

Модели случайных графов и их применения*

А.М. Райгородский

В приложении дается обзор основных современных направлений в теории случайных графов. Делается акцент на связь моделей случайного графа с транспортной проблематикой.

1 Введение

Теория графов играет огромную роль в фундаментальной и прикладной математике. Ей посвящены сотни монографий и тысячи – если не десятки тысяч – статей. Разумеется, мы не можем ставить перед собою цель дать на этих страницах сколь-нибудь подробное изложение теории графов. Нас будет интересовать лишь одно направление, которое с каждым годом становится все более актуальным. В рамках этого направления графы изучаются с вероятностной точки зрения. Типичная постановка вопроса (говоря не совсем строго) такова: *велика ли вероятность того, что граф обладает данным свойством?* Вопрос исключительно важный, и мы в этом не раз убедимся ниже. Правда, в нем ни слова не сказано о том, как именно мы понимаем термин “вероятность”. Всякий человек, имеющий представление об аксиоматике Колмогорова, хорошо знает, что можно вложить множество разных смыслов в этот термин. И его можно действительно определять по-разному. В зависимости от определения получится та или иная модель *случайного графа*. С чисто математических позиций любая такая модель имеет право на существование. Однако для приложений – в том числе приложений к транспортной проблематике – некоторые из этих моделей более интересны, некоторые – менее. Соответственно, ниже мы расскажем о двух классах моделей, каждый из которых за десятилетия, прошедшие с момента своего появления, зарекомендовал себя плодотворным как в рамках “чистой” математики, так и в рамках ее разнообразных приложений, среди которых надежность транспортной сети, рост интернета и других социальных и биологических сетей, теория алгоритмов и пр.

Не претендуя на полноту изложения (это было бы нелепо, т.к. и здесь наука разрослась безгранично), мы постараемся выделить лишь самые основные и принципиальные моменты теории случайных графов.

2 Модель Эрдеша – Реньи

Этот раздел мы посвятим описанию модели случайного графа, которая возникла исторически первой. На рубеже 50ых и 60ых годов XX века эту модель предложили классики современной комбинаторики и теории вероятностей П. Эрдеш и А. Реньи (см. [1], [2], [3]). Отметим, что Эрдеш – это, пожалуй, одна из самых ярких фигур в математике XX века. Ему принадлежат сотни статей и задач, которые оказали огромное влияние на развитие многих математических дисциплин. Реньи также сыграл значительную роль в формировании венгерской вероятностной школы, и его именем назван математический институт в Будапеште.

2.1 Формальное описание модели

Пусть дано множество $V_n = \{1, \dots, n\}$, элементы которого мы назовем *вершинами*. Именно на этом множестве мы будем “строить” случайный граф. Понятно, стало быть, что случайным будет

*Работа выполнена при финансовой поддержке гранта РФФИ 09-01-00294, гранта Президента РФ МД-8390.2010.1, гранта поддержки ведущих научных школ НШ-8784.2010.1.

множество ребер графа. Мы не хотим сейчас рассматривать графы с кратными ребрами (мультиграфы), графы с петлями (псевдографы) и ориентированные графы (орграфы). Поэтому мы считаем, что потенциальных ребер у графа не больше, чем C_n^2 штук. Будем соединять любые две вершины i и j ребром с некоторой вероятностью $p \in [0, 1]$ независимо от всех остальных $C_n^2 - 1$ пар вершин. Иными словами, ребра появляются в соответствии со стандартной схемой Бернулли, в которой C_n^2 испытаний и “вероятность успеха” p . Обозначим через E случайное множество ребер, которое возникает в результате реализации такой схемы. Положим $G = (V_n, E)$. Это и есть случайный граф в модели Эрдеша – Реньи.

Если записывать приведенное только что определение в формате аксиоматики Колмогорова, то мы имеем вероятностное пространство

$$G(n, p) = (\Omega_n, \mathcal{F}_n, P_{n,p}),$$

в котором

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n,p}(G) = p^{|E|}(1-p)^{C_n^2-|E|}. \quad (1)$$

Здесь через $|A|$ обозначена мощность множества A (количество элементов в нем), а 2^A – это совокупность всех подмножеств множества A .

Элемент сигма-алгебры \mathcal{F}_n – это набор графов. Если нам хочется найти вероятность, с которой граф на n вершинах обладает данным свойством A , то мы просто берем множество $\mathcal{A} \in \mathcal{F}_n$, состоящее из всех графов, для которых выполнено свойство A , и вычисляем

$$P_{n,p}(\mathcal{A}) = \sum_{G \in \mathcal{A}} P_{n,p}(G).$$

Таким образом, вероятность того, например, что случайный граф связан, – это величина, равная сумме вероятностей всех связных графов (на фиксированном множестве вершин). Казалось бы, все совсем просто и мы вряд ли имеем шансы наткнуться здесь на нечто особенно интересное. Однако дело обстоит прямо противоположным образом: специфика вероятностных методов, которые эффективно работают в задачах о случайных графах, позволит нам пронаблюдать весьма нетривиальные и, главное, неожиданные явления, которые возникают даже в этой простой модели и которые влекут приятные следствия для приложений.

Прежде чем двигаться дальше, сделаем еще ряд полезных замечаний. Во-первых, если $p = \frac{1}{2}$, то, как видно из формулы (1), вероятность любого графа равна $2^{-C_n^2}$. Иными словами, в этом специальном случае все графы равновероятны и всякое утверждение о вероятности какого-либо свойства – это, по сути, утверждение о доле графов, данным свойством обладающих.

В действительности, мы не только не обязаны предполагать, что $p = \frac{1}{2}$ (хотя и этот случай очень важен), мы даже можем считать, что с ростом величины n (числа вершин) вероятность p возникновения ребра изменяется. Иначе говоря, $p = p(n)$ – любая функция, значения которой заключены между нулем и единицей. Как правило, в науке о случайных графах важны даже не сами вероятности событий, но их предельные значения. Почему это так, мы скоро увидим.

Скажем, наконец, что свойство выполнено *почти всегда*, если его вероятность стремится к единице при $n \rightarrow \infty$.

2.2 Транспортная интерпретация модели

Представим себе, что в некоторой стране есть 10 городов, которые *попарно* соединены дорогами. Это довольно сильное предположение, но пока сохраним его. Допустим, каждая из дорог за определенный срок изнашивается (т.е. становится непроезжей) с известной вероятностью q . При этом износ данной дороги никак не зависит от совокупного износа остальных дорог. Спрашивается: какова максимальная вероятность q , при которой с вероятностью больше $1/2$ не исчезнет возможность перемещения между любыми двумя городами? По существу, это вопрос о *надежности* транспортной сети: чем выше искомая вероятность q , тем, разумеется, сеть надежнее.

Нетрудно видеть, что вопрос о надежности сети – это, в свою очередь, вопрос о *связности* случайного графа. В самом деле, сопоставим каждому городу вершину $i \in V_{10}$. Тогда “дорога” между

“городами” i и j – это ребро. Износ дороги – это исчезновение ребра. Значит, утверждение “дорога изнашивается с вероятностью q ” равносильно утверждению “ребро появляется с вероятностью $p = 1 - q$ ”. Таким образом, нас интересует, какова *минимальная* вероятность p , при которой в модели Эрдеша – Реньи $G(n, p)$ вероятность связности графа больше половины (граф, скорее, связан, чем несвязен).

Понятно, что если мы заменим число 10 другим числом, то соответствующее минимальное p может измениться. Этим и обусловлено наше желание рассматривать не только постоянные p , но и нетривиальные функции $p = p(n)$.

В параграфе 2.4 мы обсудим ряд строгих утверждений, касающихся сформулированного выше вопроса. Однако есть у нас и более срочное дело: все-таки предположение о том, что города связаны дорогами *попарно*, чересчур сильное, и в следующем параграфе мы приведем модификацию модели Эрдеша – Реньи, в рамках которой это предположение можно будет адекватно ослабить.

2.3 Обобщения модели Эрдеша – Реньи

Пусть по-прежнему $V_n = \{1, \dots, n\}$. Однако теперь вероятность ребра между вершинами i и j мы обозначим через p_{ij} . Иными словами, мы, как и раньше, проводим ребра независимо друг от друга, но с *разными* вероятностями. В формате аксиоматики Колмогорова мы получаем вероятностное пространство

$$G(n, p_{ij}) = (\Omega_n, \mathcal{F}_n, P_{n, p_{ij}}),$$

в котором

$$\Omega_n = \{G = (V_n, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n, p_{ij}}(G) = \prod_{(i, j) \in E} p_{ij} \cdot \prod_{(i, j) \notin E} (1 - p_{ij}).$$

Важный частный случай описанного пространства получается, коль скоро мы фиксируем некоторый граф $H_n = (V_n, E_n)$ и полагаем

$$p_{ij} = \begin{cases} p, & (i, j) \in E_n, \\ 0, & (i, j) \notin E_n. \end{cases}$$

Иначе говоря, ребра графа H_n возникают в случайном графе независимо друг от друга с одной и той же вероятностью $p = p(n) \in [0, 1]$, а ребра, которых в графе H_n нет, не возникают в случайном графе вовсе. Этот вариант модели принято обозначать $G(H_n, p)$. В ней

$$P_{n, p_{ij}}(G) = p^{|E|} (1 - p)^{|E_n| - |E|}.$$

Ясно, что модель $G(H_n, p)$ и есть та самая модель, которая вполне адекватна вопросу о надежности транспортной сети. На сей раз мы не обязаны предполагать, что города попарно соединены дорогами; мы можем с самого начала зафиксировать граф дорог H_n и следить за износом его ребер.

2.4 Некоторые математические результаты о надежности сети

Прежде всего справедлива следующая теорема Эрдеша – Реньи.

Теорема 1. *Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c \ln n}{n}$. Если $c > 1$, то почти всегда случайный граф связан. Если $c < 1$, то почти всегда случайный граф не является связным.*

Этот довольно простой с точки зрения доказательства факт мы обоснуем в параграфе 2.5. Однако при всей своей формальной простоте теорема 1 несет весьма содержательную и в каком-то смысле неожиданную информацию. Действительно, вернемся к вопросу о надежности сети. Пусть число n городов, попарно соединенных дорогами, растет. Тогда, разумеется, величина $p = \frac{c \ln n}{n}$ довольно быстро стремится к нулю. Тем не менее, теорема 1 утверждает, что вероятность сохранения связности графа при уничтожении его ребер с вероятностью $q = 1 - p$ стремится к единице. Грубо говоря, если городов 1000, то мы можем позволить дорогам разрушаться с вероятностью ≈ 0.993 , так что в результате с вероятностью, близкой к единице, перемещение между любыми двумя городами

останется возможным. Поначалу это кажется противоречащим интуиции, но, по здравом размышлении, становится понятно, в чем здесь смысл. Дорог у нас $C_n^2 = \Theta(n^2)$, вероятность износа дороги равна $1 - \Theta(\ln n/n)$ (мы пишем $f = \Theta(g)$ для функций f и g , если существуют константы $c_1, c_2 > 0$, с которыми выполнено $c_1 g \leq f \leq c_2 g$). Значит, ожидаемое количество неизношенных дорог имеет порядок $n \ln n$. Этого хватает для сохранения связности.

При определенной аккуратности в выкладках, которые мы частично проведем в параграфе 2.5, можно доказать, например, такой факт.

Теорема 1'. *Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c \ln n}{n}$. Если $c > 3$, то при $n > 100$*

$$P_{n,p}(G) \geq 1 - \frac{1}{n}.$$

Этот результат совсем замечателен своей конкретностью. Получается, что при той же тысяче городов и вероятности износа дороги $1 - \frac{3 \ln 1000}{1000} \approx 0.98$ вероятность сохранения связности не меньше, чем 0.999!

Теорема 1 любопытна еще и тем, что в ней наблюдается резкий скачок от “почти всегда связности” к “почти всегда несвязности”. Функция $p(n) = \frac{\ln n}{n}$ служит своего рода рубежом, преодоление которого означает переход от ненадежности к надежности. Такой переход принято называть *фазовым*, а соответствующую функцию $p(n)$ – *пороговой*.

Следующая теорема содержит в себе еще более глубокую информацию о природе связности-надежности. Она была доказана самими Эрдешем и Реньи (см. [1], [2], [3]).

Теорема 2. *Рассмотрим модель $G(n, p)$. Пусть $p = \frac{c}{n}$. Если $c < 1$, то найдется такая константа $\beta = \beta(c)$, что почти всегда размер каждой связной компоненты случайного графа не превосходит $\beta \ln n$. Если же $c > 1$, то найдется такая константа $\gamma = \gamma(c)$, что почти всегда в случайном графе есть ровно одна компонента размера $\geq \gamma n$.*

И снова мы имеем фазовый переход – резкое изменение свойств случайного графа при преодолении некоторого порога. В данном случае порогом служит функция $p = \frac{1}{n}$. Оказывается, что если вероятность ребра в $c > 1$ раз “ниже” порога, то все связные компоненты графа, скорее всего, крошечные – имеющие логарифмический от общего числа вершин размер; если же вероятность ребра в $c > 1$ раз “выше” порога, то, скорее всего, найдется компонента с числом вершин порядка n . Такая компонента называется *гигантской*.

Теорема 2 допускает различные уточнения. Например, можно утверждать, что при $c > 1$, помимо единственной гигантской компоненты, в случайном графе ничего сколь-нибудь крупного почти никогда не возникает: все остальные компоненты снова логарифмические. Можно еще аккуратнее описывать размер гигантской компоненты. В действительности, верно не только неравенство $\geq \gamma n$, но и асимптотика $\sim \gamma n$. А В.Е. Степанов доказал, что, опять же при $p = \frac{c}{n}$, $c > 1$, размер гигантской компоненты асимптотически нормален (см. [4], [5], [6]).

В целом, изменение свойств случайного графа при изменении вероятности ребра p принято трактовать как эволюцию графа. Нам кажется, что несколько правильнее говорить о своего рода истории мира. Сначала (при $p \ll \frac{1}{n}$) имеет место “феодализм” – весь граф поделен на несвязанные между собой логарифмические кусочки. Затем (при $p \gg \frac{1}{n}$) возникает “империя” – гигантская компонента. Наконец, при $p \gg \frac{\ln n}{n}$ империя уничтожает “окраины” и добивается Мирового господства – связности.

В терминах надежности смысл теоремы 2 также очевиден: можно еще в $\ln n$ раз уменьшить вероятность сохранности отдельной дороги, и, однако же, если не вся страна, то значительная ее часть окажется консолидированной, т.е. не лишенной инфраструктуры – возможности сообщения между любыми двумя городами.

Глубокий интерес представляет, конечно, устройство мира “внутри” фазовых переходов, т.е. при $p \sim \frac{1}{n}$ и при $p \sim \frac{\ln n}{n}$. В первом случае все совсем сложно, и мы отсылаем читателя к книгам [7], [8], [9]. Во втором случае можно сформулировать, например, следующий понятный результат.

Теорема 3. *Пусть $p = \frac{\ln n + c + o(1)}{n}$. Тогда*

$$P_{n,p}(G) \rightarrow e^{-e^{-c}}, \quad n \rightarrow \infty.$$

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

Здесь уже речь не идет о “почти всегда связности” или “почти всегда несвязности”. Здесь асимптотическая вероятность связности есть, но она лежит в строгих пределах от нуля до единицы.

Все, о чем мы говорили до сих пор, касалось модели $G(n, p)$. Естественно, модель $G(H_n, p)$, будучи более адекватной реальности, является и более сложной для изучения. Главный результат относительно этой модели принадлежит Г.А. Маргулису (см. [11]).

Теорема 4. Пусть $\{H_n\}$ – последовательность графов, реберная связность которых стремится к бесконечности при $n \rightarrow \infty$. Тогда существует пороговая функция p для свойства связности случайного графа в модели $G(H_n, p)$. Иными словами, функция p такова, что для любой функции $p_1 \leq c_1 p$, где $c_1 < 1$, случайный граф в модели $G(H_n, p_1)$ почти всегда не связан, но для любой функции $p_2 \geq c_2 p$, где $c_2 > 1$, случайный граф в модели $G(H_n, p_2)$ почти всегда связан. Под реберной связностью графа понимается минимальное количество его ребер, удаление которых влечет потерю связности.

Теорема 4 нетривиальна, и ее доказательство (а также массу ссылок на близкие результаты) можно найти в книге [8]. Разумеется, поиск пороговой функции, существование которой доказывается в теореме 4, – это всякий раз сложная задача, повязанная на специфику графов из последовательности $\{H_n\}$.

Практический смысл теоремы 4 банален: надо строить дороги так, чтобы связность получающегося графа неуклонно росла. К сожалению, в России положение, как правило, противоположное. Даже в Москве есть много улиц, перекрытие которых означает фактическую потерю связности. Например, таковы улицы, проходящие под железными дорогами: они крайне редки и служат единственными лазейками с одной стороны полотна на другую.

В следующем параграфе мы докажем теорему 1, а в параграфе 2.6 мы обсудим основные идеи доказательства теоремы 2. Отметим, что дополнительную информацию о поведении случайных графов в модели Эрдеша – Реньи можно почерпнуть из книг [7], [8], [9] и [10].

2.5 Доказательство теоремы 1

Сперва обсудим случай $c > 1$.

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

$$X_n = X_n(G) = \begin{cases} 0, & G \\ k, & G^k \end{cases}$$

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

Представим X_n в виде суммы

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

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

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

когда скоро

$$X_{n,k,i} = X_{n,k,i}(G) = \begin{cases} 1, & K_i \\ 0, & \end{cases} G.$$

В итоге

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

Очевидно,

$$\begin{aligned} MX_{n,k,i} &= P_{n,p}(K_i \cap G) \leq \\ &\leq P_{n,p}(K_i \cap V_n \setminus K_i \cap G). \end{aligned}$$

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

$$P_{n,p}(K_i \cap V_n \setminus K_i \cap G) = (1-p)^{k(n-k)},$$

и, значит,

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

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

$$n(1-p)^{n-1} \leq ne^{-p(n-1)} = ne^{-\frac{c(\ln n)(n-1)}{n}} = n \left(\frac{1}{n}\right)^{c(1+o(1))} = o(1),$$

поскольку $c > 1$.

Оставшаяся часть рассуждения состоит в доказательстве того, что слагаемые с $k > 1$ и $k < n-1$ пренебрежимо малы по сравнению с первым слагаемым. Соответствующую выкладку мы пропустим. Если же поверить в ее справедливость, то получится, что вся сумма доминируется первым и последним слагаемыми, а стало быть, и она стремится к нулю.

Теорема 1 для случая $c > 1$ доказана.

Теперь рассмотрим случай $c < 1$. Обозначим через X_n количество изолированных вершин в случайном графе. Запишем

$$X_n = X_{n,1} + \dots + X_{n,n},$$

где

$$X_{n,k} = X_{n,k}(G) = \begin{cases} 1, & k \in V_n \cap G, \\ 0, & \text{иначе.} \end{cases}$$

Тогда

$$MX_n = MX_{n,1} + \dots + MX_{n,n}.$$

В свою очередь,

$$MX_{n,k} = P_{n,p}(k \in G) = (1-p)^{n-1}.$$

Таким образом,

$$MX_n = n(1-p)^{n-1} = n(1-p)^n(1+o(1)) = (1+o(1))ne^{-c \ln n} = (1+o(1))n^{1-c}.$$

Заметим, что ввиду неравенства $c < 1$ выполнено $MX_n \rightarrow \infty$.

Посчитаем дисперсию случайной величины X_n :

$$\begin{aligned} DX_n &= MX_n^2 - (MX_n)^2 = M(X_{n,1} + \dots + X_{n,n})^2 - (MX_n)^2 = \\ &= MX_{n,1}^2 + \dots + MX_{n,n}^2 + \sum_{i \neq j} MX_{n,i}X_{n,j} - (MX_n)^2 = \\ &= MX_{n,1} + \dots + MX_{n,n} + \sum_{i \neq j} MX_{n,i}X_{n,j} - (MX_n)^2 = MX_n + \sum_{i \neq j} MX_{n,i}X_{n,j} - (MX_n)^2. \end{aligned}$$

Далее,

$$MX_{n,i}X_{n,j} = P_{n,p}(i \cap j \in G) = (1-p)^{2n-1} = (1+o(1))(1-p)^{2n},$$

т.е.

$$\sum_{i \neq j} MX_{n,i}X_{n,j} = n(n-1)(1+o(1))(1-p)^{2n} = (1+o(1))n^{2-2c} = (1+o(1))(MX_n)^2.$$

В итоге

$$DX_n = MX_n + (1 + o(1))(MX_n)^2 - (MX_n)^2 = o((MX_n)^2).$$

По неравенству Чебышёва

$$\begin{aligned} P_{n,p}(G) &\leq P_{n,p}(X_n = 0) = P_{n,p}(X_n \leq 0) = P_{n,p}(-X_n \geq 0) = \\ &= P_{n,p}(MX_n - X_n \geq MX_n) \leq \frac{DX_n}{(MX_n)^2} = o(1), \end{aligned}$$

и вторая часть теоремы доказана.

2.6 Идеи доказательства теоремы 2

Метод, о котором мы будем здесь говорить, восходит к Р. Карпу (см. [12]), и в таком виде он описан в книге [9]. Мы лишь перечислим ниже основные шаги рассуждения.

2.6.1 Простейший ветвящийся процесс

Пусть Z_1, \dots, Z_t, \dots – независимые пуассоновские величины с одним и тем же средним λ . Положим

$$Y_0 = 1, \quad Y_t = Y_{t-1} + Z_t - 1.$$

Представить себе описанный только что процесс можно так. В начальный момент времени есть одна частица. Затем она приносит Z_1 потомков и умирает. Заметим, что она может умереть, даже не принеся потомства, т.к. величина Z_1 равна нулю с положительной вероятностью. На следующем шаге все повторяется: какая-то частица (порядок роли не играет) порождает Z_2 новых частиц, а сама гибнет. И так далее. Популяция может выродиться, а может и жить вечно. Хорошо известно, что имеют место следующие результаты.

Теорема 5. Пусть $\lambda \leq 1$. Тогда с вероятностью 1 процесс Y_t вырождается, т.е. $P(\exists t : Y_t \leq 0) = 1$.

Теорема 6. Пусть $\lambda > 1$. Пусть $\gamma \in (0, 1)$ – единственное решение уравнения $1 - \gamma = e^{-\lambda\gamma}$. Тогда процесс Y_t вырождается с вероятностью $1 - \gamma$, т.е. $P(\exists t : Y_t \leq 0) = 1 - \gamma$.

Доказательства теорем 5 и 6 можно найти, например, в [9]. Впрочем, это стандартные факты теории ветвящихся процессов (см. [13]). Забегая вперед, скажем, что величина γ в теореме 6 и одноименная величина в теореме 2 суть одно и то же. Вероятность вырождения процесса Y_t и размер гигантской компоненты напрямую связаны.

2.6.2 Ветвящийся процесс на случайном графе

Пусть дан граф $G = (V, E)$: конкретный граф, не случайный. Зафиксируем какую-нибудь его вершину $v_1 \in V$. Назовем ее “живой”, а все остальные вершины – “нейтральными”. Выберем среди нейтральных вершин всех соседей вершины v_1 . После этого объявим вершину v_1 “мертвой”, ее соседей – живыми, а все остальные вершины – нейтральными. Снова зафиксируем какую-нибудь живую вершину v_2 . Выберем всех ее соседей среди *нейтральных*. Вершину v_2 отправим в царство мертвых, а в живых останутся все, кто был жив, кроме v_2 , и новорожденные “потомки” вершины v_2 . Продолжая этот ветвящийся процесс, мы в конце концов получим кладбище вершин и ни одной живой вершины. Кладбище будет в аккурат компонентой, содержащей v_1 .

Обозначим число живых вершин в момент времени t через Y_t , число нейтральных вершин – через N_t , а число потомков очередной живой вершины, отправляющейся в последний путь, – через Z_t . Тогда, очевидно,

$$Y_0 = 1, \quad Y_t = Y_{t-1} + Z_t - 1.$$

Разумеется, все введенные величины зависят от графа G и от последовательности выбираемых вершин v_1, \dots . Если граф G посчитать случайным, то при любом выборе вершин v_1, \dots получатся случайные величины Y_t, N_t, Z_t на пространстве $G(n, p)$. На самом деле, ясно, конечно, что распределения этих величин не зависят от v_1, \dots ; поэтому мы нигде зависимость от вершин и не указываем.

Сразу понятно, что Z_t не является пуассоновской. Тем не менее, она похожа на пуассоновскую. Дело в том, что она имеет биномиальное распределение с “числом испытаний” N_t и вероятностью “успеха” p . Правда, число испытаний само случайно. По счастью, это не проблема. Удастся доказать, что Y_t имеет вид

$$Y_t = \xi_t + 1 - t, \quad \xi_t \sim \text{Binom}(n - 1, 1 - (1 - p)^t).$$

Подробности можно найти в [9], и мы их здесь не касаемся.

Как известно, биномиальное распределение сходится к пуассоновскому, коль скоро вероятность успеха обратно пропорциональна числу испытаний. Нечто подобное имеет место и у нас ($p = c/n$), и ровно на этом мы сыграем в итоге.

2.6.3 Случай $c < 1$

Положим $t_0 = \lceil \beta \ln n \rceil$, где $\beta = \beta(c)$ – константа, которую мы подберем позднее. Нам хочется доказать, что с большой вероятностью каждая из компонент случайного графа имеет размер не больше t_0 . Но размер компоненты – это момент вырождения процесса Y_t на случайном графе. Значит, интересующее нас утверждение можно записать в следующем виде:

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \rightarrow 0, \quad n \rightarrow \infty.$$

Поскольку

$$P_{n,p}(\exists v_1 : Y_{t_0} > 0) \leq nP_{n,p}(Y_{t_0} > 0),$$

достаточно найти такое β , при котором

$$P_{n,p}(Y_{t_0} > 0) = o\left(\frac{1}{n}\right).$$

Дальнейшие выкладки будут слегка некураторными, но при желании их можно сделать безукоризненно строгими. Мы же хотим максимально прояснить основную суть подхода. Итак,

$$P_{n,p}(Y_{t_0} > 0) = P_{n,p}(\xi_{t_0} \geq t_0) \approx P_{n,p}(\text{Binom}(n, 1 - (1 - p)^{t_0}) \geq t_0) \approx$$

(с учетом асимптотики $1 - (1 - p)^{t_0} \sim pt_0$)

$$\approx P_{n,p}(\text{Binom}(n, pt_0) \geq t_0) \approx$$

(с учетом центральной предельной теоремы)

$$\approx \int_{\frac{t_0 - npt_0}{\sqrt{npt_0(1-pt_0)}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Поскольку $c < 1$, нижний предел интегрирования имеет порядок $\sqrt{t_0}$. Стало быть, весь интеграл не превосходит величины $e^{-\delta t_0}$. Выберем β таким, чтобы $e^{-\delta t_0}$ оказалось меньше, чем $e^{-2 \ln n} = \frac{1}{n^2}$, и в случае $c < 1$ теорема доказана.

2.6.4 Случай $c > 1$

Этот случай гораздо сложнее предыдущего. Здесь ветвящийся процесс на графе нужно “запустить” не один раз, а многократно. Только так удастся доказать, что почти наверняка хотя бы в одном запуске возникнет гигантская компонента. Подробности можно найти в [9], мы же лишь поясним, откуда в текущей ситуации появляется константа γ из формулировки теоремы и почему она совпадает с одноименной константой из теоремы 6.

Грубо говоря, идея следующая. Нам хочется доказать, что есть гигантская компонента. Тогда, как следствие, нам нужно, чтобы ветвящийся процесс на графе не вырождался даже при $t \approx \gamma n$. Иными словами, необходимо, чтобы

$$P_{n,p}(Y_t \leq 0) \rightarrow 0, \quad t \approx \gamma n, \quad n \rightarrow \infty.$$

У нас $p = \frac{c}{n}$. Значит, при $t \sim \alpha n$ выполнено

$$1 - (1 - p)^t \sim 1 - e^{-pt} \sim 1 - e^{-c\alpha}.$$

Применим центральную предельную теорему к

$$P_{n,p}(Y_t \leq 0) \approx P_{n,p}(\text{Binom}(n, 1 - e^{-c\alpha}) \leq \alpha n).$$

Интегрирование пойдет от минус бесконечности до

$$\frac{\alpha n - n(1 - e^{-c\alpha})}{\sqrt{n(1 - e^{-c\alpha})e^{-c\alpha}}}.$$

Если $\alpha < 1 - e^{-c\alpha}$, то мы получим искомое стремление вероятности к нулю. Если $\alpha > 1 - e^{-c\alpha}$, то вероятность, напротив, будет стремиться к единице. Таким образом, критическое значение α , вплоть до которого есть именно стремление к нулю, – это решение уравнения $\alpha = 1 - e^{-c\alpha}$ или, что равносильно, $1 - \alpha = e^{-c\alpha}$. А это и есть уравнение из теоремы 6, коль скоро λ мы заменяем на c .

3 Модели Барабаши – Альберт

В этом разделе мы поговорим о самых современных моделях случайных графов, которые призваны описывать рост различных сетей – социальных, биологических, транспортных. Но в первую очередь речь пойдет об интернете. В 90е годы XX века, когда интернет только зарождался, исследователи уже задались вопросом, каким законам подчиняется рост интернета и какова наиболее адекватная модель для описания свойств этой сети. Одними из первых здесь были А.-Л. Барабаши и Р. Альберт. Они нашли ряд важных эмпирических закономерностей в поведении интернета и на их основе придумали модель, которую впоследствии по-разному формализовывали многие авторы. Соответственно, мы построим наше изложение следующим образом. В первом параграфе мы расскажем о результатах Барабаши – Альберт. Во втором параграфе мы опишем модель Б. Боллобаша и О. Риордана, которая весьма неплохо ложится на статистики Барабаши – Альберт. В третьем, четвертом и пятом параграфах мы обсудим возможные уточнения модели Боллобаша – Риордана.

3.1 Наблюдения Барабаши – Альберт

В своих работах [14], [15], [16] Барабаши и Альберт, а также Х. Джеонг описали те статистики интернета, которые легли в основу науки о росте этой сети – науки, имеющей глубокие приложения как в собственно интернетской проблематике, так и в многочисленных близких дисциплинах. В действительности, большинство реальных сетей (социальные, биологические, транспортные и пр.) имеют похожую “топологию”.

Итак, сперва договоримся о том, что мы понимаем под сетью интернет. Это так называемый *веб-граф*, вершины которого суть какие-либо конкретные структурные единицы в интернете: речь может идти о страницах, сайтах, хостах, владельцах и пр. Для определенности будем считать, что вершинами веб-графа служат именно сайты. Ребрами же мы будем соединять те вершины, между которыми имеются ссылки. При этом разумно проводить столько ребер между двумя вершинами, сколько есть ссылок между соответствующими сайтами. Более того, ребра естественно считать направленными. Таким образом, веб-граф ориентирован и он может иметь кратные ребра, петли и даже кратные петли (ссылки вполне могут идти с одной страницы данного сайта на другую его страницу). Это такой “псевдомультиторграф”. Сразу понятно, что для подобного “зверя” модель Эрдеша – Реньи вряд ли подходит.

Теперь мы готовы перечислить самые основные моменты исследования Барабаши – Альберт. По существу, этих моментов всего три. Во-первых, веб-граф – это весьма разреженный граф. У него на t вершинах примерно kt ребер, где $k \geq 1$ – некоторая константа. Для сравнения, у полного графа на t вершинах $C_t^2 = \Theta(t^2)$ ребер. Однако – и это во-вторых, – диаметр веб-графа исключительно скромен.

В 1999 году он имел величину (см. [16]) 5 – 7. Это хорошо всем известное свойство любой социальной сети, которое принято в обыденной речи характеризовать выражением “мир тесен”. Например, говорят о том, что любые два человека в мире “знакомы через 5 – 6 рукопожатий”. Точно так же и сайты: “кликакая” по ссылкам, можно с любого сайта на любой другой перейти за 5 – 7 нажатий клавиши компьютерной мыши. Конечно, тут есть важная оговорка. Некоторые едва появившиеся сайты могут не быть связаны с внешним по отношению к ним миром. Несколько правильнее сказать, что в веб-графе есть гигантская компонента, и уже ее диаметр невелик. Таким образом, веб-граф очень специфичен: будучи разреженным, он, тем не менее, в известном смысле тесен.

У-третьих, у веб-графа весьма характерное распределение степеней вершин. Эмпирическая вероятность того, что вершина веб-графа имеет степень d , оценивается как c/d^λ , где $\lambda \approx 2.1$, а c – нормирующий множитель, вычисляемый из условия “сумма вероятностей равна 1”. Этот любопытный факт роднит интернет с очень многими реальными сетями – биологическими, социальными, транспортными. Все они подчиняются степенному закону, только у каждой из них свой показатель λ .

Ввиду перечисленных наблюдений, не остается никаких сомнений в том, что модель Эрдеша – Реньи не применима для описания роста интернета и подобных сетей. Если подбором вероятности p еще можно добиться разреженности и тесноты (хотя и не с теми параметрами), то степенной закон совсем уж не имеет отношения к схеме Бернулли, в рамках которой появляются ребра обычного случайного графа. В модели $G(n, p)$ степень каждой вершины случайного графа биномиальна с параметрами $n - 1$ и p , и при тех p , которые мало-мальски гарантируют разреженность (т.е. при $p = \Theta(1/n)$), указанное биномиальное распределение аппроксимируется пуассоновским, а вовсе не степенным.

Сами Барабаши и Альберт предложили очень разумный взгляд на процесс формирования интернета. Давайте считать, сказали они, что в каждый момент времени появляется новый сайт, и этот сайт ставит фиксированное количество ссылок на своих предшественников. На кого он предпочтет сослаться? Наверное, на тех, кто и так уже популярен. Можно допустить, что вероятность, с которой новый сайт поставит ссылку на один из прежних сайтов, пропорциональна числу уже имевшихся на тот сайт ссылок.

Модели случайных графов, основанные на описанной идее, называются моделями *предпочтительного присоединения*. В своих работах Барабаши и Альберт никак не конкретизировали, какую именно из этих моделей они предлагают рассматривать. А эти модели исключительно разнородны по своим свойствам. Ведь можно ставить ссылки независимо друг от друга, а можно еще и зависимости между разными ссылками с одного сайта учитывать. В итоге удается доказать даже такой забавный факт (см. [17]).

Теорема 7. Пусть $f(n)$, $n \geq 2$, – произвольная целочисленная функция, такая, что $f(2) = 0$, $f(n) \leq f(n + 1) \leq f(n) + 1$ для всех $n \geq 2$ и $f(n) \rightarrow \infty$ при $n \rightarrow \infty$. Тогда существует такая модель типа Барабаши – Альберт, что в ней с вероятностью, стремящейся к единице при $n \rightarrow \infty$, случайный граф содержит в точности $f(n)$ треугольников.

Одну из наиболее правильных спецификаций модели Барабаши – Альберт предложили в начале двухтысячных годов Б. Боллобаш и О. Риордан. В следующем параграфе мы ее обсудим.

3.2 Модель Боллобаша – Риордана

Наиболее полно эта модель описана в книге [8] и обзоре [17]. Также имеется малодоступная книга [18]. Мы представим здесь две основных и, по сути, совпадающих модификации этой модели. В одной дается динамическое, а в другой статическое описание случайности. Интуитивно более понятна динамическая модификация, с нее и начнем.

3.2.1 Динамическая модификация

Сперва построим последовательность (случайных) графов $\{G_1^n\}$, в которой у графа с номером n число вершин и ребер равно n . Затем сделаем из нее последовательность $\{G_k^n\}$, в которой у графа с

номером n число вершин равно n , а число ребер равно kn , $k \in \mathbb{N}$.

Итак, пусть $G_1^1 = (\{1\}, \{(1, 1)\})$, т.е. в начальный момент времени есть одна вершина и одна петля. Пусть теперь граф G_1^{n-1} уже построен. У него вершины образуют множество $\{1, \dots, n-1\}$, а ребер у него тоже $n-1$ штука. Добавим вершину n и ребро (n, i) , у которого $i \in \{1, \dots, n\}$. Ребро (n, n) будет появляться с вероятностью $\frac{1}{2n-1}$; ребро (n, i) возникнет с вероятностью $\frac{\deg i}{2n-1}$, где $\deg i$ – степень вершины i в графе G_1^{n-1} . Очевидно, что распределение вероятностей задано корректно, поскольку

$$\sum_{i=1}^{n-1} \frac{\deg i}{2n-1} + \frac{1}{2n-1} = \frac{2n-2}{2n-1} + \frac{1}{2n-1} = 1.$$

Случайный граф G_1^n построен, и он удовлетворяет принципу предпочтительного присоединения.

Осталось перейти к G_k^n . Берем G_1^{kn} . Это граф с kn вершинами и kn ребрами. Делим множество его вершин на последовательные куски размера k :

$$\{1, \dots, k\}, \quad \{k+1, \dots, 2k\}, \quad \dots, \quad \{k(n-1)+1, \dots, kn\}.$$

Объявляем каждый кусок “вершиной”, а ребра сохраняем, т.е. если были ребра внутри куска, то будут кратные петли, а были ребра между двумя различными кусками – будут кратные ребра. Внешне – вполне себе интернет, как мы его и представляли. Вершин стало n , а ребер – по-прежнему kn . Цель реализована.

3.2.2 Статическая модификация, или LCD-модель

Введем такой объект, который называется *линейной хордовой диаграммой*. Вообще-то, он возник в топологии и теории узлов (см., например, [19]), но его комбинаторика оказывается напрямую связана с формированием веб-графа.

Итак, зафиксируем на оси абсцисс на плоскости $2n$ точек: $1, 2, 3, \dots, 2n$. Разобьем эти точки на пары, и элементы каждой пары соединим дугой, лежащей в верхней полуплоскости. Полученный объект назовем *линейной хордовой диаграммой* (*linearized chord diagram* или, короче, LCD по-английски). Дуги в нем могут пересекаться, лежать друг под дружкой, но не могут иметь общих вершин. Количество различных LCD легко считается. Оно равно

$$l_n = \frac{(2n)!}{2^n n!}.$$

По каждой LCD построим граф с n вершинами и n ребрами. Действуем так. Идем слева направо по оси абсцисс, пока не встретим впервые правый конец какой-либо дуги. Пусть этот конец имеет номер i_1 . Объявляем набор $\{1, \dots, i_1\}$ первой вершиной будущего графа. Снова идем от i_1+1 направо до первого правого конца i_2 какой-либо дуги. Объявляем второй вершиной графа набор $\{i_1+1, \dots, i_2\}$. И так далее. Поскольку правых концов у дуг в данной диаграмме n штук, получаем всего n вершин. А ребра порождает дугами. Иными словами, две вершины соединяем ребром, коль скоро между соответствующими наборами есть дуга. Ребра ориентируем справа налево. Аналогично возникают петли. Дуг n , и ребер n .

Теперь считаем LCD случайной, т.е. полагаем вероятность каждой диаграммы равной $1/l_n$. Возникают случайные графы. Можно показать, что такие графы по своим вероятностным характеристикам практически неотличимы от графов G_1^n .

Графы с n вершинами и kn ребрами получаем тем же способом, что и в предыдущем пункте.

3.2.3 Некоторые результаты

Замечательна модель Боллобаша – Риордана не только тем, что с ее помощью наводится порядок в “каше”, которую “заварили” Барабаша и Альберт, но еще и тем, что она полностью адекватна эмпирическим наблюдениям. Прежде всего справедлива

Теорема 8. Для любого $k \geq 2$ и любого $\varepsilon > 0$

$$P \left((1 - \varepsilon) \frac{\ln n}{\ln \ln n} \leq \text{diam } G_k^n \leq (1 + \varepsilon) \frac{\ln n}{\ln \ln n} \right) \rightarrow 1, \quad n \rightarrow \infty.$$

На первый взгляд, утверждение кажется непонятным. Ну, хорошо: диаметр плотно сконцентрирован (по вероятности) около величины $\ln n / \ln \ln n$. А у нас ведь какие-то 5 – 7 были? Так ничего странного. Вершин в интернете образца 1999 года около 10^7 . Значит,

$$\frac{\ln 10^7}{\ln \ln 10^7} = \frac{7 \ln 10}{\ln 7 + \ln \ln 10} \approx 6.$$

Фантастическое попадание. Отметим, что при недавней проверке с другими цифрами эмпирика снова подтвердилась.

Теорема 8 доказана в работе [20] авторами модели. А в работе [21] была внесена ясность и в вопрос о распределении степеней вершин.

Теорема 9. Для любого $k \geq 1$ и любого $d \leq n^{1/15}$

$$M \left(\frac{\left| \left\{ i = 1, \dots, n : \text{deg}_{G_k^n} i = d \right\} \right|}{n} \right) \sim \frac{2k(k+1)}{(d+k+1)(d+k+2)(d+k+3)}. \quad (2)$$

Поскольку k – константа, выражение в правой части (2) имеет вид const/d^3 . Да это же в точности степенной закон! Правда, в формулировке теоремы написано математическое ожидание, а не вероятность, но одно из другого получается за счет мартингальных неравенств и соответствующих теорем о плотной концентрации меры около среднего (см. [21]).

У теоремы 9 есть все же два неприятных момента. Первый состоит в том, что степень d в степенном законе, который в ней устанавливается, равна не 2.1, а 3. Второй – это ограничение $d \leq n^{1/15}$, которое ставит крест на практической применимости теоремы. Даже при $n \approx 10^{12}$, чего в природе (пока) не бывает, мы имеем лишь $d \leq 10^{4/5}$, и это нелепо.

Последний недостаток недавно устранил Е.А. Гречников – исследователь-разработчик в Яндексе, который получил более точный результат практически без ограничений на d (см. [22]).

Первым же недостатком занимались много и, в частности, предлагали различные альтернативные модели. Две из таких моделей мы обсудим в параграфах 3.3 и 3.4. Но прежде скажем еще несколько слов о свойствах LCD-модели.

Пусть H – фиксированный граф. Обозначим через $\sharp(H, G_k^n)$ случайную величину, равную количеству подграфов графа G_k^n , изоморфных графу H . Как распределена эта величина? Изучали ее математическое ожидание в разных специальных случаях. Например, в работе [17] приводится громоздкая общая формула и пара ее симпатичных следствий, которые мы выпишем и здесь.

Теорема 10. Пусть $k \geq 2$. Пусть также K_3 – полный граф на трех вершинах. Тогда

$$M(\sharp(K_3, G_k^n)) = (1 + o(1)) \cdot \frac{k \cdot (k-1) \cdot (k+1)}{48} \cdot (\ln n)^3$$

при $n \rightarrow \infty$.

Теорема 11. Пусть фиксированы $k \geq 2$ и $l \geq 3$. Пусть также C_l – цикл на l вершинах. Тогда

$$M(\sharp(C_l, G_k^n)) = (1 + o(1)) \cdot c_{k,l} \cdot (\ln n)^l$$

при $n \rightarrow \infty$, где $c_{k,l}$ – это положительная константа. Более того, при $k \rightarrow \infty$ имеем $c_{k,l} = \Theta(k^l)$.

Студенты МФТИ А. Рябченко и Е. Самосват недавно (в несколько иной, но очень близкой модели) установили следующий общий факт.

Теорема 12. Пусть задан граф H , степени вершин которого равны d_1, \dots, d_s . Обозначим через $\#(d_i = m)$ число вершин в H , степень каждой из которых равна m . Тогда

$$M(\#(H, G_k^n)) = \Theta\left(n^{\#(d_i=0)} \cdot (\sqrt{n})^{\#(d_i=1)} \cdot (\ln n)^{\#(d_i=2)}\right).$$

Зависимость от k занесена в константу Θ .

Надо полагать, что нечто подобное было известно и авторам статьи [17], но мы ничего похожего в литературе не встречали. А такая запись результата очень удобна. Скажем, в теореме 10 речь идет про K_3 . Ясно, что для K_3 выполнено

$$\#(d_i = 0) = \#(d_i = 1) = 0, \quad \#(d_i = 2) = 3.$$

По теореме 12

$$M(\#(K_3, G_k^n)) = \Theta((\ln n)^3),$$

и это прекрасно согласуется теоремой 10. Аналогично можно разобраться и с циклами (теорема 11). А если взять K_4 – полный граф на четырех вершинах, – то теорема 12 скажет, что средняя его встречаемость в веб-графе постоянна. Иными словами, “тетраэдров” в веб-графах почти не бывает.

Отметим, что в реальном вебе случаются не только тетраэдры, но и клики куда большей мощности. Это связано с деятельностью спамеров, которые искусственно расставляют ссылки, желая повысить рейтинги сайтов, заплативших за раскрутку. Спам в модели Боллобаша – Риордана не учтен, и это тоже минус.

Последний в этом пункте любопытный сюжет связан с распределением не первых, а так называемых вторых степеней вершин случайного веб-графа. Речь идет о величине

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

Эта величина для данной вершины t графа G_1^n равна числу ребер, выходящих из вершин, которые являются соседями вершины t , и не ведущих в t . Граф G_k^n с $k \geq 2$ устроен сложнее и в этом контексте пока не рассматривался. Недавно Л.А. Остроумова при участии Гречникова установила следующий результат (см. [23]).

Теорема 13. Для любого $d > 1$ выполнено

$$M\left(\frac{|\{t \in G_1^n : d_2(t) = d\}|}{n}\right) = \frac{4}{d^2} \left(1 + O\left(\frac{\ln^2 d}{d}\right) + O\left(\frac{d^2}{n}\right)\right).$$

Иными словами, при полном отсутствии каких-либо ограничений мы имеем степенной закон и для вторых степеней.

3.3 Модель Бакли – Остгус

Простейшая идея о том, как можно, слегка модифицировав модель Боллобаша – Риордана, получить отличный от тройки показатель в степенном законе распределения степеней вершин, пришла в голову практически одновременно сразу двум независимым группам исследователей, которые в статьях [24], [25] предложили следующую конструкцию. Зафиксируем положительное вещественное число a . Сперва построим последовательность случайных графов $\{H_{1,a}^n\}$. Граф $H_{1,a}^1$ совпадает с графом G_1^1 из пункта 3.2.1. У графа $H_{1,a}^{n-1}$ вершины образуют множество $\{1, \dots, n-1\}$, а ребер у него $n-1$ штука. Добавляем вершину n и присоединяем ребро (n, i) , дабы получить граф $H_{1,a}^n$. Ребро (n, n) появляется с вероятностью $\frac{a}{(a+1)^{n-1}}$, а ребро (n, i) – с вероятностью $\frac{(\deg i) - 1 + a}{(a+1)^{n-1}}$. Граф $H_{k,a}^n$ получаем из графа $H_{1,a}^{kn}$ стандартной склейкой (см. п. 3.2.1).

Ясно, что при $a = 1$ мы возвращаемся к модели Боллобаша – Риордана. Вообще, число a интерпретируют как “начальную притягательность” каждой вершины. Сами авторы модели не доказали ни одного строгого утверждения относительно тех или иных ее характеристик. Зато П. Бакли и Д.

Остгус обосновали в [26] следующую важную теорему (и именно поэтому модель носить теперь имя Бакли – Остгус).

Теорема 14. Для любого $k \geq 1$, любого целого $a \geq 1$ и любого $d \leq n^{1/100(a+1)}$

$$M \left(\frac{\left| \left\{ i = 1, \dots, n : \deg_{H_{k,a}^n} i = d \right\} \right|}{n} \right) = \Theta(d^{-2-a}).$$

На самом деле, авторы теоремы 14 писали даже асимптотику математического ожидания. Однако мы не приводим ее, т.к. при всей своей важности результат не вполне удовлетворителен. Дело в том, что в нем a целое. Иными словами, мы не можем подставить вместо a величину 0.1, желая получить в итоге правильный степенной закон (с показателем 2.1). Кроме того, ограничение $d \leq n^{1/100(a+1)}$ еще ужаснее ограничения $d \leq n^{1/15}$ из теоремы 9.

Обе проблемы недавно устранил Гречников (см. [27]).

Теорема 15. Для любого $k \geq 1$, любого $a > 0$ и любого $d \geq k$

$$M \left(\frac{\left| \left\{ i = 1, \dots, n : \deg_{H_{k,a}^n} i = d \right\} \right|}{n} \right) \sim (a+1) \frac{\Gamma(ka+a+1)}{\Gamma(ka)} d^{-2-a}.$$

Таким образом, модель Бакли – Остгус и впрямь адекватнее модели Боллобаша – Риордана.

3.4 Модель копирования

Здесь мы опишем еще одну очень интересную модель, которая также призвана объяснить феномен степенного закона в реальных сетях. Эта модель возникла практически в одно время с моделью Барабаша – Альберт. Она принадлежит Р. Кумару, П. Рагхавану, С. Раджагопалану, Д. Сивакумару, А. Томкинсу и Э. Упфалу (см. [28]).

Фиксируем $\alpha \in (0, 1)$ и $d \geq 1$, $d \in \mathbb{N}$. Случайный граф будет расти, и это будет похоже на процесс из пункта 3.2.1. Однако здесь процесс будет устроен сильно по-другому.

В качестве начального графа возьмем любой d -регулярный граф (граф, у которого степень каждой вершины равна d). Пусть построен граф с номером t . Обозначим его $G_t = (V_t, E_t)$. Здесь $V_t = \{u_1, \dots, u_s\}$, где s отличается от t на число вершин начального графа, т.е. на некоторую константу, выражаемую через d . Добавим к G_t одну новую вершину u_{s+1} и d ребер, выходящих из u_{s+1} . Для этого сперва выберем случайную вершину $p \in V_t$ (все вершины в V_t равновероятны). Одно за другим строим ребра из u_{s+1} в V_t . На шаге с номером i , $i \in \{1, \dots, d\}$, разыгрываем случайную величину, которая с вероятностью α принимает значение 1 (“монетка падает решкой кверху”) и с вероятностью $1 - \alpha$ принимает значение 0 (“монетка падает орлом кверху”). Если вышла единица, то выпускаем ребро из u_{s+1} в случайную вершину из V_t (все вершины в V_t равновероятны). Если вышел ноль, то берем i -ого по номеру соседа вершины p . Последнее действие всегда возможно, т.к. по построению у каждой вершины не менее d соседей.

Интуиция за всем этим примерно такая. Появляется новый сайт. Проставляя очередную ссылку, его владелец с некоторой вероятностью будет ориентироваться на кого-то из своих предшественников. Скажем, сайт посвящен автомобилям. Вероятно, владелец возьмет один из уже существовавших сайтов про автомобили и *скопирует* оттуда ссылку (с точки зрения стороннего наблюдателя, вполне случайную). Это ситуация, когда монетка выпала орлом кверху (p – это сайт, с которого копируются ссылки). Однако при простановке ссылки владелец может и никого не копировать, а случайно (по нашему мнению) цитировать кого-то из предшественников. Это случай выпадения решки. Таким образом, $1 - \alpha$ – это вероятность копирования или, если угодно, вероятность выбора, мотивированного тематикой сайта.

Основной результат из [28] – это теорема 16.

Теорема 16. Пусть $N_{t,r}$ – это математическое ожидание числа вершин степени r в графе G_t . Тогда

$$\lim_{t \rightarrow \infty} \frac{N_{t,r}}{t} = \Theta \left(r^{-\frac{2-\alpha}{1-\alpha}} \right).$$

Пафос теоремы в том, что в ней мы снова приходим к степенному закону. Более того, если вероятность копирования близка к 1 (а величина α – к нулю), то показатель степени может равняться ожидаемой величине 2.1, и это хорошо.

В целом, распределение степеней вершин в модели копирования очень похоже на распределение степеней вершин в модели Боллобаша – Риордана. В остальном, модели сильно разнятся. Например, в модели Боллобаша – Риордана практически отсутствуют плотные двудольные подграфы (см. теорему 12); в модели копирования таких подграфов полно. Это особенно важно ввиду того, что спамерские структуры, о которых мы вскользь говорили в конце пункта 3.2.3, зачастую образуют именно двудольные графы с плотной перелинковкой.

3.5 Ориентированная модель

Еще один существенный недостаток всех ранее рассмотренных моделей состоит в том, что по факту в них отсутствует ориентация ребер. Вернее, ориентация есть, но, скажем, в модели Боллобаша – Риордана исходящих ребер у каждой вершины k , что никак не соответствует реальности. Один из вариантов решения проблемы предложили в 2003 году Б. Боллобаш, К. Боргс, Дж. Чайес и О. Риордан (см. [29]).

Идея стандартная: строить случайный граф шаг за шагом. На сей раз, однако, не на каждом шаге, вообще говоря, добавляется новая вершина. А именно, с вероятностью $\alpha \in (0, 1)$ добавляется ровно одна вершина и с вероятностью $1 - \alpha$ новые вершины не появляются. Если реализуется первый вариант (с добавлением вершины), то возможны снова два случая. В первом случае (который возникает с вероятностью β) новая вершина ссылается на одну из старых, и вероятность этого пропорциональна *входящей* степени старой вершины (плюс начальная притягательность $\delta_{in} > 0$). Во втором случае, наоборот, одна из старых вершин ссылается на новую, и вероятность этого пропорциональна *исходящей* степени старой вершины (плюс “начальная тяга к простановке ссылок” $\delta_{out} > 0$). Наконец, в рамках второго варианта (когда новая вершина не возникает) ребро проводится между двумя случайными старыми вершинами. Начало ребра выбирается пропорционально исходящим степеням с учетом δ_{out} ; конец ребра – пропорционально входящим степеням с учетом δ_{in} . Могут возникнуть и петли.

Как показали авторы модели, степенной закон в ней есть, и подбором параметров его можно сделать абсолютно адекватным. Правда, они это сделали лишь для фиксированных значений d степеней вершин. Сейчас Гречников работает над устранением этого недостатка.

Список литературы

- [1] Erdős P., Rényi A. On random graphs I // Publ. Math. Debrecen. — 1959. — V. 6. — P. 290–297.
- [2] Erdős P., Rényi A. On the evolution of random graphs. // Publ. Math. Inst. Hungar. Acad. Sci. — 1960. — V. 5. — P. 17–61.
- [3] Erdős P., Rényi A. On the evolution of random graphs // Bull. Inst. Int. Statist. Tokyo. — 1961. — V. 38. — P. 343–347.
- [4] Степанов В.Е. О вероятности связности случайного графа $g_m(t)$ // Теория вероятностей и ее применения. — 1970. — Т. 15, вып. 1. — С. 55–67.
- [5] Степанов В.Е. Фазовый переход в случайных графах // Теория вероятностей и ее применения. — 1970. — Т. 15, вып. 2. — С. 187–203.
- [6] Степанов В.Е. Структура случайных графов $g_n(x|h)$ // Теория вероятностей и ее применения. — 1972. — Т. 17, вып. 3. — С. 227–242.
- [7] Колчин В.Ф. Случайные графы. — М.: Физматлит, 2004. — 256 с.
- [8] Bollobás B. Random Graphs. — Cambridge Univ. Press, 2001. — 520 с.
- [9] Алон Н., Спенсер Дж. Вероятностный метод. — М: Бином. Лаборатория знаний, 2007. — 320 с.
- [10] Janson S., Luczak T., Rucinski A. Random graphs. — New York: Wiley, 2000. — 333 с.
- [11] Маргулис Г.А. Вероятностные характеристики графов с большой связностью // Проблемы передачи информации. — 1974. — Т. 10. — С. 101–108.
- [12] Karp R. The transitive closure of a random digraph // Random structures and algorithms. — 1990. — V. 1. — P. 73–94.
- [13] Карлин С. Основы теории случайных процессов. — М: Мир, 1971. — 536 с.
- [14] Barabási L.-A., Albert R. Emergence of scaling in random networks // Science. — 1999. — V. 286. — P. 509–512.
- [15] Barabási L.-A., Albert R., Jeong H. Scale-free characteristics of random networks: the topology of the world-wide web // Physica. — 2000. — V. A281. — P. 69–77.
- [16] Albert R., Jeong H., Barabási L.A. Diameter of the world-wide web // Nature. — 1999. — V. 401. — P. 130–131.
- [17] Bollobás B., Riordan O. Mathematical results on scale-free random graphs // Handbook of graphs and networks. — Weinheim: Wiley-VCH, 2003. — P. 1–34.
- [18] Райгородский А.М. Экстремальные задачи теории графов и анализ данных. — М: Регулярная и хаотическая динамика, 2009. — 120 с.
- [19] Stoimenow A. Enumeration of chord diagrams and an upper bound for Vassiliev invariants // J. Knot Theory Ramifications. — 1998. — V. 7, N1. — P. 93–114.
- [20] Bollobás B., Riordan O. The diameter of a scale-free random graph // Combinatorica. — 2004. — V. 24, N1. — P. 5–34.
- [21] Bollobás B., Riordan O., Spencer J., Tusnády G. The degree sequence of a scale-free random graph process // Random Structures Algorithms. — 2001. — V. 18, N3. — P. 279–290.

- [22] *Grechnikov E.A.* An estimate for the number of edges between vertices of given degrees in random graphs in the Bollobás–Riordan model // Moscow Journal of Combinatorics and Number Theory. — 2011. — V. 1, N2. — P.
- [23] *Ostroumova L.A., Grechnikov E.A.* The distribution of second degrees in the Bollobás–Riordan random graph model // arXiv:1108.5585v1.
- [24] *Drinea E., Enachescu M., Mitzenmacher M.* Variations on random graph models for the web // Technical report, Harvard University, Department of Computer Science. — 2001.
- [25] *Dorogoutsev S.N., Mendes J.F.F., Samukhin A.N.* Structure of growing networks with preferential linking // Phys. rev. lett.. — 2000. — V. 85. — P. 4633.
- [26] *Buckley P.G., Osthus D.* Popularity based random graph models leading to a scale-free degree sequence // preprint available from <http://www.informatik.hu-berlin.de/~osthus/>.
- [27] *Grechnikov E.A.* The degree distribution and the number of edges between nodes of given degrees in the Buckley–Osthus model of a random web graph // arXiv:1108.4054.
- [28] *Kumar R., Raghavan P., Rajagopalan S., Sivakumar D., Tomkins A., Upfal E.* Stochastic models for the web graph // Proc. 41st Symposium on Foundations of Computer Science. — 2000.
- [29] *Bollobás B., Borgs Ch., Chayes J., Riordan O.* Directed scale-free graphs // SODA'03 Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms. — 2003.