ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2024, том 516, с. 21–25

МАТЕМАТИКА

УДК 519.175.4

ИНДУЦИРОВАННЫЕ ЛЕСА И ДЕРЕВЬЯ В СЛУЧАЙНОМ ГРАФЕ ЭРДЁША–РЕНЬИ

© 2024 г. М. Б. Ахмеджанова[1, *], В. С. Кожевников[2, **]

Представлено академиком РАН В.В. Козловым

Поступило 10.09.2023 г.
После доработки 25.02.2024 г.
Принято к публикации 27.02.2024 г.

Доказана концентрация в интервале размера $2 + o(1 / p)$ размера максимального индуцированного леса (ограниченной и неограниченной степени) в $G(n, p)$ при $C_\varepsilon / n < p < 1 - \varepsilon$ для произвольного заданного $\varepsilon > 0$. Доказана двухточечная концентрация размера максимального индуцированного леса (а также дерева) ограниченной степени в биномиальном случайном графе Эрдёша–Реньи $G(n, p)$ при $p = \text{const}$.

Ключевые слова: случайный граф, граф Эрдёша–Реньи, индуцированный подграф, дерево, лес

DOI: 10.31857/S2686954324020041, **EDN:** XJAOZE

1. ВВЕДЕНИЕ И ФОРМУЛИРОВКА РЕЗУЛЬТАТОВ

Случайные графы играют ключевую роль в комбинаторной математике и теории графов. Наиболее распространённой и хорошо изученной моделью случайного графа является модель Эрдёша–Реньи, обозначаемая как $G(n, p)$. В таком случайном графе n вершин, и каждая пара вершин соединяется ребром с вероятностью p независимо от других пар. При анализе таких графов часто рассматривается понятие индуцированного подграфа. Под индуцированным подграфом графа G, обозначаемым $G[V']$, понимается подграф графа G, вершины которого составляют подмножество V' множества вершин V исходного графа G, и рёбра которого включают все рёбра исходного графа, соединяющие пары вершин из V'. Говоря о максимальном индуцированном подграфе, обычно имеют в виду максимальный по размеру (по числу вершин) индуцированный подграф. Еще одно ключевое понятие в анализе случайных графов — “асимптотически почти наверное” (а.п.н.). Говорят, что некоторое свойство выполняется а.п.н., если оно истинно с вероятностью, стремящейся к 1 при n стремящемся к бесконечности.

Боллобаш и Эрдёш в своей фундаментальной статье [1] доказали, что если p является константой, то существует функция $f(n) \sim 2\log_q(np)$, где $q = \dfrac{1}{1-p}$, такая что а.п.н.

$$f(n) \le \alpha\big(G(n,p)\big) \le f(n) + 1,$$

где $\alpha(G)$ обозначает число независимости графа G, оно определяется как максимальный размер множества вершин V в G, в котором никакие две вершины не соединены ребром.

После этого открытия появился ряд статей, демонстрирующих, что вышеуказанный результат справедлив не только для числа независимости, но и для размера максимального индуцированного:

- подграфа ограниченной степени: Фунтулакис, Канг, МакДиармид [2],
- подграфа с ограниченной средней степенью: Фунтулакис, Канг, МакДиармид [3],
- пути и цикла: Дутта, Субраманиан [4],
- дерева и подграфа с заданной средней степенью: Камалдинов, Скоркин, Жуковский [5],
- леса: Кривошапко, Жуковский [6].

Результаты этого типа называются результатами *2-точечной концентрации*, так как вероятностное

[1] *Научно-технологический университет имени короля Абдаллы, Кауст, Саудовская Аравия*

[2] *Московский физико-технический институт (национальный исследовательский университет), Москва, Россия*

**E-mail: margarita.akhmejanova@kaust.edu.sa*

***E-mail: vladislavkozhevnikov@gmail.com*

распределение исследуемой случайной величины (например, $\alpha(G(n,p))$) асимптотически концентрируется в не более чем двух целых значениях.

Оказывается, что если ослабить ограничение $p = \text{const}$, максимальный размер индуцированных подграфов многих типов всё равно концентрируется в некоторой окрестности величины $2\log_q(np)$. Наиболее значимый результат был получен Фризом в его статье [7]. Он доказал, что существует функция $f(n) \sim 2\log_q(np)$ такая, что если $\frac{C_\varepsilon}{n} < p = o(1)$, где C_ε – достаточно большая константа, зависящая от ε, то а.п.н.

$$\left|\alpha\left(G\left(n,p\right)\right) - f\left(n\right)\right| \leq \frac{\varepsilon}{p}.$$

Насколько нам известно, это единственный результат, показывающий концентрацию в интервале размера $o\left(\frac{1}{p}\right)$. Однако было установлено, что при ослаблении ограничений на размер интервала, а именно, при рассмотрении концентрации в интервале $(2 \pm \varepsilon)\log_q(np)$ для фиксированного $\varepsilon > 0$ и $\frac{C_\varepsilon}{n} < p = o(1)$, где C_ε – достаточно большая константа, зависящая от ε, такая концентрация наблюдается также для размера максимального индуцированного:

1. дерева [8] (хотя доказательство предоставлено для $p = \frac{C_\varepsilon}{n}$, оно также справедливо для больших значений p),
2. паросочетания [9],
3. пути и цикла [10].

Нами доказана концентрация размера наибольшего леса (неограниченной степени) в интервале размера $o\left(\frac{1}{p}\right)$ для случая $p \to 0$. Мы также показываем, что аналогичное утверждение имеет место быть для лесов ограниченной степени.

Теорема 1. *Пусть* $p = p(n) \in (0,1)$, $q = \frac{1}{1-p}$. *Пусть* X *обозначает размер максимального индуцированного леса в* $G(n,p)$. *Тогда для любого фиксированного* $\varepsilon > 0$ *существует константа* $C > 0$, *такая что если* $\frac{C}{n} < p < 1-\varepsilon$, *то а.п.н*

$$\left\lfloor 2\log_q\left(enp\left(1-\varepsilon\right)\right) + 3\right\rfloor \leq X \leq \left\lceil 2\log_q\left(enp\left(1+\varepsilon\right)\right) + 3\right\rceil.$$

Теорема 2. *Пусть* $p = p(n) \in (0,1)$, $q = \frac{1}{1-p}$. *Пусть* Y *обозначает размер максимального индуцированного леса в* $G(n,p)$ *с максимальной степенью, не превосходящей* Δ. *Тогда для любых фиксированных* $\Delta \geq 3$ *и* $\varepsilon > 0$ *существуют такие константы* $a_\Delta \in (1,e)$ *и* $C > 0$, *что если* $\frac{C}{n} < p < 1-\varepsilon$, *то а.п.н*

$$\left\lfloor 2\log_q(a_\Delta np(1-\varepsilon)) + 3\right\rfloor \leq Y \leq \left\lceil 2\log_q(a_\Delta np(1+\varepsilon)) + 3\right\rceil.$$

Нами также доказана двухточечная концентрация размера наибольшего индуцированного леса и дерева ограниченной степени в $G(n, p = \text{const})$.

Теорема 3. *Пусть* $p = \text{const}$, $q = \frac{1}{1-p}$. *Тогда для любого фиксированного* $\Delta \geq 3$ *существует константа* $a_\Delta \in (1,e)$, *такая что а.п.н. размер максимального индуцированного дерева с максимальной степенью, не превосходящей* Δ, *равен либо* $\left\lfloor 2\log_q\left(a_\Delta np\right) + 1\right\rfloor$, *либо* $\left\lfloor 2\log_q\left(a_\Delta np\right) + 2\right\rfloor$.

Теорема 4. *Пусть* $p = \text{const}$, $q = \frac{1}{1-p}$. *Тогда для любого фиксированного* $\Delta \geq 3$ *существует константа* $a_\Delta \in (1,e)$, *такая что а.п.н. размер максимального индуцированного леса с максимальной степенью, не превосходящей* Δ, *равен либо* $\left\lfloor 2\log_q\left(a_\Delta np\right) + 1\right\rfloor$, *либо* $\left\lfloor 2\log_q\left(a_\Delta np\right) + 2\right\rfloor$.

2. СХЕМЫ ДОКАЗАТЕЛЬСТВ

В этом разделе мы изложим основные шаги в доказательствах теорем 1–4.

Доказательство теоремы 1. Доказательство основано на комбинации методов первого и второго момента, а также применении неравенства Талаграна. Пусть X_k – это количество индуцированных корневых (с выделенной вершиной, корнем, в каждой компоненте) лесов в $G(n,p)$ на k вершинах. Тогда

$$\mathrm{E}X_k = \binom{n}{k}\sum_{m=1}^{k}\binom{k-1}{m-1}k^{k-m}p^{k-m}\left(1-p\right)^{\binom{k}{2}-(k-m)} =$$

$$= \binom{n}{k}q^{-\binom{k}{2}}\left(kpq+1\right)^{k-1}.$$

Положим $k_{+\varepsilon} := \lfloor 2\log_q(enp(1+\varepsilon)) + 3 \rfloor$, $k_{-\varepsilon} := \lfloor 2\log_q(enp(1-\varepsilon)) + 3 \rfloor$. Тогда $\mathrm{E}X_{k_{+\varepsilon}} \to 0$ и, пользуясь методом первого момента, мы доказываем верхнюю оценку.

Для доказательства нижней оценки сначала мы применим хорошо известный метод второго момента, суть которого заключается в том, что если дисперсия мала по сравнению с квадратом математического ожидания, то вероятность того, что исследуемая случайная величина больше нуля, велика. А именно, если X принимает только целые неотрицательные значения, то:

$$\mathbb{P}(X > 0) \ge \frac{(\mathrm{E}X)^2}{\mathrm{E}X^2} = \frac{1}{1 + \frac{\mathrm{Var}X}{(\mathrm{E}X)^2}}.$$

Следующим ключевым моментом является оценивание дисперсии $\mathrm{Var}X_k$ для $k = k_{-\varepsilon}$:

$$\mathrm{Var}X_k \le \sum_{\ell=1}^{k} F_\ell,$$

где F_ℓ – ожидание числа (упорядоченных) пар индуцированных корневых лесов на k вершинах, имеющих ℓ общих вершин. Далее мы получаем точное выражение для величины F_ℓ, перечисляя пары пересекающихся индуцированных лесов с заданным числом компонент и заданными размерами компонент в их пересечении. Для этого нам потребовалось вывести некоторое обобщение формулы Кэли: пусть F – ациклический подграф K_n с m компонентами на $f_1, f_2, \ldots, f_m$ вершинах. Тогда величина $f(k, h, \{f_1, \ldots, f_m\})$, равная количеству корневых лесов с h компонентами в K_n, таких что F является индуцированным подграфом, выражается следующим образом:

$$f\left(k, h, \{f_1, \ldots, f_m\}\right) = f_1 \cdot \ldots \cdot f_m \sum_{k_0=0}^{k-\ell} \binom{k-\ell}{k_0} \ell^{k-\ell-k_0} \times$$
$$\times \binom{k_0 + m - 1}{h - 1} (k - \ell)^{k_0 + m - h},$$

где $f_1 + f_2 + \ldots f_m = \ell$.

Заметим, что метод второго момента позволяет доказать теорему только для $p > (\ln n)^2 / \sqrt{n}$. Для остальных p мы дополнительно используем неравенство Талаграна, о методике применения которого можно прочитать в [11].

Доказательство теоремы 2. Техника доказательства остается прежней, но требует весьма точной оценки величины $F_\Delta(n, m)$, равной числу помеченных корневых лесов с n вершинами и m компонентами, максимальная степень которых не превышает Δ. Используя модифицированные коды Прюфера для таких лесов, можно получить выражение для $F_\Delta(n, m)$ через производящие функции:

$$F_\Delta(n, m) = \binom{n-1}{m-1} (n-m)! \left[x^{n-m}\right] \left(f_\Delta(x)\right)^n =$$
$$= \frac{(n-1)!}{(m-1)!} \left[x^n\right] x^m \left(f_\Delta(x)\right)^n,$$

где $f_\Delta(x) = \sum_{k=0}^{\Delta-1} \frac{x^k}{k!}$, а $[x^n]$ – оператор взятия коэффициента при x^n. Данное выражение для $F_\Delta(n, m)$ оказывается очень удобным при вычислении ожидания числа индуцированных корневых лесов ограниченной степени на k вершинах в $G(n, p)$, которое мы обозначим через Z_k. Несложно проверить, что

$$\mathrm{E}Z_k = \binom{n}{k} (1-p)^{\binom{k}{2}} \sum_{m=1}^{k} F_\Delta(k, m) p^{k-m} (1-p)^{-(k-m)} =$$
$$= \binom{n}{k} q^{-\binom{k}{2}} (pq)^{k-1} (k-1)! \left[x^k\right] x \left(f_\Delta(x)\right)^k e^{\frac{x}{pq}}.$$

Далее, для подсчета коэффициента при x^n мы применяем комплексный анализ:

$$\left[x^k\right] x \left(f_\Delta(x)\right)^k e^{\frac{x}{pq}} =$$
$$= \frac{1}{2\pi i} \oint z \left(f_\Delta(z)\right)^k e^{\frac{z}{pq}} \frac{dz}{z^{k+1}} \sim \frac{\left(f_\Delta(r)\right)^k e^{\frac{r}{pq}}}{r^{k-1}\sqrt{2\pi\beta(r) k}},$$

вычисляя асимптотику интеграла методом седловой точки. В выражении выше r – седловая точка, получаемая как решение уравнения $\frac{r f'_\Delta(r)}{f_\Delta(r)} + \frac{r}{kpq} = 1$, а $\beta(r) = \frac{r^2 f''_\Delta(r)}{f_\Delta(r)} + \frac{2r}{kpq} - \frac{r^2}{(kpq)^2}$.

Доказательство теоремы 3. Техника доказательства остается прежней, однако теперь нам потребуется оценка числа $T_\Delta(n)$ помеченных деревьев на n вершинах с максимальной степенью, не превосходящей Δ. Мы покажем, что верно следующее утверждение: для любого целого числа $\Delta \geq 3$ существует константа $a_\Delta \in [1, e)$, такая что

$$T_\Delta(n) = \Theta\left(\left(\frac{a_\Delta}{e}\right)^n n^{n-2}\right).$$

Доказательство последнего утверждение опирается на использование центральной предельной теоремы (ЦПТ). Пусть, как и выше, $f_\Delta(x) = \sum_{k=0}^{\Delta-1} \frac{x^k}{k!}$. Используя коды Прюфера и производящие функции, легко получить выражение $T_\Delta(n) = \left(\frac{d^{n-2}}{dx^{n-2}}\left((f_\Delta(x))^n\right)\right)|_{x=0}$. Введем случайную величину ξ_Δ с производящей функцией

$$\varphi_{\xi_\Delta}(x) := \mathrm{E}x^{\xi_\Delta} = \frac{f_\Delta(\alpha_\Delta x)}{f_\Delta(\alpha_\Delta)},$$

где α_Δ — параметр, определяемый ниже, и случайную величину S_n с производящей функцией

$$\varphi_{S_n}(x) = \left(\varphi_{\xi_\Delta}(x)\right)^n.$$

Тогда для любого целого неотрицательного m:

$$\mathbb{P}(S_n = m) = \frac{\alpha_\Delta^m}{m!} \frac{d^m}{dx^m}\left(\varphi_{S_n}\left(\frac{x}{\alpha_\Delta}\right)\right)\Bigg|_{x=0} =$$

$$= \frac{\alpha_\Delta^m}{m!\left(f_\Delta(\alpha_\Delta)\right)^n} \frac{d^m}{dx^m}\left(\left(f_\Delta(x)\right)^n\right)\Bigg|_{x=0}.$$

Подставляя $m = n - 2$, получаем

$$T_\Delta(n) = \frac{\left(f_\Delta(\alpha_\Delta)\right)^n}{\alpha_\Delta^{n-2}}(n-2)! \cdot \mathbb{P}(S_n = n-2) \sim$$

$$\sim \alpha_\Delta \sqrt{\frac{f_\Delta(\alpha_\Delta)}{f_{\Delta-2}(\alpha_\Delta)}} \left(\frac{f_{\Delta-1}(\alpha_\Delta)}{e}\right)^n n^{n-2},$$

вычисляя асимптотику вероятности $\mathbb{P}(S_n = n-2)$ с помощью локальной ЦПТ, применимой при условии, что $n-2$ достаточно близко к $\mathrm{E}S_n$, чего можно достигнуть подбором α_Δ, а именно, определив α_Δ через уравнение $\mathrm{E}\xi_\Delta = \frac{\alpha_\Delta f_{\Delta-1}(\alpha_\Delta)}{f_\Delta(\alpha_\Delta)} = 1$. Полагая $a_\Delta = f_{\Delta-1}(\alpha_\Delta)$, получаем желаемый результат. □

Доказательство теоремы 4. Нижняя оценка следует из теоремы 3, поскольку каждое дерево является лесом. Верхняя же оценка легко получается подсчётом математического ожидания и применением метода первого момента. □

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Исследования В. Кожевникова выполнены при поддержке Российского научного фонда (проект № 21-71-10092). Работа Ахмеджановой по теореме 3 выполнена за счет гранта Российского научного фонда No 22-21-00202, а работа по теоремам 1, 2 и 4 выполнена при финансировании Королевского университета науки и технологий имени короля Абдаллы (KAUST). Ахмеджанова была лауреатом премии “Молодая математика России” 2020 года и выражает благодарность организаторам и жюри за оказанное доверие.

СПИСОК ЛИТЕРАТУРЫ

1. *Bollobás B., Erdős P.* Cliques in random graphs // Mathematical Proceedings of the Cambridge Philosophical Society. 1976. V. 80. P. 419–427.
2. *Fountoulakis N., Kang R.J., McDiarmid C.* The t-stability number of a random graph // The Electronic Journal of Combinatorics. 2010. V. 17. P. 1–10.
3. *Fountoulakis N., Kang R.J., McDiarmid C.* Largest sparse subgraphs of random graphs // European Journal of Combinatorics. 2014. V. 35. P. 232–244.
4. *Dutta K., Subramanian C.R.* On Induced Paths, Holes and Trees in Random Graphs // 2018 Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics. 2018. P. 168–177.
5. *Kamaldinov D., Skorkin A., Zhukovskii M.* Maximum sparse induced subgraphs of the binomial random graph with given number of edges // Discrete Mathematics. 2021. V. 344. P. 112205.
6. *Krivoshapko M., Zhukovskii M.* Maximum induced forests in random graphs // Discrete Applied Mathematics. 2021. V. 305. P. 211–213.
7. *Frieze A.M.* On the independence number of random graphs // Discrete Mathematics. 1990. V. 81. P. 171–175.
8. *Fernandez de la Vega W.* The largest induced tree in a sparse random graph // Random Structures and Algorithms. 1996. V. 9. P. 93–97.
9. *Cooley O., Draganić N., Kang M., Sudakov B.* Large Induced Matchings in Random Graphs // SIAM Journal on Discrete Mathematics. 2021. V. 35. P. 267–280.
10. *Draganić N., Glock S., Krivelevich M.* The largest hole in sparse random graphs // Random Structures & Algorithms. 2022. V. 61. P. 666–677.
11. *Janson S., Łuczak T., Ruciński A.* Random Graphs. John Wiley & Sons, Inc. 2000.

INDUCED FORESTS AND TREES IN ERDÖS–RÉNYI RANDOM GRAPH

M. B. Akhmejanova[a,*], **V. S. Kozhevnikov**[b,**]

Presented by Academician of the RAS V.V. Kozlov

[a]*King Abdullah University of Science and Technology, KAUST, KSA*

[b]*Moscow Institute of Physics and Technology (National Research University), Moscow, Russia*

We prove concentration in the interval of size $o(1/p)$ for the size of the maximum induced forest (of bounded and unbounded degree) in $G(n,p)$ for $C_\varepsilon / n < p < 1-\varepsilon$ for arbitrary fixed $\varepsilon > 0$. We also show 2-point concentration of the size of the maximum induced forest (and tree) of bounded degree in the binomial random graph $G(n,p)$ for $p = \text{const}$.

Keywords: random graph, Erdös–Rényi model, induced subgraph, tree, forest