Случайные графы, модели и генераторы безмасштабных графов

М.М. Берновский. mikhail.bermovskiy@phystech.edu
Н.Н. Кузюрин nnkuz@ispras.ru

Аннотация. В статье рассматриваются различные модели случайных графов, описывающие реальные сети, возникающие в различных областях: биологии, компьтерных науках, инженерии, социологии. Особое внимание уделено моделям, описывающим социальные сети.

Ключевые слова случайный граф, безмасштабная сеть, генератор случайных графов, Эрдеш, Боллобаш, Чунг, Янсон, Лучак.

1. Введение

Теория случайных графов стала интенсивно развиваться с конца 50-х годов прошлого века после публикации статей Эрдеша-Реньи об эволюции случайных графов. В этой модели все ребра появляются случайно и независимо с одинаковой вероятностью p и под эволюцией понимается изменение свойств графов с ростом вероятности p. Оказалось, что в некоторых значениях p происходит так называемый фазовый переход и свойства графа кардинально меняются. В этом направлении было получено много интересных и глубоких результатов. Однако, в начале 2000-х выяснилось, что модель Эрдеша-Реньи плохо описывает реальные графы, возникающие в различных областях, в частности в графы таких социальных сетей как Facebook, Twitter и т.п.

Это породило много новых исследований математических моделей случайных графов, о некоторых из которых и будет идти речь в данной статье.

В задачах описания динамиики социальных сетей основное значение имеет правильный выбор математической модели. На данный момент известно множество моделей случайных графов и безмасштабных (scale-free) сетей, некоторые из которых показали удовлетворительные результаты при сравнении с экспериментальными данными. Наиболее полным справочником по моделям безмасштабных сетей является [1].

Вообще говоря, модели социальных сетей можно разделить на три класса: модели случайных графов (модель Эрдеша-Реньи [2] и ее обобщения), простейшие модели безмасштабных сетей (модель Боллобаша [9] и ее

обобщения, модель копирования [8] и др.) и более гибкие модели безмасштабных сетей (модель Чунг-Лу [3], модель Янсона-Лучака [4] и др.) На наш взгляд третий класс моделей представляет наибольший интерес при моделировании больших реальных социальных сетей, таких как Facebook.

Данный обзор не претендует на полноту и служит в основном для обозначения трендов в обозначенной теме.

2. Модель Эрдеша-Реньи

Зафиксируем натуральное число n и рассмотрим множество $V = \{1,\ldots,n\}$. Таким образом мы задали множество вершин случайного графа. Зададим полный граф K_n на множестве вершин V. Пронумеруем ребра K_n: $e_1,\ldots,e_N$, где $N = \binom{n}{2}$. Зададим некотое $p \in [0,1]$ и будем выбирать ребра из множества $e_1,\ldots,e_n$ согласно схеме Бернулли. Таким образом мы получили случайный граф $G = (V,E)$. Формально выражаясь, мы имеем вероятностное пространство $G(n,p) = (\Omega_n, F_n, P_{n,p})$, в котором:

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

Таким образом, в модели Эрдеша-Реньи каждое ребро *независимо* от других ребер входит в случайный граф с вероятностью p. Модель Эрдеша-Реньи на данный момент является самой изученной моделью случайных графов. Приведем несколько наиболее значительных с нашей точки зрения фактов о ней.

Будем далее говорить, что случайный граф обладает некоторым свойством почти наверное, если вероятность обладания этим свойством стремится к единице.

- **Треугольники в случайном графе.** Обозначим через $T_{3,n}$ случайную величину на пространстве $G(n,p)$, равную количеству треугольников в случайном графе G. Тогда верны следующие три теоремы:

Теорема 1 *Пусть* $\alpha(n) \to 0$ *при* $n \to \infty$. *Если* $p(n) = \dfrac{\alpha(n)}{n}, n \in \boldsymbol{N}$, *то почти наверное* $T_{3,n} = 0$ *(т.е. граф не содержит треугольников).*

Теорема 2 *Пусть* $p(n) \sim \frac{c}{n}$, *где* $c > 0$ – *константа. Тогда* $T_{3,n}$ *имеет асимптотически пуассоновское распределение с параметром* $\lambda = \frac{c^3}{6}$.

Теорема 3 *Пусть* $\omega \to \infty$ *при* $n \to \infty$. *Тогда если* $p(n) = \frac{\omega(n)}{n}, n \in \boldsymbol{N}$, *то* $T_{3,n} \geq 1$.

• **Связность случайного графа.** Одно из самых интересных свойств модели Эрдеша-Реньи – наличие фазового перехода:

Теорема 4 *Пусть* $p = \frac{c \ln n}{n}$. *Если* $c > 1$, *то почти наверное случайный граф связен. Если* $c < 1$, *то почти наверное случайный граф связным не является.*

Теорема 5 *Пусть* $p = \frac{c}{n}$. *Тогда при любом* $c < 1$ *существует такая константа* $\beta = \beta(c) > 0$, *что почти наверное каждая компонента случайного графа имеет не более* $\beta \ln n$ *вершин. При любом* $c > 1$ *существует такая константа* $\gamma = \gamma(c) \in (0,1)$, *что почти наверное среди компонент случайного графа есть одна (гигантская), число вершин которой не меньше* γn.

Т.о., модель Эрдеша-Реньи и ее простейшие обобщения являются слишком негибкими для моделирования больших социальных сетей, т.к., например, в них нет соответствующих фазовых переходов и распределения числа треугольников.

3. Наблюдения Барабаши-Альберт

В своих статьях [5], [6] и [7] авторы заметили следующие закономерности в веб-графе (графе, вершинами которого являются сайты, а ребра соответствуют ссылкам):

- Веб-граф разрежен (на n вершинах у него mn ребер, $m \in N$)
- Веб-граф подчиняется феномену «малого мира» (его диаметр $\approx 5-7$)
- Он подчиняется степенному закону:

$$\frac{|\{v : \deg(v) = d\}|}{n} \approx \frac{c}{d^{\lambda}}, c = const, \lambda \sim 2.1$$

На основании своих наблюдений авторы ввели понятие **предпочтительного присоединения (preferential attachment)**. Рассмотрим процесс генерации

графа. На n-ом шагу мы добавляем новую вершину n с m ребрами, инцидентыми ей, причем вероятность ребра к вершине i пропорциональна степени вершины i (см. Рисунок 1):

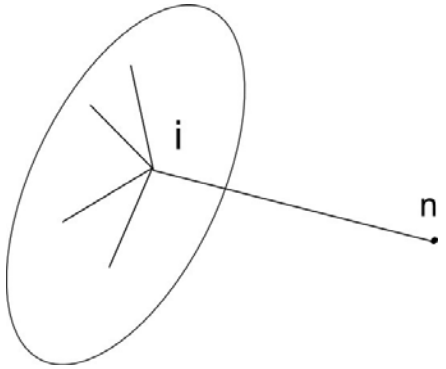

Рис. 1:Добавление вершины на n-ом шагу.

$$P(\textit{ребро из } n \textit{ в } i) = \frac{\deg(i)}{\Sigma_j \deg(j)}. \qquad (1)$$

Основных проблем со спецификацией модели Барабаши-Альберт две.

Во-первых, результирующий граф зависит от начального параметра m. Например, при $m = 1$ модель Барабаши-Альберт описывает генрацию дерева, если начальный граф – тоже дерево. Если начальный граф несвязный, то и все последующие тоже будут таковыми.

Во-вторых, трудность с предпочительным присоединением заключается в случайном выборе вершин (если $m \geq 2$), к которым присоединится новая вершина. Например, верна следующая теорема:

Теорема 1 *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 любых* $n \geq 2$ *и* $f(n) \to \infty$ *при* $n \to \infty$. *Тогда существует такой процесс генерации случайного графа* $T^{(n)}$, *удовлетвовряющий* (1), *что с вероятностью 1 в* $T^{(n)}$ *ровно* $f(n)$ *треугольников для достаточно больших* n.

Говоря менее формально, теорема **1** говорит о том, что если вы хотите иметь в графе с n вершинами $\log n$ треугольников, есть модель Барабаши-Альберт, которая выдаст такой результат.

4. Модель Боллобаша-Риордана

4.1. Генерация графа

Боллобаш и Риордан предложили следующую спецификацию модели Барабаши-Альберта. Построим последовательность случайных графов $\{G_1^n\}$, в которой у графа с номером n число вершин и ребер равно n. Преобразуем ее в последовательность $\{G_k^n\}$, в которой у графа с номером n число вершин равно n, а число ребер равно $kn, k \in N$.

Пусть $G_1^1 = (\{1\},\{(1,1)\})$. Предположим, что граф G_1^{n-1} уже построен. Ребер и вершин у него по $n-1$. Добавим вершину n и ребро (n,i), у которого $i \in \{1,\ldots,n\}$. Ребро (n,n) появится с вероятностью $\frac{1}{2n-1}$, ребро (n,i) – с вероятностью $\frac{\deg(i)}{2n-1}$, причем $\deg(i)$ – степень вершины i в графе G_1^{n-1}. Распределение вероятностей задано корректно, т.к.

$$\Sigma_{i=1}^{n-1} \frac{\deg(i)}{2n-1} + \frac{1}{2n-1} = 1.$$

Т.о., граф G_1^n построен, и он удовлетворяет принципу предпочтительного присоединения. Теперь перейдем к G_1^{kn}, у которого по kn вершин и ребер. Делим множество его вершин на последовательные куски размера k:

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

Каждый кусок примем за новую вершину, а ребра сохраним (ребра внутри куска становятся кратными петлями, ребра между разными кусками – кратными ребрами).

4.2. Основные результаты

Оказалось, что модель Боллобаша-Риордана довольно хорошо сходится с эмпирическими данными. За время изучения этой модели было получено огромное множество полезных результатов, мы же приведем только некоторые из них.

Теорема 1. Для любого $k \geq 2$ и любого $\varepsilon > 0$

$$P\left((1-\varepsilon)\frac{\ln n}{\ln\ln n} \leq diam\, G_k^n \leq (1+\varepsilon)\frac{\ln n}{\ln\ln n}\right) \to 1, n \to \infty. \qquad (2)$$

При $n : 10^7$, что соответствует Интернету образца 1999 года, имеем в (2) $\frac{\ln n}{\ln\ln n} \approx 6$, что совпадает с наблюдениями Барабаши-Альберт.

Теорема 2. Для любого $k \geq 1$ и для любого $d \leq n^{1/15}$

$$E\left(\frac{|\{i=1,\ldots,n : indeg_{G_k^n}(i) = d\}|}{n}\right) \sim \frac{2k(k+1)}{(d+k)(d+k+1)(d+k+2)}, \qquad (3)$$

где $indeg_{G_k^n}(i)$ – количество ребер, имеющих вершину i своим левым концом в графе G_k^n.

Т.е. мы получили степенной закон d^{-3}. От условия $d \leq n^{1/15}$ смогли избавиться сравнительно недавно, а для того чтобы получить степень 2.1 (соответствующую реальной степени веб-графа несколько лет назад) вместо 3, надо отойти от модели Боллобаша-Риордана.

Пусть H – фиксированный граф. Обозначим через $\#(H, G_k^n)$ случайную величину, равную количеству подграфов графа G_k^n, изоморфных графу H. В работе [9] приведены результаты о математическом ожилании этой величины.

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

$$E\left(\#(K_3, G_k^n)\right) = (1+o(1)) \cdot \frac{k(k-1)(k+1)}{48} \cdot (\ln n)^3 \qquad (4)$$

при $n \to \infty$.

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

$$E\left(\#(C_l, G_k^n)\right) = (1+o(1)) \cdot c_{k,l} \cdot (\ln n)^l \qquad (5)$$

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

А. Рябченко и Е. Самосват из Яндекса в модели, близкой к модели Боллобаша-Риордана, установили следующий факт:

Теорема 4 *Пусть задан граф H, степени вершин в котором равны $d_1,\ldots,d_s$. Обозначим через $\#(d_i = m)$ число вершин в H, степень каждой из которых равна m. Тогда*

$$E\left(\#(H, G_k^n)\right) = \Theta\left(n^{\#(d_i=0)} \cdot (\sqrt{n})^{\#(d_i=1)} \cdot (\ln n)^{\#(d_i=2)}\right). \qquad (6)$$

Зависимость от k занесена в константу Θ.

По (6)

$$E\left(\#(K_3, G_k^n)\right) = \Theta((\ln n)^3),$$

что согласуется с теоремой **2**. А для K_4 теорема **4** говорит, что его средняя частота в веб-графе постоянна. Т.о., «тетраэдров» в веб-графе почти нет.

Следует отметить, что последнее утверждение имеет мало общего со свойствами реального веба: в нем встречаются и тетраэдры, и клики большей мощности. Это связано с действием спамеров и агентств по раскрутке сайтов (групп в социальных сетях). Спам в модели Боллобаша-Риордана не учитывается.

5. Модель LCD [10]

Выделим в пространстве ось абсцисс и зафиксируем на ней $2n$ точек: 1, 2, 3, …, $2n$. Разобьем эти точки на пары, и элементы каждой пары соединим дугой, лежащей в верхней полуплоскости. Полученный объект назовем *линейной хордовой диаграммой (LCD)*. Дуги в LCD могут как пересекаться, так и лежать друг под другом, но *не* могут иметь общих вершин. Количество различных диаграмм равно

$$l_n = \frac{(2n)!}{2^n n!}. \qquad (7)$$

По каждой диаграмме построим граф с n вершинами и n ребрами. Процесс построения описан в алгоритме 1 и показан на рис. 2.

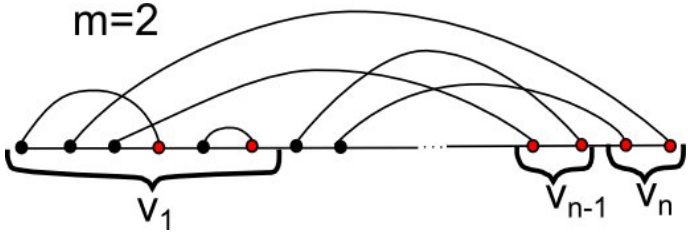

Рис. 2: LCD модель

Algorithm 1 Алгоритм построения графа по LCD

Идем слева направо по оси абсцисс, пока не встретим правый конец какой-либо дуги.
Пусть этот конец имеет номер i_1.
Объявляем набор $\{1,\dots,i_1\}$ **первой вершиной** графа.
Снова повторяем шаг 1.
Пусть новый конец имеет номер i_2.
Объявляем набор $\{i_1+1,\dots,i_2\}$ **второй вершиной** графа.
Продолжаем процедуру до прохода по всем точкам.
Ребра порождаем дугами.

Алгоритм 1

Теперь считаем LCD случайной, т.е. полагаем вероятность каждой диаграммы равной $1/l_n$, где l_n – общее число диаграмм из (7). Т.о. мы получаем случайные графы. В [7] показано, что такие графы по своим вероятностным

характеристикам почти неотличимы от графов G_1^n (см. предыдущий параграф). Графы с n вершинами и kn ребрами получаются так же, как и в модели Боллобаша-Риордана.

6. Модель Buckley-Osthus

Buckley, Osthus и другие исследователи предложили модификацию модели Барабаши-Альберт, в которой вершины обладают «*изначальной привлекательностью» («initial attractiveness»)*: вероятность того, что старая вершина будет выбрана соседом новой вершины пропорциональна ее входящей степени (in-degree) плюс константе, т.е. «изначальной привлекательности», т.е. am, где m – число ребер, входящих в новую вершину. Если $a=1$, то мы просто получаем модель Барабаши-Альберт, т.к. там используется полная степень, а каждая исходящая степень равна m. Buckley и Osthus в работе [11] обобщили результат для произвольной привлекательности a.

Для произвольного фиксированного положительного целого a граф $H_{a,1}^t s$ определяется так же, как и G_1^t в модели Боллобаша-Риордана, но при этом вероятность выбора старой вершины v_i определяется по-другому (см. статью [11] или обзор [9]). Оказывается, в этой модели распределение степеней тоже подчиняется степенному закону. Если обозначить через $\#_{a,m}^n(d)$ число вершин графа $H_{a,m}^n$ с входящей степенью d, то при определенных условиях:

$$\frac{\#_{a,m}^n(d)}{n} = \Theta(d^{-2-a}).$$

7. Модель копирования [8]

7.1. Гененерация графа

Фиксируем $\alpha \in (0,1)$ и $d \geq 1, d \in N$. В качестве начального графа возьмем любой d-регулярный граф. Пусть построен граф с номером t. Обозначим его через $G_t = (V_t, E_t)$, где $V_t = \{u_1,\dots,,u_s\}$, причем s отличается от t на число вершин начального графа, т.е. на константу, выражаемую через d. Добавим к G_t новую вершину u_{s+1} и d ребер, выходящих из u_{s+1}. Для этого сначала случайно выберем вершину $p \in V_t$. Затем строим ребра из u_{s+1} в V_t. На шаге с номером $i, 1 \leq i \leq d$, бросаем неоднородную монетку (падает решкой с вероятностью α, орлом – с вероятностью $1-\alpha$). Если выпала решка, то выпускаем ребро из u_{s+1} в случайную вершину из V_t. Если выпал орел, то

берем i-го по номеру соседа p. Последнее действие всегда возможно, т.к. исходный граф d-регулярен.

7.2. Основной результат

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

$$\lim_{t\to\infty}\frac{N_{t,r}}{t}=\Theta\left(r^{-\frac{2-\alpha}{1-\alpha}}\right). \qquad (8)$$

Если вероятность копирования близка к 1 (α близка к 0), то показатель степени в (8) близок к 2.1, что соответствует наблюдениям Барабаши-Альберт. Также в модели копирования есть плотные двудольные графы, которые соответствуют спамерским структурам, отсутствующим в модели Боллобаша-Риордана.

8. Ориентированные безмасштабные графы

Большинство безмасштабных моделей описывают только неориентированные графы. Но поскольку большинство реальных сетей ориентировано, то логично создать ориентированные модели, в которых предпочтительное присоединение зависит от входящих и исходящих степеней. Такая модель была предложена Боллобашем, Риорданом и др. в работе [12]. В этой модели тоже получается степенной закон для входящих и исходящих степеней вершин.

9. Модель Чунг-Лу

9.1. Генерация графа

Пусть нам задано некоторое конечное множество вершин $V=\{v_1,\dots,v_n\}$ и степень каждой вершины $d_i, i=\overline{1,n}$. Генерация графа $G=(V,E)$ происходит следующим образом:

• Формируем множество L, состоящее из d_i копий v_i для каждого i от 1 до n.

• Задаем случайные паросочетания на множестве L.

• Для вершин u и v из Vs количество ребер в графе G, соединяющее их, равно числу паросочетаний между копиями u и v в L.

Сгенерированный таким образом граф соответствует степенной модели $P(\alpha,\beta)$, описывающей графы, для которых:

$$|\{v\,|\,\deg(v)=x\}|=e^{\alpha}/x^{\beta}.$$

9.2. Основные результаты

• Почти наверное при $\beta<1$ граф связен.

• При $1<\beta<2$ в графе есть гигантская компонента, при этом все остальные компоненты имеют размер $O(1)$.

• При $2<\beta<\beta_0$ в графе есть гигантская компонента, при этом все остальные компоненты имеют размер $O(\log n)$. $\beta_0\approx 3.47$ – решение уравнения $\zeta(\beta-2)-2\zeta(\beta-1)=0$.

• При $\beta=2$ меньшие компоненты имеют размер $O(\log n/\operatorname{loglog} n)$

• При $\beta>\beta_0$ в графе почти наверное нет гигантской компоненты.

9.3. Сравнение с реальными сетями

Авторы проверяли гипотезу на графе звонков (максимальное число узлов 47×10^6). Качественно, прогнозы авторов верны (см. иллюстрации к статье [3] и ее последний раздел).

10. Модель Янсона-Лучака

10.1. Генерация графа

Рассмотрим упорядоченный набор $[n]s$ из ns вершин. Каждой вершине i присваивается вес W_i. Для простоты и понятности положим их независимо и одинаково распределенными случайными величинами со степенным хвостом:

$$P(W>x)=ax^{-\alpha}, x\ge x_0, \quad (9)$$

с некоторыми константами $a>0$ и $\alpha>0$, и некоторым $x_0>0$. Мы обозначаем **наибольший** вес $W_{\max}=\max_i W_i$. Из (9) следует:

$$P(W_{\max}>tn^{1/\alpha})\le nP(W>tn^{1/\alpha})=O(t^{-\alpha}). \quad (10)$$

Следует отметить, что $EW^{\beta}<\infty$, только если $\alpha>\beta$. В частности, при $\alpha<2$ в случае распределения с экспоненциально неограниченным хвостом (heavy tail case) имеем $EW^2=\infty$.

При условии, что нам дан набор весов $\{W_i\}_1^n$, соединим кажду пару вершин $\{i,j\}$ посредством E_{ij} параллельных ребер, где E_{ij} – независимые пуассоновские случайные величины с ожиданием:

$$E(E_{ij})=\lambda_{ij}=b\frac{W_iW_j}{n}, \qquad (11)$$

$b > 0$ – константа. В результате мы получим *мультиграф* $\hat{G}(n,\alpha)$. Далее мы можем стянуть параллельные ребра в одно, т.о. получив простой случайный граф $G(n,\alpha)$, в котором вершины i и j соединены с вероятностью

$$p_{ij} = 1 - \exp(-\lambda_{ij}), \qquad (12)$$

независимой для всех пар (i,j) таких, что $1 \le i < j \le n$.

10.2. Основные результаты

Обозначим за $\omega(G(n,\alpha))$ размер максимальной клики в графе $G(n,\alpha)$. В этиз обозначениях в модели Янсона-Лучака получены следующие результаты:

Теорема 5

(i) Если $0 < \alpha < 2$, то

$$\omega(G(n,\alpha)) = (c + o_p(1))n^{1-\alpha/2}(\log n)^{-\alpha/2},$$

где

$$c = ab^{\alpha/2}(1-\alpha/2)^{-\alpha/2}. \qquad (13)$$

(ii) Если $\alpha = 2$, то $\omega(G(n,\alpha)) = O_p(1)$.

(iii) Если $\alpha > 2$, то почти наверное $\omega(G(n,\alpha)) \in \{2,3\}$. Более того, при $n \to \infty$:

$$P(\omega(G(n,\alpha)) = 2) \to e^{-\frac{1}{6}(bE(W^2))^3} \qquad (14)$$

$$P(\omega(G(n,\alpha)) = 3) \to 1 - e^{-\frac{1}{6}(bE(W^2))^3} \qquad (15)$$

Основной задачей при работе с веб-графом является поиск клик в нем. Опишем несколько разных типов клик в модели Янсона-Лучака:

- Жадная клика (greedy clique)

$$\mathrm{K}_{gr} = \{i : i \sim j \text{ для всех } j \text{ с } W_j > W_i \text{ и } j \in \mathrm{K}_{gr}\}.$$

- Quasi top clique

$$\mathrm{K}_{qt} = \{i : i \sim j \text{ для всех } j \text{ с } W_j > W_i\}$$

- Full top clique

$$\mathrm{K}_{ft} = \{i : j \sim k \text{ для всех различных } j,k \text{ с } W_j, W_k \ge W_i\}.$$

Имеем $\boldsymbol{K}_{ft} \subseteq \boldsymbol{K}_{qt} \subseteq \boldsymbol{K}_{gr}$. Обозначим за $\boldsymbol{K}_{\max}$ максимальную клику в графе. Тогда

$$|\boldsymbol{K}_{ft}| \le |\boldsymbol{K}_{qt}| \le |\boldsymbol{K}_{gr}| \le |\boldsymbol{K}_{\max}| = \omega(G(n,\alpha)). \qquad (16)$$

Тогда верны следующие теоремы:

Теорема 6 *Если* $0 < \alpha < 2$, *то* $\boldsymbol{K}_{gr}$ *и* $\boldsymbol{K}_{qt}$ *имеют размер* $(1+o_p(1))\omega(G(n,\alpha))$. *С другой стороны,*

$$|\boldsymbol{K}_{ft}| / |\boldsymbol{K}_{\max}| \to^p 2^{-\alpha/2}.$$

Теорема 7 *Для любого* $\alpha > 0$ *существует алгоритм, который почти наверное находит в* $G(n,\alpha)$ *клику размером* $(1+o_p(1))\omega(G(n,\alpha))$ *за полиномиальное время.*

10.3. Основные результаты для схожих моделей

10.3.1 Детерминированные веса

Вместо того, чтобы выбирать веса независимо согласно распределению W, можно выбрать их из подходящей детерменированной последовательности W_i (как в модели Чунг-Лу), например

$$W_i = a^{1/\alpha}\frac{n^{1/\alpha}}{i^{1/\alpha}}, i = 1,\ldots,n. \qquad (17)$$

В данной модели верными остаются все результаты для модели Янсона-Лучака.

11. Получение и обработка экспериментальных данных из социальных сетей

Большие социальные сети (например, Facebook и MySpace) не публикают свой веб-граф, поэтому на текущий момент его характеристики приходится получать экспериментально. Наиболее популярный способ для этого – создание популярных приложений в социальной сети (с числом пользователей не менее нескольких миллионов) и последующий сбор данных их пользователей. Сотрудники центра RUBINET [17] написали три приложения для Facebook с общим числом пользователей около 8 миллионов. Анонимизированные данные, полученные с помощью этих приложений, доступны on-line. Проанализировав различные характеристики пользователей (их географическое расположение, связи друг с другом и другие) и как они различаются в зависимости от типа приложения, исследователи пришли к следующим выводам:

- Граф отношений (interaction graph) пользователей обладает свойством «малого мира».
- При этом пользователи неигровых приложений образуют более плотные сообщества, чем пользователи игр. При этом

популярность приложений характеризуется степенным законом с экспоненциальным угасанием.

12. Генераторы синтетических веб-графов

Поскольку мы считаем разумным использование безмасштабных моделей типа Чунг-Лу [3] и Янсона-Лучака [4], то остановимся на генераторах, способных воспроизводить эти или похожие модели. В известных нам свободно распространяемых генераторах [13], [14], [15], [16] паросочетания реализуются не напрямую, а приближенно (например через введение понятия *потенцтального соседа* в [16]). Качество работы генератора определяется в сравнении некоторых параметров (например, размера плотных подграфов) получаемого графа с их теоретическими оценками. Наиболее полный отчет приведен авторами генератора DIGG [13].

13. Заключение

Существующие модели случайных безмасштабных графов можно условно разделить на три класса. В первый класс попадают модели со строгим математическим обоснованием выполнения степенного закона для порождаемых графов, для которых также доказан ряд важных свойств (подсчитан диаметр графа, коэффициент кластеризации, и т.п.), однако, показатель степени фиксирован и не может выбираться заранее. Типичным примером такой модели является модель Боллобаша-Риордана (степень равна 3) и ее обобщения (степень не меньше 2). Ко второму классу можно отнести модели, в которых показатель степенной зависимости может задаваться произвольно, что позволяет изучать эффекты фазовых переходов при его изменении. К числу таких моделей можно отнести модель Чунг-Лу и модель Лучака-Янсона. Доказательство наличия фазового перехода в них по свойству содержать большую клику при переходе степени через значение 2 представляет несомненный интерес и стимулирует исследование других свойств случайных графов с этой точки зрения. Наконец к третьему, самому многочисленному классу, относятся модели, в которых характеристики и свойства графов определяются эмпирически. Про такие модели не доказано никаких содержательных результатов (однако, возможно это дело недалекого будущего). Примерами таких моделей могут являться модель Forest Fire [18] и ряд других.

Кроме очевидной задачи моделирования динамики социальных сетей, модели безмасштабных сетей также применимы к таким популярным сейчас практическим задачам, как *поиск сообщетсв*, *максимизация распространения влияния*, *оценка безопасности и выносливости сложных сетей* и другие. Они позволяют сравнивать качество предлагаемых алгоритмов на целых семействах случайных графов и сравнивать получаемые результаты с предсказанными теоретически.

В целом, можно резюмировать, что данное направление исследований переживает период интенсивного развития и привлекает внимание математиков, инженеров, специалистов по компьютерным наукам, социологов и других специалистов.

Список литературы

[1] Mark Newman, Albert-László Barabási, Duncan J. Watts. The Structure and dynamics of networks. Princeton University Press, 2006.

[2] P. Erdös and A. Rényi. On the evolution of random graphs // Publ. Math. Debrecen. 1959. V. 6. P. 290-297.

[3] W. Aiello, F. Chung, L. Lu. A Random Graph Model for Massive Graphs. *STOC*'2000.

[4] S. Janson, T. Łuczak, I. Norros, (2010), Large cliques in a power-law random graph, J. Appl. Probab. V. 47, N. 4 1124-1135.

[5] L.-A. Barabási, R. Albert. Emergence of scaling in random networks. Science, 1999. V.286. P. 509-512

[6] L.-A. Barabási, R. Albert, H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. Physica, 2000. V. A281. P. 69-77.

[7] L.-A. Barabási, R. Albert, H. Jeong. Diameter of the world-wide web. Nature, 1999. V. 401. P. 130-131.

[8] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal. Stochastic models for the web graph. Proc. 41st Symposium on Foundations of Computer Science. 2000.

[9] B. Bollobás, O. Riordan. Mathematical results on scale-free random graphs. // Handbook of graphs and networks. Weinheim: Wiley-VCH, 2003. P. 1-34.

[10] А.М. Райгородский. Модели случайных графов. МЦНМО, 2011.

[11] P.G. Buckley, D. Osthus. Popularity based random graph models leading to scale-free degree sequence. Discrete Mathematics. Volume 282, Issues 1–3, 6 May 2004, Pages 53–68.

[12] B. Bollobás, C. Borgs, T. Chayes, O.M. Riordan. Directed scale-free graphs. ProceedingSODA '03 Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms

[13] Arthur R. Brady. A compact routing scheme for power-law networks using empirical discoveries in power-law graph topology. http://digg.cs.tufts.edu/readings/pdf/021.pdf

[14] Deepayan Chakrabarti et al. R-MAT: A Recursive Model for Graph Mining. http://www.cs.cmu.edu/ christos/PUBLICATIONS/siam04.pdf

[15] A. Medina et al. BRITE: Universal Topology Generation from a User's Perspective. http://www.cs.bu.edu/brite/publications/usermanual.pdf

[16] Joel C. Miller, Aric Hagberg. Efficient Generation of Networks with Given Expected Degrees. WAW, 2011.

[17] Atif Nazir, Saqib Raza, Chen-Nee Chuah. Unveiling Facebook: a measurement study of social network based applications. IMC'08.

[18] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. ACM Transact. on Knowledge Discovery from Data, 1(1), 2007.