МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.Ломоносова
механико-математический факультет

О размерах компонент связности в случайном гиперграфе

Дипломная работа
студента 6-го курса 609-ой группы
кафедры теории вероятностей
Мирмоминова Руслана Мэргыязовича

Научный руководитель
доктор физико-математических наук, профессор
Шабанов Дмитрий Александрович

Москва, 2022 г.

1 Введение

Настоящая работа посвящена теории случайных гиперграфов. Напомним основные определения.

1.1 Основные определения

В дискретной математике *гиперграфом* называется пара множеств $H = (V, E)$, где $V = V(H)$ – некоторое конечное множество, называемое *множеством вершин* гиперграфа, а $E = E(H)$ есть некоторая совокупность подмножеств множества V, называемых *рёбрами* гиперграфа. Гиперграф является k*-однородным*, если каждое его ребро содержит ровно k вершин.

Пусть $v, u \in V$ — это некоторые вершины гиперграфа $H = (V, E)$. *Простым путем реберной длины* s из вершины v в вершину u называется такая чередующаяся последовательность $(A_1, v_1, \dots, v_{s-1}, A_s)$ из s различных ребер $A_1, \dots, A_s$ и $s - 1$ вершины $v_1, \dots, v_{s-1}$, что $v_{i-1}, v_i \in A_i$ для всех $i = 1, \dots, s$, где $v_0 = v$, $v_s = u$. Мы будем говорить, что вершины v и u *связаны* в гиперграфе H, если существует простой путь из v в u. Легко видеть, что данное отношение является отношением эквивалентности (считаем, что v связана сама с собой), и потому в каждом гиперграфе множество вершин разбивается на классы эквивалентности — *компоненты связности*. Если $H = (V, E)$ состоит всего из одной компоненты связности, то он называется *связным*. Если гиперграф является k-однородным и $W \subset V$ — его компонента связности, имеющая m ребер и t вершин, то *сложностью* W (также используют термин *циклический индекс*) называется величина

$$\ell(W) = (k - 1)m - t + 1.$$

Дипломная работа посвящена изучению размеров компонент и их сложности в случайном гиперграфе биномиальной модели $H_k(n, p)$. Напомним, что модель $H_k(n, p)$ можно рассматривать как схему Бернулли на ребрах полного k-однородного гиперграфа на n-вершинах: каждое k-подмножество вершин включается в качестве ребра в $H_k(n, p)$ независимо от других с вероятностью p. В работе рассматриваются последовательности случайных гиперграфов $H_k(n, p)$, $n \in \mathbb{N}$, где $k \geq 2$ фиксировано и не меняется с ростом n, а $p = p(n) \in (0, 1)$ зависит от n. Для удобства будем считать, что вся последовательность $(H_k(n, p), n \in \mathbb{N})$ задана на некотором одном вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$.

1.2 История задачи: случайные графы

В случае графов, $k = 2$, модель $H_2(n,p)$ совпадает с классической моделью Эрдеша–Реньи $G(n,p)$. Данная модель активно изучается с конца 50-х годов прошлого века, и с тех пор ей посвящено большое число работ ведущих мировых специалистов. Отметим также выдающиеся монографии [1]–[3], посвященные в первую очередь именно модели $G(n,p)$.

В одной из первых работ [4] по случайным графам Эрдешем и Реньи был обнаружен феномен фазового перехода наибольшего размера компоненты связности случайного графа $G(n,p)$. Введем следующие обозначения: пусть $C_i(n)$ обозначает i-й по величине размер компоненты случайного графа $G(n,p)$, т.е., например, $C_1(n)$ — это максимальный размер компоненты. Сформулируем в данных обозначениях результат Эрдеша и Реньи.

Теорема 1. (П. Эрдеш, А. Реньи, [4]) *Пусть $p = c/n$, где $c > 0$ — фиксировано и не зависит от n.*

1) Если $c < 1$, то имеет место следующая сходимость по вероятности

$$\frac{C_1(n)}{\ln n} \xrightarrow{\mathbb{P}} \alpha(c) = \frac{1}{c - \ln c - 1}, \quad n \to \infty.$$

2) Если же $c > 1$, то имеет место следующая сходимость по вероятности

$$\frac{C_1(n)}{n} \xrightarrow{\mathbb{P}} \beta(c), \quad n \to \infty,$$

где $\beta(c)$ — это единственное решение уравнения $\beta + e^{-\beta c} = 1$ на интервале $(0,1)$. При этом

$$\frac{C_2(n)}{\ln n} \xrightarrow{\mathbb{P}} \alpha(c) = \frac{1}{c - \ln c - 1}, \quad n \to \infty.$$

Теорема показывает, что максимальный размер компоненты случайного графа резко меняется с логарифмического на линейный при малом изменении вероятности появления ребра p. Из теоремы также вытекает следующее следствие для ситуации, когда величина np растет с ростом n.

Следствие 1. *Если $np \to +\infty$, то*

$$\frac{C_1(n)}{n} \xrightarrow{\mathbb{P}} 1, \quad n \to \infty.$$

Результаты теоремы 1 в дальнейшем были дополнительно уточнены. В частности, Степановым [5] было показано, что при $c > 1$ показана асимптотическая нормальность величины $C_1(n)$.

Теорема 2. (В. Степанов, [5]) *Пусть $p = c/n$, где $c > 1$ — фиксировано и не зависит от n. Тогда имеет место следующая сходимость по распределению*

$$\sqrt{n}\left(\frac{C_1(n)}{n} - \beta\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad n \to \infty,$$

где $\beta = \beta(c)$ — это единственное решение уравнения $\beta + e^{-\beta c} = 1$ на интервале $(0, 1)$, а $\sigma^2 = \frac{\beta(1-\beta)}{(1-c(1-\beta))^2}$.

Оригинальное доказательство Степанова весьма громоздко и опирается на точную асимптотику вероятности связности случайного графа. В 2012 году Боллобаш и Риордан предложили [6] короткое доказательства теоремы с помощью центральной предельной теоремы для мартингалов.

Случай $np \sim 1$ оказался наиболее трудным. Здесь оказалось важным насколько близко подходит np к 1. Фундаментально данную ситуацию исследовали Лучак, Питтель и Виерман [7]. Полученные ими результаты относительно нашей задачи можно сформулировать следующим образом.

Теорема 3. (Т. Лучак, Б. Питтель, Дж. Виерман, [7]) *Пусть $\alpha \in \mathbb{R}$ — фиксировано. Пусть также $p = (1 + \alpha n^{-1/3})/n$ и $L(n)$ обозначает максимальную сложность компоненты случайного графа $G(n, p)$. Тогда*

$$C_1(n) = \Theta_{\mathbb{P}}(n^{2/3}), \quad L(n) = O_{\mathbb{P}}(1).$$

Тем самым, внутри фазового перехода происходит двойной скачок максимального размера компоненты.

Однако вопрос о нахождении точного предельного распределения $C_1(n)$ оставался открытым и был решен Олдосом только в 1997 году. Для формулировки его результата нам понадобится ввести несколько обозначений, связанным с броуновским движением.

Пусть $(W_t, t \geq 0)$ — процесс броуновского движения (винеровский процесс). Тогда положим

$$W_t^\alpha = W_t + t \cdot \alpha - \frac{t^2}{2}, \ t \geq 0.$$

Далее, обозначим

$$A_t^\alpha = W_t^\alpha - \min_{s \leq t} W_s^\alpha, \ t \geq 0.$$

В силу построения процесс A_t^α будет неотрицательным, но нулевым, если в момент времени t значение W_t^α совпадает с текущим минимумом. Иногда происходят "выбросы" и процесс становится положительным на небольшом интервале I. Пусть $\Gamma = \{\gamma_j, j \in \mathbb{N}\}$ — это набор интервалов (упорядоченных по убыванию длин), на которых A_t^α положителен.

Далее, введем точечный процесс $N = N(A^\alpha) = (N_t, t \geq 0)$, который удовлетворяет уравнению

$$\mathbb{P}\left(N_t \text{ имеет точку на } [t, t+dt] | A_u^\alpha, u \leq t\right) = A_t^\alpha \cdot dt.$$

Пусть μ_j — это число точек N_t внутри интервала γ_j, $j \in \mathbb{N}$.

Пусть $\sigma_j(n)$ — это сложность j-й по размеру компоненты. Как было доказано Олдосом [8], последовательность $((C_j(n), \sigma_j(n)), j \in \mathbb{N})$ сходится по распределению к перечисленным выше функционалам от винеровского процесса.

Теорема 4. (Д. Олдос, [8]) *Выполнена сходимость*

$$\left((n^{-2/3} \cdot C_j(n), \sigma_j(n)), j \in \mathbb{N}\right) \xrightarrow{d} ((|\gamma_j|, \mu_j), j \in \mathbb{N}) \quad \textit{при } n \to +\infty,$$

где сходимость по распределению понимается, как слабая сходимость в метрическом пространстве

$$\ell_2 = \left\{(x, y) \in \mathbb{R}^{+\infty} \times \mathbb{R}^{+\infty} : \sum_{k=1}^{+\infty} x_k^2 + \sum_{k=1}^{+\infty} y_k^2 < +\infty \right\}.$$

Целью настоящей работы было обобщение ряда вышеперечисленных результатов в модели случайных гиперграфов.

1.3 История задачи: случайные гиперграфы

Случайный гиперграф $H_k(n, p)$ в общей ситуации изучен менее детально. Эволюция случайного гиперграфа была впервые исследована в работе [9] Шмидт–Прузан и Шамира. Пусть $L_i(n)$ обозначает i-й по величине размер компоненты случайного гиперграфа $H_k(n, p)$.

Теорема 5. (Дж. Шмидт–Прузан, Э. Шамир, [9]) *Пусть*

$$p = \frac{\lambda}{(k-1)C_{n-1}^{k-1}} = \lambda \cdot (k-2)! n^{1-k}(1 + O(1/n)),$$

где $\lambda > 0$ — фиксировано и не зависит от n.

1) Если $\lambda < 1$, то с вероятностью, стремящейся к 1, $L_1(n) = \Theta(\ln n)$.
2) Если $\lambda > 1$, то с вероятностью, стремящейся к 1, $L_1(n) = \Theta(n)$.
3) Если же $\lambda = 1$, то $L_1(n) = \Theta_{\mathbb{P}}(n^{2/3})$.

В ситуации $\lambda > 1$ результат из [9] был уточнен в работе Бериша, Койя-Оглана и Канг [10], в ней авторы доказали асимптотическую нормальность величины $L_1(n)$.

Теорема 6. (М. Бериш, А. Койя-Оглан, М. Канг, [10]) *Пусть $p = \lambda/[(k-1)C_{n-1}^{k-1}]$, где $\lambda > 1$ — фиксировано и не зависит от n. Тогда имеет место следующая сходимость по распределению*

$$\sqrt{n}\left(\frac{L_1(n)}{n} - (1-\rho)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad n \to \infty,$$

где $\rho = \rho(c)$ — это единственное решение уравнения $\rho = e^{\frac{\lambda}{k-1}(\rho^{k-1}-1)}$ на интервале $(0, 1)$, а

$$\sigma^2 = \frac{\rho(1-\rho+\lambda(\rho-\rho^{k-1}))}{(1-\lambda\rho^{k-1})^2}.$$

Авторами в [10] получена и локальная предельная теорема для $L_1(n)$. Позднее Боллобаш и Риордан [11] упростили доказательство с помощью мартингального подхода. Также в работе [11] авторы постулировали результат, аналогичный теореме 4 в ситуации, когда $\lambda \sim 1$, однако ими были опущены разные детали доказательств, а также не обсуждался вопрос о сложности компонент. Целью настоящей работы является закрытие данного пробела.

1.4 Мотивация исследования

Итак, пусть $H_k(n,p)$ — случайный k-однородный гиперграф. Будем рассматривать ситуацию, когда

$$p = p(n) = \lambda(k-2)!n^{1-k},$$

где $\lambda = \lambda(n) \sim 1$, $n \to +\infty$. Нас интересует асимптотическое распределение размеров компонент связности такого гиперграфа при $n \to +\infty$.

Вообще говоря, понятие компоненты связности в случае гиперграфа можно интерпретировать по-разному. Это зависит от понятия связи между ребрами. Например, можно считать, что два ребра связаны друг с другом, если множества их вершин пересекаются хотя бы по r вершинам для некоторого $1 \leq r \leq k-1$. Много работ посвящено рассмотрению классическому случаю $r = 1$, кроме того, в некоторых других работах исследуется случай $r = k-1$ (например, [12]). В текущей работе рассматривается только случай $r = 1$.

Результаты существенно зависят от асимптотического поведения $(\lambda - 1)n^{\frac{1}{3}}$ при $n \to +\infty$. В текущей работе описывается подход Боллобаша и Риордана из работы [11], в рамках которого главный результат (асимптотическое распределение размеров компонент связности) получается на стыке теории случайных процессов и теории графов. Новизна работы состоит в том, что помимо результата Боллобаша и Риордана, получено асимптотическое распределение сложности компонент, что является обобщением результата, полученного в работе [8] для графов.

Работа примечательна тем, что подобная техника рассуждений с гиперграфами достаточно проста и опирается на утверждения из теории случайных процессов и теории графов, остающиеся верными для многих моделей. В частности, одним из ключевых элементов полученной теории является алгоритм обхода в ширину, который подробно описан в следующей секции.

Помимо необычного подхода к работе с гиперграфами посредством теории случайных процессов, результат замечателен тем, что является практически значимым. Например, в гиперграфе, который соответствует крупной транспортной сети, можно оценивать «кластеризуемость» всей сети по доли дорог, тем самым принимая оптимальное решение о строительстве / деконструкции дорог.

2 Обход в ширину

Идея алгоритма: просматриваем очередную вершину графа, затем проходим по всем инцидентным рёбрам, которые не содержат предыдущих вершин. Если оказалось, что таких рёбер нет, то текущая компонента связности просмотрена, и следующей в очереди будет первая вершина новой компоненты. Формализуем данный алгоритм.

Будем говорить, что вершина может находиться в одном из трёх состояний: *рассмотренная*, *активная*, *неактивная*. Пусть Q_t – множество активных вершин, а U_t – неактивных в момент времени $t \in \overline{0,n}$. На старте алгоритма $Q_0 = \emptyset$, а $U_0 = [n]$. Рассмотрим очередной шаг $t \in \overline{1,n}$, который соответствует переходу от $t-1$ к t:

1. Если $Q_{t-1} = \emptyset$, достаём вершину из U_{t-1} и объявляем её рассмотренной. Все вершины инцидентных ей рёбер кладём в Q_{t-1}.

2. В противном случае достаём первую добавленную вершину из Q_{t-1}, и тоже объявляем рассмотренной. Возьмём рёбра, которые состоят из данной

вершины и $k-1$ вершины из U_{t-1}. Все вершины таких рёбер из U_{t-1} перемещаем в Q_{t-1}.

Отметим некоторые свойства алгоритма. По построению в Q_t лежит вершин не больше, чем размер максимальной компоненты гиперграфа. В нашем случае, соответственно, заведомо

$$|Q_t| = O(n^{\frac{2}{3}}), \quad n \to +\infty,$$

чем будем активно пользоваться в дальнейшем. Кроме того, легко видеть, что $|Q_t| + |U_t| = n - t$, потому что Q_t и U_t содержат нерассмотренные вершины, а их число уменьшается на 1 с каждым шагом. Такой алгоритм просматривает компоненты связности гиперграфа одну за другой, и $Q_t = \emptyset$ означает, что в момент времени t завершён обход очередной компоненты. Таким образом, если

$$\{t_i\}_{i=0}^{r} = \{t \in \overline{0,n} : Q_t = \emptyset\},$$

то $\{t_i - t_{i-1}\}_{i=1}^{r}$ – последовательность длин компонент. Рассмотрим случайный процесс

$$X_t = |Q_t| - C_t, \quad t \in \overline{0,n},$$

где C_t – номер компоненты рассмотренной вершины на шаге t ($C_0 = 0$). Заметим, что $t_i = \inf\{t \in \overline{0,n} : X_t = -i\}$ – это вытекает из определения t_i.

Пусть на шаге t в Q_{t-1} добавляется η_t вершин. Если перед этим шагом $Q_{t-1} = \emptyset$, то $X_t - X_{t-1} = \eta_t - 1$, потому что в таком случае $|Q_t| - |Q_{t-1}| = \eta_t$, а $C_t = C_{t-1} + 1$. Если $Q_{t-1} \neq \emptyset$, то, аналогично, $X_t - X_{t-1} = \eta_t - 1$, поскольку $|Q_t| - |Q_{t-1}| = \eta_t - 1$, ведь согласно алгоритму из Q_{t-1} в этот момент достали вершину. В свою очередь, $C_t = C_{t-1}$. Итак,

$$X_t = \sum_{i=1}^{t} (\eta_i - 1).$$

3 Свойства случайного процесса X_t

Пусть $\mathcal{F}_t \subset \mathcal{F}$ – фильтрация, порождённая случайными процессами Q_t и C_t. Попробуем выделить мартингальную составляющую процесса X_t. Для этого достаточно оценить распределение η_{t+1} при условии $\mathcal{F}_t$.

Пусть U'_t – множество вершин-кандидатов, которые могут оказаться в очереди после момента t. Тогда

$$|U'_t| = |U_t| - I(\{Q_t = \emptyset\}),$$

поскольку если очередь пустая, то одна из вершин, лежащих в U_t, сразу объявляется просмотренной. При этом, $|U_t'|$ – случайная величина, измеримая относительно $\mathcal{F}_t$.

Для каждой вершины-кандидата существует $C_{n-t-2-|Q_t|}^{k-2}$ рёбер, содержащих её вместе с рассматриваемой на шаге t вершиной и не содержащих предыдущих t просмотренных вершин. Обозначим это количество c_{t+1}. Значит, для вершины из U_t' вероятность попасть в очередь на шаге t равна

$$\pi_1 = 1-(1-p)^{c_{t+1}} = pc_{t+1} + O(p^2c_{t+1}^2) = \lambda(n-t-|Q_t|)^{k-2}n^{-k+1} + O(n^{-2})$$
$$= \lambda(n-t)^{k-2}n^{-k+1} + o(n^{-1}), \ \ n \to +\infty.$$

Таким образом,

$$\mathbb{E}(\eta_{t+1}|\mathcal{F}_t) = |U_t'|\pi_1 = |U_t'|\lambda(n-t)^{k-2}n^{-k+1} + o(1), \ \ n \to +\infty.$$

Оценим похожим образом $\mathrm{Var}(\eta_{t+1}|\mathcal{F}_t)$. Пусть $v_1, v_2 \in U_t'$. Обе вершины могут попасть в очередь в двух случаях:

1. Они оказались частью одного ребра, не содержащего предыдущих t просмотренных вершин и содержащего текущую. Всего существует $C_{n-t-3-|Q_t|}^{k-3}$ таких рёбер. Значит, соответствующая вероятность равна

$$\pi_2 = 1-(1-p)^{C_{n-t-3-|Q_t|}^{k-3}} = \lambda(k-2)(n-t-|Q_t|)^{k-3}n^{-k+1} + O(n^{-4})$$
$$= \lambda(k-2)(n-t)^{k-3}n^{-k+1} + o(n^{-2}), \ \ n \to +\infty.$$

2. Они не являются частью одного ребра, но присутствуют в двух отдельных рёбрах. Количество рёбер первого типа – $\widehat{c}_{t+1} := C_{n-t-3-|Q_t|}^{k-2}$. В свою очередь, рёбра второго типа не пересекаются с рёбрами первого типа, и их количество тоже равно $\widehat{c}_{t+1}$. Значит, соответствующая вероятность равна

$$\pi_3 = 1-2(1-p)^{\widehat{c}_{t+1}} + (1-p)^{2\widehat{c}_{t+1}}$$
$$= p^2c_{t+1}^2 + O(n^{-2}) = \pi_1^2 + O(n^{-2}), \ \ n \to +\infty.$$

Таким образом,

$$\mathrm{Var}(\eta_{t+1}|\mathcal{F}_t) = |U_t'|\pi_1 + |U_t'|(|U_t'|-1)(\pi_2+\pi_3) - |U_t'|^2\pi_1^2 \sim |U_t'|^2\pi_2 + |U_t'|\pi_1$$
$$\sim \lambda(k-2)(1-t/n)^{k-3}\frac{|U_t|^2}{n^2} + \lambda(1-t/n)^{k-2}\frac{|U_t|}{n}, \ \ n \to +\infty.$$

Полученная оценка влечёт $\max_t \sup_\Omega \mathrm{Var}(\eta_{t+1}|\mathcal{F}_t) \le C$ для некоторого C. Положим $D_t = \mathbb{E}(\eta_t - 1|\mathcal{F}_{t-1})$. Тогда

$$|U_t| = n - t - |Q_t| = n - t - X_t - C_t \Rightarrow |U_t'| = n - t - X_t - C_{t+1}$$

$$\Rightarrow D_{t+1} = \alpha_{t+1}(n - t - X_t - C_{t+1}) - 1 + O(n^{-1}), \ \ n \to +\infty,$$

где $\alpha_t = pc_t$. Пусть $\Delta_t = X_t - X_{t-1} - D_t$. Тогда Δ_t измерима относительно $\mathcal{F}_t$, и $\mathbb{E}(\Delta_t|\mathcal{F}_{t-1}) = 0$ по построению. Итак,

$$X_{t+1} = X_t + \Delta_{t+1} + D_{t+1} = (1-\alpha_{t+1})X_t + \alpha_{t+1}(n-t) - 1 + \Delta_{t+1} - \alpha_{t+1}C_{t+1} + E_{t+1},$$

где $E_{t+1} = O(n^{-1})$ – слагаемое, характеризующее погрешность. Аппроксимируем X_t суммой детерминированной последовательности $(x_t)_t$ и мартингала. Последовательность x_t зададим следующим рекуррентным соотношением:

$$x_{t+1} = (1 - \alpha_{t+1})x_t + \alpha_{t+1}(n - t) - 1.$$

Отнимая x_t от X_t, получим

$$X_{t+1} - x_{t+1} = (1 - \alpha_{t+1})(X_t - x_t) + \Delta_{t+1} - \alpha_{t+1}C_{t+1} + E_{t+1}$$

$$\Rightarrow X_t - x_t = \sum_{i=1}^{t} \frac{\beta_t}{\beta_i}(\Delta_i - \alpha_i C_i + E_i),$$

где $\beta_t = \prod_{i=1}^t (1 - \alpha_i)$. Стоит отметить, что $(\beta_t)_t$ – убывающая последовательность, поскольку $0 < \alpha_i < 1$ для всех $i \in \overline{1,n}$. Согласно полученной формуле, определим процесс $(S_t)_t$:

$$S_t = \sum_{i=1}^{t} \beta_i^{-1}\Delta_i.$$

Такой процесс является мартингалом относительно фильтрации $(\mathcal{F}_t)_t$. В качестве искомой аппроксимации процесса $(X_t)_t$ возьмём

$$\widehat{X}_t = x_t + \beta_t S_t.$$

Лемма 1. *Пусть* $p = p(n) = O(n^{-k+1}), \ \ n \to +\infty$. *Тогда*

$$|X_t - \widehat{X}_t| = O(tC_t/n), \ \ n \to +\infty,$$

равномерно по t.

Доказательство. Пользуясь определением $\widehat{X}_t$, равенством $\alpha_t = pc_t$ и полученной ранее оценкой $E_t = O(n^{-1})$, $n \to +\infty$, получаем

$$|X_t - \widehat{X}_t| \le \sum_{i=1}^{t} \frac{\beta_t}{\beta_i}|E_i - \alpha_i C_i| \le \sum_{i=1}^{t} |E_i| + \sum_{i=1}^{t} \alpha_i C_i$$

$$\le t \max_{i \in \overline{1,t}} |E_i| + tC_t \max_{i \in \overline{1,t}} \alpha_i = O(tC_t/n), \quad n \to +\infty.$$

□

Представим x_t в виде явной формулы. Для этого введём последовательность $y_t = x_t - n + t$. Тогда рекуррентная формула для x_t принимает вид $y_{t+1} = (1 - \alpha_{t+1})y_t$. Поскольку $y_0 = n$, получаем $y_t = n\beta_t$, и

$$x_t = n - t - n\beta_t.$$

Найдём порядок β_t. Согласно определению,

$$\beta_t = \prod_{i=1}^{t}(1 - \alpha_i) \Rightarrow \log \beta_t = -\sum_{i=1}^{t} \alpha_i + O(n^{-1}), \quad n \to +\infty.$$

В свою очередь, $\alpha_t = pc_t$, поэтому

$$\sum_{i=1}^{t} \alpha_i = p \sum_{i=1}^{t} C^{k-2}_{n-i-1-|Q_{i-1}|} = p \sum_{i=1}^{t} (C^{k-2}_{n-i-1-|Q_{i-1}|} - C^{k-2}_{n-i-1} + C^{k-2}_{n-i-1})$$

$$= p(C^{k-1}_{n-1} - C^{k-1}_{n-t-1}) + o(n^{-1})$$

$$= \lambda \left(\frac{(n-1)!}{(k-1)(n-k)!n^{k-1}} - \frac{(n-t-1)!}{(k-1)(n-t-k)!n^{k-1}} \right) + o(n^{-1})$$

$$= \frac{\lambda}{k-1} \left(\prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) - \prod_{i=1}^{k-1} \left(1 - \frac{t+i}{n}\right) \right) + o(n^{-1})$$

$$= \frac{\lambda}{k-1} \left(1 - \frac{k(k-1)}{2n} + \frac{(t+k)(t+k-1) - t(t+1)}{2n} \right) + o(n^{-1})$$

$$= \frac{\lambda}{k-1}(1 - (1 - t/n)^{k-1}) + o(n^{-1}), \quad n \to +\infty.$$

Определим функцию $g(\tau)$ следующим образом:

$$g(\tau) = g_{k,\lambda}(\tau) = 1 - \tau - \exp\left(-\frac{\lambda}{k-1}(1-(1-\tau)^{k-1})\right).$$

Из полученной оценки следует, что

$$x_t = n - t - n\exp\left(-\frac{\lambda}{k-1}(1-(1-t/n)^{k-1}) + O(n^{-1})\right) = f(t) + O(1), \quad n \to +\infty,$$

где $f(t)$ – функция, заданная следующей формулой:

$$f(t) = f_{n,k,\lambda}(t) = ng_{k,\lambda}(t/n).$$

Стоит отметить, что оценка $x_t = f(t) + O(1)$ является равномерной по t. Проанализируем свойства функции $g(\tau)$:

$$g'(\tau) = -1 + \lambda(1-\tau)^{k-2}\exp\left(-\frac{\lambda}{k-1}(1-(1-\tau)^{k-1})\right) \Rightarrow g'(0) = \lambda - 1.$$

В свою очередь,

$$g''(\tau) = (-\lambda(k-2)(1-\tau)^{k-3} - (\lambda(1-\tau)^{k-2})^2)\exp\left(-\frac{\lambda}{k-1}(1-(1-\tau)^{k-1})\right),$$

откуда следует, что $g'' \leq 0$. Таким образом, функция g является вогнутой, как и f. Из полученного равенства также вытекает, что $\sup_{\tau\in[0,1]}|g''(\tau)| = O(1)$ при $n \to +\infty$, соответственно, $f''(\tau) = O(1/n)$ равномерно по $0 \leq t \leq n$. Используя равномерную ограниченность g''' и формулу Тейлора, получаем

$$g(\tau) = g(0) + \tau g'(0) + \tau^2 g''(0)/2 + O(\tau^3)$$

$$= (\lambda - 1)\tau - \lambda(k-2+\lambda)\frac{\tau^2}{2} + O(\tau^3)$$

при $n \to +\infty$.

4 Сложность компоненты связности

Вернёмся к конструкции алгоритма обхода в ширину. На очередном шаге вершина достаётся из очереди (если очередь непуста), и рассматриваются все рёбра, не содержащие предыдущих просмотренных вершин. Вершины таких рёбер добавляются в очередь, что можно соотнести с добавлением рёбер, содержащих эти вершины. Соответствующий гиперграф назовём *гиперграфом обхода в ширину.*

Таким образом, на каждом шаге в текущий гиперграф обхода в ширину добавляются, фактически, только те рёбра, которые «необходимы» для дальнейшего обхода компоненты. В случае графа гиперграфом обхода в ширину является остовный лес, в котором каждой компоненте связности соответствует остовное дерево. Попробуем разобраться, что происходит в более общем случае k-однородного гиперграфа.

Утверждение 1. *Пусть*

$$p = p(n) = \lambda(n)(k-2)!n^{-k+1}, \quad \lambda \sim 1, \quad n \to +\infty.$$

Тогда математическое ожидание количества случаев, в которых алгоритм добавит в гиперграф обхода в ширину хотя бы два ребра, пересекающихся по более, чем одной вершине, асимптотически не превосходит $o(1)$ при $n \to +\infty$.

Доказательство. На шаге t вероятность взять одну и ту же вершину из U_t два раза равна

$$|U_t|C_{|U_t|-1}^{k-2}p(n)^2 = O(n^{-1}), \quad n \to +\infty.$$

Остаётся заметить, что размер компоненты связности в гиперграфе не превосходит $O(n^{\frac{2}{3}+\epsilon})$ для некоторого $\epsilon > 0$, а недревесных компонент связности не более $\log n$. □

Итак, в случае разреженного гиперграфа, рассматримаевого в работе, пары рёбер пересекаются не более, чем по одной вершине, за исключением асимптотически малого числа пар, пересекающихся по двум вершинам, и стремящегося к нулю – по трём.

Напомним, что в связном k-однородном гиперграфе на n вершинах и с m рёбрами сложностью называется величина

$$(k-1)m - n + 1.$$

Сложность, грубо говоря, соответствует количеству избыточных рёбер гиперграфа, умноженному на $k-1$. Попробуем оценить число таких рёбер в $H_k(n,p)$, основываясь на алгоритме обхода в ширину.

Заметим, что рёбра, которые не рассматриваются алгоритмом на очередном шаге, состоят из вершин, которые лежат в очереди, и содержат просматриваемую вершину. Эти рёбра и образуют сложность очередной компоненты $H_k(n,p)$.

Нетрудно видеть, что на шаге $t+1$ количество вершин, среди которых нужно выбрать $k-1$ для того, чтобы образовать ребро с просматриваемой, равно

$$\nu_t := \begin{cases} |Q_t| - 1, & |Q_t| \neq 0, \\ 0, & |Q_t| = 0, \end{cases}$$

что можно записать в виде формулы

$$\nu_t = |Q_t| - 1 + C_{t+1} - C_t = X_t - \min_{s \le t} X_s,$$

поскольку $C_t = 1 - \min_{s<t} X_s$. Таким образом, количество кандидатов на то, чтобы стать избыточным ребром, равно

$$C_{\nu_t}^{k-1} = C_{X_t - \min_{s \le t} X_s}^{k-1}.$$

Ассоциируем с процессом X_t точечный процесс N_t^n, значение которого увеличивается на 1 при появлении избыточного ребра. Тогда

$$\mathbb{E}(N_{t+1}^n - N_t^n | \mathcal{F}_t) = p(n) C_{X_t - \min_{s \le t} X_s}^{k-1},$$

что является условной интенсивностью данного процесса. Полученная формула лежит в основе результата, который будет описан в соответствующей секции.

5 Результаты

Для каждого $s \in [0, +\infty)$ положим

$$X_s^* := \frac{X_{s(k-1)^{-\frac{1}{3}} n^{\frac{2}{3}}}}{((k-1)n)^{\frac{1}{3}}},$$

линейно интерполируя значения процесса X_t в нецелочисленных точках.

Теорема 7. *Пусть $k \ge 3$, и $p = p(n) = \lambda (k-2)! n^{-k+1}$, где $\lambda = \lambda(n)$ удовлетворяет*

$$(\lambda - 1)^3 n \to (k-1)^2 \alpha, \ \ n \to +\infty,$$

для некоторого $\alpha \in \mathbb{R}$. Тогда

$$X_s^* \to^d W_s^\alpha, \ \ n \to +\infty.$$

Доказательство. Рассмотрим случайный процесс X_t при $t \leq An^{\frac{2}{3}}$, где $A > 0$ – достаточно большая положительная константа. Данный процесс является семимартингалом относительно фильтрации $\mathcal{F}_t$, и допускает разложение

$$X_t = A_t + M_t,$$

где M_t – мартингал, A_t – процесс ограниченной вариации. В нашем случае

$$M_t = \beta_t S_t, \;\; A_t = X_t - M_t.$$

Мы покажем, что масштабированные версии M_t и A_t сходится по распределению к винеровскому процессу и функции $\alpha t - \frac{t^2}{2}$ при $n \to +\infty$, а отсюда будет следовать, что $X_t^* \to^d W_t^\alpha$.

Итак, используя разложение $x_t = ng(t/n) + O(1)$, при $t = s(k-1)^{-\frac{1}{3}} n^{\frac{2}{3}}$ получаем

$$\frac{x_t}{((k-1)n)^{\frac{1}{3}}} = (k-1)^{-\frac{1}{3}} n^{\frac{2}{3}} g(s(k-1)^{-\frac{1}{3}} n^{-\frac{1}{3}}) + o(1)$$

$$= n^{\frac{1}{3}}(\lambda(n)-1)(k-1)^{-\frac{2}{3}} s - \frac{s^2}{2} + o(1) = \alpha s - \frac{s^2}{2} + o(1), \;\; n \to +\infty.$$

В свою очередь, при $t \leq An^{\frac{2}{3}}$ (вообще говоря, и при $t = o(n), \; n \to +\infty$), условная дисперсия мартингальных приращений процесса $S_t = \sum_{i=1}^t \beta_i^{-1} \Delta_i$ имеет следующий асимптотический порядок:

$$\beta_t^{-2} \mathrm{Var}(\Delta_t | \mathcal{F}_{t-1}) \sim \mathrm{Var}(\eta_t | \mathcal{F}_{t-1}) = \mathrm{Var}(\Delta_t | \mathcal{F}_{t-1})$$

$$\sim \lambda(k-2)(1-(t-1)/n)^{k-3} \frac{|U_{t-1}|^2}{n^2} + \lambda(1-(t-1)/n)^{k-2} \frac{|U_{t-1}|}{n}$$

$$\sim \lambda(k-2) + \lambda \sim k-1, \;\; n \to +\infty,$$

поскольку $|U_t| \sim n$ и $\beta_t \sim 1$ при малых t. При этом, $\lambda(n) \sim 1$ выполнено по условию. Пользуясь оценкой $|X_t - \widehat{X}_t| = O(tC_t/n), \; n \to +\infty$ при малых t, получаем

$$X_s^* \to^d W_s^\alpha, \;\; n \to +\infty,$$

что и требовалось. □

Таким образом, с точностью до масштабирования по времени, процесс X_t, соответствующий алгоритму обхода в ширину, при больших n аппроксимируется броуновским движением со смещением. Более того, оказывается, что размеры компонент соответствующего случайного гиперграфа сходятся к экскурсиям процесса W_t^α. Получить такой результат поможет следующая ([8])

Лемма 2. *Пусть* $f : [0, +\infty) \to \mathbb{R}$ *– непрерывная функция, а* $\mathcal{E} := \{(l, r)\}$ *– множество непустых интервалов, таких, что*

$$f(r) = f(l) = \min_{s \le l} f(s), \quad f(s) > f(l), \quad l < s < r.$$

Также положим, что для двух интервалов (l_1, r_1) *и* (l_2, r_2) *с* $l_1 < l_2$ *выполнено* $f(l_1) > f(l_2)$. *Обозначим* $\Xi := \{(l, r - l) : \ (l, r) \in \mathcal{E}\}$ *– точечный процесс на* $[0, +\infty) \times (0, +\infty)$. *Пусть* $f_n \to f$ *локально равномерно, и для каждого* $n \in \mathbb{N}$ *существует последовательность* $\{t_{n,i}\}_{i \in \mathbb{N}}$, *удовлетворяющая следующим свойствам:*

1. $t_{n,1} = 0$, *последовательность* $\{t_{n,i}\}_{i \in \mathbb{N}}$ *строго возрастает и стремится к бесконечности.*

2. $f_n(t_{n,i}) = \min_{u \in [0, t_{n,i}]} f_n(u)$.

3. $\max_{i:\, t_{n,i} \le s_0}(f_n(t_{n,i}) - f_n(t_{n,i+1})) \to 0$ *при* $n \to +\infty$ *для всех* $s_0 \in \mathbb{R}$.

Положим $\Xi^n = \{(t_{n,i}, t_{n,i+1} - t_{n,i})\}_{i \in \mathbb{N}}$. *Тогда*

$$\Xi^n \to \Xi,$$

где имеется в виду слабая сходимость точечных процессов на $[0, +\infty) \times (0, +\infty)$.

Применив такую лемму к процессам X_t^* и W^α, можно убедиться, что последовательность размеров компонент связности гиперграфа сходится к последовательности экскурсий. Таким образом, имеет место

Теорема 8. *Пусть* $k \ge 3$, *и* $p = p(n) = \lambda(k-2)! n^{-k+1}$, *где* $\lambda = \lambda(n)$ *удовлетворяет*

$$(\lambda - 1)^3 n \to (k-1)^2 \alpha, \quad n \to +\infty,$$

для некоторого $\alpha \in \mathbb{R}$. *Пусть* $C_i(n)$ *– размеры компонент связности* $H_k(n, p)$, *отсортированные по невозрастанию. Тогда для любого* $r \in \mathbb{N}$

$$(k-1)^{\frac{1}{3}} n^{-\frac{2}{3}} \cdot (C_i(n))_{i=1}^r \xrightarrow{d} (|\gamma_i|)_{i=1}^r, \quad n \to +\infty.$$

Итак, получено асимптотическое распределение размера компонент связности $H_k(n, p)$ при $n \to +\infty$. Похожий результат имеет место и для распределения сложностей компонент. Напомним, что в конце соответствующей секции был введён точечный процесс N_t^n, и была найдена его условная интенсивность. Если $[t_i, t_{i+1}]$ – интервал рассмотрения отдельной компоненты связности в алгоритме обхода в ширину, то её сложность (с точностью до константного числа рёбер) равна $(k-1)(N^n_{t_{i+1}} - N^n_{t_i})$. В свою очередь, интенсивность процесса $(k-1)N^n_t$ при масштабировании становится равной

$$\lambda_s = (k-1)p(n)C^{k-1}_{X^*_s - \min_{u\le s} X^*_s} \sim \lambda(k-1)!n^{-k+1}(X^*_s - \min_{u\le s} X^*_s)^{k-1} \to (A^\alpha_s)^{k-1}$$

при $n \to +\infty$. Таким образом, верна следующая

Теорема 9. *Пусть $k \ge 3$, и $p = p(n) = \lambda(k-2)!n^{-k+1}$, где $\lambda = \lambda(n)$ удовлетворяет*

$$(\lambda - 1)^3 n \to (k-1)^2\alpha, \quad n \to +\infty,$$

для некоторого $\alpha \in \mathbb{R}$. Пусть $\sigma_i(n)$ – сложности компонент связности $H_k(n,p)$, отсортированных по невозрастанию размеров. Тогда для любого $r \in \mathbb{N}$

$$(\sigma_i(n))^r_{i=1} \xrightarrow{d} (\xi_i)^r_{i=1}, \quad n \to +\infty,$$

где ξ_i – число точек $N((A^\alpha)^{k-1})$ внутри интервала γ_i, $i \in \mathbb{N}$.

Список литературы

[1] Bollobás B., *Random graphs*, Cambridge University Press, Cambridge, 2001.

[2] Janson S., Luczak T., Rucinski A., *Random Graphs*, Wiley Intersci. Ser. Discrete Math. Optim., Wiley Interscience, New York, 2000.

[3] Frieze A., Karonski M., *Introduction to random graphs*, Cambridge, Cambridge University Press, 2015.

[4] Erdos P., Renyi A., On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, **38** (1961), 343–347.

[5] Степанов В.Е., Фазовые переходы в случайных графах, *Теория вероятн. и ее примен.*, **15**:2 (1970), 200–215.

[6] Bollobás B., Riordan O. Asymptotic normality of the size of the giant component via a random walk, *Journal of Combinatorial Theory, Series B*, **102** (2012), 53–61.

[7] Łuczak T., Pittel B., Wierman J., The structure of a random graph at the point of phase transition, *Transactions of the American Mathematical Society*, **341**:2 (1994), 721–748.

[8] Aldous D., Brownian excursions, critical random graphs and the multiplicative coalescent, *The Annals of Probability*, **25**:2 (1997), 812–854.

[9] Schmidt-Pruzan J., Shamir E., Component structures in the evolution of random hypergraphs, *Combinatorica*, **5**:1 (1985), 81–94.

[10] Behrisch M., Coja-Oghlan A., Kang M. The order of the giant component of random hypergraphs, *Random Struct. Alg.*, **36** (2010), 149–184.

[11] Bollobas B., Riordan O., Asymptotic Normality of the Size of the Giant Component in a Random Hypergraph, *Random Struct. Algorithms*, **41** (2012), 441–450.

[12] I. Derényi, G. Palla, and T. Vicsek, Clique percolation in random networks, Phys Rev Lett 94 (2005), 160202 (4 pages).

[13] Aldous, D. J. (1991). The continuum random tree. II: an overview. In Stochastic Analysis (M. T. Barlow and N. H. Bingham, eds.) 23-70. Cambridge Univ. Press.

[14] Aldous, D. J. and Pitman, J. (1994). Brownian bridge asymptotics for random mappings. Random Structures and Algorithms 5 487-512.

[15] Bollobas, B. (1985). Random Graphs. Academic Press, London.

[16] Revuz, D. and Yor, M. (1991). Continuous Martingales and Brownian Motion. Springer, Berlin.

[17] M. Luczak and T. Łuczak, The phase transition in the cluster-scaled model of a random graph, Random Struct Algorithm 28 (2006), 215–246.

[18] B. Pittel and C. Wormald, Counting connected graphs inside-out, J. Comb Theory B 93 (2005), 127–172.

[19] M. Karoński and T. Łuczak, The phase transition in a random hypergraph, J Comput Appl Math 142 (2002), 125–135.

[20] K. L Chung. "Excursions in Brownian motion". Arkiv för Matematik, 1976, 14 (1): 155–177.