Исследование максимального размера плотного подграфа случайного графа

[1,2] *Н.Н. Кузюрин* <nnkuz@ispras.ru>
[2] *Д.О. Лазарев* <dennis810@mail.ru>
[1] *Институт системного программирования им. В.П. Иванникова РАН, 109004, Россия, г. Москва, ул. А. Солженицына, д. 25*
[2] *Московский физико-технический институт, 141700, Московская область, г. Долгопрудный, Институтский пер., 9*

Аннотация. Для описания случайных сетей используется модель случайного графа Эрдёша-Реньи $\boldsymbol{G(n,p)}$. При исследовании современных случайных сетей часто требуется определить размер максимальной плотной подсети. В настоящей статье приводятся оценки максимального размера c-плотного подграфа, асимптотически почти наверно содержащегося в случайном графе $G(n,\frac{1}{2})$. Было показано, что при $c<\frac{1}{2}, G(n,\frac{1}{2})$ — асимптотически почти наверно c-плотный; получены верхняя и нижняя оценки размера максимального $\frac{1}{2}$-плотного подграфа, асимптотически почти наверно. содержащегося в $G(n,\frac{1}{2})$; а при $c>\frac{1}{2}$ получена оценка сверху на размер максимального c-плотного подграфа асимптотически почти наверно содержащегося в $G(n,\frac{1}{2})$.

Ключевые слова: случайный граф; случайный граф Эрдёша-Реньи; плотность графа; максимальный плотный подграф.

DOI: 10.15514/ISPRAS-2017-29(6)-12

Для цитирования: Кузюрин Н.Н., Лазарев Д.О. Исследование максимального размера плотного подграфа случайного графа. Труды ИСП РАН, том 29, вып. 6, 2017 г., стр. 213-220. DOI: 10.15514/ISPRAS-2017-29(6)-12

1. Введение

Вероятностный метод — эффективный и широко применяемый математический метод, позволяющий доказывать существование объекта с заданными свойствами. Он заключается в оценке вероятности того, что случайный объект из заданного класса удовлетворяет нужному условию. Если доказано, что эта вероятность положительна, то объект с нужными свойствами существует. Несмотря на явную неконструктивность метода, на его основе могут быть созданы вероятностные алгоритмы построения объекта с параметрами, близкими к параметрам, оцененным с помощью вероятностного метода. Вероятностный метод — сравнительно молодой и динамически развивающийся метод, основоположником которого по праву считается выдающийся венгерский математик Пол Эрдёш. Несмотря на то, что метод применялся и до Эрдёша, например, Селёшем для доказательства существования турнира на n вершинах не менее чем с $\frac{(n-1)!}{2^n}$ гамильтоновыми циклами, именно Пол Эрдёш в полной мере осознал его потенциал, именно ему, его ученикам и соавторам принадлежит большее число доказательств, ставших впоследствии классическими.

Различные разделы науки постоянно ставят всё новые и новые задачи, так или иначе связанные со случайными графами. Впервые случайный граф был определен в рассматриваемой в работе модели Эрдёшем и Реньи в их совместной работе [1]. Данная модель используется и по сей день при исследовании свойств объектов, связанных со случайными сетями.

Широко изучается, в связи с его важной ролью в биологических и социальных сетях, также и понятие плотности графа. Например, в работе [2] приведён алгоритм нахождения плотного подграфа в гигантском графе, вроде графа цитирования в Интернете, а в работе [3] предложен алгоритм нахождения *k* самых плотных подграфов в динамическом графе.

Ниже приведены основные определения, которые будут многократно использоваться в работе.

Определение 1. Дан граф *H* с *p* (*p*>1) вершинами и *q* рёбрами. Его плотностью называется величина *D*, равная отношению количества рёбер графа к числу рёбер в полном графе с числом вершин, равным числу вершин в графе *H*:

$$D = \frac{2q}{p(p-1)}$$

Определение 2. Случайным графом в модели Эрдёша-Реньи, или просто случайным графом с *n* вершинами и вероятностью выпадения ребра, равной *p*, называется вероятностное пространство над пространством всех графов на *n* вершинах, в котором вероятность события, для любого ребра полного графа на n вершинах заключающегося в том, что оно принадлежит графу, не зависит от аналогично определённых событий для других рёбер и равна p. Данное пространство обозначается $\boldsymbol{G(n,p)}$.

Определение 3. Скажем, что некоторое свойство выполняется асимптотически почти наверно (а.п.н.) для некоторого семейства случайных графов, если при $n \to \infty$ вероятность того, что граф на n вершинах из данного семейства обладает данным свойством стремится к 1.

Многократно используется в работе следующее неравенство, доказанное русским математиком А. А. Марковым:

Если X — случайная величина, принимающая целые неотрицательные значения, то $\mathrm{P(X > 0) \leq E(X)}$, где $P\langle Y\rangle$ — вероятность события Y, а $E(X)$ — математическое ожидание случайной величины X.

В настоящей работе исследовался размер максимального с-плотного подграфа случайного графа в модели Эрдёша-Реньи на n вершинах $G(n,\frac{1}{2})$. Было доказано, что

- При $c < \frac{1}{2}, G(n,\frac{1}{2})$ — асимптотически почти наверно с-плотный;
- Получены верхняя и нижняя оценки размера максимального $\frac{1}{2}$ $\frac{1}{2}$-плотного подграфа, а.п.н. содержащегося в $G(n,\frac{1}{2})$;
- При $c > \frac{1}{2}$ получена оценка сверху на размер максимального с-плотного подграфа, а.п.н. содержащегося в $G(n,\frac{1}{2})$.

2. Максимальный размер клики в случайном графе в модели Эрдёша-Реньи

В работе [4] Эрдёшем и Боллобашем исследовался размер максимальной клики в случайном графе $G(n,\frac{1}{2})$. Было установлено, что асимптотически почти наверно данная величина имеет плотную концентрацию.

Обозначим за k(n) верхнюю целую часть решения уравнения

$$\binom{n}{k} 2^{-\binom{k}{2}} = 1;\ k(n) = (1 + o(1)) \log_2 n$$

Тогда имеет место следующее утверждение:

Теорема.(Erdos and Bollobas, 1976). Размер максимальной клики в $G(n,\frac{1}{2})$ а.п.н. равен либо $k(n)$, либо $k(n)-1$.

Идея доказательства: оценка сверху получается вычислением математического ожидания числа клик размера k графа $G(n,\frac{1}{2})$; для оценки снизу требуется воспользоваться методом второго момента.

3. Оценки размера максимального c-плотного подграфа случайного графа G(n,c) при различных значениях c

Обозначим за $k(n,c)$, при $c \geq \frac{1}{2}$ наименьшее целое решение неравенства:

$$\binom{n}{k}\Pr\left\langle Bi(\binom{k}{2},\frac{1}{2}) \geq c\binom{k}{2}\right\rangle \geq 1,$$

где $Bi(n,p)$ — биномиальное распределение с параметрами n и p, а $\Pr\langle x\rangle$ — вероятность события X.

Теорема.

1) При $c < \frac{1}{2}$ сам случайный граф а.п.н. c-плотный.

2) При $c = \frac{1}{2}$ выполняется:

а) При $k = \frac{n}{f(n)}$, где $f(n) \to \infty$ произвольно медленно при $n \to \infty$,

граф $G(n,\frac{1}{2})$ а.п.н. содержит $\frac{1}{2}$- плотный подграф размера k.

б) При $k = \frac{n}{4} - \sqrt{\frac{n\ln n}{2}} - \sqrt{\frac{n}{\ln n}}\ln\ln\ln n - \sqrt{\frac{n}{32\ln n}}\ln\ln n$

$\exists\varepsilon = \frac{1}{3}, \exists N_0 : \forall n > N_0$, верно: $\Pr\left\langle G(n,\frac{1}{2})\ \textit{содержит}\ \frac{1}{2} - \textit{плотный подграф размера}\ k\right\rangle > \varepsilon$

3) При $\frac{1}{2} < c < 1$, размер максимального c-плотного подграфа графа $G(n,\frac{1}{2})$ а.п.н. не превосходит $k(n,c)+1$.

Доказательство.

1) Оценив соотношение соседних биномиальных коэффициентов $\binom{n}{k}$ и $\binom{n}{k+1}$, можно установить верность следующей леммы:

Лемма.

$$\Pr\left\langle Bi\left(n,\frac{1}{2}\right) \geq cn\right\rangle \in \left[2^{-n}\binom{n}{cn}\frac{1-c}{c},\ 2^{-n}\binom{n}{cn}\right] \blacksquare$$

Пусть $m = \binom{n}{2}$. Число рёбер в случайном графе $G(n,p)$ имеет биномиальное распределение $Bi(m,p)$. Используя равенство $\binom{n}{k} = \binom{n}{n-k}$ и лемму, можно получить:

$$\Pr\left\langle Bi(m,\frac{1}{2}) \geq cm \right\rangle = 1 - \Pr\left\langle Bi(m,\frac{1}{2}) \geq (1-c)m \right\rangle \to 0,\, m \to \infty$$

Следовательно, $G(n,\frac{1}{2})$ — а.п.н. c-плотный.

2) а) Заметим, что

$$\Pr\left\langle \textit{плотность}\, G(n,\frac{1}{2}) \geq \frac{1}{2} \right\rangle \geq \frac{1}{2}$$

Разобьём вершины графа $G(n,\frac{1}{2})$ на $[f(n)]$ множеств размера либо $\frac{n}{f(n)}]$, либо $\left[\frac{n}{f(n)}\right] + 1$. $\left[\frac{n}{f(n)}\right]+1$. Вероятность того, что каждый из подграфов, порождённых этими множествами вершин, имеет плотность, меньшую $\frac{1}{2}$, не превосходит $\frac{1}{2}$. Стало быть, вероятность того, что ни один из подграфов не имеет плотность $D > \frac{1}{2}$ не превосходит $2^{-f(n)} \to 0,\ n \to \infty$. Следовательно, граф $G(n,\frac{1}{2})$ а.п.н. содержит $\frac{1}{2}$-плотный подграф размера k, что и требовалось доказать.

б) В работе [5] Боллобаш, в частности, показывает, что, если за A_n обозначить событие, заключающееся в том, что степень каждой вершины $G(n,\frac{1}{2})$ — не менее, чем $\frac{n}{2} - \sqrt{\frac{n \ln n}{2}} - \sqrt{\frac{n}{\ln n}} \ln\ln n - \sqrt{\frac{n}{32 \ln n}} \ln\ln n$, то $\Pr\langle A_n \rangle \to 1$ при $n \to \infty$; также известно, что

$$\Pr\left\langle D\left(G\left(n,\frac{1}{2}\right)\right) \leq \frac{1}{2} \right\rangle \geq \frac{1}{2},$$

следовательно $\exists N_0$:

$$\Pr\left\langle\left(\text{ниоднавершина}\,\mathrm{G}\left(\mathrm{n},\tfrac{1}{2}\right) \text{неимеетстепеньменьше}\, \frac{\mathrm{n}}{2} - \sqrt{\frac{\mathrm{n}\ln \mathrm{n}}{2}} - \sqrt{\frac{\mathrm{n}}{\ln \mathrm{n}}} \ln\ln\ln \mathrm{n} - \sqrt{\frac{\mathrm{n}}{32\ln \mathrm{n}}} \ln\right.\right.$$

$$\left.(\mathrm{D}\left(\mathrm{G}\left(\mathrm{n},\tfrac{1}{2}\right)\right) \leq \tfrac{1}{2})\right\rangle$$

для всех $n > N_0$. Рассмотрим подмножество H множества вершин графа $G(n,\frac{1}{2})$ размером не менее $M = \frac{3n}{4} + \sqrt{\frac{n\ln n}{2}} + \sqrt{\frac{n}{\ln n}}\ln\ln\ln n + \sqrt{\frac{n}{32\ln n}}\ln\ln n$ и дополнительное к нему множество H'. Из любой вершины H' в H идёт по крайней мере $\frac{n}{4}$ рёбер. Значит, суммарное число рёбер в подграфах $G(n,\frac{1}{2})$,порождённых множествами вершин H и H' не превосходит

$$\frac{n(n-1)}{4} - \frac{(n-M)n}{4} = \frac{n(M-1)}{4} \le \frac{(n-M-1)(n-M)}{4} + \frac{M(M-1)}{4}$$

Значит, либо плотность подграфа $G(n,\frac{1}{2})$, порождённого множеством вершин H менее $\frac{1}{2}$, либо плотность подграфа $G(n,\frac{1}{2})$, порождённого множеством вершин H'менее $\frac{1}{2}$, что и требовалось доказать.

3) Вычислим $E(n,k,c)$ – математическое ожидание количества порождённых подграфов размера k случайного графа $G(n,\frac{1}{2})$, имеющих плотность, не меньше c:

$$E(n,k,c) = \binom{n}{k}\Pr\left\langle \textit{плотность случайного графа на } k \textit{ вершинах} \ge c \right\rangle$$

$$= \binom{n}{k}\Pr\left\langle Bi\left(\binom{k}{2},\frac{1}{2}\right) \ge c\binom{k}{2}\right\rangle$$

Решая уравнение $E(n,k,c) = 1$, и учитывая, что, по Лемме(обозначим $\binom{k}{2}$за m) $\Pr\left\langle Bi\left(m,\frac{1}{2}\right) \ge cm\right\rangle \in \left[2^{-m}\binom{m}{cm}\frac{1-c}{c},\ 2^{-m}\binom{m}{cm}\right]$, получаем:

$$k(n,c) = (1+o(1))2\log_{2c^c(1-c)^{(1-c)}} n.$$

Для нахождения верхней оценки размера максимального c-плотного подграфа случайного графа, учитывая соотношение, которое может быть получено анализом выражения $E(n,k,c)$:

$$\frac{E(n,k(n,c)+1,c)}{E(n,k(n,c),c)} < \frac{1}{n},$$

выполняющееся для любого $c > \frac{1}{2}$ при любом $n > N_c$, можем получить, что $E(n,k(n,c)+1,c) < \frac{1}{n}$ $\mathrm{E(n, k(n, c) + 1, c)} < \frac{1}{\mathrm{n}}$. Стало быть, по неравенству Маркова,

$$\Pr\left\langle \textit{существует подграф } G\left(n,\frac{1}{2}\right) \textit{ на } k(n,c)+1 \textit{ вершинах}\right\rangle \to 0,$$

при $n \to \infty$. ■

Стоит отметить, что, так как $2c^c(1-c)^{1-c} \to 2$ при $c \to 1$, то при $c = 1$ настоящий результат совпадает с результатом, полученным Боллобашем в работе [4]: $k = (1+o(1))2\log_2 n$ для плотности подграфа, равной 1.

Список литературы

[1]. P.Erdos and A. Renyi. On Random Graphs. Publicationes Mathematicae (Debrecen), Vol. 6, 1959, pp. 290-297.

[2]. D.Gibson, R.Kumar and A.Tomkins. Discovering Large Dense Subgraphs in Massive Graphs. Proceedings of the 31st international conference on Very large data bases, 2005, pp. 721-732.

[3]. Valari E., Kontaki M., Papadopoulos A.N. Discovery of Top-k Dense Subgraphs in Dynamic Graph Collections. In: Ailamaki A., Bowers S. (eds) Scientific and Statistical Database Management. SSDBM 2012, Lecture Notes in Computer Science, vol 7338. Springer, Berlin, Heidelberg, pp 213-230, DOI: 10.1007/978-3-642-31235-9_14.

[4]. B. Bollobas and P. Erdos. Cliques in Random Graphs. Mathematical Proceedings of the Cambridge Philosophical Society, Volume 80, Issue 3, 1976, pp. 419-427, DOI: 10.1017/S0305004100053056.

[5]. B. Bollobas. Degree sequences in Random Graphs. Discrete Mathematics, Volume 33, Issue 1, 1981, pp. 1-19, DOI: 10.1016/0012-365X(81)90253-3.

Analysis of size of the largest dense subgraph of random hypergraph

[1,2] *N.N. Kuzyrin <nnkuz@ispras.ru>*

[2] *D.O. Lazarev <dennis810@mail.ru>*

[1] *Ivannikov Institute for System Programming of the Russian Academy of Sciences, 25, Alexander Solzhenitsyn st., Moscow, 109004, Russia*

[2] *Moscow Institute of Physics and Technology , Dolgoprudnyj, Institutskij alley, Moscow region, 141700, Russia*

Abstract. Random networks are often described using Erdos-Renyi model of random graph $\boldsymbol{G(n,p)}$. The concept of graph density is often used in random network analysis**.** In this article, the maximal size of c-dense subgraph almost surely included in random graph $G(n,\frac{1}{2})$ was evaluated. It was shown, that if $c < \frac{1}{2}$, then $G(n,\frac{1}{2})$ is almost surely c-dense; the upper and lower bounds for the size of maximal $\frac{1}{2}$-dense subgraph almost surely included in $G(n,\frac{1}{2})$ were determined**;** in case when $c > \frac{1}{2}$, the upper bound for the maximal size of c-dense subgraph almost surely included in $G(n,\frac{1}{2})$ was attained**.**

Keywords: random graph; Erdos-Renyi model; graph density; maximal dense subgraph.

DOI: 10.15514/ISPRAS-2017-29(6)-12

For citation Kuzyrin N.N., Lazarev D.O. Analysis of size of the largest dense subgraph of random hypergraph. Trudy ISP RAN/Proc. ISP RAS, vol. 29, issue 6, 2017. pp. 213-220 (in Russian). DOI: 10.15514/ISPRAS-2017-29(6)-12

References

[1]. P.Erdos and A. Renyi. On Random Graphs. Publicationes Mathematicae (Debrecen), Vol. 6, 1959, pp. 290-297.

[2]. D.Gibson, R.Kumar and A.Tomkins. Discovering Large Dense Subgraphs in Massive Graphs. Proceedings of the 31st international conference on Very large data bases, 2005, pp. 721-732**.**

[3]. Valari E., Kontaki M., Papadopoulos A.N. Discovery of Top-k Dense Subgraphs in Dynamic Graph Collections. In: Ailamaki A., Bowers S. (eds) Scientific and Statistical Database Management. SSDBM 2012, Lecture Notes in Computer Science, vol 7338. Springer, Berlin, Heidelberg, pp 213-230, DOI: 10.1007/978-3-642-31235-9_14.

[4]. B. Bollobas and P. Erdos. Cliques in Random Graphs. Mathematical Proceedings of the Cambridge Philosophical Society, Volume 80, Issue 3, 1976, pp. 419-427, DOI: 10.1017/S0305004100053056.

[5]. B. Bollobas. Degree sequences in Random Graphs. Discrete Mathematics, Volume 33, Issue 1, 1981, pp. 1-19, DOI: 10.1016/0012-365X(81)90253-3.