Проблема изоморфизма графов

Райгородский Андрей Михайлович

Берендеевы поляны, 20–25 августа 2019 г.

Определение 1. Графы $G = (V_1, E_1)$ и $H = (V_2, E_2)$ изоморфны, если

$$\exists \varphi : V_1 \to V_2 : \ \exists \varphi^{-1}, \forall v_1, v_2 \in V_1 \ (v_1, v_2) \in E_1 <=> (\varphi(v_1), \varphi(v_2)) \in E_2.$$

Пусть n — число вершин графов G и H. Мы хотим проверить за полиномиальное время от n изоморфность графов, решить «проблему изоморфизма».

Классы графов, для которых проблема изоморфизма решена:

- Деревья.
- Если максимальная степень k, то существует алгоритм, работающий за $O(n^k)$.
- Планарные графы.
- Интервальные графы (можно сопоставить каждой вершине отрезок на прямой, что отрезки имеют общую точку тогда и только тогда, когда соответственные вершины смежные).

Упражнение 1 (1 б.). Придумайте пример неинтервального графа.

Пусть степени вершин $G - g_1 \geq g_2 \geq \dots g_n$, а степени $H - h_1 \geq h_2 \geq \dots h_n$. Если последовательности различны, то графы неизоморфны, но обратное неверно (например, два треугольника неизоморфны шестиугольнику).

Определение 2. Каноническая нумерация — такая нумерация, что графы изоморфны тогда и только тогда, когда они совпадают в канонической нумерации. Например, если в графе все степени различны, то каноническая нумерация может сопоставлять каждой вершине его степень (но таких графов 2 — из 0 и 1 вершин).

Неизвестно, эквивалентна ли задача о канонической нумерации проблеме изоморфизма графов.

Тривиальный алгоритм

Проверяем все перестановки. Алгоритм работает за $n!$.

Теорема 1. $n! \geq \left(\frac{n}{e}\right)^n$.

Доказательство. Доказываем по индукции. База верна.

Шаг индукции.

$$(n+1)! = n!(n+1) \geq \left(\frac{n}{e}\right)^n (n+1) \geq^? \left(\frac{n+1}{e}\right)^{n+1}.$$

Утверждение с вопросом вытекает из определения числа e. ■

Теорема 2. Существует (явно указанный) алгоритм, работающий за $e^{c\sqrt{n \ln n}}$.

Теорема 3 (Бабаи, 2015, статья не проверена до конца). Существует алгоритм, работающий за $e^{\ln^c n}$.

Теорема 4 (Бабаи, Эрдёш, Селков, 1974). Для каждого n существует множество графов на n вершинах $\mathfrak{K}_n \subset \Omega_n$ такое, что:

- Проверка того, что $G \in \mathfrak{K}_n$, квадратична от n.
- На $\mathfrak{K}_n$ существует (тривиальная) каноническая нумерация.
- $\lim_{n\to\infty} \dfrac{|\mathfrak{K}_n|}{|\Omega_n|} = 1$. (изначально сходилось с полиномиальной скоростью — $1 - n^{-c}$, сейчас известен алгоритм с экспоненциальной скоростью сходимости — $1 - e^{-nc}$ для какой-то константы c)

Определение 3. Математическое ожидание $\mathbb{E}X$ случайной величины X — сумма $X(G)P(G)$. В частности, если все события равновероятны, $\mathbb{E}X = \frac{\sum X(G)}{|\Omega|}$.

Утверждение. $\mathbb{E}[c_1X_1 + c_2X_2] = c_1\mathbb{E}X_1 + c_2\mathbb{E}X_2$.

Пример. Пусть X — количество треугольников в графе. Тогда считать $\mathbb{E}X$ по определению очень сложно (получается $\sum\limits_i iP(X = i)$ и неочевидно, например, как считать графы без треугольников). Рассмотрим такие $\binom{n}{3}$ случайных величин:

$$Y_{ijk}(V, E) = \begin{cases} 1, (i,j) \in E, (j,k) \in E, (k,i) \in E \\ 0, \text{ иначе} \end{cases} \tag{1}$$

Тогда, с одной стороны, $\mathbb{E}Y_{ijk} = \frac{1}{8}$, с другой стороны, $\mathbb{E}X = \mathbb{E}[\sum\limits_{i,j,k} Y_{ijk}] = \sum\limits_{i,j,k} \mathbb{E}Y_{ijk} = \binom{n}{3}\cdot\frac{1}{8}$.

Определение 4. Дисперсия случайной величины — $\mathbb{D}X = \mathbb{E}[(X - \mathbb{E}X)^2]$.

Лемма 5. $\mathbb{D}X = \mathbb{E}[X^2] - \mathbb{E}^2X$.

Доказательство. Раскроем скобки и получим нужное равенство. ■

Теорема 6 (Марков). Пусть $X \geq 0$ и $a > 0$. Тогда $P(x > a) \leq \frac{\mathbb{E}X}{a}$.

Доказательство.

$$\mathbb{E}X = \sum_i y_iP(X = y_i) = \sum_{y_i > a} y_iP(X = y_i) + \sum_{y_j \leq a} y_jP(X = y_j) \geq \sum_{y_i > a} y_iP(X = y_i) > aP(x > a). \blacksquare$$

Теорема 7 (Чебышёв). Пусть $a > 0$. Тогда $P(|X - \mathbb{E}X| > a) \leq \frac{\mathbb{D}X}{a^2}$.

Доказательство. Пусть $Y = (X - \mathbb{E}X)^2$. Применим 6 для Y и a^2. Получим $P((X - \mathbb{E}X)^2 > a^2) \leq \frac{\mathbb{D}X}{a^2}$, откуда следует утверждение задачи. ■

Теорема 8. Случайный граф на n вершинах не является планарным с вероятностью, стремящейся к 1.

Доказательство. Пусть $X(G)$ — количество K_5 в графе. Хотим доказать: $P(X \geq 1) \to^? 1$.

$$P(X \geq 1) = 1 - P(X \leq 0 = 1 - P(-X \geq 0) = 1 - P(\mathbb{E}X - X \geq \mathbb{E}X) \geq 1 - P(|\mathbb{E}X - X| \geq \mathbb{E}X).$$

Применим 7. Получим $P(X \geq 1) \geq 1 - \frac{\mathbb{D}X}{\mathbb{E}^2X}$. Пусть ребро возникает с вероятностью p. Тогда $\mathbb{E}X = \binom{n}{5} \cdot p^{10}$. Посчитаем $\mathbb{E}[X^2]$:

$$\mathbb{E}X^2 = \mathbb{E}[(\sum Y_j)^2] = \mathbb{E}[Y_1^2 + \ldots + Y_{\binom{n}{5}}^2] + \sum_{i \neq j} \mathbb{E}[Y_iY_j] = \sum_{i \neq j} \mathbb{E}[Y_iY_j] + \mathbb{E}X.$$

Пятёрки i и j могут пересекаться либо по 0 вершинам, либо по 1, либо по 2, либо по 3, либо по 4, и в этих случаях коэффициенты у матожидания будут соответственно $p^{20}, p^{20}, p^{19}, p^{17}, p^{14}$ и все слагаемые, кроме p^{20}, будут стремиться к нулю при $p^2n > 1$:

$$\mathbb{E}[X^2] \sim \mathbb{E}X + \binom{n}{5}p^{20}\left(\binom{n-5}{5} + 5\binom{n-5}{4}\right).$$

Осталось посчитать $\frac{\mathbb{D}X}{\mathbb{E}^2X}$:

$$\frac{\mathbb{D}X}{\mathbb{E}^2X} = s - 1 + \frac{*}{\mathbb{E}^2X} \sim \frac{n^5p^{10}}{120} - 1 + \frac{*}{\mathbb{E}^2X} \sim \frac{n^5p^{10}}{120}.$$

Получается, что при $p > \frac{c}{\sqrt{n}}$ вероятность отсутствия K_5 стремится к нулю. ■

Случайные блуждания на прямой

Пусть есть случайные величины $X_1, X_2, \ldots$, равные 1 и -1 с вероятностью 0.5 (например, в точке 0 стоит пьяница и на каждом шаге идёт на 1 в какую-то сторону), и мы считаем $p = P(X_1 + \ldots + X_n > a)$. Заметим, что $\mathbb{E}[X_1 + \ldots + X_n] = 0$ по линейности. Применим 7:

$$p \leq \frac{\mathbb{D}[X_1 + \ldots + X_n]}{a^2} = \frac{\mathbb{E}[(X_1 + \ldots + X_n)^2] - \mathbb{E}^2[X_1 + \ldots + X_n]}{a^2} = \frac{n}{a^2}.$$

Теорема 9. $p \leq e^{-\frac{a^2}{2n}}$, причём это неравенство неулучшаемо.
Доказательство. Пусть $\lambda > 0$. Тогда

$$P(X_1 + \ldots + X_n \geq a) = P(\lambda(X_1 + \ldots + X_n) \geq \lambda a) = P(e^{\lambda(X_1 + \ldots + X_n)} \geq e^{\lambda a}).$$

Применим 6. Получим

$$p \leq \frac{\mathbb{E}[e^{\lambda(X_1 + \ldots + X_n)}]}{e^{\lambda a}} = \mathbb{E}[e^{\lambda X_1} \cdot \ldots \cdot e^{\lambda X_n}] \cdot e^{-\lambda a} = \mathbb{E}[e^{\lambda X_1}]\mathbb{E}[e^{\lambda X_2}] \ldots \mathbb{E}[e^{\lambda X_n}] \cdot e^{-\lambda a} = \left(\frac{e^{\lambda} + e^{-\lambda}}{2}\right)^n e^{-\lambda a}.$$

Заметим, что $\operatorname{ch}\lambda := \frac{e^{\lambda}+e^{-\lambda}}{2} = \sum\limits_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!}$ и $(2k)! \geq k!2^k$. Тогда:

$$p \leq \left(\sum_{k=0}^{\infty} \frac{\lambda^{2k}}{k!2^k}\right)^n \cdot e^{-\lambda a} = e^{\frac{\lambda^2 n}{2} - \lambda a}.$$

Подставим $\lambda = \frac{a}{n}$. Получим искомую оценку — $p \leq e^{-\frac{a^2}{2n}}$. ∎

Алгоритм Бабаи-Эрдёша-Селкова

1. Находим первые $r = [3\log_2 n]$ членов последовательности g_i. Назовём соответствующие вершины *большими*, остальные *маленькими*.

2. Если среди степеней больших вершин есть две одинаковых, то $G \notin \mathfrak{K}_n$.

3. Для всех маленьких вершин запишем $f(x_i) = \sum\limits_{j=1}^{r} 2^j \cdot ((i,j) \in E\ ?1 : 0)$.

4. Если среди $f(x_i)$ два одинаковых числа, то $G \notin \mathfrak{K}_n$, иначе мы задали каноническую нумерацию (можно приписать к малым вершинам 1 в начале двоичного кода, чтобы они не пересекались с большими).

Теорема 10. Для этого алгоритма вероятность ошибки стремится к 0.

Доказательство. Докажем, что $P(\overline{\mathfrak{K}_n}) \to 0$. Пусть $\mathcal{C}$ — такое событие: для $i \in \{1, 2, \ldots, r+1\}$ выполняется $d_i - d_{i+1} > 2$ (для сравнения, в первой проверке мы проверяем, что для $i \in \{1, 2, \ldots, r-1\}$ выполняется $d_i - d_{i+1} > 0$). Пусть $\mathcal{D}(i,j)$ — такое событие: $G/v_i/v_j$ прошёл 1-ю проверку для того же r. Пусть $\mathcal{A}(i,j)$ — такое событие: $\overline{\mathcal{D}(i,j)} \cup G/v_i/v_j$ не прошёл 2-ю проверку. Заметим, что:

- $\overline{\mathfrak{K}_n} \implies (\overline{\mathcal{C}} \cup G$ не прошёл 2-ю проверку$) \implies \overline{\mathcal{C}} \cup (\mathcal{C} \cap \bigcup\limits_{i,j} \mathcal{A}(i,j))$.

- $\forall i, j :\ \mathcal{C} \implies \mathcal{D}(i,j)$.

Тогда

$$P(\overline{\mathfrak{K}_n}) \leq P(\overline{\mathcal{C}}) + \sum_{i,j} P(\mathcal{A}(i,j) \cap \mathcal{C}) \leq P(\overline{\mathcal{C}}) + \sum_{i,j} P(\mathcal{A}(i,j) \cap \mathcal{D}(i,j)) \leq P(\overline{\mathcal{C}}) + \sum_{i,j} P(\mathcal{A}(i,j) | \mathcal{D}(i,j)).$$

Заметим, что каждое слагаемое вида $P(\mathcal{A}(i,j)|\mathcal{D}(i,j))$ означает «две последовательности из r нулей и единиц совпадают», т.е. у этого вероятность $\frac{1}{2^r}$. Таких слагаемых не более, чем n^2, значит:

$$P(\overline{\mathfrak{K}_n}) \leq P(\overline{\mathcal{C}}) + n^2 2^{-3\log_2 n} = P(\overline{\mathcal{C}}) + \frac{1}{n} \to \lim_{n\to\infty} P(\overline{\mathcal{C}}).$$

Обозначим $k = \frac{n-1}{2} + x\sqrt{(n-1)\ln n}$ для некоторой константы x. Пусть $Y(G)$ — число вершин G, имеющих степени k и больше, Z_j — число вершин, имеющих степень ровно j. Заметим, что

$$\begin{aligned}\overline{\mathcal{C}} \implies (Y < r) \cup (Y \geq r \cap \exists i \in \mathbb{N}_0, u, v \in V : k + i \leq \deg u \leq \deg v \leq k + i + 2)\\ P(\overline{\mathcal{C}}) \leq P(Y < r) + P((Y \geq r) \cap \exists i, u, v : k + i \leq \deg u \leq \deg v \leq k + i + 2)\\ P(Y < r) = P(-Y > -r) = P(\mathbb{E}Y - Y > \mathbb{E}Y - r)\end{aligned}$$

Оценим $\mathbb{E}Y$:

$$\mathbb{E}Y = \mathbb{E}[Z_k + \ldots + Z_{n-1}] = n \cdot \sum_{j=k}^{n-1} \left(\frac{1}{2}\right)^{n-1} \cdot \binom{n-1}{j}.$$

Введём $\alpha_1, \ldots, \alpha_{n-1}$, случайные величины, равные 0 или 1 с вероятностью $\frac{1}{2}$, тогда $\mathbb{E}Y = nP(\alpha_1 + \ldots + \alpha_{n-1} \geq k) = nP((2\alpha_1 - 1) + \ldots + (2\alpha_{n-1} - 1) \geq 2k - (n-1)) \leq ne^{-2x\ln n}$, где последнее неравенство возникает из-за 9, и при $x < \frac{\sqrt{2}}{2}$ это стремится к ∞. Применим 7:

$$P(Y < r) \leq \frac{\mathbb{D}Y}{(\mathbb{E}Y_k - r)^2} \sim \frac{\mathbb{D}Y}{n} \sim \frac{\mathbb{E}Y}{n} \to 0.$$

Осталось оценить вторую часть неравенства:

$$P((Y \geq r) \cap \exists i, u, v : k+i \leq \deg u \leq \deg v \leq k+i+2) \leq P(\exists i, u, v : k+i \leq \deg u \leq \deg v \leq k+i+2).$$

Заметим, что это не больше, чем сумма по всем i каких-то вероятностей. Оценим одно из слагаемых так:

$$P(\exists u, v : k \leq \deg u \leq \deg v \leq k+2) \leq n(n-1)P(k \leq \deg u \leq \deg v \leq k+2) =$$

$$= \frac{n(n-1)}{2} \sum_{i,j} \left(\binom{n-2}{i-1} \left(\frac{1}{2}\right)^{n-2} \binom{n-2}{j-1} \left(\frac{1}{2}\right)^{n-2} + \binom{n-2}{i} \left(\frac{1}{2}\right)^{n-2} \binom{n-2}{j} \left(\frac{1}{2}\right)^{n-2} \right) \leq$$

$$= n(n-1)\binom{n-1}{k}^2 \cdot \left(\frac{1}{2}\right)^{2n-3} \cdot 10 \leq 80n^2 \left(\frac{1}{2}\right)^{2n} \binom{n-1}{\frac{n-1}{2}}^2 e^{-\frac{4x^2(n-1)\ln n}{n-1}}.$$

Утверждение. $\binom{n}{\frac{n}{2}} \sim \frac{2^n}{\sqrt{\pi n}}$.

После применения утверждения получится, что при $x > \frac{1}{2}$ эта вероятность стремится к нулю, т.е. после подстановки x из интервала $(\frac{1}{2}, \sqrt{\frac{1}{2}})$ вся вероятность $\overline{\mathfrak{K}_n}$ стремится к нулю. ■