иметь нужную степень в графе G_1^{i+1}.

Используя равенство (1.1), лемму 1 и предположение индукции, получаем

$$\mathrm{E}N_{i+1}(1,k) = \mathrm{E}N_i(1,k)\frac{2i-k-1}{2i+1} + \frac{k\mathrm{E}N_i(1,k-1)}{2i+1} + \frac{k\,\mathrm{E}N_i(k)}{2i+1} =$$
$$= \mathrm{E}N_i(1,k)\frac{2i-k-1}{2i+1} +$$
$$+ \left(\frac{i\,k\,c(1,k-1)}{2i+1} + \frac{4\,i}{(2i+1)(k+1)(k+2)}\right)\left(1+\theta\left(k^2/i\right)\right).$$

Введем следующие обозначения:

$$a_i = \frac{2i-k-1}{2i+1},$$

$$b_i = \frac{2i}{2i+1}\left(1+\theta\left(k^2/i\right)\right),$$
$$m = \frac{c(1,k-1)\,k}{2} + \frac{2}{(k+1)(k+2)}.$$

Используя эти обозначения, получим

$$\mathrm{E}N_{i+1}(1,k) = \mathrm{E}N_i(1,k)\,a_i + m\,b_i.$$

Докажем следующее равенство индукцией по n:

$$\mathrm{E}N_n(1,k) = \frac{2mn}{k+4}\left(1+\theta(n,1,k)\right).$$

Для $n=1$ имеем $\mathrm{E}N_1(1,k)=0$. Поскольку у нас есть условие $|\theta(1,1,k)| < (k+1)^2$, мы можем взять $\theta(1,1,k)=-1$.

Теперь положим $t=k+1$ (это будет использоваться позже). Предположим, что

$$\mathrm{E}N_i(1,k) = \frac{2mi}{t+3}\left(1+\theta(i,1,t-1)\right).$$

Тогда

$$\mathrm{E}N_{i+1}(1,k) = \mathrm{E}N_i(1,k)\,a_i + m\,b_i =$$
$$= \frac{2mi(2i-t)}{(2i+1)(t+3)}\left(1+\theta(t^2/i)\right) + \frac{2mi}{2i+1}\left(1+\theta\left((t-1)^2/i\right)\right) =$$
$$= \frac{2m}{t+3}\left(i+1-\frac{1}{2i+1}+\theta\left(\frac{(2i-t)t^2}{2i+1}\right)+\theta\left(\frac{(t-1)^2(t+3)}{2i+1}\right)\right).$$

Если $t \geqslant 1$ и $2i-t \geqslant 0$, то

$$\frac{1}{2i+1} + \frac{t^2|2i-t|}{2i+1} + \frac{(t-1)^2(t+3)}{2i+1} < t^2.$$

Поэтому

$$\mathrm{E}N_{i+1}(1,k) = \frac{2m(i+1)}{t+3}\left(1+\theta\left(t^2/(i+1)\right)\right).$$

Что и требовалось показать.

Если $t \geqslant 1$ и $2i-t \leqslant -2$, то у нас нет достаточного количества ребер в графе и $\mathrm{E}N_{i+1}(l,k)=0$. В этом случае $\theta(i+1,l,k)=-1$.

Случай $2i-t=-1$ будет рассмотрен позже.

Получили

$$\mathrm{E}N_n(1,k) = \frac{2mn}{k+4}\left(1+\theta(n,1,k)\right).$$

Заметим, что

$$\frac{2m}{k+4} = \frac{4}{(k+1)(k+2)(k+4)} + \frac{2\,c(1,k-1)\,k}{2(k+4)} =$$

$$= \frac{4}{(k+1)(k+2)(k+4)} + \frac{2(k-1)^2 + 14(k-1)}{(k+1)(k+2)(k+3)(k+4)} =$$

$$= \frac{2k^2 + 14k}{(k+1)(k+2)(k+3)(k+4)} = c(1,k).$$

Это завершает доказательство для $\mathrm{E}N_n(1,k)$.

Рассмотрим случай $l \geqslant 2$. Допустим, мы уже доказали, что $\mathrm{E}N_n(i,j) = n\, c(i,j)\,(1+\theta(n,i,j))$ для всех i и j таких, что $i < l$, $j \leqslant k$ или $i \leqslant l$, $j < k$. Положим $t = 2l + k - 1$. Очевидно, $\mathrm{E}N_1(l,k) = 0$. Для $i \geqslant 1$ имеем

$$\mathrm{E}N_{i+1}(l,k) = \mathrm{E}N_i(l,k)\left(1 - \frac{2l+k}{2i+1}\right) +$$

$$+ \frac{(l-1)\,\mathrm{E}N_i(l-1,k)}{2i+1} + \frac{(l+k-1)\mathrm{E}N_i(l,k-1)}{2i+1} =$$

$$= \mathrm{E}N_i(l,k)\frac{2i-t}{2i+1} +$$

$$+ \left(\frac{(l-1)c(l-1,k)i}{2i+1} + \frac{(l+k-1)c(l,k-1)i}{2i+1}\right)\left(1 + \theta\left((t-1)^2/i\right)\right).$$

Введем обозначения:

$$a_i = \frac{2i-t}{2i+1},$$

$$b_i = \frac{2i}{2i+1}\left(1 + \theta\left((t-1)^2/i\right)\right),$$

$$m = \frac{(l-1)\,c(l-1,k)}{2} + \frac{(l+k-1)\,c(l,k-1)}{2}.$$

Имеем

$$\mathrm{E}N_{i+1}(l,k) = \mathrm{E}N_i(l,k)\,a_i + m\,b_i.$$

Остается доказать индукцией по n следующее утверждение:

$$\mathrm{E}N_n(l,k) = \frac{2mn}{t+3}\left(1 + \theta\left(t^2/n\right)\right) = \frac{2mn}{t+3}\left(1 + \theta(n,l,k)\right).$$

Доказательство похоже на доказательство в случае $l = 1$. В данном случае

$$\frac{2m}{t+3} = \frac{(l-1)\,c(l-1,k)}{2l+k+2} + \frac{(l+k-1)\,c(l,k-1)}{2l+k+2} = c(l,k).$$

Теперь нам осталось рассмотреть только случай $2i - t = -1$. Требуется показать, что $\mathrm{E}N_{i+1}(l,k) = (i+1)c(l,k)(1+\theta(i+1,l,k))$. Имеем $2(i+1) = 2l+k$. В графе G_1^{i+1} есть $i+1$ ребро. Поэтому сумма всех степеней равна $2l+k$. Предположим, что есть хотя бы одна вершина степени l и второй степени k. Мы не считаем вершины с петлями в $N_{i+1}(l,k)$, поэтому l ребер выходят из этой вершины. Есть еще $k/2$ ребер между соседями этой вершины, а больше ребер нет. Поэтому наша вершина соединена со всеми остальными вершинами в G_1^{i+1}. Следовательно, $l = i$. Таким образом, $k = 2$. Из этого следует, что мы рассматриваем вершину 2. И есть одно ребро из вершины 2 в вершину 1, а еще есть ребра из вершин $3, \ldots, i+1$ в вершину 2. Таким образом, существует всего один граф с $N_{i+1}(l, k) \neq 0$. И в этом графе есть ровно одна вершина со степенью l и второй степенью k. Поэтому вероятность этого графа равна $\mathrm{E}N_{l+1}(l, 2)$. Получили $\mathrm{E}N_{l+1}(l, 2) = \frac{2(l-1)!}{(2l+1)!!}$.

Напомним, что $l = i$ и $k = 2$. Теперь нам нужно доказать, что

$$\mathrm{E}N_{l+1}(l, 2) = (l+1)c(l, 2)(1 + \theta(l+1, l, 2)).$$

Докажем неравенство

$$c(l, 2) \geqslant \frac{24(l-1)!}{5(2l+4)!!}.$$

Из определения $c(l, k)$ следует, что

$$c(1, 2) = \frac{1}{10},$$

$$c(l, 2) \geqslant c(l-1, 2)\frac{l-1}{2l+4}, \ l \geqslant 2.$$

Очевидно, $\theta(l+1, l, 2) \geqslant -1$. Получим следующую верхнюю оценку:

$$\theta(l+1, l, 2) + 1 = \frac{\mathrm{E}N_{l+1}(l, 2)}{(l+1)c(l, 2)} \leqslant \frac{2(l-1)!5(2l+4)!!}{(2l+1)!!(l+1)24(l-1)!} =$$

$$= \frac{5(2l+4)!!}{12(2l+1)!!(l+1)} \leqslant \frac{(2l+1)^2}{(l+1)}.$$

Это завершает доказательство теоремы 4.

Доказательство теоремы 2

В теореме 4 были получены константы $c(l, k)$. Представим, что есть таблица со значениями $c(l, k)$, где l — номер строки, а k — номер столбца. Сумма всех чисел в таблице равна 1. Сумма чисел в l-ой строке равна $\frac{4}{l(l+1)(l+2)}$. Это можно легко проверить, используя определение $c(l, k)$. Но требуется оценить $\mathrm{E}X_n(k)$, поэтому нас интересует сумма чисел в k-ом столбце. А точнее,

$$\mathrm{E}X_n(k) = \sum_{l=1}^{\infty} \mathrm{E}N_n(l, k) + \sum_{l=1}^{\infty} \mathrm{E}P_n(l, k).$$

Для начала оценим $\sum_{l=1}^{\infty} c(l,k)$. Напомним, что

$$\begin{aligned} c(l,0) &= 0, \\ c(1,k) &= \frac{2k^2+14k}{(k+1)(k+2)(k+3)(k+4)}, \quad k \geqslant 1, \\ c(l,k) &= c(l,k-1)\frac{l+k-1}{2l+k+2} + c(l-1,k)\frac{l-1}{2l+k+2}, \quad k \geqslant 1, l \geqslant 2. \end{aligned}$$

Заметим, что существует функция $C(k) \geqslant 0$ такая, что для всех $l \geqslant k \geqslant 0$ и $l \geqslant 1$ неравенство

$$c(l,k) \leqslant C(k) 2^{-l} \frac{(l-1)!}{(l-k)!} \tag{1.2}$$

верно. Действительно, случай $k=0$ очевиден с $C(k)=0$. Для $k \geqslant 1$ определим $C(k)$ так, что $C(k) \geqslant C(k-1)$ и (1.2) выполнено для $l=k$. Имеем

$$(2l+k+2)\frac{c(l,k)}{C(k)} \leqslant \frac{C(k-1)}{C(k)}(l+k-1)2^{-l}\frac{(l-1)!}{(l-k+1)!} + 2^{-l+1}\frac{(l-1)!}{(l-k-1)!} \leqslant$$

$$\leqslant 2^{-l}\frac{(l-1)!}{(l-k)!}\left(\frac{l+k-1}{l-k+1} + 2(l-k)\right) \leqslant (2l+k+2)2^{-l}\frac{(l-1)!}{(l-k)!}.$$

Это доказывает (1.2).

В частности, ряд $\sum_{l=1}^{\infty} l^N c(l,k)$ сходится для всех N и k.

Сделаем некоторые преобразования:

$$(2l+k+2)c(l,k) = (l+k-1)c(l,k-1) + (l-1)c(l-1,k),$$

$$\sum_{l=2}^{\infty}(2l+k+2)c(l,k) = \sum_{l=2}^{\infty}(l+k-1)c(l,k-1) + \sum_{l=1}^{\infty} lc(l,k),$$

$$\sum_{l=2}^{\infty}(l+k+2)c(l,k) = \sum_{l=2}^{\infty}(l+k-1)c(l,k-1) + c(1,k).$$

Положим $x_k = \sum_{l=2}^{\infty} c(l,k)$. Тогда $x_0 = 0$ и для $k \geqslant 1$ имеем

$$(k+2)x_k = (k-1)x_{k-1} + c(1,k) + \sum_{l=2}^{\infty} l(c(l,k-1) - c(l,k)),$$

$$(k+2)(k+1)kx_k = (k-1)(k+1)kx_{k-1} + (k+1)kc(1,k) +$$

$$+\sum_{l=2}^{\infty} l(k(k+1)c(l,k-1) - k(k+1)c(l,k)),$$

$$(k+2)(k+1)kx_k = \sum_{s=1}^{k}(s(s+1)(s+2)x_s - (s-1)s(s+1)x_{s-1}) =$$

$$= \sum_{s=1}^{k} s(s+1)c(1,s) + \sum_{l=2}^{\infty} l \left(\sum_{s=1}^{k} (s(s+1)c(l,s-1) - s(s+1)c(l,s)) \right) =$$

$$= \sum_{s=1}^{k} s(s+1)c(1,s) +$$

$$+ \sum_{l=2}^{\infty} l \left(\sum_{s=1}^{k} ((s+1)(s+2) - s(s+1))c(l,s) - (k+1)(k+2)c(l,k) \right) =$$

$$= \sum_{s=1}^{k} s(s+1)c(1,s) + \sum_{l=2}^{\infty} l \left(\sum_{s=1}^{k} 2(s+1)c(l,s) - (k+1)(k+2)c(l,k) \right),$$

$$x_k = \frac{1}{k(k+1)(k+2)} \sum_{s=1}^{k} s(s+1)c(1,s) +$$
$$+ \frac{2}{k(k+1)(k+2)} \sum_{l=2}^{\infty} l \left(\sum_{s=1}^{k} (s+1)c(l,s) \right) - \frac{1}{k} \sum_{l=2}^{\infty} lc(l,k). \quad (1.3)$$

Положим $y_k = \sum_{l=2}^{\infty} lc(l,k)$. Тогда

$$x_k = \frac{1}{k(k+1)(k+2)} \sum_{s=1}^{k} s(s+1)c(1,s) + \frac{2}{k(k+1)(k+2)} \sum_{s=1}^{k} (s+1)y_s - \frac{1}{k} y_k.$$

Сделаем некоторые преобразования:

$$(2l+k+2)lc(l,k) = (l+k-1)lc(l,k-1) + l(l-1)c(l-1,k),$$

$$\sum_{l=2}^{\infty} (2l+k+2)lc(l,k) = \sum_{l=2}^{\infty} (l+k-1)lc(l,k-1) + \sum_{l=1}^{\infty} l(l+1)c(l,k),$$

$$\sum_{l=2}^{\infty} (l+k+1)lc(l,k) = \sum_{l=2}^{\infty} (l+k-1)lc(l,k-1) + 2c(1,k),$$

$$ky_k + \sum_{l=2}^{\infty} (l+1)lc(l,k) = (k-2)y_{k-1} + \sum_{l=2}^{\infty} l(l+1)c(l,k-1) + 2c(1,k),$$

$$k(k-1)y_k = \sum_{s=1}^{k} (s(s-1)y_s - (s-1)(s-2)y_{s-1}) =$$

$$= \sum_{s=1}^{k} \left((s-1) \sum_{l=2}^{\infty} l(l+1)c(l,s-1) - \right.$$

$$(s-1)\sum_{l=2}^{\infty}(l+1)lc(l,s)+2(s-1)c(1,s)\Bigg)=$$

$$=2\sum_{s=1}^{k}(s-1)c(1,s)+\sum_{l=2}^{\infty}l(l+1)\sum_{s=1}^{k}c(l,s)-k\sum_{l=2}^{\infty}l(l+1)c(l,k).$$

Для $k\geqslant 2$ имеем

$$y_k=\frac{2}{k(k-1)}\sum_{s=1}^{k}(s-1)c(1,s)+\frac{1}{k(k-1)}\sum_{l=2}^{\infty}l(l+1)\sum_{s=1}^{k}c(l,s)-$$

$$-\frac{1}{k-1}\sum_{l=2}^{\infty}l(l+1)c(l,k).$$

Положим $z_k=\sum_{l=2}^{\infty}l(l+1)c(l,k)$. Тогда для $k\geqslant 2$

$$y_k=\frac{2}{k(k-1)}\sum_{s=1}^{k}(s-1)c(1,s)+\frac{1}{k(k-1)}\sum_{s=1}^{k}z_s-\frac{1}{k-1}z_k.$$

Сделаем похожие преобразования

$$(2l+k+2)l(l+1)c(l,k)=(l+k-1)l(l+1)c(l,k-1)+(l+1)l(l-1)c(l-1,k),$$

$$\sum_{l=2}^{\infty}(2l+k+2)l(l+1)c(l,k)=\sum_{l=2}^{\infty}(l+k-1)l(l+1)c(l,k-1)+\sum_{l=1}^{\infty}l(l+1)(l+2)c(l,k),$$

$$\sum_{l=2}^{\infty}(l+k)l(l+1)c(l,k)=\sum_{l=2}^{\infty}(l+k-1)l(l+1)c(l,k-1)+6c(1,k),$$

$$\sum_{s=1}^{k}\sum_{l=2}^{\infty}(l+s)l(l+1)c(l,s)=\sum_{s=0}^{k-1}\sum_{l=2}^{\infty}(l+s)l(l+1)c(l,s)+\sum_{s=1}^{k}6c(1,s),$$

$$\sum_{l=2}^{\infty}(l+k)l(l+1)c(l,k)=\sum_{s=1}^{k}6c(1,s).$$

Поскольку $c(1,s)=O\left(\frac{1}{s^2}\right)$, имеем

$$0\leqslant z_k\leqslant\frac{1}{k}\sum_{l=2}^{\infty}(l+k)l(l+1)c(l,k)=O\left(\frac{1}{k}\sum_{s=1}^{k}\frac{1}{s^2}\right)=O\left(\frac{1}{k}\right),$$

$$\sum_{s=1}^{k}(s-1)c(1,s)=O\left(\sum_{s=1}^{k}\frac{1}{s}\right)=O(\ln k),$$

$$y_k = O\left(\frac{\ln k}{k^2}\right),$$

$$\sum_{s=1}^{k}(s+1)y_s = O\left(\sum_{s=1}^{k}\frac{\ln s}{s}\right) = O(\ln^2 k),$$

$$x_k = \frac{1}{k(k+1)(k+2)}\sum_{s=1}^{k} s(s+1)c(1,s) + O\left(\frac{\ln^2 k}{k^3}\right).$$

Наконец, $c(1,s) = \frac{2}{s(s+1)} + O\left(\frac{1}{s^3}\right)$, поэтому $\sum_{s=1}^{k} s(s+1)c(1,s) = 2k + O(\ln k)$ и

$$x_k = \frac{2}{(k+1)(k+2)} + O\left(\frac{\ln^2 k}{k^3}\right) = \frac{2}{k^2} + O\left(\frac{\ln^2 k}{k^3}\right),$$

$$\sum_{l=1}^{\infty} c(l,k) = c(1,k) + x_k = \frac{4}{k^2} + O\left(\frac{\ln^2 k}{k^3}\right).$$

Теперь мы можем оценить $\mathrm{E}X_n(k)$:

$$\mathrm{E}X_n(k) = \sum_{l=1}^{\infty} c(l,k)\, n(1+\theta(n,l,k)) + \sum_{l=1}^{\infty}\mathrm{E}P_n(l,k).$$

Первая сумма:

$$\sum_{l=1}^{\infty} c(l,k)\, n = \frac{4n}{k^2} + O\left(\frac{n\ln^2 k}{k^3}\right).$$

Вторая сумма:

$$\sum_{l=1}^{\infty} c(l,k)\, n\, |\theta(n,l,k)| \leqslant \sum_{l=1}^{\infty} c(l,k)\,(2l+k)^2 =$$

$$= \sum_{l=1}^{\infty} 4l^2 c(l,k) + \sum_{l=1}^{\infty} 4lkc(l,k) + \sum_{l=1}^{\infty} k^2 c(l,k) =$$

$$= 4c(1,k) + \sum_{l=2}^{\infty} 4l(l+1)c(l,k) - \sum_{l=2}^{\infty} 4lc(l,k) +$$

$$+4kc(1,k) + \sum_{l=2}^{\infty} 4lkc(l,k) + \sum_{l=1}^{\infty} k^2 c(l,k) =$$

$$= (4+4k)c(1,k) + 4z_k + (4k-4)y_k + \sum_{l=1}^{\infty} k^2 c(l,k) =$$

$$= O\left(\frac{1}{k} + \frac{1}{k} + \frac{\ln k}{k} + \frac{\ln^2 k}{k} + 1\right) = O\left(1\right).$$

Третья сумма:
$$\sum_{l=1}^{\infty} \mathrm{E} P_n(l,k) \leqslant \sum_{l=1}^{\infty} p(l,k).$$
Напомним, что
$$\begin{aligned} p(2,0) &= 1, \\ p(l,k) &= p(l,k-1)\frac{l+k-3}{2l+k-2} + p(l-1,k)\frac{l-1}{2l+k-2}, \quad k \geqslant 0, l \geqslant 3. \end{aligned}$$
Для других l и k выполнено $p(l,k) = 0$. Мы можем оценить $p(l,k)$ следующим образом:
$$p(l,k) \leqslant \frac{6}{l(l+1)}.$$
Действительно, несложно проверить, что функция $\frac{6}{l(l+1)}$ удовлетворяет рекуррентному соотношению. Поэтому для $l = 2$ и $k = 0$ мы используем неравенство $p(l,k) = 1 \leqslant \frac{6}{l(l+1)}$, а затем продолжаем по индукции. Таким образом, ряд $\sum_{l=2}^{\infty} p(l,k)$ сходится. Другими словами, $\sum_{l=2}^{\infty} p(l,k) = O(1)$. Поэтому
$$\mathrm{E} X_n(k) = \frac{4n}{k^2} + O\left(\frac{n \ln^2 k}{k^3}\right) + O(1) + O(1) = \frac{4n}{k^2}\left(1 + O\left(\frac{\ln^2 k}{k}\right) + O\left(\frac{k^2}{n}\right)\right).$$
Это завершает доказательство теоремы 2.

Осталось доказать леммы 1 и 2.

Доказательство леммы 1

В работе [21] Боллобаш и Риодан вычислили математическое ожидание количества вершин степени d. Однако они рассматривали только вершины степени $d \leqslant n^{1/15}$ и доказали, что
$$\mathrm{E} N_n(d) \sim \frac{4n}{d(d+1)(d+2)}.$$
А нам требуется оценка $\mathrm{E} N_n(d)$ для всех d. Кроме того, требуется оценить остаточный член. Именно поэтому в этой работе заново оценивается $\mathrm{E} N_n(d)$.

Доказательство проводится индукцией по d. Сначала отдельно рассмотрим два случая: $d = 1$ и $d = 2$.

Рассмотрим случай $d = 1$. Очевидно, $\mathrm{E} N_0(1) = 0$. Пусть
$$\mathrm{E} N_i(1) = \frac{2i}{3}\left(1 + \tilde{\theta}(i,1)\right).$$
Тогда
$$\mathrm{E} N_{i+1}(1) = \mathrm{E} N_i(1)\left(1 - \frac{1}{2i+1}\right) + \frac{2i}{2i+1} = \frac{2i}{3}\left(1 + \tilde{\theta}(i,1)\right)\frac{2i}{2i+1} + \frac{2i}{2i+1} =$$

$$= \frac{2}{3}\left(i+1-\frac{1}{2i+1}+\frac{2i^2}{2i+1}\tilde\theta(i,1)\right)=$$

$$= \frac{2(i+1)}{3}\left(1-\frac{1}{(2i+1)(i+1)}+\frac{2i^2}{(2i+1)(i+1)}\tilde\theta(i,1)\right). \qquad (1.4)$$

Положим $\tilde\theta(i+1,1)=\frac{2i^2}{(2i+1)(i+1)}\tilde\theta(i,1)-\frac{1}{(2i+1)(i+1)}$. Заметим, что

$$|\tilde\theta(i+1,1)|\leqslant\frac{2i}{(2i+1)(i+1)}+\frac{1}{(2i+1)(i+1)}\leqslant 1/(i+1).$$

Это завершает доказательство для $d=1$.

Случай $d=2$ немного отличается от $d=1$. Очевидно, $\mathrm{E}N_0(2)=0$. Пусть $\mathrm{E}N_i(2)=\frac{i}{6}\left(1+\tilde\theta(i,2)\right)$. Тогда

$$\mathrm{E}N_{i+1}(2)=\mathrm{E}N_i(2)\left(1-\frac{2}{2i+1}\right)+\mathrm{E}N_i(1)\frac{1}{2i+1}+\frac{1}{2i+1}=$$

$$=\frac{i}{6}\left(1+\tilde\theta(i,2)\right)\frac{2i-1}{2i+1}+\frac{2i}{3(2i+1)}\left(1+\tilde\theta(i,1)\right)+\frac{1}{2i+1}=$$

$$=\frac{1}{6}\left(i+1+\frac{5}{2i+1}+\frac{(2i-1)i}{2i+1}\tilde\theta(i,2)+\frac{4i}{2i+1}\tilde\theta(i,1)\right)=$$

$$=\frac{i+1}{6}\left(1+\frac{5}{(2i+1)(i+1)}+\frac{(2i-1)i}{(2i+1)(i+1)}\tilde\theta(i,2)+\frac{4i}{(2i+1)(i+1)}\tilde\theta(i,1)\right).$$

Положим

$$\tilde\theta(i+1,2)=\frac{5}{(2i+1)(i+1)}+\frac{(2i-1)i}{(2i+1)(i+1)}\tilde\theta(i,2)+\frac{4i}{(2i+1)(i+1)}\tilde\theta(i,1).$$

Заметим, что $\tilde\theta(i,1)<0$. Поэтому

$$|\tilde\theta(i+1,2)|\leqslant\left|\frac{(2i-1)i}{(2i+1)(i+1)}\theta(4/i)\right|+$$

$$+\max\left\{\left|\frac{5}{(2i+1)(i+1)}\right|,\left|\frac{4i}{(2i+1)(i+1)}\theta(1/i)\right|\right\}\leqslant\frac{4}{i+1}.$$

Это завершает доказательство для $d=2$.

Пусть $d\geqslant 3$ и мы уже доказали теорему для всех меньших степеней. Будем использовать индукцию по i. Для $i=0$ имеем $\mathrm{E}N_0(d)=0$. Пусть

$$\mathrm{E}N_i(d)=\frac{4i}{d(d+1)(d+2)}\left(1+\tilde\theta(i,d)\right).$$

Тогда

$$\mathrm{E}N_{i+1}(d)=\mathrm{E}N_i(d)\left(1-\frac{d}{2i+1}\right)+\mathrm{E}N_i(d-1)\frac{d-1}{2i+1}=$$

$$= \frac{4i(2i+1-d)}{d(d+1)(d+2)(2i+1)}\left(1+\tilde\theta(i,d)\right) + \frac{4i}{d(d+1)(2i+1)}\left(1+\tilde\theta(i,d-1)\right) =$$

$$= \frac{4(i+1)}{d(d+1)(d+2)}\left(1 - \frac{1}{(2i+1)(i+1)} + \right.$$

$$\left. + \frac{i(2i+1-d)}{(2i+1)(i+1)}\tilde\theta(i,d) + \frac{i(d+2)}{(2i+1)(i+1)}\tilde\theta(i,d-1)\right).$$

Если $2i+1-d \geqslant 0$, то можем взять

$$\tilde\theta(i+1,d) = -\frac{1}{(2i+1)(i+1)} + \frac{i(2i+1-d)}{(2i+1)(i+1)}\tilde\theta(i,d) + \frac{i(d+2)}{(2i+1)(i+1)}\tilde\theta(i,d-1).$$

Получили следующую оценку:

$$|\tilde\theta(i+1,d)| \leqslant \frac{1}{(2i+1)(i+1)} + \frac{i(2i+1-d)}{(2i+1)(i+1)}|\tilde\theta(i,d)| +$$

$$+ \frac{i(d+2)}{(2i+1)(i+1)}|\tilde\theta(i,d-1)| \leqslant \frac{d^2}{i+1}.$$

Если $2i+2-d \leqslant 0$, то вершин степени d нет в графе G_1^{i+1}. Действительно, в G_1^{i+1} сумма всех степеней равна $2i+2$. Если $d > 2i+2$, то ребер очевидно не хватает. Если $d = 2i+2$, то легко проверить, что не может быть вершин степени d ($d > 2$). Поэтому можем взять $\tilde\theta(i+1,d) = -1$. Это завершает доказательство.

Доказательство леммы 2

Очевидно, $\mathrm{E}P_n(0,k) = \mathrm{E}P_n(1,k) = 0$. Для всех $k > 0$ имеем $\mathrm{E}P_n(2,k) = 0$. Для $k = 0$ имеем

$$\mathrm{E}P_n(2,0) = \sum_{i=1}^{n} \frac{1}{2i-1} \prod_{j=i+1}^{n} \frac{2j-3}{2j-1} = \sum_{i=1}^{n} \frac{1}{2n-1} = \frac{n}{2n-1} \leqslant 1.$$

Далее доказываем по индукции. Рассмотрим $l \geqslant 3$, $k \geqslant 0$. Пусть уже доказано, что $\mathrm{E}P_n(i,j) \leqslant p(i,j)$ для всех i и j таких, что $i < l$, $j \leqslant k$ или $i \leqslant l$, $j < k$.

Очевидно, $P_1(l,k) = 0$. Легко показать, что $\mathrm{E}P_{i+1}(l,k) = 0$ если $2i+4 < 2l+k$.

Если $2i+4 = 2l+k$ и $P_{i+1}(l,k) \neq 0$, то $l = i+2$ и $k = 0$. И есть всего один граф с $P_{i+1}(l,k) \neq 0$. Рассуждая как в конце параграфа 1.2.2, получаем, что вероятность этого графа равна $\frac{(l-1)!}{(2l-1)!!}$. Из рекуррентного соотношения имеем $p(l,0) = \frac{1}{2^{l-2}}$. В этом случае получаем

$$\mathrm{E}P_{i+1}(l,k) = \frac{(l-1)!}{(2l-1)!!} < \frac{1}{2^{l-2}} = p(l,0).$$

Если $2i+3 \geqslant 2l+k$, то

$$\mathrm{E}P_{i+1}(l,k) = \mathrm{E}P_i(l,k)\left(1 - \frac{2l+k-2}{2i+1}\right) +$$

$$+\mathrm{E}P_i(l,k-1)\frac{l+k-3}{2i+1} + \mathrm{E}P_i(l-1,k)\frac{l-1}{2i+1}.$$

Используя рекуррентное соотношение для $p(l,k)$ и индукцию по i, легко показать, что $\mathrm{E}P_n(l,k) \leqslant p(l,k)$. Это завершает доказательство леммы 2.

Доказательство теоремы 3

Для доказательства этой теоремы нам потребуется неравенство Азумы (см. [13]).

Лемма 3. *Пусть* $(X_i)_{i=0}^n$ — *мартингал, причем* $|X_i - X_{i-1}| \leqslant c$ *для всех* i, $1 \leqslant i \leqslant n$. *Тогда*

$$\mathsf{P}\left(|X_n - X_0| \geqslant x\right) \leqslant 2e^{-\frac{x^2}{2c^2n}}$$

для любого $x > 0$.

Зафиксируем произвольное $\varepsilon > 0$. Фиксируем $n \geqslant 3$ и такое k, что $1 \leqslant k \leqslant n^{1/6-\varepsilon}$. Рассмотрим последовательность случайных величин $X^i(k) = \mathrm{E}(X_n(k)|G_1^i)$, $i = 0,\ldots,n$. Поясним обозначение $\mathrm{E}(X_n(k)|G_1^i)$. Мы строим граф $G_1^n \in \mathfrak{G}_1^n$ индуктивно. Для любого $t \leqslant n$ существует и единственный граф $G_1^t \in \mathfrak{G}_1^t$ такой, что G_1^n получен из G_1^t. Таким образом, $\mathrm{E}(X_n(k)|G_1^t)$ — математическое ожидание количества вершин второй степени k в графе G_1^n, если на шаге t мы имеем граф G_1^t.

Заметим, что $X^0(k) = \mathrm{E}X_n(k)$ и $X^n(k) = X_n(k)$. Нетрудно понять, что $X^i(k)$ — мартингал.

Далее будет доказано, что для любого $i = 1,\ldots,n$

$$|X^i(k) - X^{i-1}(k)| \leqslant 10\,k\,\ln n.$$

Теорема 3 сразу же следует из этого утверждения. Положим $c = 10\,k\ln n$. Тогда из неравенства Азумы следует, что

$$\mathsf{P}\left(|X_n(k) - \mathrm{E}X_n(k)| \geqslant k\sqrt{n}\,\ln^2 n\right) \leqslant 2\exp\left\{-\frac{n\,k^2\,\ln^4 n}{200\,n\,k^2\ln^2 n}\right\} = o(1).$$

Если $k \leqslant n^{1/6-\varepsilon}$, то величина n/k^2 растет быстрее, чем $k\,\ln^2 n\sqrt{n}$. Поэтому $\left(k\sqrt{n}\,\ln^2 n\right) / \left(n/k^2\right) = o(1)$. Что и требовалось доказать.

Остается оценить $|X^i(k) - X^{i-1}(k)|$.

Фиксируем i, $1 \leqslant i \leqslant n$, и некоторый граф G_1^{i-1}. Заметим, что

$$\left|\mathrm{E}\left(X_n(k)|G_1^i\right) - \mathrm{E}\left(X_n(k)|G_1^{i-1}\right)\right| \leqslant$$

$$\leqslant \max_{\tilde{G}_1^i \supset G_1^{i-1}} \left\{ \mathrm{E}\left(X_n(k) | \tilde{G}_1^i \right) \right\} - \min_{\tilde{G}_1^i \supset G_1^{i-1}} \left\{ \mathrm{E}\left(X_n(k) | \tilde{G}_1^i \right) \right\}.$$

Положим $\hat{G}_1^i = \arg\max \mathrm{E}(X_n(k)|\tilde{G}_1^i)$, $\bar{G}_1^i = \arg\min \mathrm{E}(X_n(k)|\tilde{G}_1^i)$. Нам нужно оценить разность $\mathrm{E}(X_n(k)|\hat{G}_1^i) - \mathrm{E}(X_n(k)|\bar{G}_1^i)$.

Используя обозначения $N_n(l,k)$ и $P_n(l,k)$, введенные в параграфе 1.2.1, мы получаем

$$\mathrm{E}(X_n(k)|\hat{G}_1^i) = \sum_{l=1}^{\infty} \mathrm{E}(N_n(l,k)|\hat{G}_1^i) + \sum_{l=1}^{\infty} \mathrm{E}(P_n(l,k)|\hat{G}_1^i),$$

$$\mathrm{E}(X_n(k)|\bar{G}_1^i) = \sum_{l=1}^{\infty} \mathrm{E}(N_n(l,k)|\bar{G}_1^i) + \sum_{l=1}^{\infty} \mathrm{E}(P_n(l,k)|\bar{G}_1^i).$$

Для t, $i \leqslant t \leqslant n$, положим

$$\delta_t(l,k) = \mathrm{E}(N_t(l,k)|\hat{G}_1^i) - \mathrm{E}(N_t(l,k)|\bar{G}_1^i), \quad \delta'_t(l,k) = \delta_t(l,k) I(\delta_t(l,k) > 0),$$

$$\epsilon_t(l,k) = \mathrm{E}(P_t(l,k)|\hat{G}_1^i) - \mathrm{E}(P_t(l,k)|\bar{G}_1^i), \quad \epsilon'_t(l,k) = \epsilon_t(l,k) I(\epsilon_t(l,k) > 0),$$

$$\delta_t(k) = \mathrm{E}(N_t(k)|\hat{G}_1^i) - \mathrm{E}(N_t(k)|\hat{G}_1^i)), \quad \delta'_t(k) = \delta_t(k) I(\delta_t(k) > 0).$$

Заметим, что

$$\mathrm{E}(X_n(k)|\hat{G}_1^i) - \mathrm{E}(X_n(k)|\bar{G}_1^i) = \sum_{l=1}^{\infty} \delta_n(l,k) + \sum_{l=1}^{\infty} \epsilon_n(l,k) \leqslant$$

$$\leqslant \sum_{l=1}^{\infty} \delta'_n(l,k) + \sum_{l=1}^{\infty} \epsilon'_n(l,k) \leqslant \sum_{l=1}^{\infty} \sum_{j=0}^{k} \left(\delta'_n(l,j) + \epsilon'_n(l,j) \right).$$

Оценим полученную двойную сумму.

Сначала предположим, что $n = i$. Фиксируем граф G_1^{i-1}. Графы $\hat{G}_1^i$ и $\bar{G}_1^i$ получены из графа G_1^{i-1}. Мы добавляем вершину i и одно ребро $i\hat{q}$ или $i\bar{q}$ соответственно. Новое ребро меняет только степень $\hat{q}$ или $\bar{q}$ и вторую степень соседей $\hat{q}$ или $\bar{q}$ соответственно. Рассмотрим граф $\hat{G}_1^i$. Фиксируем l и $j \leqslant k$. Нас интересует, насколько может вырасти количество вершин степени l и второй степени j на шаге i. Во-первых, вершина i может стать вершиной второй степени j с $j \leqslant k$. Во-вторых, вершина $\hat{q}$ может стать вершиной второй степени j с $j \leqslant k$. В-третьих, вторая степень соседа вершины $\hat{q}$ выросла. Если вершина $\hat{q}$ имеет хотя бы $k+1$ соседей в графе G_1^{i-1}, то после шага i эти вершины имеют вторую степень больше чем k, и мы не учитываем их. Если у вершины $\hat{q}$ более чем k соседей в G_1^{i-1}, то не более k вершин меняют свою вторую степень на шаге i. Рассуждая аналогично, рассмотрим $\bar{G}_1^i$. Хотим оценить, насколько могут уменьшиться величины $N_{i-1}(l,j)$ и $P_{i-1}(l,j)$. Во-первых, вершина $\bar{q}$ меняет степень на шаге i. Во-вторых, некоторые соседи

$\bar{q}$ могут иметь вторую степень $j \leqslant k$ в G_1^{i-1} (поэтому количество соседей $\bar{q}$ в G_1^{i-1} не превосходит $k+1$). Просуммируем все полученные числа. Имеем

$$\sum_{l=1}^{\infty} \sum_{j=0}^{k} (\delta'_i(l,j) + \epsilon'_i(l,j)) \leqslant 1 + 1 + k + 1 + (k+1) = 2k+4.$$

Случай $n = i$ разобран. Теперь рассмотрим t: $i \leqslant t \leqslant n-1$. Заметим, что

$$\mathrm{E}\left(N_{t+1}(1)|G_1^i\right) = \mathrm{E}\left(N_t(1)|G_1^i\right)\left(1 - \frac{1}{2t+1}\right) + \frac{2t}{2t+1},$$

$$\mathrm{E}\left(N_{t+1}(2)|G_1^i\right) = \mathrm{E}\left(N_t(2)|G_1^i\right)\left(1 - \frac{2}{2t+1}\right) + \mathrm{E}\left(N_t(1)|G_1^i\right)\frac{1}{2t+1} + \frac{1}{2t+1},$$

$$\mathrm{E}\left(N_{t+1}(j)|G_1^i\right) = \mathrm{E}\left(N_t(j)|G_1^i\right)\left(1 - \frac{j}{2t+1}\right) + \mathrm{E}\left(N_t(j-1)|G_1^i\right)\frac{j-1}{2t+1}, \quad j \geqslant 3,$$

$$\mathrm{E}\left(N_{t+1}(1,j)|G_1^i\right) = \mathrm{E}\left(N_t(1,j)|G_1^i\right)\left(1 - \frac{j+2}{2t+1}\right) +$$
$$+\frac{j\,\mathrm{E}\left(N_t(1,j-1)|G_1^i\right)}{2t+1} + \frac{j\,\mathrm{E}\left(N_t(j)|G_1^i\right)}{2t+1},$$

$$\mathrm{E}\left(N_{t+1}(l,j)|G_1^i\right) = \mathrm{E}\left(N_t(l,j)|G_1^i\right)\left(1 - \frac{2l+j}{2t+1}\right) + \frac{(l-1)\,\mathrm{E}\left(N_t(l-1,j)|G_1^i\right)}{2t+1} +$$
$$+\frac{(l+j-1)\mathrm{E}\left(N_t(l,j-1)|G_1^i\right)}{2t+1}, \quad l \geqslant 2,$$

$$\mathrm{E}\left(P_{t+1}(2,0)|G_1^i\right) = \mathrm{E}\left(P_t(2,0)|G_1^i\right)\left(1 - \frac{2}{2t+1}\right) + \frac{1}{2t+1},$$

$$\mathrm{E}\left(P_{t+1}(l,j)|G_1^i\right) = \mathrm{E}\left(P_t(l,j)|G_1^i\right)\left(1 - \frac{2l+j-2}{2t+1}\right) +$$
$$+\mathrm{E}\left(P_t(l,j-1)|G_1^i\right)\frac{l+j-3}{2t+1} + \mathrm{E}\left(P_t(l-1,j)|G_1^i\right)\frac{l-1}{2t+1}, \quad l \geqslant 3.$$

Мы получали похожие равенства в доказательствах теоремы 4, леммы 1 и леммы 2. Заменим граф G_1^i графом $\hat{G}_1^i$ или $\bar{G}_1^i$ в этих равенствах. Вычитая равенства с $\bar{G}_1^i$ из равенств с $\hat{G}_1^i$ и используя неравенство $(a+b)I(a+b > 0) \leqslant aI(a>0) + bI(b>0)$, мы получаем

$$\delta'_{t+1}(j) \leqslant \delta'_t(j)\left(1 - \frac{j}{2t+1}\right) + \delta'_t(j-1)\frac{j-1}{2t+1},$$

$$\delta'_{t+1}(1,j) \leqslant \delta'_t(1,j)\left(1 - \frac{j+2}{2t+1}\right) + \frac{j\delta'_t(1,j-1)}{2t+1} + \frac{j\delta'_t(j)}{2t+1},$$

$$\delta'_{t+1}(l,j) \leqslant \delta'_t(l,j)\left(1-\frac{2l+j}{2t+1}\right)+\frac{(l-1)\delta'_t(l-1,j)}{2t+1}+$$
$$+\frac{(l+j-1)\delta'_t(l,j-1)}{2t+1},\ \ l \geqslant 2,$$
$$\epsilon'_{t+1}(2,0) \leqslant \epsilon'_t(2,0)\left(1-\frac{2}{2t+1}\right),$$
$$\epsilon'_{t+1}(l,j) \leqslant \epsilon'_t(l,j)\left(1-\frac{2l+j-2}{2t+1}\right)+\frac{(l-1)\epsilon'_t(l-1,j)}{2t+1}+$$
$$+\frac{(l+j-3)\epsilon'_t(l,j-1)}{2t+1},\ \ l \geqslant 3.$$

Теперь мы можем оценить сумму

$$\sum_{l=1}^{\infty}\sum_{j=0}^{k}\delta'_{t+1}(l,j)+\sum_{l=2}^{\infty}\sum_{j=0}^{k}\epsilon'_{t+1}(l,j) \leqslant$$

$$\leqslant \sum_{j=1}^{k}\left(\delta'_t(1,j)\left(1-\frac{j+2}{2t+1}\right)+\frac{j\delta'_t(1,j-1)}{2t+1}+\frac{j\delta'_t(j)}{2t+1}\right)+$$
$$+\epsilon'_t(2,0)\left(1-\frac{2}{2t+1}\right)+$$
$$+\sum_{l=2}^{\infty}\sum_{j=1}^{k}\left(\delta'_t(l,j)\left(1-\frac{2l+j}{2t+1}\right)+\right.$$
$$\left.+\frac{(l-1)\delta'_t(l-1,j)}{2t+1}+\frac{(l+j-1)\delta'_t(l,j-1)}{2t+1}\right)+$$
$$+\sum_{l=3}^{\infty}\sum_{j=0}^{k}\left(\epsilon'_t(l,j)\left(1-\frac{2l+j-2}{2t+1}\right)+\right.$$
$$\left.+\frac{(l-1)\epsilon'_t(l-1,j)}{2t+1}+\frac{(l+j-3)\epsilon'_t(l,j-1)}{2t+1}\right)=$$

$$=\sum_{j=1}^{k}\delta'_t(1,j)-\sum_{j=1}^{k}\frac{(j+2)\delta'_t(1,j)}{2t+1}+\sum_{j=1}^{k-1}\frac{(j+1)\delta'_t(1,j)}{2t+1}+\sum_{j=1}^{k}\frac{j\delta'_t(j)}{2t+1}+$$

$$+\sum_{l=2}^{\infty}\sum_{j=1}^{k}\delta'_t(l,j)-\sum_{l=2}^{\infty}\sum_{j=1}^{k}\frac{(2l+j)\delta'_t(l,j)}{2t+1}+\sum_{l=1}^{\infty}\sum_{j=1}^{k}\frac{l\delta'_t(l,j)}{2t+1}+$$

$$+\sum_{l=2}^{\infty}\sum_{j=1}^{k-1}\frac{(l+j)\delta'_t(l,j)}{2t+1}+\epsilon'_t(2,0)-\frac{2\epsilon'_t(2,0)}{2t+1}+\sum_{l=3}^{\infty}\sum_{j=0}^{k}\epsilon'_t(l,j)-$$

$$-\sum_{l=3}^{\infty}\sum_{j=0}^{k}\frac{(2l+j-2)\epsilon'_t(l,j)}{2t+1}+\sum_{l=3}^{\infty}\sum_{j=0}^{k}\frac{l\epsilon'_t(l,j)}{2t+1}+\frac{2\epsilon'_t(2,0)}{2t+1}+$$

$$+\sum_{l=3}^{\infty}\sum_{j=0}^{k-1}\frac{(l+j-2)\epsilon'_t(l,j)}{2t+1}=$$

$$=\sum_{l=1}^{\infty}\sum_{j=0}^{k}(\delta'_t(l,j)+\epsilon'_t(l,j))+\sum_{j=1}^{k}\frac{j\delta'_t(j)}{2t+1}-$$

$$-\sum_{l=1}^{\infty}\delta'_t(l,k)\frac{l+k}{2t+1}-\sum_{l=3}^{\infty}\epsilon'_t(l,k)\frac{l+k-2}{2t+1}\leqslant$$

$$\leqslant\sum_{l=1}^{\infty}\sum_{j=0}^{k}(\delta'_t(l,j)+\epsilon'_t(l,j))+\sum_{j=1}^{k}\frac{j\delta'_t(j)}{2t+1}.$$

Остается оценить сумму $\sum_{j=1}^{k}\frac{j\delta'_t(j)}{2t+1}$. Заметим, что для любого $t\geqslant i$ мы имеем $\sum_{j=0}^{k}\delta'_t(j)\leqslant 3$. Это очевидно для $t=i$ (когда мы добавляем новую вершину i, мы меняем только степени $\hat{q}$ и $\bar{q}$). Если $t+1>i$, то

$$\sum_{j=1}^{k}\delta'_{t+1}(j)\leqslant\sum_{j=1}^{k}\left(\delta'_t(j)\left(1-\frac{j}{2t+1}\right)+\delta'_t(j-1)\frac{j-1}{2t+1}\right)=$$

$$=\sum_{j=1}^{k}\delta'_t(j)-\delta'_t(k)\frac{k}{2t+1}.$$

Другими словами, $\sum_{j=1}^{k}\delta'_t(j)$ не возрастает, когда t растет.

Поэтому мы получаем

$$\sum_{l=1}^{\infty}\sum_{j=0}^{k}(\delta'_{t+1}(l,j)+\epsilon'_{t+1}(l,j))\leqslant\sum_{l=1}^{\infty}\sum_{j=0}^{k}(\delta'_t(l,j)+\epsilon'_t(l,j))+\frac{3k}{2t+1}.$$

Таким образом,

$$|\mathrm{E}(X_n(k)|G_1^{i-1})-\mathrm{E}(X_n(k)|G_1^{i})|\leqslant\sum_{l=1}^{\infty}\sum_{j=0}^{k}(\delta'_i(l,j)+\epsilon'_i(l,j))+\sum_{t=i}^{n-1}\frac{3k}{2t+1}\leqslant$$

$$\leqslant 2k+4+\sum_{t=1}^{n-1}\frac{3k}{2t+1}\leqslant 2k+5+\frac{3}{2}k\ln n\leqslant 10\,k\ln n.$$

Это завершает доказательство теоремы 3.

1.3 k-е степени вершин

1.3.1 Введение обозначений и формулировка результатов

Определение k-й входящей степени вершины

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

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

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

В этом разделе, когда будем писать $ij \in G_m^n$, будем иметь в виду, что *направленное* ребро из вершины i в вершину j проведено в графе G_m^n. Таким образом, $D_k(t)$ — множество вершин i, от которых можно добраться до вершины t самое меньшее за k шагов, двигаясь все время по направленным ребрам. Далее, пусть

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

Величина $|S_k(t)| = d_k(t)$ и есть k-я входящая степень вершины. Считаем $d_0(t) = 1$. Кроме того, можно было бы назвать k-й входящей степенью вершины t количество вершин, находящихся от t на "входящем" расстоянии k, то есть $|D_k(t)|$. Заметим, что в графе G_1^n при $k > 1$ оба определения совпадают. Далее под $d(t)$ (или $d_{G_1^n}(t)$) понимаем полную степень вершины t, а под $d_k(t)$ — k-ю входящую степень вершины t.

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

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

Формулировка результатов

Сначала рассмотрим граф G_1^n. Определим функцию $f_k(t,n)$. Пусть $f_0(t,n) = 1$, а для $k > 0$

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

Теорема 5. *Для любого натурального $k \geqslant 2$ и любой вершины $t < n$ в графе G_1^n*

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

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

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

Следствие 1. *Пусть $w(n) \to \infty$ при $n \to \infty$. Пусть, далее, $\ln n \leqslant t \leqslant n - w(n)$ и $k = o(\sqrt{t\ln\frac{n}{t}})$, $k \geqslant 2$. Тогда*

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

при $n \to \infty$.

Далее от графа G_1^n перейдем к графу G_m^n с $m \geqslant 2$. Обозначим через I_{it} индикатор того, что в графе G_m^n проведено ребро it. Заметим, что для величин $g_k(t)$ в графе G_1^n выполняется соотношение

$$g_k(t) = \sum_{i=t+1}^{n} g_{k-1}(i)I_{it}$$

при $k \geqslant 3$. Для $k = 2$ это не всегда выполнено из-за возможности петли в вершине t. Для $m \geqslant 2$ введем аналогичную величину $q_k(t)$. Определим ее индуктивно. Пусть $q_1(t) = d(t)$, а для $k > 1$

$$q_k(t) = \sum_{i=t+1}^{n} q_{k-1}(i)I_{it}.$$

Ясно, что $q_k(t) \geqslant g'_k(t)$, поскольку это неравенство выполнено для $k = 1$, а для $k > 1$

$$g'_k(t) \leqslant \sum_{i=t+1}^{n} g'_{k-1}(i) I_{it}.$$

Кроме того, заметим, что $q_k(t)$ — оценка сверху на сумму степеней вершин из множества $D'_{k-1}(t)$. Для $k = 1$ имеем точное равенство, а для $k > 1$ работает рекуррентное соотношение.

Для величин $q_k(t)$ докажем следующее утверждение.

Теорема 6. *Фиксируем $m \geqslant 2$. Для любого натурального $k \geqslant 2$ и любой вершины $t < n$ в графе G_m^n*

$$\mathrm{E} q_k(t) = m^k f_k(t,n) \left(1 + \theta(k,t,n)\right) \left(1 + \theta_1(t)\right)^k,$$

где $|\theta(k,t,n)| \leqslant \tilde{C} \frac{k^2}{t \ln \sqrt{n/t}}$, $|\theta_1(t)| \leqslant \tilde{C}_1 \frac{1}{t}$, $\tilde{C}$ и $\tilde{C}_1$ — некоторые константы.

И, наконец, оценим $g'_k(t)$ снизу и получим следующую теорему.

Теорема 7. *Пусть натуральные $m \geqslant 2$ и $k \geqslant 1$ фиксированы. Тогда для $t < n$*

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

В следующем параграфе будут последовательно доказаны теоремы 5, 6 и 7.

1.3.2 Доказательства теорем

Доказательство теоремы 5

Заметим, что $g_1(t) = d(t)$ — обычная степень вершины t. Докажем, что для любого $k \geqslant 2$

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

где $|\theta(k,t,n)| \leqslant C \frac{k^2}{t \ln \sqrt{n/t}}$, $|\theta_1(t)| \leqslant C_1 \frac{1}{t}$, константы C и C_1 будут определены далее.

Математическое ожидание степени вершины t графа G_1^n известно, оно равно $\sqrt{n/t}\,(1 + O(1/t))$ (см. [21]). Это несложно вычисляется следующим образом:

$$\mathrm{E}(d_{G_1^t}(t)) = 1 + 1/(2t-1).$$

А для $s > t$ имеем

$$\mathrm{E}(d_{G_1^s}(t) | d_{G_1^{s-1}}(t)) = d_{G_1^{s-1}}(t) + d_{G_1^{s-1}}(t)/(2s-1).$$

Откуда следует, что

$$\mathrm{E}(d_{G_1^s}(t)) = \frac{2s}{2s-1}\mathrm{E}(d_{G_1^{s-1}}(t)).$$

И получаем, что для любой вершины t

$$\mathrm{E}(d_{G_1^n}(t)) = \prod_{i=t}^{n} \frac{2i}{2i-1} = \frac{4^{n-t+1} n!^2 (2t-2)!}{(2n)!(t-1)!^2} = \sqrt{n/t}\,(1+O(1/t)).$$

В последнем равенстве используем формулу Стирлинга. Это в точности рассуждения из [21], нам потребуется сослаться на них при доказательстве следующего вспомогательного утверждения.

Лемма 4. *Для любых вершин t и i, $1 \leqslant t < i \leqslant n$, при $k \geqslant 1$ и $k \neq 2$ выполнено*

$$\mathrm{E}I_{it}g_k(i) = \mathrm{E}I_{it} \cdot \mathrm{E}g_k(i) \cdot (1+\theta_2(i)),$$

где $|\theta_2(i)| \leqslant \frac{1}{i}$. Для $k=2$

$$\mathrm{E}I_{it}g_2(i) = \mathrm{E}I_{it} \cdot \left(\mathrm{E}g_2(i) - \frac{1}{2i-1}\right) \cdot (1+\theta_2(i)).$$

Доказательство. Рассмотрим случай $k=1$. Помним, что $g_1(i) = d(i)$. Имеем:

$$\mathrm{E}I_{it}d(i) = \mathrm{E}I_{it} \cdot \mathrm{E}(d(i)|I_{it}=1) = \mathrm{E}I_{it} \cdot \mathrm{E}d(i) \cdot \left(1 - \frac{1}{2i}\right).$$

Поясним, откуда взялось последнее равенство. Вспомним, как мы считали математическое ожидание степени вершины. В графе G_1^i выполняется $\mathrm{E}d_{G_1^i}(i) = \frac{2i}{2i-1}$, а далее при добавлении каждой следующей вершины s число умножается на $2s/(2s-1)$. А при наличии ребра it получаем $d_{G_1^i}(i) = 1$, т.е. $\mathrm{E}(d_{G_1^i}(i)|I_{it}=1) = 1$. Дальнейшие вычисления не изменяются. Получаем, что если в графе есть ребро it, то математическое ожидание степени вершины i изменяется всего лишь в $1-\frac{1}{2i}$ раз. Тем самым для $k=1$ лемма доказана.

Пусть теперь $k \geqslant 3$:

$$\mathrm{E}I_{it}g_k(i) = \mathrm{E}I_{it} \cdot \mathrm{E}(g_k(i)|I_{it}=1),$$

$$\mathrm{E}(g_k(i)|I_{it}=1) = \sum_{j=i+1}^{n} \mathrm{E}(g_{k-1}(j)I_{ji}|I_{it}=1) =$$

$$= \sum_{j=i+1}^{n} \mathrm{E}(I_{ji}|I_{it}=1) \cdot \mathrm{E}(g_{k-1}(j)|I_{ji}=1, I_{it}=1) =$$

$$= \sum_{j=i+1}^{n} \mathrm{E} I_{ji} \cdot \left(1 - \frac{1}{2i}\right) \cdot \mathrm{E}(g_{k-1}(j)|I_{ji} = 1) = \mathrm{E} g_k(i) \cdot (1 + \theta_2(i)).$$

Здесь воспользовались случаем $k = 1$ при вычислении $\mathrm{E}(I_{ji}|I_{it} = 1)$. Кроме того, следует пояснить равенство $\mathrm{E}(g_{k-1}(j)|I_{ji} = 1, I_{it} = 1) = \mathrm{E}(g_{k-1}(j)|I_{ji} = 1)$. Заметим, что наличие ребра ji полностью определяет вероятности ребер jj и $(j + 1)j$. Далее, по индукции, оно однозначно определяет вероятности тех или иных ребер на множестве вершин $j, \ldots, n$. Поэтому распределение ребер на вершинах $j, \ldots, n$ при условии $I_{ji} = 1$ совпадает с распределением этих ребер при условии события $(I_{ji} = 1) \cap (I_{it} = 1)$.

Осталось рассмотреть случай $k = 2$. Заметим, что равенство

$$\mathrm{E}(g_2(i)|I_{it} = 1) = \sum_{j=i+1}^{n} \mathrm{E}(g_1(j) I_{ji}|I_{it} = 1)$$

верно, хотя без условия $I_{it} = 1$ таковым не являлось. Итак,

$$\mathrm{E} I_{it} g_2(i) = \mathrm{E} I_{it} \cdot \mathrm{E}(g_2(i)|I_{it} = 1),$$

$$\mathrm{E}(g_2(i)|I_{it} = 1) = \sum_{j=i+1}^{n} \mathrm{E}(g_1(j) I_{ji}|I_{it} = 1) =$$

$$= \sum_{j=i+1}^{n} \mathrm{E}(I_{ji}|I_{it} = 1) \cdot \mathrm{E}(g_1(j)|I_{ji} = 1, I_{it} = 1) =$$

$$= \sum_{j=i+1}^{n} \mathrm{E} I_{ji} \cdot \left(1 - \frac{1}{2i}\right) \cdot \mathrm{E}(g_1(j)|I_{ji} = 1) = \left(\mathrm{E} g_2(i) - \frac{1}{2i - 1}\right) \cdot (1 + \theta_2(i)).$$

Лемма доказана. $\square$

Вернемся к доказательству теоремы. Для $k = 1$, как мы уже знаем,

$$\mathrm{E} d(t) = \sqrt{n/t}\,(1 + \varphi_1(t)),$$

где $|\varphi_1(t)| \leqslant M_1/t$, M_1 — константа.

Теорему будем доказывать по индукции. Рассмотрим сначала $k = 2$. Нетрудно видеть, что

$$\mathrm{E} g_2(t) = \mathrm{E} \sum_{i=t+1}^{n} I_{it} d(i) + \mathrm{E} I_{tt} = \sum_{i=t+1}^{n} \mathrm{E} I_{it} d(i) + \mathsf{P}(I_{tt} = 1) =$$

$$= \sum_{i=t+1}^{n} \mathrm{E} I_{it} \cdot \left(\mathrm{E} d(i) - \frac{1}{2i - 1}\right)(1 + \theta_2(i)) + \frac{1}{2t - 1}.$$

В последнем равенстве воспользовались леммой 4.

Математическое ожидание индикатора I_{it} — это вероятность ребра it в графе G_1^n:

$$\mathrm{E}I_{it} = \frac{\mathrm{E}d_{G_1^{i-1}}(t)}{2i-1} = \frac{\sqrt{(i-1)/t}}{2i-1}(1+O(1/t)) = \frac{\sqrt{i/t}}{2i}(1+\theta_3(t)),$$

где $|\theta_3(t)| \leqslant C_3/t$, C_3 — константа.

Введем обозначение: $\theta_1(t)$ — величина, модуль которой не превосходит C_1/t, где C_1 выбирается настолько большой, чтоб можно было заменить выражением $(1+\theta_1(t))$ следующие два выражения: $(1+\varphi_1(t))$ и $(1+\theta_2(i))(1+\theta_3(t))$ для $i > t$.

В итоге получаем

$$\mathrm{E}I_{it}d(i) = \frac{\sqrt{i/t}\sqrt{n/i}}{2i}(1+\varphi_1(t))(1+\theta_2(i))(1+\theta_3(t)) - \frac{\mathrm{E}I_{it}}{2i-1}(1+\theta_2(i)) =$$

$$= \frac{\sqrt{n/t}}{2i}(1+\theta_1(t))^2 + O\left(\frac{1}{i\sqrt{it}}\right).$$

Складываем по всем $i > t$ (пользуемся интегральным признаком):

$$\sum_{i=t+1}^{n} \frac{1}{2i} = \int_t^n \frac{1}{2x}\,dx + O(1/t) = \ln\sqrt{n/t} + O(1/t),$$

$$\sum_{i=t+1}^{n} O\left(\frac{1}{i\sqrt{it}}\right) = O\left(\frac{1}{t}\right).$$

Итак,

$$\mathrm{E}g_2(t) = \sqrt{n/t}(\ln\sqrt{n/t} + O(1/t))(1+\theta_1(t))^2 + O\left(\frac{1}{t}\right) =$$

$$= \sqrt{n/t}(1+\theta_1(t))^2\left(\ln\sqrt{n/t} + O\left(\frac{1}{t}\right)\right).$$

Получили выражение

$$\mathrm{E}g_2(t) = \frac{f_2(t,n)}{\ln\sqrt{n/t}}(1+\theta_1(t))^2(\ln\sqrt{n/t} + \varphi(t,n)),$$

где $|\varphi(t,n)| \leqslant \frac{M}{t}$, M — константа. Это равенство верно для любого $t \leqslant n$.

Теперь выберем константу C. Пусть $C = \max\{M, 1\}$. С такой константой, очевидно, утверждение верно в случае $k = 2$. Предположим, что $k \geqslant 2$ и

$$\mathrm{E}g_k(t) = \frac{f_k(t,n)}{\ln\sqrt{n/t}}\left(\ln\sqrt{\frac{n}{t}} + \theta(k,t,n)\ln\sqrt{\frac{n}{t}}\right)(1+\theta_1(t))^k$$

для $t \leqslant n$. Сделаем переход индукции. Поступим аналогично. Заметим, что для любого натурального $k \geqslant 2$

$$\mathrm{E}g_{k+1}(t) = \mathrm{E}\sum_{i=t+1}^{n} I_{it}g_k(i) = \sum_{i=t+1}^{n} \mathrm{E}I_{it}g_k(i) = \sum_{i=t+1}^{n} \mathrm{E}I_{it} \cdot \mathrm{E}g_k(i)(1+\theta_2(i)).$$

В последнем равенстве воспользовались леммой 4. Далее

$$\sum_{i=t+1}^{n} \mathrm{E}I_{it} \cdot \mathrm{E}g_k(i)(1+\theta_2(i)) = \sum_{i=t+1}^{n} \frac{\sqrt{i/t}}{2i}\mathrm{E}g_k(i)(1+\theta_2(i))(1+\theta_3(t)) =$$

$$= \sum_{i=t+1}^{n} \frac{\sqrt{n/t}}{2i(k-1)!}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-2}(1+\theta_1(t))^{k+1} \cdot \left(\ln\sqrt{\frac{n}{i}} + \ln\sqrt{\frac{n}{i}}\,\theta(k,i,n)\right).$$

В итоге получили

$$\mathrm{E}g_{k+1}(t) = \sum_{i=t+1}^{n} \frac{\sqrt{n/t}}{2i(k-1)!}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-2}(1+\theta_1(t))^{k+1} \cdot$$

$$\cdot\left(\ln\sqrt{\frac{n}{t}} + \ln\sqrt{\frac{n}{t}}\,\theta(k,i,n)\right) = (1+\theta_1(t))^{k+1}\frac{\sqrt{n/t}}{(k-1)!}\cdot$$

$$\cdot\left(\sum_{i=t+1}^{n}\frac{1}{2i}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-1} + \sum_{i=t+1}^{n}\frac{1}{2i}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-1}\theta(k,i,n)\right).$$

Сначала рассмотрим первую сумму в получившемся выражении:

$$\sum_{i=t+1}^{n}\frac{1}{2i}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-1} = \frac{1}{2^k}\int_t^n \frac{1}{x}\left(\ln\frac{n}{x}\right)^{k-1}dx + \theta_4(t),$$

где $|\theta_4(t)| \leqslant \frac{1}{2t}\left(\ln\sqrt{\frac{n}{t}}\right)^{k-1}$. Вычислим интеграл

$$\int_t^n \frac{1}{x}\left(\ln\frac{n}{x}\right)^{k-1}dx = -\int_t^n \left(\ln\frac{n}{x}\right)^{k-1} d\left(\ln\frac{n}{x}\right) = \frac{1}{k}\left(\ln\frac{n}{t}\right)^{k}.$$

В итоге первая сумма в выражении для $\mathrm{E}g_{k+1}(t)$ имеет вид:

$$\sum_{i=t+1}^{n}\frac{1}{2i}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-1} = \frac{1}{k}\left(\ln\sqrt{\frac{n}{t}}\right)^{k} + \theta_4\left(t\right).$$

Посмотрим на вторую сумму. Обозначим ее $\theta_5(t)$. Хотим оценить сверху модуль этой величины:

$$|\theta_5| = \left|\sum_{i=t+1}^{n}\frac{1}{2i}\left(\ln\sqrt{\frac{n}{i}}\right)^{k-1}\theta(k,i,n)\right| \leqslant$$

$$\leqslant \left| \sum_{i=t+1}^{n} \frac{1}{2i^2} \left(\ln \sqrt{\frac{n}{i}} \right)^{k-1} \frac{C\,k^2}{\ln \sqrt{n/i}} \right| \leqslant$$

$$\leqslant \int_t^n \frac{C\,k^2}{2xt} \left(\ln \sqrt{\frac{n}{x}} \right)^{k-2} dx = \frac{C\,k^2 \left(\ln \sqrt{n/t} \right)^{k-1}}{t(k-1)}.$$

Итак,

$$\mathrm{E}g_{k+1}(t) = (1+\theta_1(t))^{k+1} \frac{\sqrt{n/t}}{(k-1)!} \left(\frac{1}{k} \left(\ln \sqrt{\frac{n}{t}} \right)^k + \theta_4(t) + \theta_5(t) \right) =$$

$$= \frac{1}{k!} \sqrt{\frac{n}{t}} \left(\ln \sqrt{\frac{n}{t}} \right)^k \left(1 + \frac{k\,(\theta_4(t)+\theta_5(t))}{\left(\ln \sqrt{\frac{n}{t}} \right)^k} \right) (1+\theta_1(t))^{k+1}.$$

Положим $\theta(k+1,t,n) = \frac{k\,(\theta_4(t)+\theta_5(t))}{\left(\ln \sqrt{\frac{n}{t}}\right)^k}$. Осталось лишь проверить, что эта величина достаточно мала:

$$\left| \theta(k+1,t,n) \ln \sqrt{\frac{n}{t}} \right| = \left| \frac{k\,(\theta_4(t)+\theta_5(t))}{\left(\ln \sqrt{\frac{n}{t}} \right)^{k-1}} \right| \leqslant \frac{k}{2t} + \frac{C\,k^3}{t(k-1)} \leqslant \frac{C\,(k+1)^2}{t}.$$

Последнее неравенство выполнено для любых $k \geqslant 2$ и $C \geqslant 1$. Теорема доказана.

Доказательство теоремы 6

Доказательство этой теоремы будет аналогично доказательству теоремы 5. Фиксируем натуральное число $m \geqslant 2$. Помним, что граф G_m^n получается из графа G_1^{mn} объединением вершин в группы по m. Вершина t в G_m^n получена из вершин $m(t-1)+1, m(t-1)+2, \ldots, mt$. Положим $N = mn$.

Для начала посмотрим, что происходит с математическим ожиданием степени вершины t в графе G_m^n. Нужно сложить математические ожидания степеней вершин графа G_1^{mn}, из которых получена вершина t:

$$\mathrm{E}d(t) = \sum_{i=m(t-1)+1}^{mt} \sqrt{\frac{N}{i}}(1+O(1/i)) = \sum_{i=m(t-1)+1}^{mt} \sqrt{\frac{mn}{i}}(1+O(1/mt)) =$$

$$= m\sqrt{\frac{n}{t}}(1+O(1/t)).$$

Хотим доказать, что для любого $k \geqslant 2$

$$\mathrm{E}q_k(t) = m^k\, f_k(t,n)\,(1+\theta(k,t,n))\,(1+\theta_1(t))^k,$$

где $|\theta(k,t,n)| \leqslant \tilde{C} \frac{k^2}{t \ln \sqrt{n/t}}$, $|\theta_1(t)| \leqslant \tilde{C}_1 \frac{1}{t}$.

Докажем аналог леммы 4 в нашем случае.

Лемма 5. *Для любых вершин t и i, $1 \leqslant t < i \leqslant n$, в графе G_m^n выполнено*

$$\mathrm{E}I_{it}q_k(i) = \mathrm{E}I_{it} \cdot \mathrm{E}q_k(i) \cdot (1 + \theta_2(i)),$$

где $|\theta_2(i)| \leqslant \frac{\tilde{C}_2}{i}$.

Доказательство. При $k = 1$ получаем

$$\mathrm{E}I_{it}d(i) = \mathrm{E}I_{it} \cdot \mathrm{E}(d(i)|I_{it} = 1) = \mathrm{E}I_{it} \cdot \mathrm{E}d(i) \cdot (1 + O(1/i)).$$

Здесь надо пояснить последнее равенство. Степень вершины i в графе G_m^n — сумма степеней тех вершин в графе G_1^n, из которых получена i. Для каждой из m вершин, которые составляют i, математическое ожидание степени вершины при наличии ребра в одну из более ранних вершин изменяется в $1 + O(1/mi)$ раз. Отсюда получаем $\mathrm{E}(d(i)|I_{it} = 1) = \mathrm{E}d(i)\,(1 + O(1/i))$. Для $k = 1$ лемма доказана.

Для $k \geqslant 2$ в точности повторяем рассуждения из доказательства леммы 4. Только в этом случае доказательство для $k = 2$ не отличается от $k \geqslant 3$. Лемма доказана. □

Далее, посмотрим, чему равна вероятность ребра it в графе G_m^n для $i > t$. Вершина i получена объединением m вершин графа G_1^{mn}. Вероятность того, что хотя бы из одной из этих m вершин идет ребро в t, не превосходит:

$$\mathrm{E}I_{it} \leqslant \sum_{j=m(i-1)+1}^{mi} \frac{\mathrm{E}d_{G_m^i}(t)}{2j-1} =$$

$$= \sum_{j=m(i-1)+1}^{mi} \frac{m\sqrt{i/t}}{2j}\,(1 + O(1/t)) = \frac{m}{2\sqrt{it}}\,(1 + O(1/t)).$$

С другой стороны, вероятность ребра it можно оценить снизу (по формуле включения-исключения):

$$\mathrm{E}I_{it} \geqslant \left(\frac{m}{2\sqrt{it}} - C_m^2\frac{\mathrm{E}d_{G_m^i}(t)(\mathrm{E}d_{G_m^i}(t)+1)}{4m^2i^2}\right)(1 + O(1/t)) =$$

$$= \left(\frac{m}{2\sqrt{it}} + O\left(\frac{1}{ti}\right)\right)(1 + O(1/t)) = \frac{m}{2\sqrt{it}}\,(1 + O(1/t)).$$

Получили, что вероятность ребра it в графе G_m^n в m раз больше, чем в графе G_1^n.

Перейдем непосредственно к доказательству теоремы. Как мы уже знаем, $\mathrm{E}q_1(t) = \mathrm{E}d(t) = m\sqrt{\frac{n}{t}}(1 + O(1/t))$.

Далее можем проводить в точности рассуждения из теоремы 5 (индукцию на этот раз начинаем с $k = 1$). Единственное отличие — математическое ожидание индикатора ребра it в графе G_m^n в m раз больше, чем математическое ожидание того же индикатора в G_1^n. То есть каждый шаг индукции будет увеличивать функцию в m раз. В итоге придем к выражению:

$$\mathrm{E}q_k(t) = m^k f_k(t,n)\,(1+\theta(k,t,n))\,(1+\theta_1(t))^k\,.$$

Теорема доказана.

Доказательство теоремы 7

Заметим, что оценка сверху на величину $g'_k(t)$ уже была получена в теореме 6, и в случае фиксированного k она в точности такая, какая требуется. Осталось оценить $g'_k(t)$ снизу.

Доказываем утверждение теоремы индукцией по k. Сперва рассматриваем случай $k = 1$:

$$\mathrm{E}g'_1(t) = m + \sum_{i=t+1}^{n} \mathrm{E}I_{it} = m + \sum_{i=t+1}^{n} \frac{m}{2\sqrt{it}}\,(1+O(1/t)) =$$

$$= m\sqrt{\frac{n}{t}}\,(1+O(1/t)) = mf_1(t,n)(1+O(1/t)).$$

Далее переход индукции. Пусть доказали, что на шаге k

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

Докажем в этом случае следующую лемму.

Лемма 6. *Для любых вершин t и i, $1 \leqslant t < i \leqslant n$, в графе G_m^n для $k = 1$*

$$\mathrm{E}I_{it}g'_1(i) = \mathrm{E}I_{it}\cdot m\, f_1(i,n)\left(1+O\left(\frac{1}{i}\right)\right).$$

А для $k \geqslant 2$

$$\mathrm{E}I_{it}g'_k(i) = \mathrm{E}I_{it}\cdot m^k\, f_k(i,n)\left(1+O\left(\frac{1}{i\min\{1,\ln\frac{n}{i}\}}\right)\right).$$

Доказательство.

Здесь и далее при $i = n$ в знаменателе может появиться $\ln\frac{n}{i} = 0$. Но во всех случаях логарифм в знаменателе сократится с таким же в числителе, поскольку при $k \geqslant 2$ он содержится в функции f_k.

Докажем утверждение для $k \geqslant 2$, поскольку случай $k = 1$ рассматривается абсолютно аналогично. Имеем

$$\mathrm{E} I_{it} g'_k(i) = \mathrm{E} I_{it} \cdot \mathrm{E}(g'_k(i) | I_{it} = 1).$$

Далее, пусть A — событие, заключающееся в том, что из вершины i идет ребро в какую-нибудь из предыдущих вершин. Тогда

$$\mathrm{E}(g'_k(i)|I_{it} = 1) = \mathrm{E}(g'_k(i)|A) = \left(\mathrm{E} g'_k(i) - \mathrm{E}(g'_k(i)|\bar{A})\, \mathsf{P}(\bar{A})\right) / \mathsf{P}(A) =$$

$$= m^k\, f_k(i, n) \left(1 + O\left(\frac{1}{i \min\{1, \ln \frac{n}{i}\}}\right)\right) - \mathrm{E}(g'_k(i)|\bar{A})\, \mathsf{P}(\bar{A})\, (1 + O(1/i))\,.$$

Ясно, что $\mathsf{P}(\bar{A}) = O\left(\frac{1}{i^m}\right)$.

Далее, заметим, что

$$\mathrm{E}(g'_k(i)|\bar{A}) \leqslant \mathrm{E}(q_k(i)|\bar{A}) = O\left(f_k(i, n)\left(1 + \frac{1}{i \min\{1, \ln \frac{n}{i}\}}\right)\right).$$

Поясним последнее равенство. Заметим, что в графе G_m^i имеем $\mathrm{E}(d(i)|\bar{A}) = 2m$. В графе G_m^n получаем: $\mathrm{E}(d(i)|\bar{A}) = 2m\sqrt{\frac{n}{i}}(1+O(1/i))$. И тем самым вероятность ребра в вершину i из какой-либо из последующих вершин (при условии события $\bar{A}$) увеличивается не более чем в два раза. И в итоге $\mathrm{E}(q_k(i)|\bar{A})$ увеличится не более чем в 2^k раз по сравнению с $\mathrm{E} q_k(i)$.

С учетом всего вышесказанного, получаем, что

$$\mathrm{E}(g'_k(i)|I_{it} = 1) = m^k\, f_k(i, n) \left(1 + O\left(\frac{1}{i \min\{1, \ln \frac{n}{i}\}}\right)\right).$$

Что и требовалось.

Для $k = 1$ доказательство отличается лишь тем, что вместо $O\left(\frac{1}{i \min\{1, \ln \frac{n}{i}\}}\right)$ имеем $O\left(\frac{1}{i}\right)$. $\square$

Рассмотрим величину $\sum_{i=t+1}^{n} \mathrm{E} g'_k(i) I_{it}$. Для $k = 1$

$$\sum_{i=t+1}^{n} \mathrm{E} g'_1(i) I_{it} = \sum_{i=t+1}^{n} \frac{m}{2\sqrt{it}} \cdot m\, f_1(i, n) \left(1 + O\left(\frac{1}{i}\right)\right) =$$

$$= m^2 \sum_{i=t+1}^{n} \frac{\sqrt{n}}{2i\sqrt{t}} \left(1 + O\left(\frac{1}{i}\right)\right) = m^2\, f_2(t, n) \left(1 + O\left(\frac{1}{t \ln \frac{n}{t}}\right)\right).$$

Для $k \geqslant 2$

$$\sum_{i=t+1}^{n} \mathrm{E} g'_k(i) I_{it} = \sum_{i=t+1}^{n} \frac{m}{2\sqrt{it}} \cdot m^k\, f_k(i, n) \left(1 + O\left(\frac{1}{i \ln \frac{n}{i}}\right) + O\left(\frac{1}{i}\right)\right) =$$

$$= \frac{m^{k+1}}{(k-1)!} \sum_{i=t+1}^{n} \frac{\sqrt{n}}{2i\sqrt{t}} \left(\ln \sqrt{\frac{n}{i}} \right)^{k-1} \left(1 + O\left(\frac{1}{i \ln \frac{n}{i}} \right) + O\left(\frac{1}{i} \right) \right) =$$

$$= m^{k+1} f_{k+1}(t,n) \left(1 + O\left(\frac{1}{t \ln \frac{n}{t}} \right) + O\left(\frac{1}{t} \right) \right).$$

Такие суммы мы считали при доказательстве теоремы 5. Мы знаем, что $\sum_{i=t+1}^{n} \mathrm{E} g'_k(t) I_{it}$ не меньше, чем $\mathrm{E} g'_{k+1}(t)$. Некоторые вершины, из которых можно добраться за k и $k-1$ шаг до соседей вершины t, могут совпадать. Пусть есть вершины, из которых ведет более одного ребра в вершины из $D'_k(t)$. Тогда вычтем "лишние" ребра (все, кроме одного). Аналогично, вычтем "лишние" ребра для вершин, из которых ведет более одного ребра в вершины из $D'_{k-1}(t)$, их нужно вычесть m раз. Пусть $B_{k+1}(t)$ и $B_k(t)$ — соответственно количество этих ребер. Обозначим через $r_{k+1}(i,t)$ и $r_k(i,t)$ количество "лишних" ребер, выходящих из вершины i. Для начала оценим $\mathrm{E} B_k(t)$ (аналогично оценивается $\mathrm{E} B_{k+1}(t)$):

$$\mathrm{E} B_k(t) = \sum_{i=t+1}^{n} \mathrm{E} r_k(i,t).$$

Далее появятся константы c_1, c_2 и c_3, возможно зависящие от k и m. Значения этих констант нам не важны. Помним, что $q_k(t)$ является оценкой сверху на сумму степеней множества $D'_{k-1}(t)$. Обозначим через $q_k^i(t)$ — ту же величину $q_k(t)$, но в графе G_1^i. Вероятность того, что из конкретных двух вершин, составляющих i, проведены ребра в $D'_{k-1}(t)$, равна $O(\mathrm{E} q_k^i(t)(\mathrm{E} q_k^i(t)+1)/i^2)$. Поэтому при $k=1$

$$\mathrm{E} r_k(i,t) \leqslant \frac{c_1}{it}, \;\; \mathrm{E} B_1(t) \leqslant c_1 \sum_{i=t+1}^{n} \frac{1}{it} = O\left(\frac{\ln n/t}{t} \right) = O\left(\frac{f_2(t,n)}{t} \right).$$

Для $k \geqslant 2$

$$\mathrm{E} r_k(i,t) \leqslant c_2 \frac{\mathrm{E} q_k^i(t)(\mathrm{E} q_k^i(t)+1)}{i^2} \leqslant$$

$$\leqslant c_2 \left(\frac{\sqrt{\frac{i}{t}} \left(\ln \sqrt{\frac{i}{t}} \right)^{k-1}}{i} \left(1 + O\left(\frac{1}{t \min\{1, \ln \frac{i}{t}\}} \right) \right) \right)^2 +$$

$$+ \frac{c_2 \sqrt{\frac{i}{t}} \left(\ln \sqrt{\frac{i}{t}} \right)^{k-1}}{i^2} \left(1 + O\left(\frac{1}{t \min\{1, \ln \frac{i}{t}\}} \right) \right) \leqslant$$

$$\leqslant c_3 \left(\frac{\left(\ln \frac{i}{t} \right)^{2k-2}}{it} + \frac{\left(\ln \frac{i}{t} \right)^{2k-2}}{it^2 \min\{1, \ln^2 \frac{i}{t}\}} + \frac{\left(\ln \frac{i}{t} \right)^{k-1}}{i\sqrt{it}} + \frac{\left(\ln \frac{i}{t} \right)^{k-1}}{it\sqrt{it} \min\{1, \ln \frac{i}{t}\}} \right).$$

Итак,
$$\mathrm{E}B_k(t) = O\left(\sum_{i=t+1}^{n} \frac{\left(\ln\frac{i}{t}\right)^{2k-2}}{it} + \frac{\left(\ln\frac{i}{t}\right)^{2k-4}}{it^2} + \frac{\left(\ln\frac{i}{t}\right)^{k-1}}{it} + \frac{\left(\ln\frac{i}{t}\right)^{k-2}}{it^2}\right).$$
Оценим сверху сумму
$$S(l) = \sum_{i=t+1}^{n} \frac{\left(\ln\frac{i}{t}\right)^{l}}{it}.$$
Докажем, что она достаточно мала. Рассматриваем $l > 0$. Нетрудно убедиться, что функция $\frac{\left(\ln\frac{x}{t}\right)^{l}}{xt}$ на отрезке $[t, \infty]$ монотонно растет до точки $x = te^l$ и далее монотонно убывает. В точке $x = te^l$ она равна $\frac{l^l}{t^2e^l}$. Положим $\varphi_l = 1/t^2$, если $te^l < n$, и $\varphi_l = \frac{\left(\ln\frac{n}{t}\right)^{l}}{nt}$ иначе. Сумму $S(l)$ можно вычислить следующим образом (интегральный признак):
$$\sum_{i=t+1}^{n}\left(\frac{\ln(i/t)^l}{it}\right) = \int\limits_t^n \frac{\ln(x/t)^l}{xt}\,dx + O(\varphi_l) = \frac{\ln^{l+1}(n/t)}{t(l+1)} + O(\varphi_l).$$
При $l = 0$ имеем $S(0) = O\left(\frac{\ln n/t}{t}\right)$. Будем считать $\varphi_0 = 0$.

Итак, мы получили оценку на $\mathrm{E}B_k(t)$:
$$\mathrm{E}B_k(t) = O\left(\frac{\ln(n/t)^{2k-1}}{t} + \varphi_{2k-2} + \frac{\ln(n/t)^{2k-3}}{t^2} + \frac{\varphi_{2k-4}}{t} + \right.$$
$$\left. + \frac{\ln(n/t)^{k}}{t} + \varphi_{k-1} + \frac{\left(\ln n/t\right)^{k-1}}{t^2} + \frac{\varphi_{k-2}}{t}\right).$$
Аналогично
$$\mathrm{E}B_{k+1}(t) = O\left(\frac{\ln(n/t)^{2k+1}}{t} + \varphi_{2k} + \frac{\ln(n/t)^{2k-1}}{t^2} + \frac{\varphi_{2k-2}}{t} + \right.$$
$$\left. + \frac{\ln(n/t)^{k+1}}{t} + \varphi_{k} + \frac{\left(\ln n/t\right)^{k}}{t^2} + \frac{\varphi_{k-1}}{t}\right).$$
Нужно проверить, что эти величины малы по сравнению с $f_{k+1}(t,n)$. Нетрудно убедиться, что
$$\frac{\mathrm{E}B_k(t)}{\ln^k(n/t)\sqrt{n/t}} = O\left(\frac{1}{t\min\{1, \ln\frac{n}{t}\}}\right),$$
$$\frac{\mathrm{E}B_{k+1}(t)}{\ln^k(n/t)\sqrt{n/t}} = O\left(\frac{1}{t\min\{1, \ln\frac{n}{t}\}}\right).$$
Это можно проверить по очереди для каждого слагаемого.

Итак, мы вычли все “лишние” ребра и результат от этого не изменился.

Теорема доказана.

1.4 r-диаметр

1.4.1 Введение обозначений и формулировка результатов

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 [20] Боллобаш и Риордан доказали следующую теорему.

Теорема 8. *Для любого $\varepsilon > 0$ и любого натурального $m \geqslant 2$*

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

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

Теорема 9. *Пусть $\varepsilon > 0$ и натуральное $m \geqslant 2$ фиксированы. Пусть, далее, $r(n)$ — произвольная последовательность целых чисел, лежащих в пределах от 2 до n. Тогда*

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

Как и ожидалось, чем больше вершин в рассматриваемых множествах, тем меньше r-диаметр. Отметим, что если для любого $\alpha > 0$ выполнено $r(n) = o\left(n^{\alpha}\right)$, т.е. $r(n) = n^{o(1)}$, то результаты теорем 8 и 9, по существу, идентичны. Если же, например, $r(n) = n^{\alpha}$, $\alpha > 0$, то константа в асимптотике из теоремы 9 принципиально другая: $1 \pm \varepsilon$ превращается в $1 - \alpha \pm \varepsilon$. Если, наконец, $r(n) = n^{1 - o(1)}$, то асимптотики в теореме 9 нет.

Доказательство теоремы 9 естественно разбивается на два шага — оценка снизу и оценка сверху. Нижняя оценка будет доказана в параграфе 1.4.2, верхняя — в параграфе 1.4.3. При этом мы существенно будем опираться на работу [20], особенно в параграфе 1.4.3.

1.4.2 Оценка снизу

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 10. *Пусть* $m \geqslant 1$ *фиксировано. Тогда* $\operatorname{diam}_r(G_m^n) \geqslant \frac{\ln n - \ln r}{\ln(17m^2 \ln n)} - 1$ *с асимптотической вероятностью* 1.

Ясно, что требуемая оценка сразу следует из этой теоремы.

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

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 [20] доказано следующее утверждение.

Лемма 7. *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*

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

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, $\mathcal{F}_n = 2^{\Omega_n}$ — стандартная сигма-алгебра, состоящая из всех подмножеств Ω_n, а вероятностная мера задается следующим образом: для любого $G = (V, E) \in \Omega_n$

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

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

Лемма 8. *Пусть $r < \frac{n}{\ln n}$. Пусть, кроме того, для индикаторных случайных величин ξ_{ij} ($i < j$, $1 \leqslant i, j \leqslant n$) выполнено:* $\mathsf{P}(\xi_{ij} = 1) = O\left(\frac{1}{r \ln n}\right)$. *Тогда с асимптотической вероятностью 1 в графе $G \in \mathfrak{G}([n/2], \xi_{ij})$ есть независимое множество размера хотя бы r.*

Далее будет сначала доказана лемма 8, а затем непосредственно оценка снизу.

Доказательство леммы 8

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

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

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

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

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

Заметим, что проведенное выше рассуждение — это один из способов доказательства известной теоремы Турана в экстремальной теории графов (см. [7]). Мы воспроизвели его прежде всего ради большей замкнутости работы.

Доказательство теоремы 10

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

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

Рассмотрим две вершины $i, j \in A$. Оценим вероятность того, что между этими двумя вершинами найдется путь длины l, $1 \leqslant l \leqslant L$. 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 8. 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 8 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.

1.4.3 Оценка сверху

Формулировка результата и описание модели

Справедлива следующая теорема.

Теорема 11. *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 11 мы будем использовать другое, эквивалентное определение модели G_m^n — *хордовую диаграмму*, которая была описана в работе [20]. 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. [20]):

Лемма 9. *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 $W_1, \dots, W_n$ фиксированы и выполняются события $E_1, \dots, E_4$. 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 $\mathsf{P}(l_{i,j} = k) = \frac{w_k}{W_i}$ для $k \leqslant i$.

Далее мы сперва докажем некоторую техническую лемму, а затем и саму теорему 11.

Техническая лемма

Лемма 10. *Пусть* $\varepsilon > 0$, $r \geqslant \ln^{10} n$, $K = \frac{1}{2}\frac{(1+\varepsilon)\ln n - \ln r}{\ln\ln n} - 1$. *Пусть, далее,* $f_0, f_1, \dots$ *— 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 $f_k > \frac{1}{\left(\ln^3 n\right)\sqrt{n}}$. 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

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

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

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

$$(l-1)\ln(l-1) - (l-1)\ln 500e - \ln 4 = (l-1)\ln(l-1)(1+o(1)) >$$

$$> \frac{1+\varepsilon/2-\log_n r}{2}\frac{\ln n}{\ln\ln n}\left(\ln\frac{1+\varepsilon/2-\log_n r}{2} + \ln\ln n - \ln\ln\ln n\right)(1+o(1)) =$$

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

$$> \ln\sqrt{n} - \ln\sqrt{r} - \ln\ln n.$$

Воспользовались тем, что $\ln\frac{1+\varepsilon/2-\log_n r}{2} = o(\ln\ln n)$, 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. □

Доказательство теоремы 11

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 $\mathrm{diam}_r(G_m^n) \leqslant \frac{(1+\varepsilon)\ln n}{\ln\ln n}$ с асимптотической вероятностью 1. Но в работе [20] 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 [20].

Лемма 11. *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 12. *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 $\frac{1}{2}\frac{(1+\varepsilon)\ln n - \ln r}{\ln\ln n}$ 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 $w_i \geqslant \frac{1}{10\sqrt{in}}$. 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. [20]).

Лемма 13. *Пусть множества $\Gamma_0, \ldots, \Gamma_k$ 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 10 говорит нам о том, что $K \leqslant \frac{1}{2}\frac{(1+\varepsilon)\ln n - \ln r}{\ln\ln n} - 1$.

Если ни на каком шаге (до K) никакой интервал не становится полным, то (из леммы 13) 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 k, $1 \leqslant k \leqslant K$ какой-то из интервалов I_t стал полным. Тогда в множестве $N_k \setminus \{v\}$ 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

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

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

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

$$\leqslant \exp\left(-2 \ln n\right) = n^{-2}.$$

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 11 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 12 (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 11 доказана.

Глава 2

Анализ модели Бакли–Остхуса

В данной главе будет рассмотрена модель, которая является обобщением модели Боллобаша и Риордана. С целью сделать модель более гибкой, две группы авторов (см. [27] и [29]) предложили добавить в модель новый параметр — "привлекательность" вершины (константа, не зависящая от степени). В работе [23] Бакли и Остхус дали строгое определение этой модели. Модель позволяет получить степенное распределение степеней вершин с любым параметром из $(2;\infty)$. В данной главе будут проанализировано распределение вторых степеней вершин в модели Бакли–Остхуса.

Глава основана на работе автора [55].

2.1 Определение модели и известные результаты

Пусть n — количество вершин в графе, а $m \in \mathbb{N}$ и $a \in \mathbb{R}_+$ — фиксированные параметры.

Сначала рассмотрим случай $m = 1$ и индуктивно построим граф $H_{a,1}^n$. Граф $H_{a,1}^1$ состоит из одной вершины и одной петли. Предположим, что граф $H_{a,1}^{t-1}$ уже построен. На шаге t в граф добавляется одна вершина t и одно ребро между вершинами t и i, где i выбрана случайно следующим образом:

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

Здесь $d_{H_{a,1}^t}(s)$ — степень вершины s в графе $H_{a,1}^t$. В дальнейшем мы будем использовать обозначение $d(s) := d_{H_{a,1}^n}(s)$.

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

Заметим, что граф $H^n_{a,m}$ — неориентированный, однако, часто при рассуждениях бывает удобно использовать естественную ориентацию ребер в графе: от вершин с большим номером к вершинам с меньшим. В этом случае мы будем использовать термин *исходящее ребро.*

В работе [23] Бакли и Остхус доказали, что распределение степеней в графе $H^n_{a,m}$ подчиняется степенному закону с параметром $-2-a$, если a — натуральное число. Позже, Е. Гречников значительно улучшил этот результат, доказав следующую теорему.

Теорема 12. (Гречников, [35]) *Пусть $a \in \mathbb{R}_+$. Если $d = d(n) \geqslant m$ и $\psi(n) \to \infty$ при $n \to \infty$, то*

$$\left| R(d,n) - \frac{\mathrm{B}(d-m+ma, a+2)n}{\mathrm{B}(ma, a+1)} \right| \leqslant \left(\sqrt{d^{-2-a}n} + d^{-1} \right) \psi(n)\,,$$

с вероятностью, стремящейся к 1 при $n \to \infty$. Здесь $R(d,n)$ — количество вершин степени d в $H^n_{a,m}$, а $\mathrm{B}(x,y)$ — бета-функция.

Распределение ребер (то есть количество ребер между вершинами заданных степеней) в модели Бакли–Остхуса рассматривалось в работе [35]. Основываясь на этом результате, авторы работы [51] показали, что и распределение степеней и распределение ребер в веб-графе успешно моделируется графом Бакли–Остхуса.

2.2 Введение обозначений и формулировка результатов

2.2.1 Определения

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

$$d_2(t) = \left|\{ij : i \neq t, j \neq t, it \in H^n_{a,1}, ij \in H^n_{a,1}\}\right|,$$

то есть количество ребер, смежных с соседями вершины t, кроме тех, кроме смежны с самой вершиной t. Вершину t назовем *k-вершиной*, если $d_2(t) \geqslant k$.

Через $N_n(d)$ обозначим количество вершин степени d в графе $H^n_{a,1}$. Через $Y_n(k)$ обозначим количество k-вершин в $H^n_{a,1}$.

Кроме того, мы будем рассматривать величину $X_n(k)$, равную количеству вершин второй степени k в $H^n_{a,1}$. Очевидно, что $Y_n(k) = \sum_{i \geqslant k} X_n(i)$.

В данной главе будет доказано, что $Y_n(k)$ убывает как k^{-a}. Это значит, что распределение вторых степеней также подчиняется степенному закону.

Чтобы доказать это, мы сначала оценим математическое ожидание величины $Y_n(k)$, а затем покажем концентрацию этой величины около математического ожидания, используя неравенство Талаграна. Применение неравенства Талаграна в данном случае нетривиально, в частности, мы должны переопределить вероятностное пространство модели Бакли–Остхуса в терминах независимых случайных величин. После этого, мы применим неравенство Талаграна в его общей несимметричной форме. Далее, проверка условий применения неравенства Талаграна является нетривиальной комбинаторной задачей.

2.2.2 Математическое ожидание

Теорема 13. *Для любого $k > 1$ имеем*

$$\mathrm{E}Y_n(k) = \frac{(a+1)\Gamma(2a+1)}{\Gamma(a+1)k^a} n \left(1 + O\left(\frac{(\ln k)^{\lceil a+1 \rceil}}{k}\right) + O\left(\frac{k^{1+a}}{n}\right)\right).$$

Из теоремы 13 несложно получить следующее следствие.

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

Используя ту же технику, что и в доказательстве теоремы 13, можно доказать следующий результат.

Теорема 14. *Для любого $k \geqslant 1$ имеем*

$$\mathrm{E}X_n(k) = \frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)k^{a+1}} \left(1 + O\left(\frac{(\ln k)^{\lceil a+1 \rceil}}{k}\right) + O\left(\frac{k^{1+a}}{n}\right)\right).$$

Из этой теоремы не сложно получить следующее следствие.

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

В этой главе, как и в разделе 1.2, нам потребуется следующее определение. Пусть $N_n(l,k)$ — количество вершин без петель степени l и второй степени k в графе $H^n_{a,1}$:

$$N_n(l,k) = \left|\{t \in H^n_{a,1} : d(t) = l, d_2(t) = k, tt \notin H^n_{a,1}\}\right|.$$

При доказательстве теоремы 13 мы будем использовать следующую вспомогательную теорему.

Теорема 15. *В $H^n_{a,1}$ имеем*

$$\mathrm{E}N_n(l,k) = c(l,k)\,(n + \theta(n,l,k)),$$

где $\theta(n,l,k)$ — некоторая функция от n,l,k такая, что $|\theta(n,l,k)| < C(l+k)^{1+a}$ с некоторой положительной константой C. При этом константы $c(l,k)$ определены следующим образом:

$$\begin{aligned}
c(l,0) &= c(0,k) = 0, \quad l \geqslant 0, k \geqslant 0,\\
c(1,k) &= c(1,k-1)\frac{a+k-1}{k+3a+1} + c(k)\frac{a+k-1}{k+3a+1}, \quad k \geqslant 1,\\
c(l,k) &= \frac{c(l,k-1)(al+k-1)}{l(1+a)+k+2a} + \frac{c(l-1,k)(l-2+a)}{l(1+a)+k+2a}, \quad l \geqslant 2, k \geqslant 1.
\end{aligned}$$

Здесь $c(k) = \frac{\mathrm{B}(k-1+a,a+2)}{\mathrm{B}(a,a+1)}$.

При доказательстве теоремы мы будем использовать две леммы. В работе [35] Е. Гречников получил следующий результат.

Лемма 14. *Пусть $k \geqslant 1$ — натуральное число. Тогда*

$$\mathrm{E}N_n(k) = \frac{\mathrm{B}(k-1+a,a+2)n}{\mathrm{B}(a,a+1)} + \tilde{\theta}(n,k),$$

где $\tilde{\theta}(n,k)$ — некоторая функция от n и k такая, что $|\tilde{\theta}(n,k)| < \tilde{C}/k$ с некоторой константой $\tilde{C}$.

Через $P_n(l,k)$ обозначим количество вершин с петлей степени l и второй степени k в графе $H^n_{a,1}$.

Лемма 15. *Для любого n имеем*

$$\mathrm{E}P_n(l,k) \leqslant p(l,k),$$

где

$$\begin{aligned}
p(2,0) &= P,\\
p(l,0) &= p(l-1,0)\frac{l-2+a}{l(1+a)-2-a}, \quad l \geqslant 3,\\
p(l,k) &= p(l,k-1)\frac{al+k-2a-1}{l(1+a)+k-1-a} + p(l-1,k)\frac{l-2+a}{l(1+a)+k-1-a},\\
&\quad l \geqslant 3, k \geqslant 1.
\end{aligned}$$

Здесь P — некоторая константа. Для других значений l и k имеем $p(l,k) = 0$.

2.2.3 Концентрация

Теорема 16. *Пусть $\delta > 0$ и $k = O\left(n^{\frac{1}{2+a+\delta}}\right)$. Тогда для некоторого $\epsilon > 0$ имеем*

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

Эта теорема показывает концентрацию, то есть тот факт, что распределение вторых степеней действительно подчиняется (асимптотически) степенному закону.

Эта теорема является нетривиальным применением неравенства Талаграна (см. [49]). Вместо неравенства Талаграна можно было бы применить неравенство Азумы, как это было сделано в разделе 1.2, но, как будет показано позже, получился бы более слабый результат.

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

Теорема 17. *Пусть $\delta > 0$ и $k = O\left(n^{\frac{1}{4+a+\delta}}\right)$. Тогда для некоторого $\epsilon > 0$ имеем*

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

Если мы рассмотрим модель Бакли–Остхуса с параметром $a = 1$, то мы получим в точности модель Боллобаша–Риордана [20], которая была рассмотрена в первой главе данной работы. Концентрация для вторых степеней (см. теорему 3) была доказана с использованием неравенства Азумы. Там получена концентрация величины $X_n(k)$ около математического ожидания для всех $k = O\left(n^{\frac{1}{6+\delta}}\right)$ (с некоторой положительной константой δ). Согласно теореме 17, неравенство Талаграна дает более сильный результат: для модели Боллобаша–Риордана получаем концентрацию для всех $k = O\left(n^{\frac{1}{5+\delta}}\right)$. Мы получили это улучшение несмотря на то, что доказательство концентрации $X_n(k)$ в теореме 17 использует концентрацию величины $Y_n(k)$ из теоремы 16, что в некотором смысле не оптимально.

Теорему 16 (и теорему 17) можно обобщить на случай произвольного $m > 1$. Единственная проблемы состоит в том, что не удалось доказать аналог теоремы 13 (или следствия 2) для $m > 1$, поскольку это требует очень большого количества вычислений. Однако можно сделать следующую гипотезу.

Предположение 1. *Для любого $m > 1$ и $k = O\left(n^{\min\left\{\frac{1}{2+a}, \frac{1}{2a}\right\}}\right)$ имеем $\mathrm{E}Y_n^m(k) = \Theta\left(\frac{n}{k^a}\right)$, где $Y_n^m(k)$ — количество k-вершин в графе $H_{a,m}^n$.*

Теорему 16 можно обобщить следующим образом.

Теорема 18. *Допустим, предположение 1 верно. Пусть $m \in \mathbb{N}$, $\delta > 0$ и $k = O\left(n^{\min\left\{\frac{1}{2+a+\delta}, \frac{1}{2a+\delta}\right\}}\right)$. Тогда для некоторого $\epsilon > 0$ имеем*

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

В параграфах 2.3.1 – 2.3.4 будет доказана теорема 16 (с использованием следствия 2). В параграфе 2.3.5 будет доказана теорема 17. В параграфе 2.3.6

изложена идея доказательства теоремы 18. В разделе 2.4 доказаны результаты, сформулированные в параграфе 2.2.2 (теорема 13, теорема 15 и лемма 15 доказаны в параграфах 2.4.2, 2.4.1 и 2.4.3 соответственно). Наконец, теорема 14 доказана в параграфе 2.4.4.

2.3 Доказательство концентрации

2.3.1 Интерпретация модели Бакли–Остхуса в терминах независимых случайных величин

Рассмотрим следующую последовательность:

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

где $\xi_1, \ldots, \xi_n$ — независимые случайные величины. Для каждого i имеем $\xi_i : \Omega_i \to \{1, \ldots, 2i-1\}$ (где $(\Omega_i, \mathcal{F}_i, \mathsf{P}_i)$ — некоторое вероятностное пространство) и

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

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

Полученную последовательность можно интерпретировать следующим образом. Каждое натуральное число i — это вершина графа. Каждая случайная величина ξ_i — второй конец ребра, выходящего из вершины i. Если $\xi_i = 2j-1$, то ребро ведет в вершину j. Если $\xi_i = 2j$, то ребро из вершины i ведет в ту же вершину, что и ребро из вершины j. Значение случайной величины ξ_j тоже может быть четным (то есть $\xi_j = 2j_1$ для некоторого целого j_1), тогда ребро из вершины i далее перенаправляется в соответствии с величиной ξ_{j_1}. Этот процесс останавливается на некотором нечетном значении $2v-1$, и в этом случае мы говорим, что ξ_i (а также ξ_j и ξ_{j_1}) *ведут* в вершину v. Мы также будем говорить, что ξ_i ведет в ξ_j.

Нетрудно проверить, что полученная модель в точности соответствует модели Бакли–Остхуса. Действительно, на каждом шаге i входящая степень каждой вершины $j \in \{1, \ldots, i-1\}$ равна количеству величин, которые ведут (непосредственно или нет) в вершину j.

Приведем еще одну интерпретацию описанной выше модели. Рассмотрим вершину v полученного графа. Можно представить себе, что все величины, которые ведут в v, образуют дерево с корнем v. Пусть $X = \{\xi_{i_1}, \ldots, \xi_{i_d}\}$ — множество величин, которые ведут в v. Построим индуктивно соответствующее дерево на d вершинах $i_1, \ldots, i_d$. Сначала рассмотрим такие величины

$\xi_{i_1^1}, \ldots, \xi_{i_{l_1}^1}$ из X, которые ведут непосредственно в v. Соответствующие вершины $i_1^1, \ldots, i_{l_1}^1$ соединены с вершиной v дерева. Предположим, что мы выбрали все вершины на расстоянии не более s от v и $i_1^s, \ldots, i_{l_s}^s$ — вершины на расстоянии s от v. Рассмотрим множество $\{\xi_{i_1^{s+1}}, \ldots, \xi_{i_{l_{s+1}}^{s+1}}\}$ величин, которые ведут в некоторые величины из $\xi_{i_1^s}, \ldots, \xi_{i_{l_s}^s}$. Мы соединяем каждую вершину $i_1^{s+1}, \ldots, i_{l_{s+1}}^{s+1}$ с соответствующей вершиной из $\{i_1^s, \ldots, i_{l_s}^s\}$. Таким образом получаем множество вершин на расстоянии $s+1$ от v.

2.3.2 Уменьшение количества k-вершин

Фиксируем значение $\mathbf{x} = (x_1, \ldots, x_n)$ случайного вектора $\xi = (\xi_1, \ldots, \xi_n)$ из вероятностного пространства $\Omega = \prod_{i=1}^n \Omega_i$. Величина $Y_n(k)$ — это функция из Ω в $\mathbb{N}$. Нас интересует следующий вопрос. Насколько сильно может уменьшиться величина $Y_n(k) = Y_n(k, \mathbf{x})$, если мы меняем ровно одну координату x_i вектора $\mathbf{x}$? Другими словами, мы хотим найти $c(i, \mathbf{x}) = \max_{\mathbf{x}'}(Y_n(k, \mathbf{x}) - Y_n(k, \mathbf{x}'))$, где $\mathbf{x}'$ — произвольный вектор, который отличается от $\mathbf{x}$ в i-ой координате.

Лемма 16. *Для любых* $\mathbf{x} = (x_1, \ldots, x_n)$ *и* $i \in \{1, \ldots, n\}$ *имеем* $c(i, \mathbf{x}) \leqslant 2k+1$.

Доказательство. При доказательстве этого утверждения удобно работать с интерпретацией случайных величин в виде дерева. Если мы меняем значение x_i величины ξ_i на другое значение x_i', то все величины, которые вели в x_i, перенаправляются в x_i'. Иными словами, мы берем ветвь дерева, в которой содержатся все ребра, ведущие в x_i: если x_i' — нечетно, то мы присоединяем эту ветвь к вершине с номером $(x_i'+1)/2$; если x_i' — четно, то мы присоединяем эту ветвь к величине $\xi_{x_i'/2}$.

Теперь посмотрим, как изменение одной координаты влияет на граф $H_{a,1}^n$. Предположим, что x_i ведет в вершину v. Тогда все величины, которые ведут в x_i, ведут также и в v. Если мы меняем x_i на x_i' и x_i' ведет в v', то мы меняем направление всех величин от v к v'. Или, в терминах графа, мы берем некоторую группу ребер, смежных с v, и перемещаем в вершину v'. А точнее, если сначала в графе были ребра $(i_1, v), \ldots, (i_d, v)$, то после изменения получаем ребра $(i_1, v'), \ldots, (i_d, v')$. Остальной граф не меняется.

Теперь продолжим доказательство. Требуется показать, что после изменения i-ой координаты число k-вершин, которые мы "портим", не превышает $2k+1$. Предположим, что мы переместили группу ребер $(i_1, v), \ldots, (i_d, v)$. Легко видеть, что мы могли испортить только те k-вершины, у которых есть общее ребро с v, либо саму вершину v. Заметим, что мы не могли испортить k-вершины, являющиеся соседями v'.

Разделим множество N_v вершин, смежных с v, на две части: $I = \bigcup_{j=1}^{d}\{i_j\}$ и $N_v \backslash I$. Если $|N_v \backslash I| \geqslant k+1$, то после перемещения ребер все k-вершины из $N_v \backslash I$ остаются k-вершинами. Действительно, все ребра в вершине v, кроме одного, являются 2-смежными ребрами для любого соседа v, поэтому есть хотя бы k таких ребер для каждой вершины из $N_v \backslash I$. Аналогично, если $|I| \geqslant k+1$, то никакая k-вершина из $i_1, \ldots, i_d$ не портится (кроме, возможно, одной), т.к. все смежны с вершиной v'. Единственный случай, в котором некоторые из $i_1, \ldots, i_d$ портятся, — это $i_j = v'$, и поэтому мы не считаем ребра $(i_1, v'), \ldots, (i_d, v')$ во второй степени вершины i_j.

Наконец, количество k-вершин, которые мы портим, не превосходит величины $\min\{|I|, k\} + \min\{|N_v \backslash I|, k\} + 1 \leqslant 2k+1$.

$\square$

Теперь мы хотим оценить влияние каждой случайной величины более аккуратно. Предположим, что $Y_n(k, \mathbf{x}) = q$. Для каждой k-вершины v_j, $j = 1, \ldots, q$, рассмотрим подмножество координат $K_j = K_{v_j}(\mathbf{x}) = \{i_1^j, \ldots, i_{d_j}^j\}$ таких, что v_j является k-вершиной для любого $\mathbf{y}$, который совпадает с $\mathbf{x}$ на координатах из K_j. Ясно, что K_j зависит от $\mathbf{x}$, но не определяется им однозначно. Для каждого выбора множеств $K_1, \ldots, K_q$, обозначим их множество через $\mathcal{K} = \mathcal{K}(\mathbf{x})$. Ясно, что $Y_n(k, \mathbf{y}) \geqslant q$ для любого $\mathbf{y}$, который совпадает с $\mathbf{x}$ на всех координатах из всех множеств $K_j \in \mathcal{K}$.

Для каждой координаты i определим ее кратность $C(i, \mathbf{x}, \mathcal{K}) = |\{j : i \in K_j\}|$.

Легко видеть, что для любого $\mathbf{x}$ и любого $\mathcal{K}$ имеем $c(i, \mathbf{x}) \leqslant C(i, \mathbf{x}, \mathcal{K})$. Поэтому

$$c(i, \mathbf{x}) \leqslant \min\{2k+1, C(i, \mathbf{x}, \mathcal{K})\} =: c_i(\mathbf{x}, \mathcal{K}).$$

Назовем коллекцию множеств $\mathcal{K}$ *устойчивой*, если для любой k-вершины v_i мы строим все множества K_i согласно следующему правилу: если K_i содержит некоторую координату, ведущую в вершину w, то K_i содержит все координаты, ведущие в w.

Неудивительно, что для такой системы мы можем доказать аналог леммы 16. А именно, рассмотрим вектор $\mathbf{x}$, соответствующий k-вершинам $v_j, j = 1, \ldots, q$, и некоторую устойчивую коллекцию $\mathcal{K}$. Пусть $i \in \{1, \ldots, n\}$ и $\mathcal{K} \setminus \{i\} := \{K_j \setminus \{i\}, j = 1, \ldots, q\}$. Для любого i существует хотя бы $q - c_i(\mathbf{x}, \mathcal{K})$ таких k-вершин, которые являются k-вершинами для любого $\mathbf{x}'$ с $x'_s = x_s$ для всех $s \in K_j \setminus \{i\}, j = 1, \ldots, q$. Можно доказать этот факт, следуя доказательству леммы 16 и убеждаясь, что доказательство работает и в этом случае тоже. Единственное дополнительное рассуждение — количество k-вершин, которые мы теряем, не может превосходить кратность $C(j, \mathbf{x}, \mathcal{K})$.

Лемма 17. *Пусть $\mathcal{K}$ — некоторая устойчивая коллекция. И пусть J — некоторое подмножество $\{1, \ldots, n\}$. Тогда выполнено неравенство $Y_n(k, \mathbf{x}) - Y_n(k, \mathbf{x}') \leqslant \sum_{j \in J} c_j(\mathbf{x}, \mathcal{K})$ для любого такого вектора $\mathbf{x}'$, что $x'_i = x_i$ для всех $i \in \{1, \ldots, n\} \backslash J$.*

Доказательство. Предположим, что мы изменили одну координату j вектора $\mathbf{x}$ и получили вектор $\hat{\mathbf{x}}$. Рассмотрим $d_j := C(j, \mathbf{x}, \mathcal{K})$ k-вершин $w_1, \ldots, w_{d_j}$ таких, что $j \in K_{w_i}(\mathbf{x})$. Удалим j из каждого такого множества и для каждого $i = 1, \ldots, d_j$ посмотрим, гарантирует ли полученная коллекция то, что w_i является k-вершиной. Если w_i является k-вершиной, то определим $K_{w_i}(\hat{\mathbf{x}}) = K_{w_i}(\mathbf{x}) \backslash \{j\}$. Если w_i не является k-вершиной, то исключим множество $K_{w_i}(\mathbf{x})$ из $\mathcal{K}$. В итоге мы получим новую коллекцию $\hat{\mathcal{K}}$. Чтобы доказать лемму, требуется следующее рассуждение. А именно, вместо изменения ребер $(i_1, v), \ldots, (i_d, v)$ на $(i_1, v'), \ldots, (i_d, v')$, мы можем создать новую, вымышленную, вершину w и изменить ребра на $(i_1, w), \ldots, (i_d, w)$. Обозначим полученный граф G_w. Мы не считаем вершину w за k-вершину, даже если у нее $\geqslant k$ 2-смежных ребер. Легко проверить, что для такого графа коллекция $\hat{\mathcal{K}}$ является устойчивой.

Количество k-вершин (кроме w) в графе G_w заведомо не превосходит аналогичного количества в графе, соответствующем $\hat{\mathbf{x}}$. Более того, кратность каждой координаты в $\hat{\mathcal{K}}$ меньше либо равна соответствующей кратности в $\mathcal{K}$. Кроме того, имеем $Y_n(k, \mathbf{x}) - Y_n(k, G_w) \leqslant c_j(\mathbf{x}, \mathcal{K})$. Аналогично, если $\mathbf{x}'$ отличается от $\mathbf{x}$ в l координатах, то граф, соответствующий $\mathbf{x}'$, имеет хотя бы столько k-вершин, сколько граф G, полученный добавлением l воображаемых вершин. Более того, на каждом шаге (если мы меняем координату j' и формируем соответствующий граф G') мы портим не более $\min\{2k+1, C(j', \mathbf{x}, \mathcal{K})\}$ k-вершин и получаем устойчивую систему множеств.

Таким образом, $Y_n(k, \mathbf{x}) - Y_n(k, \mathbf{x}') \leqslant Y_n(k, \mathbf{x}) - Y_n(k, G) \leqslant \sum_{j \in J} c_j(\mathbf{x}, \mathcal{K})$. □

2.3.3 Построение подходящей системы множеств $\mathcal{K}$

Лемма 18. *Пусть $Y_n(k, \mathbf{x}) = q$ для некоторого вектора $\mathbf{x}$ в соответствующем графе $G_{\mathbf{x}}$. Тогда можно построить устойчивую систему множеств $\mathcal{K} = \{K_1, \ldots, K_q\}$ такую, что $\sum_{i=1}^{n} c_i(\mathbf{x}, \mathcal{K}) \leqslant (4k+5)q$.*

Доказательство. Сначала рассмотрим множество V вершин степени хотя бы $k+2$. Положим $NV = \{u : u$ — сосед $v \in V\}$. Заметим, что вершина из V может также принадлежать NV. Пусть $|NV| = z$. Все вершины из NV являются k-вершинами. Пусть BV — множество вершин из NV, которые не имеют исходящего ребра, ведущего в V. Имеем $|BV| \leqslant z/(k+1)$, поскольку каждая вершина имеет не более одного исходящего ребра.

Через $L_v, L_v \subset \{1, \ldots, n\}$ обозначим множество координат, которые ведут в v. Далее, положим $LV = \cup_{v \in V} L_v$ и $LBV = \cup_{v \in BV} L_v$. Для любого $u \in NV$ положим $K_u = LV \cup LBV$. Легко видеть, что для любого $\mathbf{x}'$ такого, что $x_i = x'_i$ для всех $i \in LV \cup LBV$, вершина u является k-вершиной в графе, соответствующем $\mathbf{x}'$.

Для $i \in LV \cup LBV$ оценим $c_i(\mathbf{x}, \mathcal{K})$ величиной $2k+1$. Заметим, что $|LV| \leqslant z + \frac{z}{k}$. Мы добавляем $\frac{z}{k}$ величин, поскольку вершины из V могут иметь петли. Имеем $\deg w \leqslant k+1$ для $w \in BV \backslash V$ и $|BV \backslash V| \leqslant z/(k+1)$, поскольку $|LBV \backslash LV| \leqslant z$. Поэтому

$$\sum_{i \in LV \cup LBV} c_i(\mathbf{x}, \mathcal{K}) \leqslant (2k+1)(|LBV \backslash LV| + |LV|) \leqslant (2k+1)\left(2z + \frac{z}{k}\right) \leqslant (4k+5)z.$$

Далее, рассмотрим множество W оставшихся k-вершин. Имеем $|W| = q - z$. По определению, для любого $w \in W$ все соседи N_w вершины w имеют степени, меньшие либо равные $k+1$.

Для каждого $w \in W$ рассмотрим $V_w = \{v_1, \ldots, v_w\}$, $V_w \subset N_w$, такое, что количество ребер, смежных хотя бы с одной вершиной $v_i \in N_w$ и не смежных с w, находится между k и $2k$. Такое V_w можно найти, поскольку w является k-вершиной. Мы можем выбирать $v_i \in N_w$ по одной до тех пор, пока общее число 2-смежных ребер не станет больше k. В то же время, оно не может стать больше $2k$, поскольку $\deg v_i \leqslant k+1$ для $v_i \in N_w$. Через LV_w обозначим все величины, которые ведут в V_w.

Теперь для любого $w \in W$ положим $K_w = LV_w \cup L_w$. Заметим, что $|LV_w| \leqslant 2k$, и для $w \in W \backslash V$ имеем $|L_w| \leqslant k+1$.

Теперь мы можем получить итоговую оценку:

$$\sum_{i=1}^{n} c_i(\mathbf{x}, \mathcal{K}) = \sum_{i \in LV \cup LBV} c_i(\mathbf{x}, \mathcal{K}) + \sum_{i \in \{1,\ldots,n\} \backslash LV \cup LBV} c_i(\mathbf{x}, \mathcal{K}) \leqslant (4k+5)z +$$

$$+ \sum_{w \in W} |LV_w| + \sum_{w \in W \backslash V} |L_w| \leqslant (4k+5)z + 2k(q-z) + (k+1)(q-z) \leqslant (4k+5)q.$$

Тот факт, что $\mathcal{K}$ — устойчивая коллекция, следует из построения. $\square$

2.3.4 Применение неравенства Талаграна

Для начала сформулируем неравенство Талаграна (см., например, [11]).

Пусть $\Omega = \prod_{i=1}^{n} \Omega_i$ является прямым произведением вероятностных пространств. Пусть, кроме того, мера на Ω является прямым произведением мер на Ω_i. Далее, пусть $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\sum_{i=1}^{n} \alpha_i^2 = 1$. Определим расстояние между множеством $A \subset \Omega$ и точкой $\mathbf{x} \in \Omega$ следующим образом:

$$\operatorname{dist}(A, \mathbf{x}) = \max_{\alpha} \min_{\mathbf{y} \in A} \sum_{i \in I_{\mathbf{xy}}} \alpha_i,$$

где $I_{\mathbf{xy}} = \{i : x_i \neq y_i\}$.

Для $t > 0$ через A_t обозначим множество $\{\mathbf{x} : \operatorname{dist}(A, \mathbf{x}) \leqslant t\}$.

Теорема 19. (Неравенство Талаграна) *Для любого $t > 0$ и $A \subset \Omega$ выполнено* $\mathsf{P}(A)(1 - \mathsf{P}(A_t)) \leqslant e^{-\frac{t^2}{4}}$.

Используя это неравенство, мы получим следующую теорему.

Теорема 20. *Для $t > 0$, $k, s \in \mathbb{N}$ и $f(s)$, удовлетворяющей условию $f^2(s) > (2k+1)(4k+5)s$, выполнено* $\mathsf{P}\big(Y_n(k) \leqslant s - tf(s)\big)\, \mathsf{P}\big(Y_n(k) \geqslant s\big) \leqslant e^{-\frac{t^2}{4}}$.

Доказательство. Неравенство очевидно для $tf(s) > s$, поэтом далее предполагаем, что $tf(s) \leqslant s$. Поскольку величины ξ_i независимы, мы можем применить неравенство Талаграна к точкам $\mathbf{x}$ вероятностного пространства Ω. Нужно доказать, что для любого вектора $\mathbf{x}$ такого, что $Y_n(k, \mathbf{x}) \geqslant s$, выполнено $\mathbf{x} \notin A_t$, где $A = \{\mathbf{y} : Y_n(k, \mathbf{y}) \leqslant s - tf(s)\}$.

Пусть $Y_n(k, \mathbf{x}) = q \geqslant s$. Для вектора $\mathbf{x}$ мы фиксируем систему множеств $\mathcal{K}$ из леммы 18. Далее, по лемме 17 для любого $\mathbf{y} \in A$ выполнено $q - s + tf(s) \leqslant Y_n(k, \mathbf{x}) - Y_n(k, \mathbf{y}) \leqslant \sum_{j \in I_{\mathbf{xy}}} c_j(\mathbf{x}, \mathcal{K})$.

Построим подходящий вектор $\alpha = \alpha(\mathbf{x})$. А именно, $\alpha_i = \frac{c_i(\mathbf{x}, \mathcal{K})}{\sqrt{\sum_{j=1}^n c_j^2(\mathbf{x}, \mathcal{K})}}$. Легко видеть, что $\sum \alpha_j^2 = 1$.

Имеем

$$\sum_{j=1}^n c_j^2(\mathbf{x}, \mathcal{K}) \leqslant \max_j c_j(\mathbf{x}, \mathcal{K}) \sum_{j=1}^n c_j(\mathbf{x}, \mathcal{K}) \leqslant (2k+1)(4k+5)q.$$

В последнем неравенстве использовалась лемма 18 и определение $c_j(\mathbf{x}, \mathcal{K})$.

Теперь покажем, что $\sum_{i \in I_{\mathbf{xy}}} \alpha_i > t$ для любого $\mathbf{y} \in A$.

$$\sum_{i \in I_{\mathbf{xy}}} \alpha_i = \frac{\sum_{i \in I_{\mathbf{xy}}} c_i(\mathbf{x}, \mathcal{K})}{\sqrt{\sum_{j=1}^n c_j^2(\mathbf{x}, \mathcal{K})}} \geqslant$$

$$\geqslant \frac{q - s + tf(s)}{\sqrt{(2k+1)(4k+5)q}} \geqslant \frac{tf(s)}{\sqrt{(2k+1)(4k+5)s}} > t. \qquad (2.1)$$

Второе неравенство выполнено, поскольку для $q \geqslant s, tf(s) \leqslant s$ имеем $\frac{q-s+tf(s)}{q} \geqslant \frac{tf(s)}{s}$. Последнее неравенство следует из условия теоремы.

Из (2.1) получаем, что $\mathrm{dist}(A, \mathbf{x}) > t$, иными словами, $\mathbf{x} \notin A_t$. $\square$

Мы применим теорему 20 с $t = 2\ln n$, $s = m(Y_n(k)) + t(\mathrm{E}Y_n(k))^{1-\varepsilon}$, $f(s) = (\mathrm{E}Y_n(k))^{1-\varepsilon}$. Здесь $m(Y_n(k))$ — медиана значений $Y_n(k)$, и, следовательно, $\mathsf{P}\big(Y_n(k) \leqslant s - tf(s)\big) \geqslant 1/2$. Поскольку для любой случайной величины Z выполнено $m(Z) \leqslant 2\mathrm{E}Z$, легко видеть, что условия теоремы 20 выполнены, если

$$(\mathrm{E}Y_n(k))^{1-2\varepsilon} \geqslant 12(2k+1)^2 \ln n.$$

Для достаточно маленького ε это неравенство является следствием следствия 2 и условий теоремы 16.

Мы получили, что

$$\mathsf{P}\Big(Y_n(k) \geqslant m(Y_n(k)) + 2\ln n(\mathrm{E}Y_n(k))^{1-\varepsilon}\Big) \leqslant 2e^{-\frac{t^2}{4}} = o(1/n),$$

и, поскольку $Y_n(k) \leqslant n$ для всех k, мы имеем

$$\mathrm{E}Y_n(k) \leqslant m(Y_n(k)) + 2\ln n(\mathrm{E}Y_n(k))^{1-\varepsilon} + o(1).$$

Подобным образом мы можем получить, что

$$\mathsf{P}\Big(Y_n(k) \leqslant m(Y_n(k)) - 2\ln n(\mathrm{E}Y_n(k))^{1-\varepsilon}\Big) \leqslant 2e^{-\frac{t^2}{4}} = o(1/n)$$

и

$$\mathrm{E}Y_n(k) \geqslant m(Y_n(k)) - 2\ln n(\mathrm{E}Y_n(k))^{1-\varepsilon} - o(1).$$

Следовательно, для некоторого $\delta > 0$ и всех достаточно больших n мы имеем $|\mathrm{E}Y_n(k) - m(Y_n(k))| \leqslant (\mathrm{E}Y_n(k))^{1-\delta}$. Поэтому для некоторого $\varepsilon' > 0$

$$\mathsf{P}\left(|Y_n(k) - \mathrm{E}(Y_n(k))| > \mathrm{E}(Y_n(k))^{1-\varepsilon'}\right) = o(1).$$

Это завершает доказательство теоремы 16.

2.3.5 Концентрация для $X_n(k)$

Докажем теорему 17. Очевидно, что $X_n(k) = Y_n(k) - Y_n(k+1)$. Зафиксируем некоторое число $\epsilon' > 0$. Сначала мы хотим применить теорему 20 к $Y_n(k)$ и $Y_n(k+1)$. Будем рассуждать, как после доказательства теоремы 20. Положим $f(s) = \frac{n^{1-\epsilon'}}{k^{1+a}}$, $t = 2\ln n$ и $s_1 = m(Y_n(k)) + tf(s), s_2 = m(Y_n(k)), s_3 = m(Y_n(k+1)) + tf(s), s_4 = m(Y_n(k+1))$.

Применим теорему 20 к $Y_n(k)$ с s_1 и s_2 и к $Y_n(k+1)$ с s_3 и s_4 и получим

$$\mathsf{P}\left(|Y_n(k) - \mathrm{E}(Y_n(k))| > \frac{n^{1-\epsilon'+o(1)}}{k^{1+a}}\right) = o(1),$$

$$\mathsf{P}\left(|Y_n(k+1) - \mathrm{E}(Y_n(k+1))| > \frac{n^{1-\epsilon'+o(1)}}{k^{1+a}}\right) = o(1),$$

если

$$\frac{n^{2-2\epsilon'}}{k^{2+2a}} \geqslant \Theta\left(nk^{2-a}\right)\ln n.$$

Легко видеть, что это выполнено, если условия теоремы 17 выполнены для некоторого $\delta > 0$. Имеем

$$|X_n(k) - \mathrm{E}(X_n(k))| \leqslant |Y_n(k) - \mathrm{E}(Y_n(k))| + |Y_n(k+1) - \mathrm{E}(Y_n(k+1))|,$$

поэтому

$$\mathsf{P}\left(|X_n(k) - \mathrm{E}(X_n(k))| > \frac{n^{1-\epsilon'+o(1)}}{k^{1+a}}\right) = o(1).$$

Поскольку $\frac{n^{1-\epsilon'+o(1)}}{k^{1+a}} = \mathrm{E}(X_n(k))^{1-\epsilon}$ для некоторого $\epsilon > 0$, это завершает доказательство теоремы 17.

2.3.6 Обобщение на случай произвольного m

Доказательство теоремы 20 может быть адаптировано к графу $H^n_{a,m}$. Чтобы не повторять все рассуждения, здесь мы представляем только схему доказательства. Пусть $m > 1$ фиксировано. Количество случайных величин теперь не n, а mn. Интерпретация в терминах независимых случайных величин работает и этом случае тоже. Леммы 17, 18 выполнены для $m > 1$, но с некоторыми незначительными изменениями.

Когда мы берем группу ребер, смежных с вершиной v, и перемещаем ее в некоторую вершину v', мы можем испортить не только соседей v, но и вершину v'. А именно, пусть мы изменили ребра $(v, w_1), \ldots, (v, w_l)$ на $(v', w_1), \ldots, (v', w_l)$. Если v' была k-вершиной и ребра $(v, w_1), \ldots, (v, w_l)$ входили во вторую степень v', то вторая степень v' может уменьшиться (что невозможно в графе $H^n_{a,1}$, поскольку $H^n_{a,1}$ — дерево). Нетрудно видеть, что мы не можем испортить другие вершины.

Теперь мы можем сформулировать аналоги лемм 16 и 17.

Лемма 19. *Для любого $\mathbf{x} = (x_1, \ldots, x_n)$ и $i \in \{1, \ldots, n\}$ выполнено $c(i, \mathbf{x}) \leqslant 2k + 2$.*

Кроме того,

$$c(i, \mathbf{x}) - 1 \leqslant \min\{2k+1, C(i, \mathbf{x}, \mathcal{K})\} =: c_i(\mathbf{x}, \mathcal{K}).$$

Лемма 20. *Пусть $\mathcal{K}$ — устойчивая система множеств. Выполнено $Y_n(k, \mathbf{x}) - Y_n(k, \mathbf{x}') \leqslant \sum_{j \in J}(c_j(\mathbf{x}, \mathcal{K}) + 1)$ для любого вектора $\mathbf{x}'$ такого, что $x'_i = x_i$ для всех $i \in \{1, \ldots, n\} \backslash J$.*

Лемма 18 выполнена для $c_j(\mathbf{x},\mathcal{K})$ и $m>1$ без изменений.

Осталось адаптировать доказательство теоремы 20. Положим

$$\alpha_i = \frac{c_i(\mathbf{x},\mathcal{K})+1}{\sqrt{\sum_{j=1}^{mn}(c_j(\mathbf{x},\mathcal{K})+1)^2}}.$$

Тогда

$$\sum_{j=1}^{mn}(c_j(\mathbf{x},\mathcal{K})+1)^2 \leqslant \left(\max_j c_j(\mathbf{x},\mathcal{K})+2\right)\sum_{j=1}^{mn} c_j(\mathbf{x},\mathcal{K})+mn \leqslant$$
$$\leqslant (2k+3)(4k+5)q+mn.$$

В итоге

$$\sum_{i\in I_{\mathbf{xy}}}\alpha_i = \frac{\sum_{i\in I_{\mathbf{xy}}}(c_i(\mathbf{x},\mathcal{K})+1)}{\sqrt{\sum_{j=1}^{mn}(c_j(\mathbf{x},\mathcal{K})+1)^2}} \geqslant$$
$$\geqslant \frac{q-s+tf(s)}{\sqrt{(2k+3)(4k+5)q+mn}} \geqslant \frac{tf(s)}{\sqrt{(2k+3)(4k+5)s+mn}} > t,$$

если $f^2(s)>(2k+3)(4k+5)s+mn$. Теперь мы можем сформулировать аналог теоремы 20.

Теорема 21. *Для $t>0$, $m,k,s\in\mathbb{N}$ и $f(s)$, удовлетворяющей условию $f^2(s)>(2k+3)(4k+5)s+mn$, выполнено*

$$\mathsf{P}\big(Y_n^m(k)\leqslant s-tf(s)\big)\,\mathsf{P}\big(Y_n^m(k)\geqslant s\big)\leqslant e^{-\frac{t^2}{4}}.$$

Наконец, рассуждая как после теоремы 20, можно увидеть, что теорема 18 следует из теоремы 21 и предположения 1.

2.4 Оценка $\mathrm{E}Y_n(k)$

2.4.1 Доказательство теоремы 15

Из определения графа $H_{a,1}^n$ следует, что $N_n(l,0)=N_n(0,k)=0$. Действительно, поскольку нет вершин степени 0, выполнено $N_n(0,k)=0$. Поскольку вершины с петлями не учитываются в $N_n(l,k)$, у нас нет вершин второй степени 0 и $N_n(l,0)=0$. Поэтому $\mathrm{E}N_n(l,0)=\mathrm{E}N_n(0,k)=0$. Мы хотим доказать, что существует такая константа C, что

$$\mathrm{E}N_n(l,k)=c(l,k)\,(n+\theta(n,l,k)),$$

где $|\theta(n,l,k)|<C(l+k)^{1+a}$.

Покажем, что $\mathrm{E}N_n(1,k) = c(1,k)\ \left(n + \theta\left(Ck^{1+a}\right)\right)$. Будем использовать индукцию по k. Для $k = 0$ утверждение очевидно.

Предположим, что для $j < k$ и для всех n выполнено

$$\mathrm{E}N_n(1,j) = c(1,j)\left(n + \theta\left(Cj^{1+a}\right)\right).$$

Через $N_n(l)$ обозначим количество вершин степени l в $H^n_{a,1}$. Чтобы доказать, что $\mathrm{E}N_n(1,k) = c(1,k)\ \left(n + \theta\left(Ck^{1+a}\right)\right)$, мы используем индукцию по числу вершин и равенство

$$\begin{aligned} &\mathrm{E}(N_{i+1}(1,k)|N_i(1,k), N_i(1,k-1), N_i(k)) = \\ &= N_i(1,k)\left(1 - \frac{k+2a}{(a+1)i+a}\right) + \frac{(k-1+a)N_i(1,k-1)}{(a+1)i+a} + \frac{(k-1+a)\,N_i(k)}{(a+1)i+a}. \end{aligned} \tag{2.2}$$

Поясним это равенство. Пусть мы построили граф $H^i_{a,1}$. Мы добавляем одну вершину и одно ребро. В графе $H^i_{a,1}$ есть $N_i(1,k)$ вершин степени 1 и второй степени k. Вероятность того, что мы "портим" одну из таких вершин, равна $\frac{k+2a}{(a+1)i+a}$. Далее, есть $N_i(1,k-1)$ вершин степени 1 и второй степени $k-1$. Вероятность того, что одна такая вершина будет иметь степень 1 и вторую степень k в $H^{i+1}_{a,1}$, равна $\frac{k-1+a}{(a+1)i+a}$. Наконец, с вероятностью $\frac{k-1+a}{(a+1)i+a}$ вершина $i+1$ будет иметь нужные степени в $H^{i+1}_{a,1}$. Из равенства (2.2) получаем:

$$\begin{aligned} \mathrm{E}N_{i+1}(1,k) = \mathrm{E}N_i(1,k)&\left(1 - \frac{k+2a}{(a+1)i+a}\right) + \\ &+ \frac{(k-1+a)\mathrm{E}N_i(1,k-1)}{(a+1)i+a} + \frac{(k-1+a)\,\mathrm{E}N_i(k)}{(a+1)i+a}. \end{aligned} \tag{2.3}$$

Заметим, что если есть хотя бы одна вершина степени 1 и второй степени k в $H^i_{a,1}$, то в этом графе есть хотя бы k ребер. Поэтому $\mathrm{E}N_i(1,k) = 0$ для $i < k$. Рассмотрим случай $i = k$. Во-первых, заметим, что

$$\mathrm{E}N_k(1,k) \geqslant c(1,k)\left(k + \theta\left(Ck^{1+a}\right)\right)$$

с некоторым C, поскольку мы можем положить $\theta\left(Ck^{1+a}\right) = -k$ для достаточно большого C. Для конечного числа маленьких k мы можем найти константу C такую, что

$$\mathrm{E}N_k(1,k) = c(1,k)\left(k + \theta\left(Ck^{1+a}\right)\right).$$

Используя (2.3), лемму 14 и предположения теоремы, мы получаем

$$\mathrm{E}N_k(1,k) = \mathrm{E}N_{k-1}(1,k-1)\frac{k-1+a}{ak+k-1} + \mathrm{E}N_{k-1}(k)\frac{k-1+a}{ak+k-1} =$$
$$= c(1,k-1)\frac{k-1+a}{ak+k-1}\left(k-1+\theta\left(C\left(k-1\right)^{1+a}\right)\right)+$$

$$+c(k)\frac{k-1+a}{ak+k-1}\left(k-1+\theta\left(C_1k^{1+a}\right)\right) =$$
$$= c(1,k)\frac{(k+3a+1)(k-1)}{ak+k-1} + c(1,k-1)\frac{k-1+a}{ak+k-1}\theta\left(C\left(k-1\right)^{1+a}\right)+$$
$$+c(k)\frac{k-1+a}{ak+k-1}\theta\left(C_1k^{1+a}\right) = kc(1,k)+$$
$$+\frac{3ak+k-3a-1-ak^2}{ak+k-1}c(1,k) + c(1,k-1)\frac{k-1+a}{ak+k-1}\theta\left(C\left(k-1\right)^{1+a}\right)+$$
$$+c(k)\frac{k-1+a}{ak+k-1}\theta\left(C_1k^{1+a}\right) \leqslant$$
$$\leqslant kc(1,k) + \frac{(3ak+k-3a-1-ak^2)(a+k-1)}{(ak+k-1)(k+3a+1)}c(1,k-1)+$$
$$+\frac{(3ak+k-3a-1-ak^2)(a+k-1)}{(ak+k-1)(k+3a+1)}c(k)+$$
$$+c(1,k-1)\frac{C(k-1+a)}{ak+k-1}\left(k-1\right)^{1+a} + c(k)\frac{k-1+a}{ak+k-1}C_1k^{1+a} \leqslant$$
$$\leqslant kc(1,k) + \frac{C(a+k-1)}{k+3a+1}c(1,k-1)k^{1+a} + \frac{C(a+k-1)}{k+3a+1}c(k)k^{1+a}.$$

Это верно для больших значений k. Действительно,

$$\frac{(3ak+k-3a-1-ak^2)}{(ak+k-1)(k+3a+1)} + \frac{C\left(k-1\right)^{1+a}}{ak+k-1} \leqslant \frac{C}{k+3a+1}k^{1+a},$$

если k и C достаточно большие.

Рассмотрим случай $i \geqslant k$. Используя (2.3), лемму 14 и предположение индукции, мы получим

$$\mathrm{E}N_{i+1}(1,k) = \mathrm{E}N_i(1,k)\left(1-\frac{k+2a}{(a+1)i+a}\right)+$$
$$+\mathrm{E}N_i(1,k-1)\frac{k-1+a}{(a+1)i+a} + \mathrm{E}N_i(k)\frac{k-1+a}{(a+1)i+a} =$$
$$= c(1,k)\left(i+\theta\left(Ck^{1+a}\right)\right)\left(1-\frac{k+2a}{(a+1)i+a}\right)+$$

$$+c(1,k-1)\left(i+\theta\left(C(k-1)^{1+a}\right)\right)\frac{k-1+a}{(a+1)i+a}+$$

$$+c(k)\left(i+\theta_1\left(C_1k^{1+a}\right)\right)\frac{k-1+a}{(a+1)i+a}=c(1,k)(i+1)-$$

$$-c(1,k)\frac{i(k+3a+1)+a}{(a+1)i+a}+c(1,k)\theta\left(Ck^{1+a}\right)\left(1-\frac{k+2a}{(a+1)i+a}\right)+$$

$$+c(1,k-1)i\frac{k-1+a}{(a+1)i+a}+c(1,k-1)\theta\left(C(k-1)^{1+a}\right)\frac{k-1+a}{(a+1)i+a}+$$

$$+c(k)i\frac{k-1+a}{(a+1)i+a}+c(k)\theta_1\left(C_1k^{a+1}\right)\frac{k-1+a}{(a+1)i+a}=$$

$$=c(1,k)(i+1)+c(1,k)\theta\left(Ck^{1+a}\right)\left(1-\frac{k+2a}{(a+1)i+a}\right)-$$

$$-\frac{(k-1+a)ac(1,k-1)}{((a+1)i+a)(k+3a+1)}-\frac{(k-1+a)ac(k)}{((a+1)i+a)(k+3a+1)}+$$

$$+c(1,k-1)\theta\left(C(k-1)^{1+a}\right)\frac{k-1+a}{(a+1)i+a}+c(k)\theta_1\left(C_1k^{a+1}\right)\frac{k-1+a}{(a+1)i+a}.$$

Мы хотим доказать, что существует константа C такая, что

$$c(1,k)Ck^{1+a}\frac{k+2a}{(a+1)i+a}\geqslant\frac{(k-1+a)ac(1,k-1)}{((a+1)i+a)(k+3a+1)}+$$

$$+\frac{(k-1+a)ac(k)}{((a+1)i+a)(k+3a+1)}+$$

$$+c(1,k-1)C(k-1)^{1+a}\frac{k-1+a}{(a+1)i+a}+c(k)C_1k^{1+a}\frac{k-1+a}{(a+1)i+a}.$$

Достаточно доказать, что выполнены следующие неравенства:

$$\frac{c(1,k-1)Ck^{1+a}(k+2a)(k-1+a)}{((a+1)i+a)(k+3a+1)}\geqslant\frac{(k-1+a)ac(1,k-1)}{((a+1)i+a)(k+3a+1)}+$$

$$+\frac{c(1,k-1)C(k-1)^{1+a}(k-1+a)}{(a+1)i+a}$$

и

$$\frac{c(k)Ck^{1+a}(k+2a)(k-1+a)}{((a+1)i+a)(k+3a+1)}\geqslant\frac{(k-1+a)ac(k)}{((a+1)i+a)(k+3a+1)}+$$

$$+\frac{c(k)C_1k^{1+a}(k-1+a)}{(a+1)i+a}.$$

Или

$$Ck^{1+a}(k+2a)\geqslant a+C(k-1)^{1+a}(k+3a+1),$$

$$Ck^{1+a}(k+2a) \geqslant a + C_1 k^{1+a}(k+3a+1).$$

Заметим, что

$$k^{1+a}(k+2a) - (k-1)^{1+a}(k+3a+1) =$$
$$= k^{1+a}(k+2a) - (k^{1+a} - (1+a)k^a + \frac{a(a+1)}{2}k^{a-1} +$$
$$+O(k^{a-2}))(k+3a+1) = \frac{(5a+2)(a+1)}{2}k^a + O(k^{a-1}).$$

Для больших значений k существует такая константа C, что

$$C(k^{1+a}(k+2a) - (k-1)^{1+a}(k+3a+1)) \geqslant a.$$

Но мы не можем выбрать константу C, если $k^{1+a}(k+2a) \leqslant (k-1)^{1+a}(k+3a+1)$. Существует конечное число значений k с $\frac{(5a+2)(a+1)}{2}k^a + O(k^{a-1}) \leqslant 0$. Для таких k мы можем доказать, что $\mathrm{E}N_n(1,k) = c(1,k)\ (n + O\left(f(k)\right))$ с некоторой функцией $f(k)$. Используя те же методы, что и выше, мы получаем следующие неравенства:

$$f(k)(k+2a) \geqslant a + f(k-1)(k+3a+1),$$
$$f(k)(k+2a) \geqslant a + C_1 k^{1+a}(k+3a+1).$$

Существует такая функция f, что неравенства выполнены. Это завершает доказательство для $\mathrm{E}N_n(1,k)$.

Рассмотрим случай $l \geqslant 2$. Пусть мы уже доказали, что

$$\mathrm{E}N_n(i,j) = c(i,j)\ \left(n + \theta\left(C(i+j)^{1+a}\right)\right)$$

для всех i и j таких, что $i < l$, $j \leqslant k$ или $i \leqslant l$, $j < k$, и для всех n.

Мы используем следующее равенство, похожее на (2.3):

$$\mathrm{E}N_{i+1}(l,k) = \mathrm{E}N_i(l,k)\left(1 - \frac{l(1+a)+k+a-1}{(a+1)i+a}\right) + \\ + \frac{(l-2+a)\,\mathrm{E}N_i(l-1,k)}{(a+1)i+a} + \frac{(k+al-1)\mathrm{E}N_i(l,k-1)}{(a+1)i+a}. \quad (2.4)$$

Заметим, что если в графе $H^i_{a,1}$ есть хотя бы одна вершина степени l и второй степени k (без петли), то в этом графе есть хотя бы $l+k-1$ ребро. Поэтому $\mathrm{E}N_i(l,k) = 0$ для $i < l+k-1$. Рассмотрим случай $i = l+k-1$. Достаточно доказать, что

$$\mathrm{E}N_{l+k-1}(l,k) \leqslant Cc(l,k)(l+k)$$

с некоторым C. Для любого конечного количества маленьких l и k мы легко можем найти такую константу C, что

$$\mathrm{E}N_{l+k-1}(l,k) \leqslant Cc(l,k)(l+k).$$

Используя (2.4), получаем

$$\mathrm{E}N_{l+k-1}(l,k) = \frac{(l-2+a)\,\mathrm{E}N_{l+k-2}(l-1,k)}{(a+1)(l+k-2)+a} + \frac{(k+al-1)\mathrm{E}N_{l+k-2}(l,k-1)}{(a+1)(l+k-2)+a} \leqslant$$

$$\leqslant Cc(l-1,k)\frac{(l-2+a)\,(l+k-1)}{(a+1)(l+k-2)+a} + Cc(l,k-1)\frac{(k+al-1)\,(l+k-1)}{(a+1)(l+k-2)+a} \leqslant$$

$$\leqslant Cc(l-1,k)\frac{(l-2+a)(l+k)}{l+al+k+2a} + Cc(l,k-1)\frac{(al+k-1)(l+k)}{l+al+k+2a}.$$

Последнее неравенство выполнено для достаточно больших k.

Осталось рассмотреть конечное количество маленьких k. Сначала покажем, что для любого конечного k

$$c(l,k) = \Omega\left(\frac{l^{k-4+\frac{a^2}{a+1}}}{(1+a)^l}\right).$$

Действительно, из рекуррентного соотношения получаем:

$$c(l,1) = c(l-1,1)\frac{l-2+a}{(a+1)(l+1+a/(a+1))}.$$

Поэтому

$$c(l,1) = \Omega\left(\frac{\Gamma(l-1+a)}{(a+1)^l\Gamma(l+2+a/(a+1))}\right) = \Omega\left(\frac{l^{-3+\frac{a^2}{a+1}}}{(1+a)^l}\right).$$

Здесь мы использовали утверждение 1 из параграфа 2.4.2. Для $k \geqslant 2$ выполнено:

$$c(l,k) = c(l,k-1)\frac{al+k-1}{l(1+a)+k+2a} + c(l-1,k)\frac{l-2+a}{l(1+a)+k+2a}.$$

Достаточно доказать, что существует такая положительная функция $f(k)$, что

$$f(k)(l(1+a)+k+2a)l^{k-4+\frac{a^2}{a+1}} \leqslant f(k-1)(al+k-1)l^{k-5+\frac{a^2}{a+1}} +$$

$$+ f(k)(l-2+a)(a+1)(l-1)^{k-4+\frac{a^2}{a+1}},$$

$$f(k)(l(1+a)+k+2a)\left(l^{k-4+\frac{a^2}{a+1}} - (l-1)^{k-4+\frac{a^2}{a+1}}\right) +$$

$$+ f(k)(3a+k-a^2+2)(l-1)^{k-4+\frac{a^2}{a+1}} \leqslant f(k-1)(al+k-1)l^{k-5+\frac{a^2}{a+1}}.$$

Последнее неравенство выполнено для некоторой положительной функции $f(k)$. Мы можем взять произвольную функцию $f(k)$, поскольку рассматриваем только конечное количество маленьких значений k.

Мы хотим доказать, что

$$\mathrm{E}N_{l+k-1}(l,k) = O\left(\frac{l^{k-3+\frac{a^2}{a+1}}}{(1+a)^l}\right).$$

Пусть у нас есть некоторый граф на $l+k-1$ вершинах и вершина t имеет степень l и вторую степень k. Тогда одно ребро из этой вершины идет в вершину 1, ребра из $l-1$-й вершины ведут в t и ребра из $k-2$-х вершин ведут в соседей t. Есть C_{l+k-2}^{k-2} способов выбрать вершину t и ее соседей. В каждом случае вероятность того, что из соседей ведут ребра в вершину t, равна

$$O\left(\frac{(a(a+1)\dots(a+l-2))}{(3a+2)(4a+3)\dots(l(a+1)+a)}\right) =$$

$$= O\left(\frac{\Gamma(a+l-1)}{(a+1)^l\Gamma(l+1+a/(a+1))}\right) = O\left(\frac{l^{-2+\frac{a^2}{a+1}}}{(a+1)^l}\right),$$

поэтому

$$\mathrm{E}N_{l+k-1}(l,k) = O\left(\frac{l^{k-4+\frac{a^2}{a+1}}}{(a+1)^l}\right).$$

Это завершает доказательство для случая $i = l+k-1$.

Для $i \geqslant l+k-1$ выполнено

$$\mathrm{E}N_{i+1}(l,k) = \mathrm{E}N_i(l,k)\left(1 - \frac{l(1+a)+k+a-1}{(a+1)i+a}\right) +$$

$$+\frac{(l-2+a)\,\mathrm{E}N_i(l-1,k)}{(a+1)i+a} + \frac{(k+al-1)\mathrm{E}N_i(l,k-1)}{(a+1)i+a} =$$

$$= c(l,k)\left(i+\theta\left(C(l+k)^{1+a}\right)\right)\left(1 - \frac{l(1+a)+k+a-1}{(a+1)i+a}\right) +$$

$$+c(l-1,k)\left(i+\theta\left(C(l+k-1)^{1+a}\right)\right)\frac{(l-2+a)}{(a+1)i+a}+$$

$$+c(l,k-1)\left(i+\theta\left(C(l+k-1)^{1+a}\right)\right)\frac{(k+al-1)}{(a+1)i+a} =$$

$$= c(l,k)i - c(l,k)i\frac{l(1+a)+k+a-1}{(a+1)i+a}+$$

$$+c(l,k)\theta\left(C(l+k)^{1+a}\right)\left(1 - \frac{l(1+a)+k+a-1}{(a+1)i+a}\right) +$$

$$+c(l-1,k)i\frac{(l-2+a)}{(a+1)i+a} + c(l,k-1)i\frac{(k+al-1)}{(a+1)i+a}+$$

$$+c(l-1,k)\theta\left(C(l+k-1)^{1+a}\right)\frac{(l-2+a)}{(a+1)i+a}+$$

$$+c(l,k-1)\theta\left(C(l+k-1)^{1+a}\right)\frac{(k+al-1)}{(a+1)i+a}=$$

$$=c(l,k)(i+1)-c(l,k)\frac{il(1+a)+ik+2ia+a}{(a+1)i+a}+c(l-1,k)i\frac{(l-2+a)}{(a+1)i+a}+$$

$$+c(l,k-1)i\frac{(k+al-1)}{(a+1)i+a}+c(l,k)\theta\left(C(l+k)^{1+a}\right)\left(1-\frac{l(1+a)+k+a-1}{(a+1)i+a}\right)+$$

$$+c(l-1,k)\theta\left(C(l+k-1)^{1+a}\right)\frac{(l-2+a)}{(a+1)i+a}+$$

$$+c(l,k-1)\theta\left(C(l+k-1)^{1+a}\right)\frac{(k+al-1)}{(a+1)i+a}=c(l,k)(i+1)-$$

$$-\frac{a(k+al-1)c(l,k-1)}{((a+1)i+a)(l(1+a)+k+2a)}-\frac{a(l-2+a)c(l-1,k)}{((a+1)i+a)(l(1+a)+k+2a)}+$$

$$+c(l,k)\theta\left(C(l+k)^{1+a}\right)\left(1-\frac{l(1+a)+k+a-1}{(a+1)i+a}\right)+$$

$$+c(l-1,k)\theta\left(C(l+k-1)^{1+a}\right)\frac{(l-2+a)}{(a+1)i+a}+$$

$$+c(l,k-1)\theta\left(C(l+k-1)^{1+a}\right)\frac{(k+al-1)}{(a+1)i+a}.$$

Нам нужно, чтобы остаточный член не превосходил $Cc(l,k)(l+k)^{1+a}$, поэтому достаточно доказать следующее неравенство:

$$Cc(l,k)\left((l+k)^{1+a}\right)\frac{l(1+a)+k+a-1}{(a+1)i+a}\geqslant$$

$$\geqslant\frac{a(k+al-1)c(l,k-1)}{((a+1)i+a)(l(1+a)+k+2a)}+\frac{a(l-2+a)c(l-1,k)}{((a+1)i+a)(l(1+a)+k+2a)}+$$

$$+Cc(l-1,k)(l+k-1)^{1+a}\frac{(l-2+a)}{(a+1)i+a}+Cc(l,k-1)(l+k-1)^{1+a}\frac{(k+al-1)}{(a+1)i+a}.$$

Достаточно показать, что выполнены следующие неравенства:

$$Cc(l,k-1)(l+k)^{1+a}\frac{(l(1+a)+k+a-1)(k+al-1)}{((a+1)i+a)(l(1+a)+k+2a)}\geqslant$$

$$\geqslant\frac{a(k+al-1)c(l,k-1)}{((a+1)i+a)(l(1+a)+k+2a)}+Cc(l,k-1)(l+k-1)^{1+a}\frac{(k+al-1)}{(a+1)i+a}$$

и
$$Cc(l-1,k)(l+k)^{1+a}\frac{(l(1+a)+k+a-1)(l-2+a)}{((a+1)i+a)(l(1+a)+k+2a)} \geqslant$$
$$\geqslant \frac{a(l-2+a)c(l-1,k)}{((a+1)i+a)(l(1+a)+k+2a)} + Cc(l-1,k)(l+k-1)^{1+a}\frac{(l-2+a)}{(a+1)i+a}.$$
Другими словами,
$$C(l+k)^{1+a}(l(1+a)+k+a-1)(k+al-1) \geqslant$$
$$\geqslant a(k+al-1) + C(l+k-1)^{1+a}(k+al-1)(l(1+a)+k+2a)$$
и
$$C(l+k)^{1+a}(l(1+a)+k+a-1)(l-2+a) \geqslant$$
$$\geqslant a(l-2+a) + C(l+k-1)^{1+a}(l-2+a)(l(1+a)+k+2a).$$
Чтобы доказать оба неравенства, сделаем следующие преобразования:
$$(l+k)^{1+a}(l(1+a)+k+a-1) - (l+k-1)^{1+a}(l(1+a)+k+2a) =$$
$$= (l+k)^{1+a}(l(1+a)+k+a-1) - ((l+k)^{1+a} - (1+a)(l+k)^a + \frac{a(1+a)}{2}(l+k)^{a-1} +$$
$$+O\left((l+k)^{a-2}\right))\left(l(1+a)+k+2a\right) =$$
$$= -(l+k)^{1+a}(1+a) + (1+a)(l+k)^a(l(1+a)+k+2a) -$$
$$-\frac{a(1+a)}{2}(l+k)^{a-1}(l(1+a)+k+2a) + O\left((l+k)^{a-2}\right)(l(1+a)+k+2a) =$$
$$= (l+k)^{a-1}(1+a)\left(al^2+alk+2al+2ak-\frac{a(1+a)}{2}l-\frac{a}{2}k-\frac{2a^2}{2}\right)+$$
$$+O\left((l+k)^{a-2}\right)(l(1+a)+k+2a) =$$
$$= (l+k)^{a-1}(1+a)\left(al^2+alk+\frac{3}{2}al+\frac{3}{2}ak-\frac{1}{2}a^2l-a^2\right)+$$
$$+O\left((l+k)^{a-2}\right)(l(1+a)+k+2a).$$
Если l или k достаточно велико, то существует такая константа C, что
$$C(l+k)^{a-1}(1+a)\left(al^2+alk+\frac{3}{2}al+\frac{3}{2}ak-\frac{1}{2}a^2l-a^2\right)+$$
$$+O\left((l+k)^{a-2}\right)(l(1+a)+k+2a) \geqslant a.$$

Осталось рассмотреть конечное количество маленьких l и k. Хотим найти такую функцию $f(l,k)$, что
$$f(l,k)c(l,k)\left((l+k)^{1+a}\right)\frac{l(1+a)+k+a-1}{(a+1)i+a} \geqslant$$

$$\geqslant \frac{a(k+al-1)c(l,k-1)}{((a+1)i+a)(l(1+a)+k+2a)} + \frac{a(l-2+a)c(l-1,k)}{((a+1)i+a)(l(1+a)+k+2a)} +$$
$$+ f(l-1,k)c(l-1,k)\frac{(l-2+a)}{(a+1)i+a} + f(l,k-1)c(l,k-1)\frac{(k+al-1)}{(a+1)i+a}.$$

Такая функция $f(l,k)$ существует, поскольку мы можем взять ее растущей настолько быстро, насколько нужно. Это завершает доказательство теоремы 15.

2.4.2 Доказательство теоремы 13

В этом доказательстве мы будем использовать следующее утверждение.

Утверждение 1. *Для $t > 0$ и фиксированного $a > 0$*

$$\frac{\Gamma(t+a)}{\Gamma(t)} = t^a \left(1 + O(1/t)\right).$$

Доказательство. Из формулы Стирлинга получаем

$$\frac{\Gamma(t+a)}{\Gamma(t)} = \sqrt{\frac{t}{t+a}}\frac{(t+a)^a}{e^a}\left(\frac{t+a}{t}\right)^t \left(1 + O(1/t)\right).$$

Легко проверить, что

$$t \ln\left(1 + \frac{a}{t}\right) = a + O(1/t).$$

Поэтому

$$\left(1 + \frac{a}{t}\right)^t = e^a(1 + O(1/t)).$$

Получили

$$\frac{\Gamma(t+a)}{\Gamma(t)} = \sqrt{\frac{t}{t+a}}(t+a)^a \left(1 + O(1/t)\right) = t^a \left(1 + O(1/t)\right).$$

□

Оценка $c(1,k)$

Лемма 21.

$$c(1,k) = \frac{\Gamma(2a+1)(1 + O(1/k))}{\Gamma(a)k^{a+1}}.$$

Доказательство. Выполнено

$$c(k) = \frac{\mathrm{B}(k-1+a,a+2)}{\mathrm{B}(a,a+1)} = \frac{(a+1)\Gamma(2a+1)\Gamma(k-1+a)}{\Gamma(a)\Gamma(k+1+2a)}.$$

Используя рекуррентное соотношение

$$c(1,k) = c(1,k-1)\frac{a+k-1}{k+3a+1} + c(k)\frac{a+k-1}{k+3a+1},$$

получаем

$$c(1,k) = \sum_{j=1}^{k} \frac{c(j)(a+j-1)\dots(a+k-1)}{(j+3a+1)\dots(k+3a+1)} =$$

$$= \frac{(a+1)\Gamma(2a+1)}{\Gamma(a)} \sum_{j=1}^{k} \frac{\Gamma(j-1+a)(a+j-1)\dots(a+k-1)}{\Gamma(j+1+2a)(j+3a+1)\dots(k+3a+1)} =$$

$$= \frac{(a+1)\Gamma(2a+1)\Gamma(a+k)}{\Gamma(a)\Gamma(k+3a+2)} \sum_{j=1}^{k} \frac{\Gamma(j+3a+1)}{\Gamma(j+1+2a)} =$$

$$= \frac{(a+1)\Gamma(2a+1)\Gamma(a+k)}{\Gamma(a)\Gamma(k+3a+2)} \sum_{j=1}^{k} j^a(1+O(1/j)) = \frac{\Gamma(2a+1)k^{a+1}(1+O(1/k))}{\Gamma(a)k^{2a+2}} =$$

$$= \frac{\Gamma(2a+1)(1+O(1/k))}{\Gamma(a)k^{a+1}}.$$

$\square$

Сумма $c(l,k)$

Хотим оценить сумму $\sum_{l=1}^{\infty} c(l,k)$. Сначала покажем, что ряд $\sum_{l=1}^{\infty} l^N c(l,k)$ сходится для всех N и k.

Неравенство

$$c(l,k) \leqslant \tilde{C}\frac{p^k}{(1+q)^l}$$

выполнено для любых $p > 1$ и $q = \min\{a,1\}\frac{(p-1)}{p}$. Здесь мы выбираем $\tilde{C}$ так, чтобы $\tilde{C}\frac{p^k}{1+ap} \geqslant c(1,k)$ для любого k. Нужно доказать, что

$$\frac{p^k}{(1+q)^l}(l+al+k+2a) \geqslant \frac{p^{k-1}}{(1+q)^l}(al+k-1) + \frac{p^k}{(1+q)^{l-1}}(l-2+a).$$

Сделаем следующие преобразования:

$$p(l+al+k+2a) \geqslant (al+k-1) + p(1+q)(l-2+a),$$

$$p(al+k+a+2) \geqslant (al+k-1) + pq(l-2+a),$$

$$al+k+a+2 \geqslant \min\{a,1\}(l-2+a).$$

Последнее неравенство выполнено. Поэтому $c(l,k) \leqslant \tilde{C}\frac{p^k}{(1+q)^l}$, и $\sum_{l=1}^{\infty} l^N c(l,k)$ сходится.

Для $l \geqslant 2$ и любого $c \geqslant 0$ выполнено

$$c(l,k)(l(1+a)+k+2a)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)} =$$

$$= c(l,k-1)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}(al+k-1) + c(l-1,k)(l-2+a)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}.$$

Поэтому

$$\sum_{l=2}^{\infty} c(l,k)(l(1+a)+k+2a)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)} =$$

$$= \sum_{l=2}^{\infty} c(l,k-1)(al+k-1)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)} + \sum_{l=1}^{\infty} c(l,k)\frac{\Gamma(l+a+c+1)}{\Gamma(l+a-1)},$$

$$\sum_{l=2}^{\infty} c(l,k)(al+k+a-c)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)} =$$

$$= \sum_{l=2}^{\infty} c(l,k-1)(al+k-1)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)} + c(1,k)\frac{\Gamma(a+c+2)}{\Gamma(a)}.$$

Рассмотрим функцию

$$f_c(k) = \frac{\Gamma(k+a-c)}{\Gamma(k)} = k^{a-c}(1+O(1/k)).$$

Легко видеть, что

$$\frac{f_c(k+1)}{f_c(k)} = 1 + \frac{a-c}{k}.$$

Выполнено

$$\sum_{j=1}^{k}\sum_{l=2}^{\infty} c(l,j)(al+j+a-c)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(j) =$$

$$= \sum_{j=1}^{k}\sum_{l=2}^{\infty} c(l,j-1)(al+j-1)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(j) + \sum_{j=1}^{k} c(1,j)\frac{\Gamma(a+c+2)}{\Gamma(a)}f_c(j),$$

$$\sum_{j=1}^{k}\sum_{l=2}^{\infty} c(l,j)(al+j+a-c)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(j) =$$

$$= \sum_{j=1}^{k-1}\sum_{l=2}^{\infty} c(l,j)(al+j)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(j)\left(1+\frac{a-c}{j}\right) +$$

$$+\sum_{j=1}^{k} c(1,j)\frac{\Gamma(a+c+2)}{\Gamma(a)}f_c(j),$$

$$\sum_{l=2}^{\infty} c(l,k)(al+k+a-c)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(k) =$$

$$= \sum_{j=1}^{k-1}\sum_{l=2}^{\infty} c(l,j)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}\frac{al(a-c)}{j}f_c(j) + \sum_{j=1}^{k} c(1,j)\frac{\Gamma(a+c+2)}{\Gamma(a)}f_c(j).$$

Если $c \geqslant a$, то, используя лемму 21 и асимптотику $f_c(j)$, получаем

$$\sum_{l=2}^{\infty} c(l,k)(al+k+a-c)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(k) \leqslant \sum_{j=1}^{k} c(1,j)\frac{\Gamma(a+c+2)}{\Gamma(a)}f_c(j) = O(1).$$

Поэтому

$$\sum_{l=2}^{\infty} c(l,k)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(k) = O(1/k).$$

Мы хотим доказать, что для любого $0 \leqslant c < a+1$ выполнено следующее равенство:

$$\sum_{l=2}^{\infty} c(l,k)\frac{\Gamma(l+a+c)}{\Gamma(l+a-1)}f_c(k) = O\left(\frac{(\ln k)^{\lceil a-c\rceil}}{k}\right). \tag{2.5}$$

Мы уже доказали это утверждение для $a \leqslant c < a+1$.

Пусть для $c' \geqslant 1$ выполнено

$$\sum_{l=2}^{\infty} c(l,k)\frac{\Gamma(l+a+c')}{\Gamma(l+a-1)}f_{c'}(k) = O\left(\frac{(\ln k)^{\lceil a-c'\rceil}}{k}\right).$$

Тогда

$$\sum_{l=2}^{\infty} c(l,k)(al+k+a-c'+1)\frac{\Gamma(l+a+c'-1)}{\Gamma(l+a-1)}f_{c'-1}(k) =$$

$$= \sum_{j=1}^{k-1}\sum_{l=2}^{\infty} c(l,j)\frac{\Gamma(l+a+c'-1)}{\Gamma(l+a-1)}\frac{al(a-c'+1)}{j}(j+a-c')f_{c'}(j) +$$

$$+\sum_{j=1}^{k} c(1,j)\frac{\Gamma(a+c'+1)}{\Gamma(a)}f_{c'-1}(j) =$$

$$= O\left(\sum_{j=1}^{k-1}\frac{(\ln k)^{\lceil a-c'\rceil}}{j}\right) = O\left((\ln k)^{\lceil a-c'+1\rceil}\right).$$

Мы доказали (2.5). В частности,

$$\sum_{l=2}^{\infty} c(l,k)\frac{\Gamma(l+a)}{\Gamma(l+a-1)}f_0(k) = \sum_{l=2}^{\infty} c(l,k)(l+a-1)f_0(k) = O\left(\frac{(\ln k)^{\lceil a\rceil}}{k}\right). \quad (2.6)$$

Положим $x_k = \sum_{l=2}^{\infty} c(l,k)$. Для $l \geqslant 2$

$$c(l,k)(l(1+a)+k+2a) = c(l,k-1)(al+k-1) + c(l-1,k)(l-2+a).$$

Поэтому

$$\sum_{l=2}^{\infty} c(l,k)(l(1+a)+k+2a) = \sum_{l=2}^{\infty} c(l,k-1)(al+k-1) + \sum_{l=1}^{\infty} c(l,k)(l-1+a),$$

$$\sum_{l=2}^{\infty} c(l,k)(al+k+a+1) = \sum_{l=2}^{\infty} c(l,k-1)(al+k-1) + ac(1,k),$$

$$(k+a+1)x_k = (k-1)x_{k-1} + ac(1,k) + a\sum_{l=2}^{\infty} l(c(l,k-1)-c(l,k)).$$

Выполнено

$$(k+a+1)f_{-1}(k)x_k = (k-1)f_{-1}(k)x_{k-1} + af_{-1}(k)c(1,k)+$$

$$+af_{-1}(k)\sum_{l=2}^{\infty} l(c(l,k-1)-c(l,k)),$$

$$\sum_{j=1}^{k}(j+a+1)f_{-1}(j)x_j = \sum_{j=1}^{k-1} f_{-1}(j)(j+a+1)x_j + \sum_{j=1}^{k} af_{-1}(j)c(1,j)+$$

$$+\sum_{j=1}^{k} af_{-1}(j)\sum_{l=2}^{\infty} l(c(l,j-1)-c(l,j)),$$

$$(k+a+1)f_{-1}(k)x_k = a\sum_{j=1}^{k} f_{-1}(j)c(1,j) + a\sum_{j=1}^{k} f_{-1}(j)\sum_{l=2}^{\infty} l(c(l,j-1)-c(l,j)).$$

$$f_{-1}(k)(k+a+1)x_k = a\sum_{j=1}^{k} f_{-1}(j)c(1,j)+$$

$$+a(a+1)\sum_{j=1}^{k-1}\frac{f_{-1}(j)}{j}\sum_{l=2}^{\infty} lc(l,j) - af_{-1}(k)\sum_{l=2}^{\infty} lc(l,k) =$$

$$= a\sum_{j=1}^{k} j^{a+1}\frac{\Gamma(2a+1)}{\Gamma(a)j^{a+1}}(1+O(1/j)) + \sum_{j=1}^{k-1} O\left(\frac{(\ln j)^{\lceil a\rceil}}{j}\right) + O\left((\ln k)^{\lceil a\rceil}\right) =$$

$$= ak\frac{\Gamma(2a+1)}{\Gamma(a)} + \sum_{j=1}^{k-1} O\left(\frac{(\ln j)^{\lceil a \rceil}}{j}\right) + O\left((\ln k)^{\lceil a \rceil}\right) =$$

$$= ak\frac{\Gamma(2a+1)}{\Gamma(a)}\left(1 + O\left(\frac{(\ln k)^{\lceil a+1 \rceil}}{k}\right)\right).$$

Здесь мы использовали (2.6) и лемму 21. Получаем

$$x_k = \frac{a\Gamma(2a+1)}{\Gamma(a)k^{a+1}}\left(1 + O\left(\frac{(\ln k)^{\lceil a+1 \rceil}}{k}\right)\right)$$

и

$$\sum_{l=1}^{\infty} c(l,k) = c(1,k) + x_k = \frac{(a+1)\Gamma(2a+1)}{\Gamma(a)k^{a+1}}\left(1 + O\left(\frac{(\ln k)^{\lceil a+1 \rceil}}{k}\right)\right).$$

Оценка $\mathrm{E}Y_n(k)$

Заметим, что

$$\sum_{l \geqslant 1}\sum_{j \geqslant k} \mathrm{E}N_{i+1}(l,j) = \sum_{l \geqslant 1}\sum_{j \geqslant k} \mathrm{E}N_i(l,j) +$$

$$+ \sum_{l \geqslant 1} \frac{(al+k-1)\mathrm{E}N_i(l,k-1)}{(a+1)i+a} + \sum_{j \geqslant k} \frac{(j-1+a)\,\mathrm{E}N_i(j)}{(a+1)i+a}.$$

Поэтому мы получаем

$$\sum_{l \geqslant 1}\sum_{j \geqslant k} \mathrm{E}N_n(l,j) = \sum_{i=1}^{n-1}\sum_{l \geqslant 1} \frac{(al+k-1)\mathrm{E}N_i(l,k-1)}{(a+1)i+a} +$$

$$+ \sum_{i=1}^{n-1}\sum_{j \geqslant k} \frac{(j-1+a)\,\mathrm{E}N_i(j)}{(a+1)i+a}.$$

Оценим сумму

$$\sum_{i=1}^{n-1}\sum_{j \geqslant k} \frac{(j-1+a)\,\mathrm{E}N_i(j)}{(a+1)i+a}.$$

Сначала вычислим

$$F_t(k) = \sum_{j \geqslant k} (j-1+a)\,\mathrm{E}N_t(j).$$

Докажем индукцией по k, что

$$F_n(k) = \frac{(a+1)\Gamma(2a+1)\Gamma(k+a)}{\Gamma(a+1)\Gamma(k+2a)} n\left(1 + \theta\left(\frac{C(k-1)^{1+a}}{n}\right)\right)$$

с некоторой константой C. Для $k=1$ и $k=2$ имеем

$$F_n(1)=\sum_{j\geqslant 1}(j-1+a)\,\mathrm{E}N_n(j)=n(1+a),$$

$$F_n(2)=F_n(1)-a\,\mathrm{E}N_n(1)=$$

$$=n(1+a)-an\frac{(1+a)}{(2a+1)}+O\left(1\right)=\frac{(1+a)^2n}{2a+1}\left(1+O\left(1/n\right)\right).$$

Для $k\geqslant 3$ выполнено

$$\mathrm{E}N_{i+1}(j)=\mathrm{E}N_i(j)\left(1-\frac{j-1+a}{(a+1)i+a}\right)+\mathrm{E}N_i(j-1)\frac{j-2+a}{(a+1)i+a}.$$

Умножим это равенство на $(j-1+a)$ и просуммируем по всем $j\geqslant k$:

$$F_{i+1}(k)=\sum_{j\geqslant k}(j-1+a)\,\mathrm{E}N_{i+1}(j)=$$

$$=\sum_{j\geqslant k}(j-1+a)\mathrm{E}N_i(j)-\sum_{j\geqslant k}\mathrm{E}N_i(j)\frac{(j-1+a)(j-1+a)}{(a+1)i+a}+$$

$$+\sum_{j\geqslant k-1}\mathrm{E}N_i(j)\frac{(j+a)(j-1+a)}{(a+1)i+a}=$$

$$=F_i(k)+\sum_{j\geqslant k}\mathrm{E}N_i(j)\frac{(j-1+a)}{(a+1)i+a}+\mathrm{E}N_i(k-1)\frac{(k-1+a)(k-2+a)}{(a+1)i+a}=$$

$$=F_i(k)\left(1+\frac{1}{(a+1)i+a}\right)+\left(F_i(k-1)-F_i(k)\right)\frac{(k-1+a)}{(a+1)i+a}=$$

$$=F_i(k)\left(1-\frac{k-2+a}{(a+1)i+a}\right)+F_i(k-1)\frac{(k-1+a)}{(a+1)i+a}.$$

Заметим, что для $i+1<k-1$ имеем $F_{i+1}(k)=0$. Рассмотрим $i+1\geqslant k-1$. Используя предположение индукции, получаем

$$F_{i+1}(k)=$$

$$=\frac{(a+1)\Gamma(2a+1)\Gamma(k+a)}{\Gamma(a+1)\Gamma(k+2a)}i\left(1-\frac{k-2+a}{(a+1)i+a}\right)\left(1+\theta\left(\frac{C(k-1)^{1+a}}{i}\right)\right)+$$

$$+\frac{(a+1)\Gamma(2a+1)\Gamma(k-1+a)i}{\Gamma(a+1)\Gamma(k-1+2a)}\frac{(k-1+a)}{(a+1)i+a}\left(1+\theta\left(\frac{C(k-2)^{1+a}}{i}\right)\right)=$$

$$=\frac{(a+1)\Gamma(2a+1)\Gamma(k+a)}{\Gamma(a+1)\Gamma(k+2a)}\Bigg(i+1-\frac{a}{(a+1)i+a}+$$

$$+i\left(1-\frac{k-2+a}{(a+1)i+a}\right)\theta\left(\frac{C(k-1)^{1+a}}{i}\right)+\frac{(k-1+2a)i}{(a+1)i+a}\theta\left(\frac{C(k-2)^{1+a}}{i}\right)\Bigg).$$

Нужно показать, что для некоторой константы C

$$\frac{a}{(a+1)i+a}+\frac{(k-1+2a)}{(a+1)i+a}C(k-2)^{1+a}\leqslant\frac{k-2+a}{(a+1)i+a}C(k-1)^{1+a}.$$

Это неравенство выполнено для достаточно больших C.

Имеем

$$\sum_{i=1}^{n-1}\frac{F_i(k)}{(a+1)i+a}=\sum_{i=1}^{n-1}\frac{\Gamma(2a+1)\Gamma(k+a)}{\Gamma(a+1)\Gamma(k+2a)}\left(1+O\left(\frac{k^{1+a}}{i}\right)\right)=$$

$$=\frac{\Gamma(2a+1)\Gamma(k+a)}{\Gamma(a+1)\Gamma(k+2a)}n\left(1+O\left(\frac{k^{1+a}}{n}\right)\right).$$

Оценим сумму

$$\sum_{i=1}^{n-1}\sum_{l\geqslant 1}\frac{(al+k-1)\mathrm{E}N_i(l,k-1)}{(a+1)i+a}.$$

Для начала оценим сумму

$$\sum_{l\geqslant 1}(al+k-1)\mathrm{E}N_i(l,k-1).$$

Легко видеть, что

$$\mathrm{E}N_i(l,k)=O(c(l,k)i).$$

Можно проверить это, следуя доказательству теоремы 15 и убеждаясь в том, что оно работает для неравенства

$$\mathrm{E}N_i(l,k)<\tilde{C}c(l,k)((a+1)i+a)$$

с некоторой константой $\tilde{C}$. Потребуется также аналог леммы 14.

Поэтому

$$\sum_{l\geqslant 1}(al-1)\mathrm{E}N_i(l,k-1)=O\left(\sum_{l\geqslant 1}(al-1)c(l,k-1)i\right)=O\left(\frac{(\ln k)^{\lceil a\rceil}i}{k^{a+1}}\right).$$

Используя (2.5), получаем

$$\sum_{l\geqslant 1}k\mathrm{E}N_i(l,k-1)=\sum_{l\geqslant 1}kc(l,k-1)i\left(1+O\left((l+k)^{1+a}/i\right)\right)=$$

$$=\frac{(a+1)\Gamma(2a+1)}{\Gamma(a)k^a}i\left(1+O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)+O\left(\frac{k^{1+a}}{i}\right)\right).$$

Здесь мы использовали следующую оценку:

$$\sum_{l\geqslant 1} kc(l,k-1)(l+k)^{1+a} = O\left(\sum_{l=1}^{k} k^{2+a}c(l,k-1) + \sum_{l\geqslant k} kc(l,k-1)l^{1+a}\right) =$$
$$= O(k) + O(1) = O(k).$$

Поэтому

$$\sum_{i=1}^{n-1}\sum_{l\geqslant 1}\frac{(al+k-1)\mathrm{E}N_i(l,k-1)}{(a+1)i+a} =$$
$$= \frac{\Gamma(2a+1)}{\Gamma(a)k^a}n\left(1+O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)+O\left(\frac{k^{1+a}}{n}\right)\right).$$

Следовательно,

$$\sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}N_n(l,j) = \frac{a\Gamma(2a+1)}{\Gamma(a+1)k^a}n\left(1+O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)\right)+$$
$$+\frac{\Gamma(2a+1)\Gamma(k+a)}{\Gamma(a+1)\Gamma(k+2a)}n\left(1+O\left(\frac{k^{1+a}}{n}\right)\right) =$$
$$= \frac{(a+1)\Gamma(2a+1)}{\Gamma(a+1)k^a}n\left(1+O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)+O\left(\frac{k^{1+a}}{n}\right)\right).$$

Теперь рассмотрим вершины с петлями. Для $k=0$ имеем

$$\sum_{l\geqslant 1}\sum_{j\geqslant 0}\mathrm{E}P_n(l,j) = \sum_{i=1}^{n}\frac{a}{(1+a)i-1} = O(\ln n).$$

Для $k\geqslant 2$:

$$\sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}P_{i+1}(l,j) = \sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}P_i(l,j) + \sum_{l\geqslant 1}\frac{(al+k-2a-1)\mathrm{E}P_i(l,k-1)}{(a+1)i+a}.$$

Поэтому

$$\sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}P_n(l,j) = \sum_{i=1}^{n-1}\sum_{l\geqslant 1}\frac{(al+k-2a-1)\mathrm{E}P_i(l,k-1)}{(a+1)i+a} \leqslant$$
$$\leqslant \sum_{i=1}^{n-1}\sum_{l\geqslant 1}\frac{(al+k-2a-1)p(l,k-1)}{(a+1)i}.$$

Из рекуррентного соотношения для $p(l,k)$ следует:

$$p(l,k) = O\left(\frac{1}{l^2}\right)$$

и

$$p(l,k) = O\left(\frac{k}{l^3}\right).$$

Чтобы получить вторую оценку, рассмотрим $q(l,k) = p(l,k)/k$. Для $k \geqslant 1$:

$$q(l,k)(l+al+k-1-a) = q(l,k-1)\frac{(k-1)(al+k-2a-1)}{k} + q(l-1,k)(l-2+a),$$

$$q(l,k)(l+al+k-1-a) - q(l,k-1)\left(al+k-2a-2-\frac{al-2a-1}{k}\right) =$$

$$= q(l-1,k)(l-2+a).$$

Таким образом, $q(l,k) = O(q(l))$, где

$$q(l)(l+a+1+(al-2a-1)) = q(l-1)(l-2+a).$$

Из этого равенства следует, что $q(l) = O\left(\frac{1}{l^3}\right)$.

Мы можем оценить следующую сумму:

$$\sum_{l\geqslant 1}\frac{(al+k-2a-1)p(l,k-1)}{(a+1)} = O\left(k\right).$$

Поэтому

$$\sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}P_n(l,j) = O(k\ln n).$$

Теперь мы наконец можем оценить $\mathrm{E}Y_n(k)$:

$$\mathrm{E}Y_n(k) = \sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}N_n(l,j) + \sum_{l\geqslant 1}\sum_{j\geqslant k}\mathrm{E}P_n(l,k) =$$

$$= \frac{(a+1)\Gamma(2a+1)}{\Gamma(a+1)k^a}n\left(1+O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)+O\left(\frac{k^{1+a}}{n}\right)\right)+O(k\ln n) =$$

$$= \frac{(a+1)\Gamma(2a+1)}{\Gamma(a+1)k^a}n\left(1+O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)+O\left(\frac{k^{1+a}}{n}\right)\right).$$

Это завершает доказательство теоремы 13.

2.4.3 Доказательство леммы 15

Легко видеть, что $\mathrm{E}P_n(0,k) = \mathrm{E}P_n(1,k) = 0$. Для $k > 0$ выполнено $\mathrm{E}P_n(2,k) = 0$. Для $k = 0$:

$$\mathrm{E}P_n(2,0) = \sum_{i=1}^{n}\frac{a}{(a+1)i-1}\prod_{j=i+1}^{n}\frac{(1+a)j-2-a}{(1+a)j-1} =$$

$$= \sum_{i=1}^{n} \frac{a}{(a+1)i-1} \frac{\Gamma\left(n - \frac{1}{a+1}\right)\Gamma\left(i + \frac{a}{a+1}\right)}{\Gamma\left(n + \frac{a}{a+1}\right)\Gamma\left(i - \frac{1}{a+1}\right)} =$$

$$= \frac{1}{n}(1 + O(1/n)) \sum_{i=1}^{n} \frac{ai}{(a+1)i-1}(1 + O(1/i)) = O(1).$$

Далее проводим доказательство по индукции. Рассмотрим $l \geqslant 3$, $k \geqslant 1$. Пусть мы уже доказали, что $\mathrm{E}P_n(i,j) \leqslant p(i,j)$ для всех таких i и j, что $i < l$, $j \leqslant k$ или $i \leqslant l$, $j < k$ и для всех n. Теперь воспользуемся следующим равенством:

$$\mathrm{E}P_{i+1}(l,k) = \mathrm{E}P_i(l,k)\left(1 - \frac{l(a+1)+k-a-1}{(a+1)i+a}\right) +$$
$$+ \mathrm{E}P_i(l,k-1)\frac{al+k-2a-1}{(a+1)i+a} + \mathrm{E}P_i(l-1,k)\frac{l-2+a}{(a+1)i+a}. \quad (2.7)$$

Заметим, что если в графе $H^i_{a,1}$ есть хотя бы одна вершина с петлей степени l и второй степени k, то в этом графе есть хотя бы $l+k-1$ ребро. Поэтому $\mathrm{E}P_i(l,k) = 0$ для $i < l+k-1$. Рассмотрим случай $i = l+k-1$. Используя (2.7), получаем (для $k \geqslant 1$)

$$\mathrm{E}P_{l+k-1}(l,k) = \frac{(l-2+a)\,\mathrm{E}P_{l+k-2}(l-1,k)}{(a+1)(l+k-2)+a} +$$

$$+ \frac{(al+k-2a-1)\mathrm{E}P_{l+k-2}(l,k-1)}{(a+1)(l+k-2)+a} \leqslant \frac{(l-2+a)p(l-1,k)}{(a+1)(l+k-2)+a} +$$

$$+ \frac{(al+k-2a-1)p(l,k-1)}{(a+1)(l+k-2)+a} = \frac{(l+al+k-1-a)p(l,k)}{(a+1)(l+k-2)+a} \leqslant p(l,k).$$

Последнее неравенство выполнено для $k \geqslant 1/a$. Рассмотрим случай $k < 1/a$. Как и при доказательстве теоремы 15, сначала оценим $p(l,k)$:

$$p(l,k) = \Omega\left(\frac{l^{k+\frac{a^2}{a+1}}}{(1+a)^l}\right).$$

Для $k = 0$:

$$p(l,0) = p(l-1,0)\frac{l-2+a}{(1+a)(l-1-\frac{1}{a+1})}.$$

Поэтому

$$p(l,0) = \Omega\left(\frac{l^{\frac{a^2}{a+1}}}{(1+a)^l}\right).$$

Для $k \geqslant 1$:

$$p(l,k) = p(l,k-1)\frac{al+k-2a-1}{l(1+a)+k-1-a} + p(l-1,k)\frac{l-2+a}{l(1+a)+k-1-a}.$$

Достаточно показать, что существует такая положительная функция $f(k)$, что для больших l:

$$f(k)(l(1+a)+k-1-a)l^{k+\frac{a^2}{a+1}} \leqslant$$

$$\leqslant f(k-1)(al+k-2a-1)l^{k+\frac{a^2}{a+1}-1} + f(k)(l-2+a)(a+1)(l-1)^{k+\frac{a^2}{a+1}},$$

$$f(k)(l(1+a)+k-1-a)\left(l^{k+\frac{a^2}{a+1}} - (l-1)^{k+\frac{a^2}{a+1}}\right) +$$

$$+f(k)(k-a^2+1)(l-1)^{k+\frac{a^2}{a+1}} \leqslant f(k-1)(al+k-2a-1)l^{k+\frac{a^2}{a+1}-1}.$$

Последнее неравенство выполнено для некоторой функции $f(k)$.

Мы хотим доказать, что

$$\mathrm{E}P_{l+k-1}(l,k) = O\left(\frac{l^{k+\frac{a^2}{a+1}}}{(1+a)^l}\right).$$

Есть l^k графов на $l+k-1$ вершине, в которых есть вершина без петли степени l и второй степени k. И эта вершина — в точности вершина 1. Вероятность того, что вершина 1 имеет степень l и вторую степень k, равна:

$$O\left(\frac{l^k((a+1)\ldots(a+l-2))}{(a+2)\ldots((l-1)(a+1)-1)}\right) = O\left(\frac{l^{k+\frac{a^2}{a+1}}}{(a+1)^l}\right).$$

Это завершает доказательство для $i = l+k-1$.

Если $i \geqslant l+k-1$, то

$$\mathrm{E}P_{i+1}(l,k) = \mathrm{E}P_i(l,k)\left(1 - \frac{l(a+1)+k-a-1}{(a+1)i+a}\right) +$$

$$+\mathrm{E}P_i(l,k-1)\frac{al+k-2a-1}{(a+1)i+a} + \mathrm{E}P_i(l-1,k)\frac{l-2+a}{(a+1)i+a}.$$

Используя рекуррентное соотношение для $p(l,k)$ и индукцию по i, легко показать, что $\mathrm{E}P_n(l,k) \leqslant p(l,k)$. Это завершает доказательство леммы 15.

2.4.4 Доказательство теоремы 14

Оценим математическое ожидание величины $X_n(k)$ следующим образом:

$$\mathrm{E}X_n(k) = \sum_{l=1}^{\infty} \mathrm{E}N_n(l,k) + \sum_{l=1}^{\infty} \mathrm{E}P_n(l,k) =$$

$$= \sum_{l=1}^{\infty} c(l,k)n + O\left(\sum_{l=1}^{\infty} c(l,k)(l+k)^{1+a}\right) + O\left(\sum_{l=1}^{\infty} p(l,k)\right) =$$

$$= \frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)k^{a+1}}\left(1 + O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right)\right) + O(1) + O(1) =$$

$$= \frac{(a+1)\Gamma(2a+1)n}{\Gamma(a)k^{a+1}}\left(1 + O\left(\frac{(\ln k)^{\lceil a+1\rceil}}{k}\right) + O\left(\frac{k^{1+a}}{n}\right)\right).$$

Глава 3

Предпочтительное присоединение с устареванием

В данной главе вводится новый класс моделей, основанных на идее предпочтительного присоединения, а именно — модели с устареванием.

Глава основана на статье автора [54].

3.1 Классические модели и свойство устаревания

Несмотря на то, что модели предпочтительного присоединения отражают такие важные свойства реальных сетей, как маленький диаметр и степенное распределение степеней вершин, они не соответствуют реальности в других важных аспектах.

В этой главе будет обсуждаться так называемое *свойство устаревания*. Свойство устаревания отражает тот факт, что в реальных сетях вершины чаще соединены с другими вершинами, близкими к ним по возрасту (времени появления). В случае моделей предпочтительного присоединения под временем появления понимается шаг n, на котором вершина добавилась в граф. Нетрудно понять, что свойство устаревания не выполняется для моделей предпочтительного присоединения: новые вершины стремятся присоединиться к уже популярным вершинам, а наиболее популярные вершины — "старые" вершины.

В этой главе будет обсуждаться новый класс моделей, которые обладают свойством устаревания. Свойство устаревания было предложено в работе [6]. А именно, рассматривается следующая величина: $e(T)$ — доля ребер в графе, которые соединяют вершины с разницей "возрастов", большей чем T, то есть вершины i и j с $|i-j| > T$. В работе [40] было показано, что в некоторых реальных сетях $e(T)$ убывает экспоненциально с ростом T.

Класс моделей, который обсуждается в данной работе, является обобщением модели Бианкони и Барабаши [16], которая, в свою очередь, обобщает модели предпочтительного присоединениям добавлением качества (fitness) вершин. В модели Бианкони и Барабаши, когда новая вершина добавляется в граф, она выбирает вершину, в которую проведет ребро, с вероятностью, пропорциональной произведению качества этой вершины и степени. В данной работе в функцию привлекательности вершины добавляется множитель, соответствующий устареванию вершины. Таким образом, вероятность того, что в данную вершину будет проведено новое ребро, определяется *привлекательностью* вершины, которая, в свою очередь, зависит от входящей степени вершины (текущая популярность), ее качества (некоторая константа) и возраста (новые страницы обладают большей привлекательностью). Идея добавить устаревание в функцию привлекательности была предложена в работе [6]. В данной работе, во-первых, предложена новая, в сравнении с [6], модель. Во-вторых, новое строгое определение позволяет теоретически исследовать распределение степеней и поведение функции $e(T)$ в графе.

3.2 Модель с устареванием

В этом разделе будет дано формальное определение модели с устареванием. Строим последовательность случайных графов $\{G_n\}$. У этой последовательности есть следующие параметры: натуральное число m (количество ребер, добавляемых вместе с новой вершиной) и функция $N(n)$, принимающая целочисленные значения. Нам также потребуется последовательность независимых положительных случайных величин $\zeta_1, \zeta_2, \dots$ с некоторым распределением. Каждый граф G_n определяется в соответствии со своей собственной процедурой построения, которая основана на идее предпочтительного присоединения.

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

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

где

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

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

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

Согласно приведенному выше определению, есть 12 различных моделей. Если $\mathrm{attr}_t(i) = d_t(i)$, то мы получаем известную модель предпочтительного присоединения. Если $\mathrm{attr}_t(i) = q(i)\, d_t(i)$, то мы получаем модель Бианкони и Барабаши [16]. Поскольку модели без *фактора устаревания* (то есть без $I[i > t - N]$, $e^{-\frac{t-i}{N}}$, или других функций, убывающих с возрастом) в функции привлекательности уже изучались, нас интересуют модели с фактором устаревания. Поэтому для нас интересны следующие функции привлекательности:

(1) $q(i)\, I[i > t - N]$,

(2) $q(i)\, d_t(i)\, I[i > t - N]$,

(3) $q(i)\, e^{-\frac{t-i}{N}}$,

(4) $q(i)\, d_t(i)\, e^{-\frac{t-i}{N}}$.

В работах [6, 40] было экспериментально показано, что функция привлекательности, которая содержит качество и фактор устаревания, лучше отражает поведение некоторых частей интернета, чем функция привлекательности, которая содержит степень, качество и фактор устаревания. Поэтому далее мы рассматриваем функции привлекательности (1) и (3).

Заметим, что в [40] рассматривался только фактор устаревания $e^{-\frac{t-i}{N}}$. Мы также рассматриваем фактор $I[i > t-N]$ по двум причинам. Во-первых, как будет показано далее, оба фактора устаревания ведут себя похожим образом, если речь идет о распределении степеней, однако теоретический анализ в случае $I[i > t - N]$ значительно проще. Во-вторых, фактор устаревания $I[i > t - N]$ обладает следующей естественной интерпретацией. Ссылки на многие новые (в основном новостные) страницы могут быть найдены не некоторых страницах, которые называют “источниками контента”. И новые страницы популярны до тех пор, пока их можно найти на источнике контента. Но через некоторое время после появления страницы заменяются более новыми страницами. Поэтому естественно предположить, что после некоторого промежутка времени старые страницы перестают быть популярными.

3.3 Функция привлекательности $q(i)I[i > t - N]$

В этом разделе мы предполагаем, что функция привлекательности вершины i имеет вид $\mathrm{attr}_t(i) = q(i)\, I[i > t - N]$. Повторим, что фактор устаревания в данном случае означает то, что вершина i может получить входящие ребра только в течение следующих N шагов после появления, и этот период мы называем *жизнью* вершины. Также будем говорить, что в течение этого периода вершина является *живой*, а после — *умирает*.

В соответствии с [40], мы также предполагаем, что случайные величины $\zeta_1, \zeta_2, \ldots$ имеют распределение Парето с плотностью $f(x) = \frac{\gamma a^\gamma I[x > a]}{x^{\gamma+1}}$, где $\gamma > 1$, $a > 0$. Далее через ζ мы обозначаем случайную величину с таким Парето распределением.

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

3.3.1 Распределение степеней

В этом разделе будет показано, что степенное распределение качества ведет к степенному распределению степеней вершин в графе.

Результаты

Через $\#(d)$ обозначим количество вершин степени d в графе G_n. Мы докажем следующую теорему.

Теорема 22. *Пусть $d = d(n)$ растет с ростом n и $d = o\left(\left(\frac{n}{N}\right)^{\frac{1}{\gamma+1}}\right)$. Если $\gamma > 2$ и $d = o\left(N^{\frac{1}{\gamma+3}}\right)$ или $1 < \gamma \leqslant 2$ и $d = o\left(N^{\frac{\alpha-1}{\gamma+\alpha+1}}\right)$ для некоторого α, $1 < \alpha < \gamma$, то*

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

Теорема 22 говорит о том, что математическое ожидание количества вершин степени d убывает как $d^{-\gamma-1}$. Чтобы показать степенное распределение степеней вершин, необходимо также доказать концентрацию количества вершин степени d около своего математического ожидания.

Теорема 23. *Для любого d выполнено:*

$$\mathsf{P}\left(|\#(d) - \mathrm{E}\#(d)| \geqslant \sqrt{Nn\ln n}\right) \leqslant \frac{2}{\ln n}.$$

Заметим, что для $d = o\left(\left(\frac{n}{N \ln n}\right)^{1/2(\gamma+1)}\right)$ имеем $\sqrt{Nn \ln n} = o\left(n/d^{\gamma+1}\right)$, поэтому теорема 23 действительно дает концентрацию.

В следующих двух параграфах мы докажем теорему 22, далее будет доказана теорема 23.

Концентрация веса

Зафиксируем n и $N = N(n)$. В этом разделе мы рассматриваем только такие вершины p, что $N \leqslant p \leqslant n - N + 1$.

Через $\bar{d}(p)$ обозначим степень вершины p в итоговом графе и через $\bar{d}_{in}(p)$ — входящую степень вершины p. Таким образом, $\bar{d}_{in}(p) = \bar{d}(p) - m$. Через $Q(t)$ обозначим сумму качеств живых вершин на шаге t, то есть

$$Q(t) = \sum_{k=t-N}^{t-1} q(k)\,.$$

Будем также говорить, что $Q(t)$ — вес вершин на шаге t. Заметим, что

$$\mathrm{E}\left(\bar{d}_{in}(p) \mid q(p-N+1), \ldots, q(p+N-1)\right) = \sum_{i=1}^{N} \frac{mq(p)}{Q(p+i)}\,.$$

Действительно, для каждого i, $1 \leqslant i \leqslant N$, вероятность ребра $(p+i, p)$ равна $\frac{mq(p)}{Q(p+i)}$ в соответствии с определением модели, поскольку $Q(p+i)$ — общая привлекательность всех вершин на шаге $(p+i)$.

Рассмотрим жизнь вершины p с качеством $q(p)$. Имеем $\mathrm{E}(Q(p+i)|q(p)) = q(p) + (N-1)\mathrm{E}\zeta$ для $1 \leqslant i \leqslant N$. Мы хотим оценить вероятность того, что вес $\mathrm{E}(Q(p+i)|q(p))$ отклоняется от среднего значения $N\mathrm{E}\zeta$.

Пусть $\xi_1, \ldots \xi_{N-1}$ — веса вершин $p-N+1, \ldots, p-1$ и $\eta_1, \ldots \eta_{N-1}$ — веса вершин $p+1, \ldots, p+N-1$. Пусть $W_p^q(i)$ — общий вес всех живых страниц в тот момент, когда возраст вершины p равен i при условии, что вершина p имеет качество q, то есть

$$W_p^q(i) = \sum_{k=1}^{i-1} \eta_k + q + \sum_{k=i}^{N-1} \xi_k\,.$$

Нам потребуется следующая лемма.

Лемма 22. *Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины,* $\mathrm{E}\xi_i = 0$, $\mathrm{E}|\xi_i|^\alpha < \infty$, $1 \leqslant \alpha \leqslant 2$. *Тогда*

$$\mathrm{E}|\xi_1 + \ldots + \xi_n|^\alpha \leqslant 2^\alpha \left(\mathrm{E}|\xi_1|^\alpha + \ldots + \mathrm{E}|\xi_n|^\alpha .\right)$$

Доказательство.

Будем использовать два вспомогательных факта.

Факт 1. Если ξ и η — независимые случайные величины и η имеет симметричное распределение, то для любого α, $1 \leqslant \alpha \leqslant 2$, выполнено

$$\mathrm{E}|\xi+\eta|^\alpha \leqslant \mathrm{E}|\xi|^\alpha + \mathrm{E}|\eta|^\alpha .$$

Доказательство.

$$\mathrm{E}|\xi+\eta|^\alpha = \frac{1}{2}\left(\mathrm{E}|\xi+\eta|^\alpha + \mathrm{E}|\xi-\eta|^\alpha\right)$$

осталось показать, что для любых x, y и $1 \leqslant \alpha \leqslant 2$ выполнено:

$$\frac{1}{2}\left(|x+y|^\alpha + |x-y|^\alpha\right) \leqslant |x|^\alpha + |y|^\alpha .$$

Без ограничения общности предположим, что $x \geqslant y \geqslant 0$ и рассмотрим функцию $f(x,y) = \frac{1}{2}\left((x+y)^\alpha + (x-y)^\alpha\right) - x^\alpha - y^\alpha$. Чтобы показать, что $f(x,y) \leqslant 0$, заметим, что $f(x,0) = 0$ и $\frac{\partial f(x,y)}{\partial y} \leqslant 0$. В свою очередь, $\frac{\partial f(x,y)}{\partial y} \leqslant 0$, поскольку $\frac{\partial f(x,y)}{\partial y}\Big|_{x=y} \leqslant 0$ и $\frac{\partial^2 f(x,y)}{\partial x\,\partial y} \leqslant 0$. □

Факт 2. Если $\alpha \geqslant 1$, ξ и η — независимые случайные величины, $\mathrm{E}\eta = 0$, $\mathrm{E}|\xi|^\alpha < \infty$, $\mathrm{E}|\eta|^\alpha < \infty$, то

$$\mathrm{E}|\xi+\eta|^\alpha \geqslant \mathrm{E}|\xi|^\alpha .$$

Доказательство. Доказательство следует из неравенства Йенсена. □

Теперь докажем лемму 22. Рассмотрим случайные величины $\xi_1', \ldots, \xi_n'$ такие, что ξ_i' имеет такое же распределение, как и ξ_i, и $\xi_1, \ldots, \xi_n, \xi_1', \ldots, \xi_n'$ — независимы. Тогда из факта 1 следует, что

$$\mathrm{E}|\xi_1+\ldots+\xi_n|^\alpha \leqslant \mathrm{E}|\xi_1-\xi_1'+\ldots+\xi_n-\xi_n'|^\alpha \leqslant \mathrm{E}|\xi_1-\xi_1'|^\alpha + \ldots + |\xi_n-\xi_n'|^\alpha .$$

Осталось заметить, что $\mathrm{E}|\xi_i-\xi_i'|^\alpha \leqslant 2^\alpha \mathrm{E}|\xi_i|^\alpha$.

□

Теорема 24. *Фиксируем вершину p, $N \leqslant p \leqslant n-N+1$, и некоторую такую константу $c > 0$, что $|q(p) - \mathrm{E}\zeta| \leqslant N^c/2$. Имеется две возможности:*
1) если $\gamma > 2$, то

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N} |W_p^q(i) - N\mathrm{E}\zeta| \geqslant N^c\right) \leqslant \frac{48\,\mathrm{D}(\zeta)}{N^{2c-1}};$$

2) если $1 < \gamma \leqslant 2$ и $1 < \alpha < \gamma$, то

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N} |W_p^q(i) - N\mathrm{E}\zeta| \geqslant N^c\right) \leqslant \frac{320\,\mathrm{E}|\zeta - \mathrm{E}\zeta|^\alpha}{N^{\alpha c-1}} .$$

Доказательство.

Заметим, что
$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}|W_p^q(i)-\mathrm{E}W_p^q(1)|\geqslant x\right)\leqslant$$
$$\leqslant \mathsf{P}\left(|W_p^q(1)-\mathrm{E}W_p^q(1)|\geqslant x/2\right)+\mathsf{P}\left(\max_{2\leqslant i\leqslant N}|W_p^q(i)-W_p^q(1)|\geqslant x/2\right).$$
Действительно,
$$\max_{1\leqslant i\leqslant N}|W_p^q(i)-\mathrm{E}W_p^q(1)|\leqslant|W_p^q(1)-\mathrm{E}W_p^q(1)|+\max_{2\leqslant i\leqslant N}|W_p^q(i)-W_p^q(1)|$$
и если $\max_{1\leqslant i\leqslant N}|W_p^q(i)-\mathrm{E}W_p^q(1)|\geqslant x$, то либо $|W_p^q(1)-\mathrm{E}W_p^q(1)|\geqslant x/2$, либо $\max_{2\leqslant i\leqslant N}|W_p^q(i)-W_p^q(1)|\geqslant x/2$.

В случае $\gamma>2$ случайные величины имеют конечные дисперсии, поэтому можно применить неравенства Чебышева и Колмогорова.

Неравенство Чебышева дает неравенство
$$\mathsf{P}(|W_p^q(1)-\mathrm{E}W_p^q(1)|\geqslant x/2)\leqslant\frac{4N\mathrm{D}(\zeta)}{x^2}.$$
Из неравенства Колмогорова следует, что
$$\mathsf{P}\left(\max_{2\leqslant i\leqslant N}|W_p^q(i)-W_p^q(1)|\geqslant x/2\right)=$$
$$=\mathsf{P}\left(\max_{1\leqslant i\leqslant N-1}\left|\sum_{k=1}^{i}(\eta_k-\xi_k)\right|\geqslant x/2\right)\leqslant\frac{8N\,\mathrm{D}(\zeta)}{x^2}.$$
Поэтому в итоге
$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}|W_p^q(i)-\mathrm{E}W_p^q(1)|\geqslant x\right)\leqslant\frac{12N\,\mathrm{D}(\zeta)}{x^2}.$$
Возьмем $x=N^c/2$ и заметим, что $|\mathrm{E}W_p^q(1)-N\mathrm{E}\zeta|=|q(p)-\mathrm{E}\zeta|\leqslant N^c/2$. Поэтому мы получаем
$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}|W_p^q(i)-N\mathrm{E}\zeta|\geqslant N^c\right)\leqslant$$
$$\leqslant\mathsf{P}\left(\max_{1\leqslant i\leqslant N}|W_p^q(i)-\mathrm{E}W_p^q(1)|\geqslant N^c/2\right)\leqslant\frac{48\,\mathrm{D}(\zeta)}{N^{2c-1}}.$$

Теперь рассмотрим случай $1<\gamma\leqslant 2$. Зафиксируем некоторую константу α, $1<\alpha<\gamma$. Вместо неравенства Чебышева, мы можем применить неравенство Маркова и лемму 22:

$$\mathsf{P}(|W_p^q(1) - \mathrm{E}W_p^q(1)| \geqslant x/2) = \mathsf{P}\left(|W_p^q(1) - \mathrm{E}W_p^q(1)|^\alpha \geqslant (x/2)^\alpha\right) \leqslant$$

$$\leqslant \frac{\mathrm{E}|W_p^q(1) - \mathrm{E}W_p^q(1)|^\alpha}{(x/2)^\alpha} \leqslant \frac{4^\alpha N \mathrm{E}|\zeta - \mathrm{E}\zeta|^\alpha}{x^\alpha}.$$

А вместо неравенства Колмогорова можно использовать неравенство Дуба для мартингалов и лемму 22. Заметим, что $S_i = \left|\sum_{j=1}^{i}(\eta_j - \xi_j)\right|$ является субмартингалом (как выпуклая функция от мартингала). Таким образом,

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N-1}\left|\sum_{j=1}^{i}(\eta_j - \xi_j)\right| \geqslant x/2\right) \leqslant \frac{\mathrm{E}\left|\sum_{j=1}^{N-1}(\eta_j - \xi_j)\right|^\alpha}{(x/2)^\alpha} \leqslant$$

$$\leqslant \frac{4^\alpha N \mathrm{E}\left|\eta_1 - \xi_1\right|^\alpha}{x^\alpha} \leqslant \frac{8^\alpha N \mathrm{E}|\zeta - \mathrm{E}\zeta|^\alpha}{x^\alpha}.$$

Поэтому в итоге получаем

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}|W_p^q(i) - \mathrm{E}W_p^q(1)| \geqslant x\right) \leqslant \frac{4^\alpha\,(2^\alpha + 1)\,N\,\mathrm{E}|\zeta - \mathrm{E}\zeta|^\alpha}{x^\alpha}.$$

Теперь возьмем $x = N^c/2$ и заметим, что $|\mathrm{E}W_p^q(1) - N\mathrm{E}\zeta| \leqslant N^c/2$. Как и ранее, можем написать оценку

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N}|W_p^q(i) - N\mathrm{E}\zeta| \geqslant N^c\right) \leqslant \frac{8^\alpha\,(2^\alpha + 1)\,N\,\mathrm{E}|\zeta - \mathrm{E}\zeta|^\alpha}{N^{\alpha c}} \leqslant \frac{320\,\mathrm{E}|\zeta - \mathrm{E}\zeta|^\alpha}{N^{\alpha c - 1}}.$$

$\square$

Математическое ожидание

Через $\rho(d,q)$ обозначим условную вероятность того, что вершина p, $N \leqslant p \leqslant n - N + 1$, имеет входящую степень d, при условии того, что ее качество равно q, то есть $\rho(d,q) = \mathsf{P}(\bar{d}_{in}(p) = d | q(p) = q)$. Заметим, что $\rho(d,q)$ не зависит от p. Через $\#_{in}(d)$ обозначим количество вершин с входящей степенью d. Выполнено $\#_{in}(d) = \#(d+m)$. Математическое ожидание $\#_{in}(d)$ равно

$$\mathrm{E}\#_{in}(d) = (n - 2N)\int_a^\infty f(q)\rho(d,q)dq + r(N), \tag{3.1}$$

где $f(q)$ — плотность распределения Парето, определенная ранее, и $r(N)$, $0 \leqslant r(N) \leqslant 2N$, — остаточный член. Этот остаточный член появляется, поскольку первые и последние N вершин графа отличаются от всех остальных.

Пусть c — это некоторая положительная константа. Мы оценим следующий интеграл:

$$I = \int_a^\infty f(q)\rho(d,q)dq = \int_a^{N^c/2} f(q)\rho(d,q)dq + \int_{N^c/2}^\infty f(q)\rho(d,q)dq = I_1 + I_2.$$

Заметим, что

$$I_2 = \int_{N^c/2}^\infty f(q)\rho(d,q)dq \leqslant \int_{N^c/2}^\infty f(q)dq = \int_{N^c/2}^\infty \frac{\gamma a^\gamma}{q^{\gamma+1}}dq = \frac{(2a)^\gamma}{N^{c\gamma}}\,. \tag{3.2}$$

Рассмотрим событие

$$A = \left\{ \max_{1\leqslant i\leqslant N} |Q(p+i) - N\mathrm{E}\zeta| \leqslant N^c \right\}$$

и следующие условные вероятности:

$$\rho_A(d,q) = \mathsf{P}(\bar d_{in}(p) = d | q(p) = q, A)\,,$$

$$\rho_{\bar A}(d,q) = \mathsf{P}(\bar d_{in}(p) = d | q(p) = q, \bar A)\,.$$

Выполнено:

$$\rho(d,q) = \rho_A(d,q)\,\mathsf{P}(A|q(p)=q) + \rho_{\bar A}(d,q)\,\mathsf{P}(\bar A|q(p)=q)\,. \tag{3.3}$$

Разложим интеграл I_1 в сумму интегралов, применяя (3.3):

$$I_1 = \int_a^{N^c/2} f(q)\rho_A(d,q)\,\mathsf{P}(A_q)dq + \int_a^{N^c/2} f(q)\rho_{\bar A}(d,q)\,\mathsf{P}(\bar A_q)dq = I_1^1 + I_1^2\,,$$

где использованы следующие обозначения:

$$A_q = [A|q(p)=q] = \left\{ \max_{1\leqslant i\leqslant N} |W_p^q(i) - N\mathrm{E}\zeta| \leqslant N^c \right\},$$

$$\bar A_q = [\bar A|q(p)=q] = \left\{ \max_{1\leqslant i\leqslant N} |W_p^q(i) - N\mathrm{E}\zeta| > N^c \right\}.$$

Допустим $N^c/2 > \mathrm{E}\zeta$. Это выполнено, если N достаточно велико (тот факт, что N растет, следует из утверждения теоремы 22, а $\mathrm{E}\zeta$ — константа). Заметим, что

$$I_1^2 \leqslant \max_{q\leqslant N^c/2} \mathsf{P}(\bar A_q) \tag{3.4}$$

и, поскольку $q \leqslant N^c/2$, теорема 24 дает верхнюю оценку, то есть

$$\max_{q\leqslant N^c/2} \mathsf{P}(\bar A_q) = O\left(N^{1-2c}\right) \quad \text{если } \gamma > 2\,, \tag{3.5}$$

$$\max_{q \leqslant N^c/2} \mathsf{P}(\bar{A}_q) = O\left(N^{1-\alpha c}\right) \quad \text{если } 1 < \gamma \leqslant 2\,, \tag{3.6}$$

где $1 < \alpha < \gamma$.

Поэтому рассмотрим теперь I_1^1. Сначала оценим $\rho_A(d,q)$. Напомним, что $\rho_A(d,q) = \mathsf{P}(\bar{d}_{in}(p) = d | q(p) = q, A)$. Заметим, что в течение жизни вершины p, mN независимых ребер могут быть проведены в p. Для ребра из вершины $p+i$ вероятность того, что оно будет проведено в p, равна $\frac{q}{W_p^q(i)}$. При условии события A_q выполнено $N\mathrm{E}\zeta - N^c \leqslant W_p^q(i) \leqslant N\mathrm{E}\zeta + N^c$. Поэтому получаем следующую оценку на $\rho_A(q,d)$:

$$C_{mN}^d \left(\frac{q}{N\mathrm{E}\zeta + N^c}\right)^d \left(1 - \frac{q}{N\mathrm{E}\zeta - N^c}\right)^{mN-d} \leqslant$$

$$\leqslant \rho_A(d,q) \leqslant$$

$$\leqslant C_{mN}^d \left(\frac{q}{N\mathrm{E}\zeta - N^c}\right)^d \left(1 - \frac{d}{N\mathrm{E}\zeta + N^c}\right)^{mN-d}.$$

Таким образом, $\left(1 - \max_{q \leqslant N^c/2} \mathsf{P}(\bar{A}_q)\right) S_- \leqslant I_1^1 \leqslant S_+$, где

$$S_{\mp} = \int_a^{N^c/2} f(q) C_{mN}^d \left(\frac{q}{N\mathrm{E}\zeta \pm N^c}\right)^d \left(1 - \frac{q}{N\mathrm{E}\zeta \mp N^c}\right)^{mN-d} dq\,.$$

Нам потребуется следующая лемма.

Лемма 23. *Пусть N и d растут, но $d = o(N^{1-c})$. Если $1/2 \leqslant c \leqslant 1$, то*

$$S_{\mp} = \frac{\gamma}{d^{\gamma+1}} \left(\frac{(\gamma-1)m}{\gamma}\right)^{\gamma} (1 + o(1))\,.$$

Эта техническая лемма доказывается в разделе 3.5. Теперь, используя эту лемму, мы можем доказать теорему. Используя равенства (3.1), (3.2) и (3.4), мы получаем следующие оценки для $\frac{\mathrm{E}\#(d+m)}{n}$:

$$\left(1 - \frac{2N}{n}\right) \underbrace{\left(1 - \max_{q \leqslant N^c/2} \mathsf{P}(\bar{A}_q)\right) S_-}_{\leqslant I_1^1} \leqslant \frac{\mathrm{E}\#(d+m)}{n} \leqslant$$

$$\leqslant \underbrace{S_+}_{\geqslant I_1^1} + \underbrace{\max_{q \leqslant N^c/2} \mathsf{P}(\bar{A}_q)}_{\geqslant I_1^2} + \underbrace{\frac{(2a)^{\gamma}}{N^{c\gamma}}}_{\geqslant I_2} + \frac{2N}{n}\,. \tag{3.7}$$

Теперь покажем, что для некоторого параметра c все остаточные члены в равенстве (3.7) незначительны по сравнению с главным членом $d^{-\gamma-1}$ из леммы 23.

Сначала рассмотрим случай $\gamma > 2$. Возьмем $c = \frac{\gamma+2}{\gamma+3}$. Заметим, что мы можем применить лемму 23, поскольку $d = o(N^{1-c})$ из условия теоремы 22.

1. $\frac{(2a)^\gamma}{N^{c\gamma}} = o(d^{-\gamma-1})$, если $d = o(N^{c\gamma/(\gamma+1)})$. Это выполнено для $c = \frac{\gamma+2}{\gamma+3}$ и $d = o\left(N^{\frac{1}{\gamma+3}}\right)$.

2. Для $\gamma > 2$, $\mathsf{P}\left(\bar{A}_q\right) = O\left(N^{1-2c}\right) = o(d^{-\gamma-1})$, если $d = o(N^{(2c-1)/(\gamma+1)})$, то есть $d = o(N^{1/(\gamma+3)})$. Здесь мы использовали равенство (3.5).

3. $N = o\left(nd^{-\gamma-1}\right)$, поскольку $d = o\left(\left(\frac{n}{N}\right)^{\frac{1}{\gamma+1}}\right)$.

Рассмотрим случай $\gamma \leqslant 2$. Возьмем $c = \frac{\gamma+2}{\gamma+\alpha+1}$. Заметим, что мы можем применить лемму 23, поскольку $d = o(N^{1-c})$ из условия теоремы 22.

1. $\frac{(2a)^\gamma}{N^{c\gamma}} = o(d^{-\gamma-1})$, если $d = o(N^{c\gamma/(\gamma+1)})$. Это выполнено для $c = \frac{\gamma+2}{\gamma+\alpha+1}$ и $d = o\left(N^{\frac{\alpha-1}{\gamma+\alpha+1}}\right)$.

2. Для $\gamma \leqslant 2$, $\mathsf{P}\left(\bar{A}_q\right) = O\left(N^{1-c\alpha}\right) = o(d^{-\gamma-1})$, если $d = o(N^{(c\alpha-1)/(\gamma+1)})$, то есть $d = o\left(N^{\frac{\alpha-1}{\gamma+\alpha+1}}\right)$. Здесь мы использовали равенство (3.6).

3. $N = o\left(nd^{-\gamma-1}\right)$, поскольку $d = o\left(\left(\frac{n}{N}\right)^{\frac{1}{\gamma+1}}\right)$.

Осталось лишь заметить, что асимптотика для $\#(d)$ совпадает с асимптотикой для $\#(d+m)$. Это завершает доказательство теоремы 22.

Концентрация

Чтобы показать концентрацию, воспользуемся неравенством Чебышева. Для этого мы сначала оценим $\mathrm{D}(\#(d))$. Заметим, что если $|i-j| \geqslant N$, то степени вершин i и j независимы. Поэтому

$$\mathrm{D}(\#(d)) = \sum_{i,j=1}^{n} \left(\mathsf{P}(d_n(i)=d, d_n(j)=d) - \mathsf{P}(d_n(i)=d)\,\mathsf{P}(d_n(j)=d)\right) \leqslant 2nN\,.$$

Применяя неравенство Чебышева, получаем

$$\mathsf{P}\left(|\#(d) - \mathrm{E}\#(d)| > \sqrt{Nn\ln n}\right) \leqslant \frac{\mathrm{D}(\#(d))}{Nn\ln n} \leqslant \frac{2}{\ln n}\,.$$

Замечание. *Вместо неравенства Чебышева можно использовать неравенство Азумы, поскольку* $|\mathrm{E}(\#(d)|G_i) - \mathrm{E}(\#(d)|G_{i-1})| \leqslant (N+1)m$. *В этом случае получим*

$$\mathsf{P}\left(|\#(d) - \mathrm{E}\#(d)| \geqslant \sqrt{n\ln n}(N+1)\right) \leqslant 2n^{-1/2m^2}.$$

Таким образом, с одной стороны, мы получим меньший диапазон степеней, для которых показана концентрация. А именно, концентрация получена для $d = o\left(\left(\frac{\sqrt{n}}{N\sqrt{\ln n}}\right)^{1/(\gamma+1)}\right)$. С другой стороны, получаем более плотную концентрацию. Например, мы можем сказать, что для всех d из этого диапазона количество вершин степени d близко к своему математическому ожиданию.

3.3.2 Свойство устаревания

Напомним, что $e(T)$ — доля ребер в графе, которые соединяют вершины с разницей возрастов, большей чем T, то есть такие вершины i и j, что $|i-j| > T$. В работе [40] было введено свойство устаревания, которое отражает тот факт, что новые страницы в основном соединены со страницами, близкими к ним по возрасту. А именно, в так называемой медийной части веба величина $e(T)$ убывает экспоненциально. В этом параграфе будет показано, что в рассматриваемой модели $e(T)$ убывает линейно.

Теорема 25. *Для любого натурального T*

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

Доказательство. Рассмотрим произвольную вершину $n > N$ и произвольное ребро ni, $i < n$, проведенное из этой вершины. Вероятность того, что $n - i > T$ — это вероятность выбрать одну вершину из $n-N, \ldots, n-T-1$. Поскольку качества вершин — независимые одинаково распределенные случайные величины, вероятности равны $\frac{N-T}{N}$. Из этого следует утверждение теоремы. □

Теорема 26.

$$\mathsf{P}\left(|e(T) - \mathrm{E}e(T)| \geqslant \sqrt{\frac{\ln n}{n}}\right) \leqslant \frac{1}{m \ln n}.$$

Доказательство. Используем неравенство Чебышева. Пусть e_1 и e_2 — любые два различных ребра в графе. Пусть, далее, $l(e_i)$ — разница в возрастах между концами ребраe_i. Легко видеть, что

$$\mathsf{P}(l(e_1) > T, l(e_2) > T) - \mathsf{P}(l(e_1) > T)\,\mathsf{P}(l(e_2) > T) = 0\,.$$

Действительно, если все качества фиксированы, то ребра независимы. После этого можем просто проинтегрировать по всем качествам.

Из этого мы получаем $\mathrm{D}(m\,n\,e(T)) \leqslant mn$, поскольку нужно учитывать только слагаемые, соответствующие совпадающим ребрам, и

$$\mathsf{P}(l(e_1) > T, l(e_1) > T) - \mathsf{P}(l(e_1) > T)\,\mathsf{P}(l(e_1) > T) \leqslant 1\,.$$

Поэтому

$$\mathsf{P}\left(m\,n\,|e(T) - \mathrm{E}e(T)| \geqslant m\sqrt{n\ln n}\right) \leqslant \frac{\mathrm{D}(m\,n\,e(T))}{m^2 n\ln n} \leqslant \frac{1}{m\,\ln n}\,.$$

□

3.4 Функция привлекательности $q(i)e^{-\frac{t-i}{N}}$

Теперь будем рассматривать функцию привлекательности $q(i)e^{-\frac{t-i}{N}}$. В этом случае популярность вершины убывает экспоненциально с возрастом. Опять же, предполагаем, что случайные величины $\zeta_1, \zeta_2, \ldots$ имеют распределение Парето с плотностью $f(x) = \frac{\gamma a^\gamma I[x>a]}{x^{\gamma+1}}$, где $\gamma > 1$, $a > 0$. И ζ — случайная величина с таким распределением Парето.

3.4.1 Распределение степеней

Результаты

Результаты для модели с экспоненциальным фактором устаревания похожи на результаты для модели с индикатором (см. параграф 3.3.1).

Теорема 27. *Пусть $d = d(n)$ растет с ростом n и $d = o\left(\left(\frac{n}{N\ln N}\right)^{\frac{1}{\gamma+1}}\right)$. Если $\gamma > 2$ и $d = o\left(N^{\frac{1}{3\gamma+5}}\right)$ или $1 < \gamma \leqslant 2$ и $d = o\left(N^{\frac{\alpha-1}{\alpha+(\gamma+1)(\alpha+1)}}\right)$ для некоторого α, $1 < \alpha < \gamma$, то*

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

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

Теорема 28. *Для любого d выполнено следующее неравенство:*

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

Как и ранее, для $d = o\left(\left(\frac{n}{N\ln n}\right)^{\frac{1}{2(\gamma+1)}}\right)$ выполнено $\sqrt{Nn\ln n} = o\left(\mathrm{E}\#(d)\right)$ и теорема 28 дает концентрацию.

Далее в этом параграфе будет сначала доказана теорема 27, а затем теорема 28.

Концентрация качества

Фиксируем n и $N = N(n)$. Обозначим через $Q(t)$ суммарное качество всех вершин на шаге t, то есть

$$Q(t) = \sum_{k=1}^{t-1} q(k) e^{-\frac{t-k-1}{N}}\,.$$

Среднее значение $Q(t)$ равно

$$\mathrm{E}Q(t) = \mathrm{E}\zeta \sum_{k=0}^{t-2} e^{-\frac{k}{N}} = \mathrm{E}\zeta \frac{1-e^{-\frac{t-1}{N}}}{1-e^{-\frac{1}{N}}} = N\mathrm{E}\zeta \left(1 + O\left(e^{-t/N}\right) + O\left(1/N\right)\right)\,.$$

Если $t > N \ln N$, то

$$\mathrm{E}Q(t) = N\mathrm{E}\zeta\left(1 + O(1/N)\right).$$

Через $W_p^q(i)$ обозначим суммарное качество всех вершин в тот момент, когда возраст вершины p равен i, при условии того, что качество вершины p равно q.

Теорема 29. *Фиксируем положительную константу c. Пусть $\varphi(N)$ — произвольная функция, такая что $\varphi(N) > \ln(CN)$ для некоторого $C > 0$. Тогда для любого $p > N\varphi(N)$ с $\mid q(p) - \mathrm{E}\zeta \mid \leqslant N^c/3$ выполнено:*
1) если $\gamma > 2$, то

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)} |W_p^q(i) - N\mathrm{E}\zeta| \geqslant N^c\right) = O\left(e^{2\varphi(N)}N^{1-2c}\right);$$

2) если $1 < \gamma \leqslant 2$ и $1 < \alpha < \gamma$, то

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)} |W_p^q(i) - N\mathrm{E}\zeta| \geqslant N^c\right) = O\left(e^{\alpha\varphi(N)}N^{1-\alpha c}\right).$$

Доказательство.
Заметим, что $\mathrm{E}\left(Q(p+i+1) \mid Q(p+i)\right) = Q(p+i)e^{-\frac{1}{N}} + \mathrm{E}\zeta$. Поэтому $X_i = e^{\frac{i}{N}}\left(Q(p+i) - \frac{\mathrm{E}\zeta}{1-e^{-\frac{1}{N}}}\right)$ — мартингал. Действительно,

$$\mathrm{E}\left(X_{i+1} \mid X_i\right) = \left(Q(p+i)e^{-\frac{1}{N}} + \mathrm{E}\zeta\right)e^{\frac{i+1}{N}} - \frac{e^{\frac{i+1}{N}}\mathrm{E}\zeta}{1-e^{-\frac{1}{N}}} = X_i\,.$$

Значит, можем применить неравенство Дуба к субмартингалу $|X_i|$:

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)} \left|e^{\frac{i}{N}}\left(Q(p+i) - \frac{\mathrm{E}\zeta}{1-e^{-1/N}}\right)\right| \geqslant x\right) \leqslant$$

$$\leqslant \frac{\mathrm{E}\left|e^{\frac{N\varphi(N)}{N}}\left(Q(p+N\varphi(N))-\frac{\mathrm{E}\zeta}{1-e^{-1/N}}\right)\right|^{\beta}}{x^{\beta}},$$

где $\beta > 1$.

Если $\gamma > 2$, то возьмем $\beta = 2$ и получим

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)}\left|Q(p+i)-\frac{\mathrm{E}\zeta}{1-e^{-1/N}}\right| \geqslant N^{c}/3\right) \leqslant$$

$$\leqslant \frac{9e^{2\varphi(N)}\mathrm{E}\left(Q(p+N\varphi(N))-\frac{\mathrm{E}\zeta}{1-e^{-1/N}}\right)^{2}}{N^{2c}}.$$

Используя

$$\frac{\mathrm{E}\zeta}{1-e^{-1/N}} = \mathrm{E}Q(p+N\varphi(N)) + \mathrm{E}\zeta \sum_{k=p+N\varphi(N)-1}^{\infty} e^{-\frac{k}{N}},$$

получаем

$$\mathrm{E}\left(Q(p+N\varphi(N))-\frac{\mathrm{E}\zeta}{1-e^{-1/N}}\right)^{2} =$$

$$= \mathrm{E}\left(Q(p+N\varphi(N))-\mathrm{E}Q(p+N\varphi(N))\right)^{2} + (\mathrm{E}\zeta)^{2}\left(\frac{e^{-\frac{p+N\varphi(N)-1}{N}}}{1-e^{-1/N}}\right)^{2} =$$

$$= \mathrm{D}(\zeta)\sum_{k=0}^{p+N\varphi(N)-2} e^{-\frac{2k}{N}} + O\left(\frac{e^{-\frac{2(N\ln(CN)+N\varphi(N))}{N}}}{(1-e^{-1/N})^{2}}\right) =$$

$$= O\left(\frac{1}{1-e^{-2/N}} + e^{-2\varphi(N)}\right) = O\left(N\right).$$

Поэтому

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)}\left|Q(p+i)-\frac{\mathrm{E}\zeta}{1-e^{-1/N}}\right| \geqslant N^{c}/3\right) = O\left(e^{2\varphi(N)}N^{1-2c}\right).$$

Теперь мы можем оценить $W_p^q(i)$, то есть $Q(p+i)$ при условии качества $q(p)$ вершины p. Выполнены неравенства $\mid q(p)-\mathrm{E}\zeta \mid \leqslant N^{c}/3$ и $\left|\frac{\mathrm{E}\zeta}{1-e^{-1/N}} - N\mathrm{E}\zeta\right| \leqslant N^{c}/3$ для больших N, поэтому

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)}|W_p^q(i)-N\mathrm{E}\zeta| \geqslant N^{c}\right) = O\left(e^{2\varphi(N)}N^{1-2c}\right).$$

Аналогично, если $\gamma \leqslant 2$, то возьмем $\beta = \alpha$ и, используя лемму 22, получим

$$\mathsf{P}\left(\max_{1\leqslant i\leqslant N\varphi(N)}|W_p^q(i)-N\mathrm{E}\zeta| \geqslant N^{c}\right) = O\left(e^{\alpha\varphi(N)}N^{1-\alpha c}\right).$$

$\square$

Математическое ожидание

Рассмотрим произвольную функцию $\varphi(N)$, для которой выполняется неравенство $\varphi(N) > \ln(CN)$ с некоторой константой $C > 0$. Обозначим через $\rho(d,q)$ условную вероятность того, что вершина p, $N\varphi(N) \leqslant p \leqslant n - N\varphi(N)+1$, имеет входящую степень d при условии, что ее качество равно q, то есть $\rho(d,q) = \mathsf{P}(\bar{d}_{in}(p) = d | q(p) = q)$. Мы опускаем n и p в обозначении $\rho(d,q)$, поскольку, как будет видно далее, полученные оценки для $\rho(d,q)$ не зависят от этих параметров, если $N\varphi(N) \leqslant p \leqslant n - N\varphi(N) + 1$. Используя это обозначение, получаем следующее равенство:

$$\mathrm{E}\#_{in}(d) = (n - 2N\varphi(N)) \int_a^\infty f(q)\rho(d,q)dq + r(N)\,, \tag{3.8}$$

где $f(q)$ — плотность распределения Парето и $r(N)$, $0 \leqslant r(N) \leqslant 2N\varphi(N)$, — остаточный член.

Фиксируем некоторые константы r и c такие, что $0 < r < 1/2$ и $1/2 < c < 1$. Как и в параграфе 3.3.1, разобьем интеграл на два:

$$I = \int_a^\infty f(q)\rho(d,q)dq = \int_a^{N^r} f(q)\rho(d,q)dq + \int_{N^r}^\infty f(q)\rho(d,q)dq = I_1 + I_2,$$

и

$$I_2 \leqslant \int_{N^r}^\infty f(q)dq = \frac{a^\gamma}{N^{r\gamma}}\,. \tag{3.9}$$

Событие A определяется так же, как и в параграфе 3.3.1:

$$A = \left\{ \max_{1\leqslant i\leqslant N\varphi(N)} |Q(p+i) - N\mathrm{E}\zeta| \leqslant N^c \right\}.$$

Разобьем интеграл I_1 на два интеграла:

$$I_1 = \int_a^{N^r} f(q)\rho_A(d,q)\,\mathsf{P}(A_q)dq + \int_a^{N^r} f(q)\rho_{\bar{A}}(d,q)\,\mathsf{P}(\bar{A}_q)dq = I_1^1 + I_1^2\,,$$

где

$$A_q = [A|q(p)=q] = \left\{ \max_{1\leqslant i\leqslant N\varphi(N)} |W_p^q(i) - N\mathrm{E}\zeta| \leqslant N^c \right\},$$

$$\bar{A}_q = [\bar{A}|q(p)=q] = \left\{ \max_{1\leqslant i\leqslant N\varphi(N)} |W_p^q(i) - N\mathrm{E}\zeta| > N^c \right\}.$$

Можно сделать оценку

$$I_1^2 \leqslant \max_{q\leqslant N^r} \mathsf{P}(\bar{A}_q), \tag{3.10}$$

и для $q \leqslant N^r$ теорема 29 дает верхнюю оценку на $\mathsf{P}(\bar{A}_q)$ (поскольку $N^r < N^c/3$ и $|q(p) - \mathrm{E}\zeta| \leqslant N^r$ для больших N):

$$\max_{q\leqslant N^r} \mathsf{P}(\bar{A}_q) = O\left(e^{2\varphi(N)} N^{1-2c}\right) \quad \text{если } \gamma > 2\,, \tag{3.11}$$

$$\max_{q\leqslant N^r} \mathsf{P}(\bar{A}_q) = O\left(e^{\alpha\varphi(N)}N^{1-\alpha c}\right) \quad \text{если } 1<\gamma\leqslant 2\,, \tag{3.12}$$

где $1<\alpha<\gamma$.

Далее будем оценивать I_1^1. Рассмотрим событие $R_p^q(k)$, состоящее в том, что есть ребро из хотя бы одной вершины $p+i$ с $i\geqslant k$ в вершину p качества q. Тогда для $k>N$ условная вероятность $R_p^q(k)$ при условии A_q может быть оценена следующим образом:

$$\mathsf{P}\left(R_p^q(k)\mid A_q\right) \leqslant \sum_{i=k}^{\infty} \mathsf{P}\left(\text{ребро } (p+i,p) \text{ принадлежит графу } G_n \mid A_q\right) \leqslant$$

$$\leqslant \sum_{i=k}^{\infty} \frac{m\,q\,e^{-\frac{i-1}{N}}}{\sum_{j=0}^{N} ae^{\frac{-j}{N}}} \leqslant \sum_{i=k}^{\infty} \frac{m\,q\,e^{-\frac{i-1}{N}}}{aN/2} = O\left(qe^{-\frac{k}{N}}\right)\,.$$

Эта оценка говорит о том, что наибольший вклад в итоговую степень вершины дают первые несколько шагов после ее появления. Получаем следующую оценку на $\rho_A(d,q)$:

$$\rho_A(d,q) = \rho_{\mp}(d,q) + O\left(qe^{-\varphi(N)}\right),$$

где $\rho_{\mp}(d,q)$ — нижняя и верхняя границы для вероятности того, что вершина p с качеством q имеет входящую степень d в графе $\tilde{G}^n_{p+N\varphi(N)}$ при условии A_q. Можно оценить $\rho_{\mp}(d,q)$ следующим образом. Вершина p имеет входящую степень d в $\tilde{G}^n_{p+N\varphi(N)}$, если d из $m\varphi(N)N$ ребер соединены с этой вершиной, а другие — нет. Для каждого множества индексов $0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N$ нужно перемножить вероятности того, что соответствующие ребра идут в вершину p. При условии A_q эти вероятности могут быть оценены как $\frac{qe^{\frac{-[i_j/m]}{N}}}{N\mathrm{E}\zeta\mp N^c}$. И еще нужно умножить полученное произведение на вероятность того, что другие ребра не ведут в p, то есть $\left(1-\frac{qe^{\frac{-[i/m]}{N}}}{N\mathrm{E}\zeta\pm N^c}\right)$ для соответствующих индексов i. В итоге получаем

$$\rho_{\mp}(d,q) = \prod_{i=0}^{m\varphi(N)N}\left(1-\frac{qe^{\frac{-[i/m]}{N}}}{N\mathrm{E}\zeta\pm N^c}\right)\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}\prod_{j=1}^{d}\frac{\frac{qe^{\frac{-[i_j/m]}{N}}}{N\mathrm{E}\zeta\mp N^c}}{1-\frac{qe^{\frac{-[i_j/m]}{N}}}{N\mathrm{E}\zeta\pm N^c}}\,.$$

Теперь положим

$$S_{\mp}(d,q) := \int_a^{N^r} f(q)\rho_{\mp}(d,q)dq.$$

Используя это обозначение, мы можем оценить I_1^1 следующим образом:

$$I_1^1 \leqslant \int_a^{N^r} f(q)\rho_A(d,q)dq \leqslant S_+ + O\left(\int_a^\infty f(q)qe^{-\varphi(N)}dq\right), \tag{3.13}$$

$$I_1^1 \geqslant \left(1-\max_{q\leqslant N^r}\mathsf{P}(\bar{A}_q)\right)S_- + O\left(\int_a^\infty f(q)qe^{-\varphi(N)}dq\right). \tag{3.14}$$

В следующей лемме оценивается величина $S_\mp$.

Лемма 24. *Пусть d и N растут, $d=o(N^{1-c})$, $d=o(e^{\varphi(N)})$ и $q\leqslant N^r$. Тогда*

$$S_\mp(d,q)=\frac{\gamma}{d^{\gamma+1}}\left(\frac{(\gamma-1)m}{\gamma}\right)^\gamma(1+o(1)).$$

Эта техническая лемма доказывается в разделе 3.5.

Теперь, используя равенства (3.8), (3.9), (3.10), (3.13) и (3.14), получаем

$$\left(1-\frac{2N\varphi(N)}{n}\right)\underbrace{\left(1-\max_{q\leqslant N^r}\mathsf{P}(\bar{A}_q)\right)S_- + O\left(\int_a^\infty f(q)qe^{-\varphi(N)}dq\right)}_{\leqslant I_1^1}\leqslant$$

$$\leqslant\frac{\mathrm{E}\#(d+m)}{n}\leqslant$$

$$\leqslant\underbrace{S_+ + O\left(\int_a^\infty f(q)qe^{-\varphi(N)}dq\right)}_{\geqslant I_1^1}+\underbrace{\max_{q\leqslant N^r}\mathsf{P}(\bar{A}_q)}_{\geqslant I_1^2}+\underbrace{\frac{a^\gamma}{N^{r\gamma}}}_{\geqslant I_2}+\frac{2N\varphi(N)}{n}.$$

Нужно показать, что все остаточные члены равны $o\left(d^{-\gamma-1}\right)$. Для этого нужно найти подходящие значения c и $\varphi(n)$. Напомним, что мы уже предположили, что $d=o\left(N^{1-c}\right)$ и $d=o\left(e^{\varphi(N)}\right)$.

1. $\int_a^\infty f(q)O\left(qe^{-\varphi(N)}\right)dq=O\left(e^{-\varphi(N)}\right)=o\left(d^{-\gamma-1}\right)$, если $d^{\gamma+1}=o\left(e^\varphi(N)\right)$.
2. $\frac{a^\gamma}{N^{r\gamma}}=o(d^{-\gamma-1})$, если $d=o(N^{r\gamma/(\gamma+1)})$. Положим $r=\frac{9}{22}$. Тогда $d=o(N^{r\gamma/(\gamma+1)})$ при выполнении условий теоремы, поскольку $N^{9\gamma/22(\gamma+1)}\geqslant N^{1/(3\gamma+5)}$ для $\gamma>2$ и $N^{9\gamma/22(\gamma+1)}\geqslant N^{\frac{\alpha-1}{\alpha+(\gamma+1)(\alpha+1)}}$ для $1\leqslant\gamma\leqslant 2$.
3. Для $\gamma>2$ выполнено $\max_{q\leqslant N^r}\mathsf{P}(\bar{A}_q)=O\left(e^{2\varphi(N)}N^{1-2c}\right)=o(d^{-\gamma-1})$, если $e^{2\varphi(N)}=o\left(N^{2c-1}d^{-\gamma-1}\right)$. Здесь использовано равенство (3.11).
   Для $\gamma\leqslant 2$ выполнено $\max_{q\leqslant N^r}\mathsf{P}(\bar{A}_q)=O\left(e^{\alpha\varphi(N)}N^{1-\alpha c}\right)=o(d^{-\gamma-1})$, если $e^{\alpha\varphi(N)}=o\left(N^{\alpha c-1}d^{-\gamma-1}\right)$. Здесь использовано равенство (3.12).
4. $N\varphi(N)/n=o(d^{-\gamma-1})$, если $d=o\left(\left(\frac{n}{N\varphi(N)}\right)^{\frac{1}{\gamma+1}}\right)$. При выполнении условий теоремы это верно.

В случае $\gamma > 2$ возьмем $c = \frac{3\gamma+4}{3\gamma+5}$, $\varphi(N) = \ln \frac{N(\gamma+1)}{3\gamma+5}$. Тогда для $d = o\left(N^{1/(3\gamma+5)}\right)$ все условия выполнены.

В случае $\gamma \leqslant 2$ возьмем $c = \frac{1+(\gamma+1)(\alpha+1)}{\alpha+(\gamma+1)(\alpha+1)}$, $\varphi(N) = \ln \frac{N(\alpha-1)(\gamma+1)}{\alpha+(\gamma+1)(\alpha+1)}$. Тогда для $d = o\left(N^{\frac{\alpha-1}{\alpha+(\gamma+1)(\alpha+1)}}\right)$ все условия выполнены.

Концентрация

Докажем теорему 28, используя неравенство Чебышева. Чтобы применить это неравенство, нужно сначала оценить $\mathrm{D}(\#(d))$:

$$\mathrm{D}(\#(d)) = \sum_{i,j=1}^{n} \left(\mathsf{P}(d_n(i) = d, d_n(j) = d) - \mathsf{P}(d_n(i) = d)\,\mathsf{P}(d_n(j) = d)\right).$$

Оценим разность $\mathsf{P}(d_n(i) = d, d_n(j) = d) - \mathsf{P}(d_n(i) = d)\,\mathsf{P}(d_n(j) = d)$ при $i < j$. Заметим, что

$$\mathsf{P}(d_j(i) = d, d_n(j) = d) = \mathsf{P}(d_j(i) = d)\,\mathsf{P}(d_n(j) = d)\,. \tag{3.15}$$

Чтобы доказать это, сначала покажем, что (3.15) выполнено для фиксированных качеств $q_1, \ldots, q_n$, и затем проинтегрируем по всем качествам. Если качество для всех вершин фиксировано, то величина $\mathsf{P}(d_j(i) = d, d_n(j) = d)$ равна сумме по всем таким $i_1, \ldots, i_d$, что $mi < i_1 < \ldots < i_d \leqslant mj$, $mj < j_1 < \ldots < j_d \leqslant mn$, вероятностей того, что соответствующие ребра $([i_k/m], i)$ и $([j_k/m], j)$ проведены, а все другие ребра (i', i) с $i < i' \leqslant j$ и (j', j) с $j < j' \leqslant n$ отсутствуют. Поскольку качества фиксированы, это события независимы и $\mathsf{P}(d_j(i) = d, d_n(j) = d) = \mathsf{P}(d_j(i) = d)\,\mathsf{P}(d_n(j) = d)$.

Пусть $R_p(k)$ — событие, состоящее в том, что есть хотя бы одно ребро из вершины $p + i$ с $i \geqslant k$ в вершину p. Тогда

$$\begin{gathered}
\mathsf{P}(d_n(i) = d, d_n(j) = d) - \mathsf{P}(d_n(i) = d)\,\mathsf{P}(d_n(j) = d) \leqslant \\
\leqslant \mathsf{P}(d_j(i) = d, d_n(j) = d) + \mathsf{P}(R_i(j-i)) - \mathsf{P}(d_n(i) = d)\,\mathsf{P}(d_n(j) = d) = \\
= \mathsf{P}(d_j(i) = d)\,\mathsf{P}(d_n(j) = d) + \mathsf{P}(R_i(j-i)) - \mathsf{P}(d_n(i) = d)\,\mathsf{P}(d_n(j) = d) \leqslant \\
\leqslant \mathsf{P}(d_j(i) = d)\,\mathsf{P}(d_n(j) = d) + \mathsf{P}(R_i(j-i))\,\mathsf{P}(d_n(j) = d) + \\
+ \mathsf{P}(R_i(j-i)) - \mathsf{P}(d_n(i) = d)\,\mathsf{P}(d_n(j) = d) \leqslant \\
\leqslant 2\,\mathsf{P}(R_i(j-i)) = 2\int_a^{\infty} R_i^q(j-i) f(q) dq = \\
= O\left(\int_a^{\infty} q^{-\gamma-1} e^{-\frac{j-i}{N}} dq\right) = O\left(e^{-\frac{j-i}{N}}\right).
\end{gathered}$$

Получили оценку:

$$\mathrm{D}(\#(d)) = O\left(\sum_{1\leqslant i\leqslant j\leqslant n} e^{-\frac{j-i}{N}}\right) = O\left(Nn\right).$$

Теперь можем применить неравенство Чебышева:

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

3.4.2 Свойство устаревания

В этом параграфе мы покажем, что поведение $e(T)$ для модели с экспоненциальным убыванием популярности совпадает с реальностью. В работе [40] было показано, что в некоторых реальных сетях $e(T)$ убывает экспоненциально с ростом T.

Сначала вычислим математическое ожидание $e(T)$. Верна следующая теорема.

Теорема 30. *Для любого натурального* T

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

Действительно, вероятность того, что ребро из вершины k идет в вершину i с $k-i > T$, равна $e^{-\frac{T}{N}} + O\left(e^{-\frac{k}{N}}\right)$. Отсюда следует теорема 30.

В точности как в параграфе 3.3.2, мы можем использовать неравенство Чебышева, чтобы обосновать концентрацию.

Теорема 31. *Для любого натурального* T

$$\mathsf{P}\left(|e(T) - \mathrm{E}e(T)| \geqslant \sqrt{\frac{\ln n}{n}}\right) \leqslant \frac{1}{m\ln n}.$$

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

3.5 Доказательства вспомогательных лемм

3.5.1 Доказательство леммы 23

Сначала докажем следующую лемму.

Лемма 25. *Обозначим* $\frac{N^c}{N\mathrm{E}\zeta}$ *через* ε*. Тогда*

$$S_{\mp} = \frac{\gamma a^{\gamma} C_{mN}^{d}}{(N\mathrm{E}\zeta)^{\gamma}} B\left(d-\gamma, mN-d+1\right) \frac{(1\mp\varepsilon)^{d-\gamma}}{(1\pm\varepsilon)^{d}} \cdot$$
$$\cdot \left(1 + O\left(\frac{\left(1-\frac{\varepsilon}{2\mp 2\varepsilon}\right)^{mN-d+1}(mN)^{d-\gamma}}{\Gamma(d-\gamma)(mN-d+1)}\right) + O\left(\frac{\left(\frac{am}{\mathrm{E}\zeta(1\mp\varepsilon)}\right)^{d-\gamma}}{\Gamma(d-\gamma+1)}\right)\right). \quad (3.16)$$

Доказательство. Запишем $S_{\mp}$ через неполные бета-функции $B(x;a,b)$:

$$S_{\mp} = \int_{a}^{N^c/2} \frac{\gamma a^{\gamma}}{q^{\gamma+1}} C_{mN}^{d} \left(\frac{q}{N\mathrm{E}\zeta \pm N^c}\right)^{d} \left(1-\frac{q}{N\mathrm{E}\zeta \mp N^c}\right)^{mN-d} dq =$$

$$= \frac{\gamma a^{\gamma} C_{mN}^{d} \left(N\mathrm{E}\zeta \mp N^c\right)^{d-\gamma-1}}{\left(N\mathrm{E}\zeta \pm N^c\right)^{d}} \cdot$$
$$\cdot \int_{a}^{N^c/2} \left(\frac{q}{N\mathrm{E}\zeta \mp N^c}\right)^{d-\gamma-1} \left(1-\frac{q}{N\mathrm{E}\zeta \mp N^c}\right)^{mN-d} dq =$$

$$= \frac{\gamma a^{\gamma} C_{mN}^{d} \left(N\mathrm{E}\zeta \mp N^c\right)^{d-\gamma}}{\left(N\mathrm{E}\zeta \pm N^c\right)^{d}} \cdot \int_{\frac{a}{N\mathrm{E}\zeta \mp N^c}}^{\frac{N^c/2}{N\mathrm{E}\zeta \mp N^c}} x^{d-\gamma-1}(1-x)^{mN-d} dx =$$

$$= \frac{\gamma a^{\gamma} C_{mN}^{d} \left(N\mathrm{E}\zeta \mp N^c\right)^{d-\gamma}}{\left(N\mathrm{E}\zeta \pm N^c\right)^{d}} \left(B\left(\frac{N^c/2}{N\mathrm{E}\zeta \mp N^c}; d-\gamma, mN-d+1\right) - \right.$$
$$\left. - B\left(\frac{a}{N\mathrm{E}\zeta \mp N^c}; d-\gamma, mN-d+1\right)\right).$$

Теперь заменим $\frac{N^c}{N\mathrm{E}\zeta}$ на ε:

$$S_{\mp} = \frac{\gamma a^{\gamma} C_{mN}^{d}}{(N\mathrm{E}\zeta)^{\gamma}} \frac{(1\mp\varepsilon)^{d-\gamma}}{(1\pm\varepsilon)^{d}} \left(B\left(\frac{\varepsilon}{2\mp 2\varepsilon}; d-\gamma, mN-d+1\right) - \right.$$
$$\left. - B\left(\frac{a}{N\mathrm{E}\zeta(1\mp\varepsilon)}; d-\gamma, mN-d+1\right)\right).$$

Будем использовать следующие оценки на неполную бета-функцию:

$$B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt = O\left(\int_0^x t^{a-1}dt\right) = O\left(\frac{x^a}{a}\right) ,$$

$$B(x;a,b) = B(a,b) - \int_x^1 (1-t)^{b-1}dt = B(a,b) + O\left(\frac{(1-x)^b}{b}\right) .$$

Эти оценки дают нам

$$S_{\mp} = \frac{\gamma a^{\gamma} C_{mN}^d}{(N\mathrm{E}\zeta)^{\gamma}} \frac{(1\mp\varepsilon)^{d-\gamma}}{(1\pm\varepsilon)^d} \left(B\left(d-\gamma, mN-d+1\right) + \right.$$
$$\left. + O\left(\frac{\left(1-\frac{\varepsilon}{2\mp 2\varepsilon}\right)^{mN-d+1}}{mN-d+1}\right) - O\left(\frac{\left(\frac{a}{N\mathrm{E}\zeta(1\mp\varepsilon)}\right)^{d-\gamma}}{d-\gamma}\right)\right) .$$

Чтобы завершить доказательство, нужно вынести бета-функцию. Будем использовать следующий факт:

$$\frac{1}{B\left(d-\gamma, mN-d+1\right)} = \frac{\Gamma(mN+1-\gamma)}{\Gamma(d-\gamma)\Gamma(mN-d+1)} = O\left(\frac{(mN)^{d-\gamma}}{\Gamma(d-\gamma)}\right) .$$

Получаем, что

$$S_{\mp} = \frac{\gamma a^{\gamma} C_{mN}^d}{(N\mathrm{E}\zeta)^{\gamma}} B\left(d-\gamma, mN-d+1\right) \frac{(1\mp\varepsilon)^{d-\gamma}}{(1\pm\varepsilon)^d} \cdot$$
$$\cdot \left(1 + O\left(\frac{\left(1-\frac{\varepsilon}{2\mp 2\varepsilon}\right)^{mN-d+1}(mN)^{d-\gamma}}{\Gamma(d-\gamma)(mN-d+1)}\right) + O\left(\frac{\left(\frac{am}{\mathrm{E}\zeta(1\mp\varepsilon)}\right)^{d-\gamma}}{\Gamma(d-\gamma+1)}\right)\right) .$$

□

Напомним, что $\varepsilon = N^{c-1}/E\zeta$. Упростим равенство (3.16):

1. $\frac{(1\mp\varepsilon)^{d-\gamma}}{(1\pm\varepsilon)^d} = \frac{\left(1\mp N^{c-1}/E\zeta\right)^{d-\gamma}}{(1\pm N^{c-1}/E\zeta)^d} = 1 + o(1)$, если $d = o(N^{1-c})$.

2. $O\left(\frac{\left(1-\frac{\varepsilon}{2\mp 2\varepsilon}\right)^{mN-d+1}(mN)^{d-\gamma}}{\Gamma(d-\gamma)(mN-d+1)}\right) = O\left(\frac{e^{(mN-d+1)\ln\left(1-\frac{N^{c-1}}{2\mathrm{E}\zeta\mp 2N^{c-1}}\right)+(d-\gamma-1)\ln(mN)}}{\Gamma(d-\gamma)(1-\frac{d-1}{mN})}\right)$

$$= O\left(\frac{e^{\frac{-mN^c\left(1-\frac{d-1}{mN}\right)}{2E\zeta\mp 2N^{c-1}}+(d-\gamma-1)\ln(mN)}}{\Gamma(d-\gamma)(1-\frac{d-1}{mN})}\right) = O\left(\frac{e^{\frac{-mN^c(1-o(1))}{2E\zeta\mp o(1)}+d\ln(mN)}}{\Gamma(d-\gamma)(1-o(1))}\right) = o(1),$$ если $d < \frac{mN^c}{2E\zeta\ln(mN)}$, что верно для достаточно больших N, если $d = o(N^{1-c})$ и $1/2 < c < 1$.

3. $O\left(\frac{\left(\frac{am}{\mathrm{E}\zeta(1\mp\varepsilon)}\right)^{d-\gamma}}{\Gamma(d-\gamma+1)}\right) = o(1)$, если d и N растут.

4. $\frac{\gamma a^\gamma C_{mN}^d}{(N\mathrm{E}\zeta)^\gamma} B\left(d-\gamma, mN-d+1\right) = \frac{\gamma a^\gamma}{(N\mathrm{E}\zeta)^\gamma}\frac{\Gamma(mN+1)}{\Gamma(d+1)\Gamma(mN-d+1)}\frac{\Gamma(d-\gamma)\Gamma(mN-d+1)}{\Gamma(mN+1-\gamma)} =$
$= \frac{\gamma}{d^{\gamma+1}}\left(\frac{am}{\mathrm{E}\zeta}\right)^\gamma(1+o(1))$, если d и N растут.

Осталось заметить, что $\mathrm{E}\zeta = \frac{\gamma a}{\gamma-1}$, поэтому $\frac{\gamma}{d^{\gamma+1}}\left(\frac{am}{\mathrm{E}\zeta}\right)^\gamma = \frac{\gamma}{d^{\gamma+1}}\left(\frac{(\gamma-1)m}{\gamma}\right)^\gamma$.

3.5.2 Доказательство леммы 24

Сначала докажем следующую лемму

Лемма 26. *Если выполнены условия леммы 24 (то есть $d = o(N^{1-c})$, $d = o(e^{\varphi(N)})$ и $q \leqslant N^r$), то*

$$\rho_\mp(d,q) = \left(1 + O\left(\frac{q^2}{N}\right) + o(1)\right)\left(\frac{qm}{\mathrm{E}\zeta}\right)^d \frac{e^{\frac{-qm}{\mathrm{E}\zeta}}}{d!}.$$

Доказательство. Напомним, что

$$\rho_\mp(d,q) = \prod_{i=0}^{m\varphi(N)N}\left(1 - \frac{qe^{\frac{-[i/m]}{N}}}{N\mathrm{E}\zeta \pm N^c}\right)\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}\prod_{j=1}^{d}\frac{\frac{qe^{\frac{-[i_j/m]}{N}}}{N\mathrm{E}\zeta\mp N^c}}{1-\frac{qe^{\frac{-[i_j/m]}{N}}}{N\mathrm{E}\zeta\pm N^c}}. \quad (3.17)$$

Заметим, что

$$\prod_{i=0}^{m\varphi(N)N}\left(1 - \frac{qe^{\frac{-[i/m]}{N}}}{N\mathrm{E}\zeta \pm N^c}\right) = \prod_{i=0}^{m\varphi(N)N}\left(1 - \frac{qe^{\frac{-i/m}{N}}e^{O\left(\frac{1}{N}\right)}}{N\mathrm{E}\zeta \pm N^c}\right) =$$

$$= \exp\left(\sum_{i=0}^{m\varphi(N)N}\ln\left(1 - \frac{qe^{\frac{-i/m}{N}}\left(1+O\left(\frac{1}{N}\right)\right)}{N\mathrm{E}\zeta \pm N^c}\right)\right) =$$

$$= \exp\left(-\sum_{i=0}^{m\varphi(N)N}\left(\frac{qe^{\frac{-i}{mN}}\left(1+O\left(\frac{1}{N}\right)\right)}{N\mathrm{E}\zeta \pm N^c} + O\left(\frac{q^2e^{\frac{-2i}{mN}}}{N^2}\right)\right)\right) =$$

$$
= \exp\left(-\frac{q\left(1-e^{-\varphi(N)}\right)}{(1-e^{\frac{-1}{mN}})(N\mathrm{E}\zeta \pm N^c)}+\right.
$$

$$
\left.+O\left(\frac{q\left(1-e^{-\varphi(N)}\right)}{N^2(1-e^{\frac{-1}{mN}})}\right)+O\left(\frac{q^2\left(1-e^{-2\varphi(N)}\right)}{N^2(1-e^{\frac{-2}{mN}})}\right)\right)=
$$

$$
=\left(1+O\left(\frac{q^2}{N}\right)+O\left(N^{c-1}\right)+O\left(e^{-\varphi(N)}\right)\right)\exp\left(\frac{-qm}{\mathrm{E}\zeta}\right)=
$$

$$
=(1+o(1))\exp\left(\frac{-qm}{\mathrm{E}\zeta}\right). \tag{3.18}
$$

Здесь мы использовали то, что $q \leqslant N^r < N^{1/2}$, поскольку это позволяет нам оценить $\exp\left(O\left(\frac{q^2}{N}\right)\right)$ как $1+O\left(\frac{q^2}{N}\right)$.

Далее,

$$
\prod_{j=1}^{d}\frac{q}{\left(1-\frac{qe^{\frac{-[i_j/m]}{N}}}{N\mathrm{E}\zeta\mp N^c}\right)(N\mathrm{E}\zeta\mp N^c)}=
$$

$$
=\left(\frac{q}{N\mathrm{E}\zeta}\right)^d\left(1+O\left(\frac{d\,q}{N}\right)+O\left(dN^{c-1}\right)\right)=\left(\frac{q}{N\mathrm{E}\zeta}\right)^d(1+o(1)), \tag{3.19}
$$

$$
\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}\prod_{j=1}^{d}e^{\frac{-[i_j/m]}{N}}=\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}e^{\frac{-i_1-\ldots-i_d}{mN}}\left(1+O\left(\frac{d}{N}\right)\right)=
$$

$$
=\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}e^{\frac{-i_1-\ldots-i_d}{mN}}\left(1+o\left(1\right)\right). \tag{3.20}
$$

Осталось оценить $\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}e^{\frac{-i_1-\ldots-i_d}{mN}}$. Используем следующее обозначение:

$$
F(k,d)=\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N}e^{\frac{-i_1-\ldots-i_{d-1}-k\,i_d}{mN}}.
$$

Лемма 27. *Если* $d\,(k+d)=o(N)$ *и* $k+d=o\left(e^{\varphi(N)}\right)$, *то*

$$
F(k,d)=\frac{(mN)^d(k-1)!}{(k+d-1)!}(1+o(1)).
$$

Доказательство. Заметим, что

$$
F(k,1)=\sum_{0\leqslant i_1\leqslant m\varphi(N)N}e^{\frac{-k\,i_1}{mN}}=\frac{1-e^{-k\varphi(N)-\frac{k}{mN}}}{1-e^{-\frac{k}{mN}}}.
$$

Получим рекуррентную формулу для $F(k,d)$:

$$F(k,d)=\sum_{0\leqslant i_1<\ldots<i_d\leqslant m\varphi(N)N} e^{\frac{-i_1-\ldots-i_{d-1}-k\,i_d}{mN}}=$$

$$=\sum_{0\leqslant i_1<\ldots<i_{d-1}\leqslant m\varphi(N)N} e^{\frac{-i_1-\ldots-i_{d-1}}{mN}}\sum_{i_d=i_{d-1}+1}^{m\varphi(N)N} e^{\frac{-k\,i_d}{mN}}=$$

$$=\sum_{0\leqslant i_1<\ldots<i_{d-1}\leqslant m\varphi(N)N} e^{\frac{-i_1-\ldots-i_{d-1}}{mN}}\frac{e^{-\frac{k(i_{d-1}+1)}{mN}}-e^{-k\varphi(N)-\frac{k}{mN}}}{1-e^{-\frac{k}{mN}}}=$$

$$=\frac{e^{-\frac{k}{mN}}}{1-e^{-\frac{k}{mN}}}\sum_{0\leqslant i_1<\ldots<i_{d-1}\leqslant N\varphi(N)}\left(e^{\frac{-i_1-\ldots-(k+1)i_{d-1}}{mN}}-e^{-k\varphi(N)}e^{\frac{-i_1-\ldots-i_{d-1}}{mN}}\right)=$$

$$=\frac{e^{-\frac{k}{mN}}}{1-e^{-\frac{k}{mN}}}\left(F(k+1,d-1)-e^{-k\varphi(N)}F(1,d-1)\right).$$

Несложно получить верхнюю оценку на $F(k,d)$:

$$F(k,d)\leqslant\frac{e^{-\frac{k}{mN}}}{1-e^{-\frac{k}{mN}}}F(k+1,d-1)\leqslant\ldots\leqslant$$

$$\leqslant\frac{e^{-\frac{(2k+d-2)(d-1)}{2mN}}F(k+d-1,1)}{\left(1-e^{-\frac{k}{mN}}\right)\ldots\left(1-e^{-\frac{k+d-2}{mN}}\right)}\leqslant\frac{e^{-\frac{(2k+d-2)(d-1)}{2mN}}}{\left(1-e^{-\frac{k}{mN}}\right)\ldots\left(1-e^{-\frac{k+d-1}{mN}}\right)}=$$

$$=\frac{e^{-\frac{(2k+d-2)(d-1)}{2mN}}}{\left(1-e^{-\frac{k}{mN}}\right)\ldots\left(1-e^{-\frac{k+d-1}{mN}}\right)}=\frac{(mN)^d(k-1)!}{(k+d-1)!}\left(1+O\left(\frac{(k+d)d}{N}\right)\right).$$

Используя эту верхнюю оценку и рекуррентное соотношения, мы можем найти нижнюю оценку. Пусть

$$F(k,d)=\frac{e^{-\frac{k}{mN}}}{1-e^{-\frac{k}{mN}}}\left(F(k+1,d-1)-e^{-k\varphi(N)}F(1,d-1)\right)=\ldots=$$

$$=\frac{e^{-\frac{(2k+d-2)(d-1)}{2mN}}F(k+d-1,1)}{\left(1-e^{-\frac{k}{mN}}\right)\ldots\left(1-e^{-\frac{k+d-2}{mN}}\right)}-$$

$$-\sum_{i=1}^{d-1}\frac{e^{-\frac{(2k+i-1)i}{2mN}}e^{-(k+i-1)\varphi(N)}F(1,d-i)}{\left(1-e^{-\frac{k}{mN}}\right)\ldots\left(1-e^{-\frac{k+i-1}{mN}}\right)}=$$

$$=\frac{(mN)^d(k-1)!}{(k+d-1)!}\left(1+O\left(\frac{(k+d)d}{N}\right)\right)-$$

$$
-\sum_{i=1}^{d-1} e^{-(k+i-1)\varphi(N)}\frac{(mN)^{d-i}}{(d-i)!}\frac{(mN)^i(k-1)!}{(k+i-1)!}\cdot
$$

$$
\cdot\left(1+O\left(\frac{(d-i)^2}{N}\right)+O\left(\frac{(k+i)i}{N}\right)\right)=
$$

$$
=\frac{(mN)^d(k-1)!}{(k+d-1)!}\left(1+o(1)-\sum_{i=1}^{d-1}\frac{e^{-(k+i-1)\varphi(N)}(k+d-1)!}{(d-i)!(k+i-1)!}\left(1+o(1)\right)\right)\geqslant
$$

$$
\geqslant\frac{(mN)^d(k-1)!}{(k+d-1)!}\left(1+o(1)-(1+o(1))\sum_{i=1}^{d-1}\frac{\left((k+d-1)e^{-\varphi(N)}\right)^{k+i-1}}{(k+i-1)!}\right)=
$$

$$
=\frac{(mN)^d(k-1)!}{(k+d-1)!}\left(1+o(1)+O\left(\frac{\left((k+d-1)e^{-\varphi(N)}\right)^{k}}{k!}\right)\right)=
$$

$$
=\frac{(mN)^d(k-1)!}{(k+d-1)!}\left(1+o(1)\right).
$$

□

Наконец, из равенств (3.17)-(3.20) и леммы 27 получаем

$$
\rho_{\mp}(d,q)=e^{\frac{-qm}{\mathrm{E}\zeta}}\left(\frac{q}{N\mathrm{E}\zeta}\right)^d F(1,d)(1+o(1))=
$$

$$
=e^{\frac{-qm}{\mathrm{E}\zeta}}\frac{(mN)^d}{d!}\left(\frac{q}{N\mathrm{E}\zeta}\right)^d(1+o(1))=(1+o(1))\left(\frac{qm}{\mathrm{E}\zeta}\right)^d\frac{e^{\frac{-qm}{\mathrm{E}\zeta}}}{d!}.
$$

□

Теперь можем написать оценку

$$
S_{\mp}(d,q)=\int_a^{N^r} f(q)\rho_{\mp}(d,q)dq=
$$

$$
=\int_a^{N^r}(1+o(1))\frac{\gamma a^\gamma}{q^{\gamma+1}}\left(\frac{qm}{\mathrm{E}\zeta}\right)^d\frac{e^{\frac{-qm}{\mathrm{E}\zeta}}}{d!}dq.
$$

Получаем неполную гамма-функцию:

$$
S_{\mp}(d,q)=(1+o(1))\frac{\gamma a^\gamma m^\gamma}{d!(\mathrm{E}\zeta)^\gamma}\int_{\frac{am}{\mathrm{E}\zeta}}^{\frac{N^r m}{\mathrm{E}\zeta}} x^{d-\gamma-1}e^{-x}dx=
$$

$$
=\frac{\gamma(1+o(1))}{\Gamma(d+1)}\left(\frac{am}{\mathrm{E}\zeta}\right)^\gamma\left(\Gamma\left(d-\gamma,\frac{am}{\mathrm{E}\zeta}\right)-\Gamma\left(d-\gamma,\frac{N^r m}{\mathrm{E}\zeta}\right)\right)=
$$

$$= (1+o(1))\,\frac{\gamma\Gamma(d-\gamma)}{\Gamma(d+1)}\left(\frac{am}{\mathrm{E}\zeta}\right)^{\gamma} = \frac{\gamma}{d^{\gamma+1}}\left(\frac{am}{\mathrm{E}\zeta}\right)^{\gamma}(1+o(1)) =$$

$$= \frac{\gamma}{d^{\gamma+1}}\left(\frac{(\gamma-1)m}{\gamma}\right)^{\gamma}(1+o(1))\,.$$

Заключение

Данная работа посвящена изучению моделей веб-графов, в основу которых положена идея предпочтительного присоединения.

Первые две главы посвящены изучению наиболее известных моделей веб-графов. В первой главе рассмотрена модель Боллобаша–Риордана. В разделе 1.2 вводится определение второй степени вершины и исследуется распределение вторых степеней. Интересный результат состоит в том, что вторые степени вершин, так же, как и обычные степени, распределены в соответствии со степенным законом. Но в случае вторых степеней степенное распределение имеет параметр 2, в отличие от параметра 3 для распределения степеней. В разделе 1.3 изучаются степени больших порядков, а именно — k-е входящие степени. Удается оценить математическое ожидание k-й входящей степени для вершин графа. В разделе 1.4 изучается обобщение диаметра графа — r-диаметр. Оказывается, с ростом n эта величина ведет себя как $\frac{\ln n - \ln r}{\ln \ln n}$.

Во второй главе исследуется более сложная модель — модель Бакли и Остхуса. Для этой модели удалось проанализировать распределение вторых степеней вершин. В данном случае оценка математического ожидания количества вершин заданной второй степени требует более аккуратной техники. А для доказательства концентрации удалось применить неравенство Талаграна. Здесь важным направлением для дальнейших исследований является изучение распределения вторых степеней как в других моделях веб-графов, так и в реальных сетях.

В третьей главе предложена новая модель, основанная на предпочтительном присоединении, для которой, в отличие от ранее изученных моделей, удается строго показать так называемое свойство устаревания, присущее многим реальным сетям. Одним из направлений будущей работы является обобщение этой модели, которое позволить моделировать другие важнейшие свойства сложных сетей, такие как, например, кластерная структура и малый диаметр.

Список литературы

[1] Ю. Л. Павлов и М. М. Степанов. Предельные распределения числа петель случайного конфигурационного графа. *Тр. МИАН*, 283:212–230, 2013.

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

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

[4] А. М. Райгородский. Модели случайных графов. *МЦНМО*, 2011.

[5] А. М. Райгородский. Модели Интернета. *Интеллект*, 2013.

[6] Е. А. Самосват. Моделирование интернета с помощью случайных графов. *Диссертация на соискание ученой степени кандидата физико-математических наук*, 2014.

[7] Ф. Харари. Теория графов. *Москва, "Мир"*, 1973.

[8] W. Aiello, F. Chung, L. Lu. A random graph model for power law graphs. *Experiment. Math.*, 10, 53-66, 2001.

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

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

[11] N. Alon and J. H. Spencer. Probabilistic method. *Wiley–Interscience, New York*, 2002.

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

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

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

[15] L.-A. Barabási, R. Albert, H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica*, A281, 69–77, 2000.

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

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

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

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

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

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

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

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

[24] C. Cooper. Distribution of vertex degree in web-graphs. *Combinatorics, Probability and Computing*, 15:637–661, 2006.

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

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

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

[28] N. Durak, A. Pinar, T.G. Kolda, C. Seshadhri. Degree Relations of Triangles in Real-world Networks and Graph Models. In *Proceedings of CIKM*, 2012.

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

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

[31] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM Conference*, 1999.

[32] D. Gibson, R. Kumar, A. Tomkins. Discovering Large Dense Subgraphs in Massive Graphs. In *Proceedings of the 31st VLDB Conference*, 2005.

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

[34] E. A. Grechnikov. An estimate for the number of edges between vertices of given degrees in random graphs in the Bollobás–Riordan model. *Moscow Journal of Combinatorics and Number Theory*, 1(2):40–73, 2011.

[35] E. A. Grechnikov. The degree distribution and the number of edges between nodes of given degrees in the Buckley–Osthus model of a random web graph. *Internet Mathematics*, 8(3), 2012.

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

[37] M. O. Jackson. Social and economic networks. Princeton University Press, 2010.

[38] M. O. Jackson and A. Watts. The evolution of social and economic networks. *Journal of Economic Theory*, 106(2):265–295, 2002.

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

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

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

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

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

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

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

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

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

[48] S. H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[49] M. Talagrand. Concentration of measures and isoperimetric inequalities in product spaces. *Publications Mathématiques de I'l.H.E.S.*, 81:73–205, 1996.

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

[51] M. Zhukovskiy, D. Vinogradov, Y. Pritykin, L. Ostroumova, E. Grechnikov, G. Gusev, P. Serdyukov, and A. Raigorodskii. Empirical validation of the Buckley–Osthus model for the web host graph: degree and edge distributions. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 1577–1581, 2012.

Публикации автора

[52] Л. А. Остроумова. Математические ожидания k-х входящих степеней вершин в случайных графах в модели Боллобаша–Риордана. *Труды МФТИ*, 13:29–40, 2012.

[53] Л. А. Остроумова. Об r-диаметрах случайных графов в модели Боллобаша–Риордана. *Труды МФТИ*, 13:41–59, 2012.

[54] Л. А. Прохоренкова и Е. А. Самосват. Модели предпочтительного присоединения с устареванием. *Деп. в ВИНИТИ 20.11.14, № 319*, 2014. Расширенная версия статьи находится в печати в журнале Internet Mathematics. [Е. Самосват предложил идею устаревания. Л. Прохоренковой была формализована модель, а также сформулированы и доказаны теоремы.].

[55] A. Kupavskii, L. Ostroumova, D. Shabanov, and P. Tetali. The distribution of second degrees in the Buckley–Osthus random graph model. *Internet Mathematics*, 9(4):297–335, 2013. [А. Купавский доказал лемму 5 в теореме о концентрации. Д. Шабанов и П. Тетали помогали в редактировании текста. Все основные результаты получены Л. Остроумовой.].

[56] L. Ostroumova and E. Grechnikov. The distribution of second degrees in the Bollobás–Riordan random graph model. *Moscow Journal of Combinatorics and Number Theory*, 2(2):85–110, 2012. [Л. Остроумовой были сформулированы и доказаны основные результаты работы. А именно, введено понятие второй степени вершины и получено распределение вторых степеней в модели. Вычислено математическое ожидание количества вершин второй степени d и доказана концентрация этой величины около своего математического ожидания. Е. Гречников помог свернуть сумму в доказательстве теоремы 2.].