

Министерство образования и науки Российской Федерации

Государственное образовательное учреждение  
высшего профессионального образования  
«Московский физико-технический институт (государственный  
университет)»  
Факультет инноваций и высоких технологий  
Кафедра анализа данных

## Магистерская диссертация

Тема: **Название моей работы (TODO)**

Направление: 010400  
Прикладные математика и информатика

Выполнил:  
студент 093 группы \_\_\_\_\_ Попов М.В.

Научный руководитель:  
д.физ.-мат.н., проф.(todo) \_\_\_\_\_ Ромашенко А.Е.

г. Москва 2016

# Содержание

|  |           |
|--|-----------|
| <b>Введение</b> . . . . .                                | <b>2</b>  |
| <b>Коммуникационная сложность</b> . . . . .              | <b>2</b>  |
| 1.1 Постановка задачи . . . . .                          | 2         |
| 1.2 Одноцветные комбинаторные прямоугольники . . . . .   | 3         |
| 1.3 Графовая интерпретация . . . . .                     | 4         |
| <b>Оценивание <math>bcc(G)</math></b> . . . . .          | <b>6</b>  |
| 2.1 Метод трудного множества . . . . .                   | 7         |
| 2.2 Метод Куликова-Юкны . . . . .                        | 8         |
| 2.3 Метод энтропийных неравенств . . . . .               | 9         |
| <b>Геометрические конфигурации</b> . . . . .             | <b>11</b> |
| 3.1 Описание двудольного графа . . . . .                 | 11        |
| 3.2 Оценки для проективных плоскостей . . . . .          | 11        |
| 3.3 Сравнение методов оценивания . . . . .               | 13        |
| <b>Трудное множество vs Куликов-Юкна</b> . . . . .       | <b>15</b> |
| 4.1 Теорема Турановского типа . . . . .                  | 16        |
| 4.2 Конструкция прямоугольного графа . . . . .           | 17        |
| 4.3 Случайные графы Эрдеша-Реньи . . . . .               | 20        |
| 4.4 Неравенство Хефдинга . . . . .                       | 21        |
| 4.5 Размер максимального паросочетания . . . . .         | 23        |
| 4.6 Трудное множество и Куликов-Юкна на случайных графах | 24        |
| 4.7 Количество трудных множеств размера $k$ . . . . .    | 26        |
| <b>Список литературы</b> . . . . .                       | <b>28</b> |

# Введение

(ToDo) Актуальность, новизна, краткая выжимка.

## Коммуникационная сложность

### 1.1 Постановка задачи

Мы будем рассматривать задачи следующего вида: пусть имеются два человека, которые хотят совместно вычислить значение некоторой функции от двух переменных  $f(x, y)$ . По традиции мы будем называть первого участника игры Алисой, а второго – Бобом. Сложность у этой задачи в том, что Алиса знает только значение аргумента  $x$ , а Боб значение аргумента  $y$ . Алиса и Боб могут обмениваться сообщениями по каналу связи. Требуется вычислить значение  $f(x, y)$ , переслав по каналу связи минимальное количество информации.

Мы предполагаем, что Алиса и Боб заранее (до того, как им станут известны значения  $x$  и  $y$ ) договариваются о коммуникационном протоколе — о наборе соглашений, какие именно данные и в каком порядке они будут пересылать друг другу при тех или иных значениях  $x$  и  $y$ .

Опишем теперь всю задачу более формально. Пусть имеются конечные множества  $X, Y, Z$  и задана некоторая функция  $f : X \times Y \rightarrow Z$ .

**Определение.** *Коммуникационным протоколом для вычисления некоторой функции  $f : X \times Y \rightarrow Z$  называется ориентированное двоичное дерево со следующей разметкой на вершинах и ребрах:*

- каждая нелистовая вершина помечена буквой  $A$  или  $B$ ;
  - у вершин с пометкой  $A$  определена функция  $g_i : X \rightarrow \{0, 1\}$ ;
  - у вершин с пометкой  $B$  определена функция  $f_j : Y \rightarrow \{0, 1\}$ ;
- каждой листовой вершине сопоставлен элемент множества  $Z$ ;
- каждое ребро помечено 0 или 1.

Пусть Алиса и Боб договорились, что будут действовать по некоторому протоколу  $\mathcal{P}$ . Затем Алиса получила  $x \in X$ , а Боб получил  $y \in Y$ .

Поместим фишку в корневую вершину нашего протокола  $\mathcal{P}$  и будем перемещать ее вниз по дереву, последовательно удаляясь от корня, пока она не попадет в один из листьев. Перемещение фишки выполняется следующим образом. Если текущая вершина помечена буквой  $A$ , то это означает, что сейчас очередь Алисы. Она применяет функцию  $g_i$  текущей вершины к своему значению  $x$ . Алиса отправляет по каналу связи бит равный  $g_i(x)$  и перемещает фишку по ребру, помеченному как  $g_i(x)$ . Боб получает отправленный бит и понимает куда была сдвинута фишка. Для вершин помеченных буквой  $B$  эту же процедуру выполняет Боб. Когда фишка попадает в лист дерева, записанное там значение  $z \in Z$  объявляется результатом выполнения протокола.

Мы говорим, что протокол  $\mathcal{P}$  вычисляет функцию  $f : X \times Y \rightarrow Z$ , если для любого  $x \in X$  и любого  $y \in Y$  при движении из корня по пути, соответствующему заданным  $x$  и  $y$ , мы попадаем в лист, помеченный  $z = f(x, y)$ .

**Определение.** *Сложностью коммуникационного протокола называется его глубина. Коммуникационной сложностью функции  $f$  называется минимальная сложность протокола, вычисляющего  $f$ . Мы будем обозначать её  $CC(f)$ .*

## 1.2 Одноцветные комбинаторные прямоугольники

**Определение.** *Множество  $S \subset X \times Y$  называется комбинаторным прямоугольником (или просто прямоугольным множеством), если существуют такие  $A \subset X$  и  $B \subset Y$ , что  $S = A \times B$ .*

Пусть  $\mathcal{P}$  – некоторый коммуникационный протокол для вычисления функции  $f : X \times Y \rightarrow Z$  и  $l$  – один из листьев протокола. Определим  $S_l$  как множество пар  $(x, y) \in X \times Y$  таких, что на входе  $(x, y)$  Алиса и Боб, следуя протоколу  $\mathcal{P}$ , приходят в лист  $l$ .

**Утверждение.** *Для всякого коммуникационного протокола  $\mathcal{P}$  и для всякого листа  $l$  множество  $S_l$  является комбинаторным прямоугольником.*

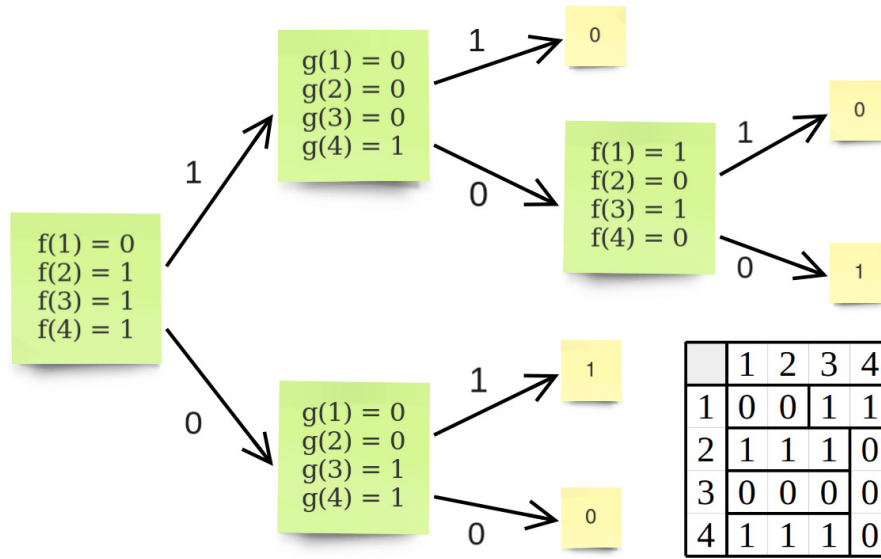


Рис. 1: Пример протокола и разбиения таблицы значений.

Доказательство этого утверждения можно прочитать, например, в [1]. В итоге мы получаем, что коммуникационный протокол для вычисления функции  $f$  задаёт разбиение  $X \times Y$  - таблицы значений  $f$  на прямоугольные множества, соответствующие листьям. Поскольку каждому листу протокола приписано одно значение функции  $f$ , эти прямоугольные множества являются одноцветными, то есть во всех точках такого прямоугольного множества функция  $f$  принимает одно и то же значение. Например, для  $X = Y = \{1, 2, 3, 4\}$ ,  $Z = \{0, 1\}$  и протокола  $\mathcal{P}$  (рис. 1) получаем разбиение на 5 одноцветных прямоугольных множеств.

Подведем промежуточные итоги: всякий протокол с  $l$  листьями (вычисляющий функцию  $f$ ) задаёт разбиение таблицы значений  $f$  на  $l$  одноцветных прямоугольных множеств. Значит, чтобы доказать, что коммуникационная сложность  $CC(f)$  не меньше  $n$ , достаточно показать, что таблицу значений невозможно разбить на менее, чем  $2^n$  одноцветных прямоугольных множеств.

### 1.3 Графовая интерпретация

Давайте теперь посмотрим на другое представление множества значений функции  $f$ . Рассмотрим полный двудольный граф  $G = (X, Y, E)$ , ребра которого раскрашены в  $|Z|$  цветов. Вершины левой доли соот-

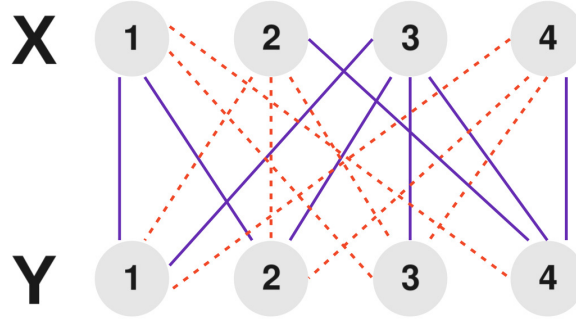


Рис. 2: Графовая интерпретация: синие – 0, красные – 1.

ветствуют элементам множества  $X$ , вершины правой доли – элементам множества  $Y$ . Ребро  $(x, y) \in X \times Y$  имеет цвет  $z \in Z$ , если  $f(x, y) = z$ .

Из определения комбинаторного прямоугольника видно, что в графовой интерпретации он является ничем иным, как полным двудольным подграфом. А разбиение таблицы значений  $f$  на одноцветные прямоугольные множества – это разбиение нашего полного двудольного графа  $G$  на одноцветные непересекающиеся биклики (полные двудольные подграфы). Для нашего примера графовую интерпретацию можно посмотреть на рис. 2.

**Определение.** Бикликовым разбиением  $bcp(G)$  двудольного графа  $G$  будем называть наименьшее число непересекающихся биклик, которыми можно покрыть все ребра графа  $G$ .

**Определение.** Бикликовым покрытием  $bcc(G)$  двудольного графа  $G$  будем называть наименьшее число, возможно, пересекающихся биклик, которыми можно покрыть все ребра графа  $G$ .

**Утверждение.** Для произвольного двудольного графа  $G$  верно

$$bcp(G) \geq bcc(G)$$

Для каждого  $z \in Z$  определим двудольный граф  $G_z = (X, Y, E_z)$ , как граф, получающийся из  $G$  выкидыванием всех ребер цвета, отличного от  $z$ . Иначе говоря  $E_z = \{(x, y) \in X \times Y \mid f(x, y) = z\}$ .

Величины  $bcp(G_z)$  и  $bcc(G_z)$  дают некоторую нижнюю оценку на коммуникационную сложность функции  $f$ , с которой намного удобнее рабо-

татья:

$$2^{CC(f)} \geq \sum_{z \in Z} bcp(G_z) \geq \sum_{z \in Z} bcc(G_z)$$

**Замечание.** На самом деле величины  $bcc(G_z)$  тесно связаны с неопределенной коммуникационной сложностью  $NCC(f)$ . Для произвольного множества  $Z$  верно:

$$NCC(f) \leq \lceil \log_2(\sum_{z \in Z} bcc(G_z)) \rceil + 1$$

а для  $Z = \{0, 1\}$ :

$$NCC(f) = \lceil \log_2(bcc(G_1)) \rceil$$

Подробнее про это можно прочитать, например, в [2].

В итоге мы получили мощный инструмент для доказательства нижних оценок коммуникационной сложности. К сожалению, задача нахождения величины  $bcc(G)$  является PSPACE-полной [3], а точное значение известно только для очень скудного класса графов (например, для "crown graphs" [4]), поэтому напрямую мы не можем использовать эту оценку. В следующей главе я рассмотрю несколько методов, позволяющих для произвольного двудольного графа оценивать снизу величину  $bcc(G)$ .

## Оценивание $bcc(G)$

В этой главе я опишу три различных метода оценивания бикликового покрытия:

- метод трудного множества ("fooling set");
- метод Куликова-Юкны;
- метод энтропийных неравенств.

Первые два метода работают для произвольных графов (необязательно двудольных), а третий применим к большому классу двудольных графов.

## 2.1 Метод трудного множества

Данный метод тесно связан с одноцветными прямоугольными множествами. Классическое определение трудного множества выглядит следующим образом:

**Определение.** Для функции  $f : X \times Y \rightarrow Z$  и элемента  $z \in Z$  будем называть множество  $S_z \subset X \times Y$  трудным (в англоязычной литературе *fooling set*), если верно:

- для всякой пары  $(x, y) \in S_z$  имеем  $f(x, y) = z$ ;
- для любых двух несовпадающих пар  $(x, y) \in S_z$  и  $(x', y') \in S_z$  имеем  $f(x, y') \neq z$  или  $f(x', y) \neq z$ .

Нас будет интересовать немного более общее определение трудного множества (графовая интерпретация):

**Определение.** Пусть  $G = (V, E)$  произвольный неориентированный граф. Будем называть подмножество ребер  $S \subseteq E$  трудным, если для любых двух различных ребер  $(x, y) \in S$  и  $(x', y') \in S$  имеем  $(x, y') \notin E$  или  $(x', y) \notin E$ . Обозначение  $\text{fool}(G)$  - размер максимального по мощности трудного множества.

**Замечание.** Классическое определение получается из графового, применением к двудольному графу  $G_z = (X, Y, E_z)$ , который строится по функции  $f : X \times Y \rightarrow Z$ .

**Теорема.** Для произвольного неориентированного графа  $G = (V, E)$ , если подмножество ребер  $S \subseteq E$  является трудным, то  $\text{bcs}(G) \geq |S|$ .

**Доказательство.** Достаточно доказать, что два ребра, лежащие одновременно в одном трудном множестве, не могут попасть в одну биклику. Пусть не так, значит существуют два ребра  $(x, y) \in B \cap S$  и  $(x', y') \in B \cap S$ , где  $B$  - биклика, а  $S$  - трудное подмножество ребер. Но тогда ребра  $(x, y')$  и  $(x', y)$  также принадлежат биклике  $B$ , а значит лежат и в нашем множестве ребер  $E$ . Противоречие. ■

**Замечание.** На практике нахождение максимального по мощности трудного множества применяют редко, потому что эта задача является  $PSPACE$ -полной [3].



## 2.2 Метод Куликова-Юкны

Следующий метод был впервые описан в статье [5], и работает он для произвольного неориентированного графа.

**Теорема.** Для произвольного неориентированного графа  $G = (V, E)$  верно:

$$bcc(G) \geq \left\lceil \frac{v(G)^2}{|E|} \right\rceil$$

где  $v(G)$  – размер максимального паросочетания графа  $G$ .

**Доказательство.** Пусть  $M \subseteq E$  – это максимальное паросочетание, тогда рассмотрим бикликовое покрытие, на котором достигается минимум  $E = B_1 \cup B_2 \cup \dots \cup B_{bcc(G)}$ . Определим отображение  $g : M \rightarrow \{1, \dots, bcc(G)\}$ , как  $g(e) = \min\{i \mid e \in B_i\}$  и пусть  $M_i = \{e \in M \mid g(e) = i\}$ . Иначе говоря  $M_i$  содержит только те ребра максимального паросочетания  $M$ , которые покрываются бикликой  $B_i$  в первый раз.

Пусть  $F_i \subseteq B_i$  биклика, индуцированная вершинами ребер из  $M_i$ . Пусть  $F = F_1 \sqcup F_2 \sqcup \dots \sqcup F_{bcc(G)}$  (биклики  $F_i$  не пересекаются по построению).

Очевидно, что  $F_i$  – биклика размера  $r_i \times r_i$ , где  $r_i = |M_i|$ . Получаем следующие соотношения:

$$r_1 + r_2 + \dots + r_{bcc(G)} = |M| = v(G)$$

и

$$r_1^2 + r_2^2 + \dots + r_{bcc(G)}^2 = |F|$$

Из неравенства Коши-Буняковского получаем

$$v(G)^2 = (r_1 + r_2 + \dots + r_{bcc(G)})^2 \leq bcc(G) \cdot (r_1^2 + r_2^2 + \dots + r_{bcc(G)}^2) = bcc(G) \cdot |F|$$

А так как  $F \subseteq E$ , то

$$v(G)^2 \leq bcc(G) \cdot |F| \leq bcc(G) \cdot |E| \blacksquare$$

В этой же статье [5] этот метод сравнивался с другой оценкой: пусть

$bcl(G) = \max_{K_{r,r} \subseteq G} \{r\}$ , тогда

$$bcc(G) \geq \left\lceil \frac{v(G)}{bcl(G)} \right\rceil \quad (*)$$

Данная оценка очевидным образом следует из того, что любая биклика  $K_{r,s}$  содержит как максимум  $\min\{r, s\}$  ребер максимального паросочетания.

Приведем примеры графов, на которых метод Куликова-Юкны работает намного лучше, чем оценка (\*), и наоборот:

- пусть двудольный граф  $G = (L, R, E)$  состоит из совершенного паросочетания размера  $n = |L| = |R|$  и еще некоторого константного числа непересекающихся биклик  $K_{r,r}$ . К тому же, пусть  $r = \Theta(\sqrt{n})$ , тогда

$$\left\lceil \frac{v(G)^2}{|E|} \right\rceil = \left\lceil \frac{n^2}{cr^2 + n} \right\rceil = \Theta(n) \gg \Theta(\sqrt{n}) = \left\lceil \frac{n}{r} \right\rceil = \left\lceil \frac{v(G)}{bcl(G)} \right\rceil$$

- рассмотрим двудольный граф Леви, построенный при помощи конечной проективной плоскости порядка  $p \in \mathbb{P}$ . В каждой доле этого графа содержится  $n = p^2 + p + 1$  вершин, причем степень каждой  $p + 1$ . Этот граф не содержит  $K_{2,2}$  (любые две прямые пересекаются максимум в одной точке). А так как в регулярных двудольных графах обязательно найдется совершенное паросочетание, то

$$\begin{aligned} \left\lceil \frac{v(G)^2}{|E|} \right\rceil &= \left\lceil \frac{(p^2 + p + 1)^2}{(p^2 + p + 1)(p + 1)} \right\rceil = \Theta(\sqrt{n}) \ll \\ &\ll \Theta(n) = \left\lceil \frac{p^2 + p + 1}{1} \right\rceil = \left\lceil \frac{v(G)}{bcl(G)} \right\rceil \end{aligned}$$

## 2.3 Метод энтропийных неравенств

Следующий метод оценивания бикликового покрытия был описан в статье [6], как результат применения энтропийного неравенства:

$$H(A|B, X) + H(A|B, Y) \leq H(A|B)$$

К сожалению, это неравенство выполняется не для произвольного совместного распределения случайных величин  $A, B, X, Y$ , и соответственно на двудольный граф будет накладываться дополнительное условие (\*).

**Теорема.** Пусть ребра двудольного графа  $G = (L, R, E)$  раскрашены следующим образом:

(\*) для произвольной биклики  $C \subseteq E$  и для произвольной пары ребер  $(x, y')$  и  $(x', y)$  из  $C$ , покрашенных в цвет  $a$ , цвет ребер  $(x, y)$  и  $(x', y')$  тоже  $a$ .

Пусть также на ребрах этого графа задано произвольное вероятностное распределение. Определим случайные величины  $(X, Y, A)$  следующим образом:

- $X = [\text{левый конец ребра}]$ ,
- $Y = [\text{правый конец ребра}]$ ,
- $A = [\text{цвет ребра}]$ .

Тогда выполняется неравенство:

$$b_{cc}(G) \geq 2^{\frac{1}{2}(H(A|X)+H(A|Y)-H(A))}$$

**Пример.** Определим двудольный граф  $G_{n,k} = (L, R, E)$  следующим образом:

- $L$  и  $R$  всевозможные  $k$ -элементные подмножества  $\{1, 2, \dots, n\}$ ,
- $E \subseteq L \times R$  состоит из пар непересекающихся множеств.

Определим цвет ребра  $(x, y) \in E$ , как  $x \sqcup y$ , и пусть на ребрах задано равномерное распределение. Условие (\*) выполнено, потому что любые два одноцветных ребра не могут лежать в одной биклике. А так как  $H(A|X) = H(A|Y) = \log_2 \binom{n-k}{k}$  и  $H(A) = \log_2 \binom{n}{2k}$ , то

$$b_{cc}(G_{n,k}) \geq \sqrt{\binom{n-k}{k}^2 / \binom{n}{2k}}$$

Если  $n \gg k$ , то  $\binom{n-k}{k}^2 / \binom{n}{2k}$  близко к  $\binom{2k}{k} \approx 2^{2k}$  и мы получаем нижнюю оценку  $b_{cc}(G_{n,k}) \geq 2^k$ .

# Геометрические конфигурации

В этой главе мы приведем класс двудольных графов, построенных при помощи некоторой геометрической конфигурации  $\Gamma$ . Далее мы увидим, что к этим двудольным графам применимы все наши оценки, и поэтому, изменяя  $\Gamma$ , мы можем сравнить какие методы работают лучше, а какие хуже.

## 3.1 Описание двудольного графа

**Определение.** Геометрической конфигурацией  $\Gamma$  (*Partial Linear Space*) будем называть конечное множество прямых  $A$  и конечное множество точек  $V$  на них, что выполняются следующие две аксиомы:

- Любые две точки лежат как максимум на одной прямой.
- На каждой прямой лежит хотя бы две точки.

**Определение.** Проективной плоскостью с параметрами  $(p_\gamma, l_\pi)$  называется геометрическая конфигурация, состоящая из  $p$  точек и  $l$  прямых, причем через каждую точку проходит ровно  $\gamma$  прямых и на каждой прямой лежит ровно  $\pi$  точек.

Пусть у нас имеется некоторая геометрическая конфигурация  $\Gamma = (V, A)$ , тогда определим двудольный граф  $G_{n,\Gamma} = (L, R, E)$  следующим образом:

- $L = R = V^n$
- $E = \{(x, y) \in L \times R \mid \forall i : x_i \neq y_i \text{ и лежат на одной прямой из } A\}$

## 3.2 Оценки для проективных плоскостей

Для произвольной проективной плоскости  $\Gamma$  с параметрами  $(p_\gamma, l_\pi)$  найдем какие оценки на  $bcc(G_{n,\Gamma})$  дают наши методы:

– Метод трудного множества:

**Лемма.** Если в  $\Gamma$  имеется цикл нечетной длины, то в  $G_{1,\Gamma}$  можно найти трудное множество размера 3.

**Доказательство.** Рассмотрим нечетный цикл минимальной длины  $\{v_1, v_2, \dots, v_{2k+1}\}$ , где  $k \geq 1$ . Заметим, что прямые могут проходить только через соседние точки этого цикла, иначе бы мы нашли нечетный цикл меньшей длины. Тогда, если  $k > 1$ , то множество ребер  $\{(v_1, v_2), (v_2, v_3), (v_3, v_4)\}$  образует трудное множество, а если  $k = 1$ , то  $\{(v_1, v_2), (v_2, v_3), (v_3, v_1)\}$  образует трудное множество.  $\square$

**Замечание.** Если на какой-нибудь прямой лежит по крайней мере три точки, то мы уже имеем цикл длины 3.

Если нечетных циклов в  $\Gamma$  нет, то мы получаем геометрическую конфигурацию, аналогичную двудольному графу. Если этот двудольный граф полный, то наибольшее трудное множество имеет размер 2, а если неполный, то мы можем найти трудное множество размера 3.

**Лемма.** Если в  $G_{1,\Gamma}$  существует трудное множество размера  $k$ , то в  $G_{n,\Gamma}$  существует трудное множество размера  $k^n$

**Доказательство.** Докажем вначале, что если в графе  $G_1$  имеется трудное множество размера  $n_1$ , а в графе  $G_2$  – трудное множество размера  $n_2$ , тогда в  $G_1 \otimes G_2$  можно найти трудное множество размера  $n_1 \cdot n_2$  (где  $\otimes$  – произведение Кронекера). Пусть  $\{v_{i,j}\}$  трудное множество в графе  $G_1$ , тогда в каждой подматрице  $v_{i,j} \cdot G_2$  матрицы графа  $G_1 \otimes G_2$  рассмотрим клетки, соответствующие трудному множеству графа  $G_2$ . Всего мы получили  $n_1 \cdot n_2$  клеток, образующих трудное множество графа  $G_1 \otimes G_2$  по построению.

Вернемся к доказательству леммы. Так как матрица графа  $G_{n,\Gamma}$  есть не что иное, как Кронекерово произведение  $n$  матриц графа  $G_{1,\Gamma}$ , то мы можем найти трудное множество размера  $k^n$ .  $\square$

В итоге мы получили, что если  $\Gamma$  является аналогом полного двудольного графа, то

$$b_{cc}(G_{n,\Gamma}) \geq 2^n$$

иначе

$$b_{cc}(G_{n,\Gamma}) \geq 3^n$$

– Метод Куликова-Юкны:

Так как  $\Gamma$  имеет параметры  $(p_\gamma, l_\pi)$ , то каждая вершина графа  $G_{1,\Gamma}$  соединена с  $\gamma \cdot (\pi - 1)$  другими, а значит всего ребер  $\gamma \cdot (\pi - 1) \cdot p$ . Тогда в графе  $G_{n,\Gamma}$  всего ребер  $\gamma^n \cdot (\pi - 1)^n \cdot p^n$ . Так как у нас однородный двудольный граф, то имеется совершенное паросочетание, а значит  $v(G_{n,\Gamma}) = p^n$ . В итоге получаем оценку:

$$bcc(G_{n,\Gamma}) \geq \frac{p^{2n}}{\gamma^n \cdot (\pi - 1)^n \cdot p^n} = \left( \frac{p}{\gamma \cdot (\pi - 1)} \right)^n$$

– Метод энтропийных неравенств:

Определим раскраску ребер нашего графа  $G_{n,\Gamma} = (L, R, E)$ : сопоставим каждой прямой конфигурации  $\Gamma$  свой цвет, тогда цвет ребра  $(x, y) \in E$  равен  $n$ -мерному вектору цветов прямых проходящих через  $x_i$  и  $y_i$ .

Проверим свойство (\*): пусть  $(x, y')$  и  $(x', y)$  одного цвета и лежат в одной биклике  $C$ , значит для любого  $i$  точки  $x_i, y'_i, x'_i, y_i$  лежат на одной прямой (некоторые точки могут совпадать), но тогда очевидно, что ребро  $(x, y)$  такого же цвета.

Пусть на ребрах графа задано равномерное распределение, тогда  $H(A) = \log_2 l^n = n \cdot \log_2 l$  и  $H(A|X) = H(A|Y) = \log_2 \gamma^n = n \cdot \log_2 \gamma$ . В итоге получаем оценку:

$$bcc(G_{n,\Gamma}) \geq 2^{n \cdot \log_2 \gamma - \frac{n}{2} \cdot \log_2 l} = \left( \frac{\gamma}{\sqrt{l}} \right)^n$$

### 3.3 Сравнение методов оценивания

Рассмотрим какие оценки получаются на известных геометрических конфигурациях. Симметричные конфигурации ( $p = l$  и  $\gamma = \pi$ ) будем обозначать сокращенно  $(p_\gamma)$ .

| Название   | FS         | KJ                                      | EI  | Результат      |
|--|------------|---|---|----------------|
| Треугольник<br>(3 <sub>2</sub> )                                   | $\geq 3^n$ | $\left(\frac{3}{2}\right)^n$            | $\left(\frac{2}{\sqrt{3}}\right)^n$         | $FS > KJ > EI$ |
| Полный четырех-<br>сторонник<br>(4 <sub>3</sub> , 6 <sub>2</sub> ) | $\geq 3^n$ | $\left(\frac{4}{3}\right)^n$            | $\left(\frac{3}{\sqrt{6}}\right)^n$         | $FS > KJ > EI$ |
| $K_m$ при $m > 4$<br>( $m_{m-1}$ , $\binom{m}{2}_2$ )              | $\geq 3^n$ | $\left(\frac{m}{m-1}\right)^n$          | $\left(\sqrt{\frac{2(m-1)}{m}}\right)^n$    | $FS > EI > KJ$ |
| $K_{m,m}$ при $m > 0$<br>( $2m_m$ , $m_2^2$ )                      | $\geq 2^n$ | $2^n$                                   | $1^n$                                       | $FS = KJ > EI$ |
| Плоскость Фано<br>(7 <sub>3</sub> )                                | $\geq 3^n$ | $\left(\frac{7}{6}\right)^n$            | $\left(\frac{3}{\sqrt{7}}\right)^n$         | $FS > KJ > EI$ |
| Конфигурация<br>Мёбиуса-Кантора<br>(8 <sub>3</sub> )               | $\geq 3^n$ | $\left(\frac{4}{3}\right)^n$            | $\left(\frac{3}{\sqrt{8}}\right)^n$         | $FS > KJ > EI$ |
| Конфигурация<br>Дезарга (10 <sub>3</sub> )                         | $\geq 3^n$ | $\left(\frac{5}{3}\right)^n$            | $\left(\frac{3}{\sqrt{10}}\right)^n$        | $FS > KJ > EI$ |
| Конфигурация<br>Гессе (9 <sub>4</sub> , 12 <sub>3</sub> )          | $\geq 3^n$ | $\left(\frac{9}{8}\right)^n$            | $\left(\frac{2}{\sqrt{3}}\right)^n$         | $FS > EI > KJ$ |
| Конфигурация<br>Шлефли<br>(12 <sub>5</sub> , 30 <sub>2</sub> )     | $\geq 3^n$ | $\left(\frac{12}{5}\right)^n$           | $\left(\frac{5}{\sqrt{30}}\right)^n$        | $FS > KJ > EI$ |
| Проективная<br>плоскость<br>( $(m^2 + m + 1)_{m+1}$ )              | $\geq 3^n$ | $\left(\frac{m^2+m+1}{m(m+1)}\right)^n$ | $\left(\frac{m+1}{\sqrt{m^2+m+1}}\right)^n$ | $FS > EI > KJ$ |
| Конфигурация<br>Кокса ( $(2^{m-1})_m$ )                            | $\geq 3^n$ | $\left(\frac{2^{m-1}}{m(m-1)}\right)^n$ | $\left(\frac{m}{\sqrt{2^{m-1}}}\right)^n$   | $FS > KJ > EI$ |

Из таблицы видно, что метод трудного множества всегда работает лучше, чем остальные. В данном случае точная оценка по методу трудного множества может превосходить величину  $3^n$  на некоторых графах, в результате мы не можем доказать, что метод Куликова-Юкны работает всегда хуже. Но зато величины  $3^n$  достаточно для метода энтропийных неравенств, а значит можно сформулировать следующее утверждение:

**Утверждение.** Для произвольной геометрической конфигурации  $(p_\gamma, l_\pi)$  оценка, получаемая по методу трудного множества, превосходит оценку метода энтропийных неравенств. Иначе говоря

$$2 \geq \frac{\gamma}{\sqrt{l}}$$

**Доказательство.** Условия

$$\begin{cases} p \cdot \gamma = l \cdot \pi, \\ p \geq \gamma \cdot (\pi - 1) + 1. \end{cases}$$

должны обязательно выполняться для того, чтобы геометрическая конфигурация была корректно определена.

Используя эти ограничения, получаем

$$\frac{\gamma^2}{l} = \frac{\pi \cdot \gamma}{p} \leq \frac{p + \gamma - 1}{p} = 1 + \frac{\gamma - 1}{p} < 4$$

Что и требовалось доказать.  $\square$

Теперь давайте сравним метод Куликова-Юкны и метод энтропийных неравенств. Рассмотрим два случая:

- Пусть выполняется условие  $l \geq \gamma^2$ , тогда

$$\gamma^2 \cdot (\pi - 1) \leq l \cdot (\pi - 1) < p \cdot \gamma \leq p \cdot \sqrt{l}$$

То есть получаем, что  $KJ > EI$ .

- Пусть теперь верно  $l \leq \gamma^2$ , тогда

$$\gamma^2 \cdot (\pi - 1) \geq l \cdot (\pi - 1) = p \cdot \gamma - l \geq p \cdot \sqrt{l} - l$$

Поделив обе части на  $\sqrt{l} \cdot \gamma \cdot (\pi - 1)$ , получаем

$$\frac{\gamma}{\sqrt{l}} \geq \frac{p}{\gamma \cdot (\pi - 1)} - \frac{\sqrt{l}}{\gamma \cdot (\pi - 1)}$$

В итоге получаем, что с некоторой небольшой точностью  $EI \gtrsim KJ$

## Трудное множество vs Куликов-Юкна

В данном разделе мы докажем, что метод трудного множества всегда дает оценку лучше, чем метод Куликова-Юкны. Также рассмотрим



вопрос о величине разницы в получаемых оценках: всегда ли она невелика или может быть, что для некоторых графов оценки из этих методов будут отличаться очень сильно.

## 4.1 Теорема Турановского типа

Как уже видно из названия, для дальнейших изысканий нам потребуется классическая теорема Турана:

**Теорема (Туран).** Пусть дан неориентированный граф  $G = (V, E)$ , где  $|V| = n$  и число независимости равно  $\alpha$ . Тогда в графе выполняется следующая оценка

$$|E| \geq n \cdot \left\lfloor \frac{n}{\alpha} \right\rfloor - \alpha \cdot \frac{\left\lfloor \frac{n}{\alpha} \right\rfloor \cdot \left( \left\lfloor \frac{n}{\alpha} \right\rfloor + 1 \right)}{2}$$

Доказательство этой теоремы можно найти в книге [7]. Используя эту теорему, мы можем с легкостью доказать следующее:

**Теорема.** Пусть имеется неориентированный граф  $G = (V, E)$ , тогда среди ребер максимального паросочетания можно найти трудное множество размера

$$\left\lceil \frac{v(G)^2}{|E|} \right\rceil$$

**Доказательство.** Давайте вместо графа  $G = (V, E)$  рассмотрим граф  $\tilde{G} = (\tilde{V}, \tilde{E})$ , в котором останутся только вершины из максимального паросочетания. Так как  $|E| \geq |\tilde{E}|$ , то достаточно найти трудное множество размера

$$\left\lceil \frac{v(G)^2}{|\tilde{E}|} \right\rceil$$

Пусть  $(v_1, v'_1), (v_2, v'_2), \dots, (v_m, v'_m)$  – ребра максимального паросочетания. Построим граф  $\hat{G} = (\hat{V}, \hat{E})$  такой, что вершин в нем ровно  $m$ . Обозначим вершины  $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_m\}$ , причем  $\hat{v}_i \leftrightarrow (v_i, v'_i)$ . Определим множество ребер  $\hat{E}$  следующим образом

$$(\hat{v}_i, \hat{v}_j) \in \hat{E} \text{ если } (v_i, v'_i) \notin \tilde{E} \text{ или } (v_j, v'_j) \notin \tilde{E}$$

Очевидно, что трудное множество на ребрах максимального паросочетания соответствует клике в  $\widehat{G}$  такого же размера. Пусть число независимости дополнения графа  $\widehat{G}$  равно  $\alpha$ , тогда мы можем представить трудное множество размера  $\alpha$ . Используя теорему Турана для дополнения графа  $\widehat{G}$ , получаем

$$|\widetilde{E}| \geq m \cdot \left\lfloor \frac{m}{\alpha} \right\rfloor - \alpha \cdot \frac{\left\lfloor \frac{m}{\alpha} \right\rfloor \cdot (\left\lfloor \frac{m}{\alpha} \right\rfloor + 1)}{2} =$$

Пусть  $m = k \cdot \alpha + r$ , где  $r < \alpha$

$$= (k \cdot \alpha + r) \cdot k - \alpha \cdot \frac{k \cdot (k + 1)}{2} = \frac{\alpha \cdot k^2}{2} + r \cdot k - \frac{\alpha \cdot k}{2}$$

Так как каждое ребро из дополнения графа  $\widehat{G}$  порождает два ребра в  $\widetilde{G}$ , а также еще имеется  $m$  ребер самого паросочетания, то получаем

$$\begin{aligned} |\widetilde{E}| &\geq m + 2 \cdot \left( \frac{\alpha \cdot k^2}{2} + r \cdot k - \frac{\alpha \cdot k}{2} \right) = \\ &= \alpha \cdot k^2 + 2r \cdot k + r \geq \alpha \cdot k^2 + 2r \cdot k + \left\lceil \frac{r^2}{\alpha} \right\rceil = \left\lceil \frac{m^2}{\alpha} \right\rceil \end{aligned}$$

В итоге получили, что

$$|\widetilde{E}| \geq \left\lceil \frac{m^2}{\alpha} \right\rceil \iff \alpha \geq \left\lceil \frac{m^2}{|\widetilde{E}|} \right\rceil = \left\lceil \frac{v(G)^2}{|\widetilde{E}|} \right\rceil \blacksquare$$

Мы доказали, что на любом неориентированном графе точная оценка по методу трудного множества лучше, чем оценка Куликова-Юкны.

## 4.2 Конструкция прямоугольного графа

В доказательстве предыдущей теоремы мы использовали некоторый модифицированный граф  $\widetilde{G} = (\widetilde{V}, \widetilde{E})$ . Оказывается, можно рассмотреть более общую конструкцию. Такие графы мы будем называть прямоугольными графами.

**Определение.** Пусть имеется двудольный неориентированный граф  $G = (L, R, E)$ . Определим прямоугольный граф  $\widetilde{G} = (\widetilde{V}, \widetilde{E})$  следующим

образом:

- $e_{i,j} \in E \leftrightarrow v_{i,j} \in \tilde{V}$ , значит  $|E| = |\tilde{V}|$ .
- $(v_{i,j}, v_{k,l}) \in \tilde{E}$  тогда и только тогда, когда  $v_{i,l} \notin \tilde{E}$  или  $v_{k,j} \notin \tilde{E}$

Введем также понятия хроматического, кликового и антикликового чисел:

**Определение.** Хроматическое число графа  $G$  – минимальное число  $k$  такое, что множество вершин графа можно покрасить в  $k$  цветов, причем любое ребро графа соединяет разноцветные вершины. Обозначение  $\chi(G)$ .

**Определение.** Кликовое число графа  $G$  – максимальное число  $k$  такое, что в нашем графе содержится полный граф на  $k$  вершинах ( $k$ -клика). Обозначение  $w(G)$ .

**Определение.** Антикликовое число графа  $G$  – максимальное число  $k$  такое, что в графе дополнения содержится полный граф на  $k$  вершинах ( $k$ -антиклика). Обозначение  $\alpha(G)$ .

Используя конструкцию прямоугольного графа, мы можем сформулировать следующую теорему:

**Теорема.** Для любого двудольного графа  $G = (L, R, E)$  верно:

$$1) \text{ fool}(G) = w(\tilde{G})$$

$$2) \max_{K_{r,s} \subseteq G} \{r \cdot s\} = \alpha(\tilde{G})$$

$$3) \text{ bcc}(G) = \chi(\tilde{G})$$

**Доказательство.** Так как каждому трудному множеству размера  $k$  в  $G$  соответствует  $k$ -клика в  $\tilde{G}$  и наоборот, то  $\text{fool}(G) = w(\tilde{G})$ .

Очевидно, что биклике  $K_{r,s}$  в  $G$ , соответствует антиклика размера  $r \cdot s$  в  $\tilde{G}$ . Обратно, если  $(v_{i,j}, v_{k,l}) \notin \tilde{E}$ , то вершины  $v_{i,l}$  и  $v_{k,j}$  определены, и между ними нет ребра. И следовательно, если мы рассмотрим какую-нибудь антиклику в  $\tilde{G}$ , мы ее можем расширить до

"прямоугольной" антиклики, которой будет соответствовать биклика в  $G$ .

Последняя часть сразу следует из того, что все вершины антиклики мы можем красить в один цвет. Имея произвольное покрытие  $bcc(G)$ , мы получаем покрытие вершин графа  $\tilde{G}$  антикликами. Пусть каждой биклике из  $bcc(G)$  сопоставлен свой цвет. Красим вершину в тот цвет, который соответствует покрывающей ее биклике (если таких биклик несколько, то в любой из них). В итоге получаем правильную раскраску графа в  $bcc(G)$  цветов. Обратно, правильная покраска графа  $\tilde{G}$  порождает покрытие антикликами, которые мы расширяем до "прямоугольных". В итоге этим антикликам соответствуют биклики в  $G$ , следовательно мы получили покрытие бикликами (возможно пересекающимися) размера  $\chi(\tilde{G})$ . ■

Эта теорема позволяет переформулировать известные оценки для хроматического числа:

$$\chi(\tilde{G}) \geq w(\tilde{G}) \iff bcc(G) \geq fool(G)$$

и

$$\chi(\tilde{G}) \geq \left\lceil \max_{U \subseteq \tilde{V}} \frac{|U|}{\alpha(\tilde{G}(U))} \right\rceil \iff bcc(G) \geq \left\lceil \max_{\mathcal{E} \subseteq E} \frac{|\mathcal{E}|}{\max_{K_{r,s} \subseteq G(\mathcal{E})} |K_{r,s} \cap \mathcal{E}|} \right\rceil$$

где  $G(\mathcal{E})$  наименьший подграф  $G$ , содержащий все ребра  $\mathcal{E}$ .

Если в качестве  $\mathcal{E}$  рассмотреть максимальное паросочетание, тогда  $\max_{K_{r,s} \subseteq G(\mathcal{E})} |K_{r,s} \cap \mathcal{E}| = \max_{K_{r,r} \subseteq G(\mathcal{E})} |K_{r,r} \cap \mathcal{E}| = \max_{K_{r,r} \subseteq G(\mathcal{E})} \{r\} \leq \max_{K_{r,r} \subseteq G} \{r\} = bcl(G)$ . В итоге получаем оценку, которую мы уже раньше встречали:

$$bcc(G) \geq \left\lceil \max_{\mathcal{E} \subseteq E} \frac{|\mathcal{E}|}{\max_{K_{r,s} \subseteq G(\mathcal{E})} |K_{r,s} \cap \mathcal{E}|} \right\rceil \geq \left\lceil \frac{v(G)}{bcl(G)} \right\rceil$$

Если же в качестве  $\mathcal{E}$  взять вообще все ребра, то

$$bcc(G) \geq \left\lceil \max_{\mathcal{E} \subseteq E} \frac{|\mathcal{E}|}{\max_{K_{r,s} \subseteq G(\mathcal{E})} |K_{r,s} \cap \mathcal{E}|} \right\rceil \geq \left\lceil \frac{|E|}{\max_{K_{r,s} \subseteq G} \{r \cdot s\}} \right\rceil$$

### 4.3 Случайные графы Эрдеша-Реньи

В этом разделе мы хотим понять насколько различаются оценки из этих двух методов в применении к "типичным" и "неэкзотичным" графам. В качестве уточнения слова "типичности" мы рассмотрим случайные графы Эрдеша-Реньи при разумном выборе параметров.

Пусть у нас имеются два  $n$ -элементных множества  $L$  и  $R$ , элементы которого будем называть вершинами левой и правой долей графа. Понятно, что случайным будет множество ребер графа. Мы не хотим рассматривать графы с кратными ребрами (мультиграфы), графы с петлями (псевдографы) и ориентированные графы. Поэтому мы считаем, что потенциальных ребер у графа не больше, чем  $n^2$  штук. Будем соединять любые две вершины  $i \in L$  и  $j \in R$  ребром с некоторой вероятностью  $p \in [0, 1]$  независимо от всех остальных  $n^2 - 1$  пар вершин. Иными словами, ребра появляются в соответствии со стандартной схемой Бернулли, в которой  $n^2$  испытаний и "вероятность успеха"  $p$ . Обозначим через  $E$  случайное множество ребер, которое возникает в результате реализации такой схемы. Положим  $G = (L, R, E)$ . Это и есть случайный двудольный граф в модели Эрдеша – Реньи.

Если записывать приведенное только что определение в формате аксиоматики Колмогорова, то мы имеем вероятностное пространство

$$G(n, p) = (\Omega_n, \mathcal{F}_n, P_{n,p})$$

в котором

$$\Omega_n = \{G = (L, R, E)\}, \quad \mathcal{F}_n = 2^{\Omega_n}, \quad P_{n,p}(G) = p^{|E|} \cdot (1 - p)^{n^2 - |E|}$$

Если нам хочется найти вероятность, с которой двудольный граф на  $2n$  вершинах обладает данным свойством  $A$ , то мы просто берем множе-

ство  $\mathcal{A} \in \mathcal{F}_n$ , состоящее из всех графов, для которых выполнено свойство  $A$ , и вычисляем

$$P_{n,p}(\mathcal{A}) = \sum_{G \in \mathcal{A}} P_{n,p}(G)$$

Далее будем рассматривать не фиксированное  $p$ , а некоторую функцию  $p(n)$ , заключенную между нулем и единицей. Скажем, наконец, что свойство выполнено почти всегда, если его вероятность стремится к единице при  $n \rightarrow \infty$ .

## 4.4 Неравенство Хефдинга

Пусть  $\xi_1, \xi_2, \dots, \xi_n$  — последовательность независимых случайных величин таких, что для любого  $i = 1, 2, \dots, n$  верно  $\xi_i \in [a_i, b_i]$  с вероятностью 1 для некоторых  $a_i, b_i \in \mathbb{R}$ . Введем обозначение  $S_n = \sum_{i=1}^n \xi_i$ . Мы хотим изучать отклонение случайной величины  $S_n$  от ее среднего значения  $\mathbb{E}[S_n]$ . Иначе говоря, получить неравенство концентрации для  $\xi = S_n - \mathbb{E}[S_n]$ . Воспользовавшись для этого неравенством Чернова получим, что для любого  $\lambda > 0$  верно

$$\begin{aligned} P\{S_n - \mathbb{E}[S_n] \geq \varepsilon\} &= P\{e^{\lambda(S_n - \mathbb{E}[S_n])} \geq e^{\lambda\varepsilon}\} \leq \frac{\mathbb{E}[e^{\lambda(S_n - \mathbb{E}[S_n])}]}{e^{\lambda\varepsilon}} = \\ &= \frac{\mathbb{E}[e^{\lambda \sum_{i=1}^n (\xi_i - \mathbb{E}[\xi_i])}]}{e^{\lambda\varepsilon}} = \frac{\mathbb{E}[\prod_{i=1}^n e^{\lambda(\xi_i - \mathbb{E}[\xi_i])}]}{e^{\lambda\varepsilon}} = \frac{\prod_{i=1}^n \mathbb{E}[e^{\lambda(\xi_i - \mathbb{E}[\xi_i])}]}{e^{\lambda\varepsilon}} \end{aligned}$$

Нам остается построить верхние оценки для производящих функций  $\varphi_{\xi_i}(\lambda)$ . Следующий результат дает нам такие оценки в тех случаях, когда случайные величины  $\xi_i$  принимают значения из ограниченных интервалов.

**Лемма** (Хефдинг). *Для произвольной случайной величины  $\xi$  такой, что  $\mathbb{E}[\xi] = 0$  и  $\xi \in [a, b]$  с вероятностью 1 для любого  $\lambda > 0$  справедливо*

$$\mathbb{E}[e^{\lambda\xi}] \leq e^{\frac{\lambda^2(b-a)^2}{8}}$$

Доказательство основано на выпуклости экспоненты.

Применив эту лемму к нашей цепочке неравенств для случайных величин  $\xi_i - \mathbb{E}[\xi_i]$ , которые почти наверное лежат в интервалах  $[a_i - \mathbb{E}[\xi_i], b_i - \mathbb{E}[\xi_i]]$ , мы получаем

$$P\{S_n - \mathbb{E}[S_n] \geq \varepsilon\} \leq \frac{\prod_{i=1}^n \mathbb{E}[e^{\lambda(\xi_i - \mathbb{E}[\xi_i])}]}{e^{\lambda\varepsilon}} \leq \frac{\prod_{i=1}^n e^{\lambda^2(b_i - a_i)^2/8}}{e^{\lambda\varepsilon}} = \frac{e^{\lambda^2 \sum_{i=1}^n (b_i - a_i)^2/8}}{e^{\lambda\varepsilon}}$$

Остается минимизировать правую часть по  $\lambda \geq 0$ . Выбирая

$$\lambda = \frac{4\varepsilon}{\sum_{i=1}^n (b_i - a_i)^2}$$

мы получаем следующий результат

**Теорема** (Неравенство Хефдинга). Пусть  $\xi_1, \xi_2, \dots, \xi_n$  — последовательность независимых случайных величин, таких что для любого  $i = 1, 2, \dots, n$  верно  $\xi_i \in [a_i, b_i]$  с вероятностью 1 для некоторых  $a_i, b_i \in \mathbb{R}$ . Тогда для любого  $\varepsilon > 0$  верно

$$P\{S_n - \mathbb{E}[S_n] \geq \varepsilon\} \leq \exp\left(\frac{-2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Аналогичное неравенство верно для  $P\{\mathbb{E}[S_n] - S_n \geq \varepsilon\}$ , поскольку условия теоремы инвариантны относительно замены знака слагаемых. Применив дважды неравенство Хефдинга, мы получаем

$$\begin{aligned} P\{|S_n - \mathbb{E}[S_n]| \geq \varepsilon\} &\leq P\{S_n - \mathbb{E}[S_n] \geq \varepsilon\} + P\{\mathbb{E}[S_n] - S_n \geq \varepsilon\} \leq \\ &\leq 2 \cdot \exp\left(\frac{-2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \end{aligned}$$

Воспользуемся этим неравенством для того, чтобы изучить отклонение числа ребер в случайном двудольном графе Эрдеша-Реньи. Определим индикаторные случайные величины  $X_{i,j} = I\{e_{i,j} \in E\}$ . Так как

случайная величина  $|E| = \sum_{i,j} X_{i,j}$ , то получаем

$$\mathbb{E}[|E|] = \sum_{i,j} \mathbb{E}[X_{i,j}] = \sum_{i,j} P\{e_{i,j} \in E\} = n^2 \cdot p$$

Наконец, воспользуемся неравенством Хефдинга для  $\varepsilon = 1$

$$P\{|E| \neq \mathbb{E}[|E|]\} = P\{||E| - \mathbb{E}[|E|]| \geq 1\} \leq 2 \cdot e^{-\frac{2}{n^2}}$$

В итоге мы получили, что почти наверное (при  $n \rightarrow \infty$ ) число ребер в графе равно  $|E| = n^2 \cdot p$ .

## 4.5 Размер максимального паросочетания

Для изучения отклонения величины максимального паросочетания нам потребуется теорема Холла.

**Теорема (Холл).** Пусть имеется неориентированный двудольный граф  $G = (L, R, E)$ . Для произвольного  $A \subseteq L$  определим множество соседей

$$N(A) = \{y \in R \mid (x, y) \in E, x \in A\}$$

В двудольном графе существует совершенное паросочетание тогда и только тогда, когда для любого  $A \subseteq L$  выполнено  $|A| \leq |N(A)|$ .

Пусть имеется двудольный граф Эрдеша-Реньи  $G = (L, R, E)$ , где  $|L| = |R| = n$ . Множество  $S \subseteq L$  не удовлетворяет условию теоремы Холла, если существует множество  $T \subseteq R$  такое, что  $|S| + |T| = n + 1$  и  $N(S) \cap T = \emptyset$  (нет ребер между множествами  $S$  и  $T$ ).

Очевидно, что

$$P\{N(S) \cap T = \emptyset\} = (1 - p)^{|S| \cdot |T|}$$

тогда

$$P\{\text{нет совершенного паросочетания}\} \leq \sum_S \sum_T (1 - p)^{|S| \cdot |T|} =$$



$$\begin{aligned}
&= \sum_{k=1}^n \binom{n}{k} \binom{n}{n-k+1} (1-p)^{k(n-k+1)} \leq \sum_{k=1}^{(n+1)/2} \binom{n}{k} \binom{n}{k-1} (1-p)^{kn/2} \leq \\
&\leq \sum_{k=1}^{(n+1)/2} n^{2k} (1-p)^{kn/2}
\end{aligned}$$

Если предположить, что  $p = p(n) = n^{-\alpha}$  при  $\alpha < 1$ , то получаем

$$\begin{aligned}
P\{\text{нет совершенного паросочетания}\} &\leq \sum_{k=1}^{(n+1)/2} n^{2k} (1-p)^{kn/2} = \\
&= \sum_{k=1}^{(n+1)/2} n^{2k} e^{-\frac{kn}{2} \cdot n^{-\alpha} + O(n^{1-2\alpha}) \cdot k} = \sum_{k=1}^{(n+1)/2} \left( n^2 e^{-\frac{1}{2} n^{1-\alpha} + O(n^{1-2\alpha})} \right)^k \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

Последнее утверждение верно потому, что с некоторого момента величина стоящая под степенью будет меньше 1, а значит первое слагаемое будет больше всех остальных

$$\sum_{k=1}^{(n+1)/2} \left( n^2 e^{-\frac{1}{2} n^{1-\alpha} + O(n^{1-2\alpha})} \right)^k \leq \frac{n+1}{2} \left( n^2 e^{-\frac{1}{2} n^{1-\alpha} + O(n^{1-2\alpha})} \right) \xrightarrow{n \rightarrow \infty} 0$$

В итоге мы получили, что почти наверное (при  $n \rightarrow \infty$ ) в нашем графе будет совершенное паросочетание.

## 4.6 Трудное множество и Куликов-Юкна на случайных графах

Если предположить, что  $p = p(n) = n^{-\alpha}$  ( $\alpha < 1$ ), то можно сделать вывод, что оценка Куликова-Юкны почти наверное (при  $n \rightarrow \infty$ ) равна

$$\left\lceil \frac{v(G)^2}{|E|} \right\rceil = \left\lceil \frac{n^2}{n^2 p} \right\rceil = \lceil n^\alpha \rceil$$

причем

$$P \left\{ \left\lceil \frac{v(G)^2}{|E|} \right\rceil \neq \lceil n^\alpha \rceil \right\} \leq P\{|E| \neq \mathbb{E}[|E|]\} + P\{v(G) \neq n\} \leq$$

$$\leq 2 \cdot e^{-\frac{2}{n^2}} + \sum_{k=1}^{(n+1)/2} \left( n^2 e^{-\frac{1}{2}n^{1-\alpha} + O(n^{1-2\alpha})} \right)^k \xrightarrow{n \rightarrow \infty} 0$$

Теперь посчитаем вероятность того, что в случайном графе найдется трудное множество размера  $k$ . Обозначим  $f_k(G)$  - число различных трудных множеств размера  $k$  в графе  $G$ . Нас интересует вероятность  $P\{|f_k(G)| > 0\}$ , которая превосходит вероятность того, что фиксированные  $k$  пар вершин образуют трудное множество. Иначе говоря

$$P\{|f_k(G)| > 0\} \geq p^k (1 - p^2)^{\binom{k}{2}}$$

Предположим теперь, что  $p = p(n) = n^{-\alpha}$  и  $k = k(n) = n^\beta$ , тогда

$$P\{|f_k(G)| > 0\} \geq n^{-\alpha k} e^{\binom{k}{2} \cdot (-n^{-2\alpha} + O(n^{-4\alpha}))} = n^{-\alpha n^\beta} e^{-\frac{1}{2}n^{2\beta-2\alpha} + \frac{1}{2}n^{\beta-2\alpha} + O(n^{2\beta-4\alpha})}$$

Заметим, что если мы докажем, что при  $n \rightarrow \infty$

$$P\{|f_k(G)| > 0\} > P\left\{\left\lceil \frac{v(G)^2}{|E|} \right\rceil \neq \lceil n^\alpha \rceil\right\}$$

то получим, что существует граф, у которого имеется трудное множество размера  $n^\beta$  и оценка Куликова-Юкны равна  $n^\alpha$ . Чтобы это было верно, достаточно показать, что

$$n^{-\alpha n^\beta} e^{-\frac{1}{2}n^{2\beta-2\alpha} + \frac{1}{2}n^{\beta-2\alpha} + O(n^{2\beta-4\alpha})} > 2 \cdot e^{-\frac{2}{n^2}} + \sum_{i=1}^{(n+1)/2} \left( n^2 e^{-\frac{1}{2}n^{1-\alpha} + O(n^{1-2\alpha})} \right)^i$$

Сравним левую часть с каждым слагаемым из правой части по отдельности:

1) Если  $2 > \beta > \alpha$  и  $2 > 2\beta - 2\alpha$ , то

$$-\alpha \cdot \ln n \cdot n^\beta - \frac{1}{2} \cdot n^{2\beta-2\alpha} \gg -2 \ln 2 \cdot n^{-2}$$

2) Поделим левую часть на  $n$  и сравним с первым членом суммы, заранее

прологарифмировав. При  $1 - \alpha > \beta > \alpha$  и  $1 + \alpha > 2\beta$  верно

$$-\alpha \cdot \ln n \cdot n^\beta - 1 - \frac{1}{2} \cdot n^{2\beta-2\alpha} \gg 2 \ln n - \frac{1}{2} n^{1-\alpha}$$

что верно, так как

$$n^{1-\alpha} \gg \ln n, \quad n^{1-\alpha} \gg \ln n \cdot n^\beta, \quad n^{1-\alpha} \gg n^{2\beta-2\alpha}$$

В итоге получаем такое утверждение

**Утверждение.** Если выполняется условие  $\min \left\{ \frac{1+\alpha}{2}, 1-\alpha \right\} > \beta > \alpha$ , тогда существует двудольный граф  $G = (L, R, E)$  ( $|L| = |R| = n$ ), в котором метод трудного множества дает оценку  $n^\beta$ , а метод Куликова-Юкны —  $n^\alpha$

К сожалению, мы смогли показать лишь существование такого графа, но не смогли доказать, что это верно для почти всех графов. Чтобы преодолеть эту трудность, нужно исследовать отклонение числа трудных множеств размера  $k$ .

## 4.7 Количество трудных множеств размера $k$

Вспомним, что  $f_k(G)$  — число трудных множеств размера  $k$  в графе  $G$ . Давайте оценим матожидание этой случайной величины

$$\mathbb{E}[f_k(G)] = \binom{n}{k}^2 \cdot k! \cdot \mathbb{E}[I_k(G)] = \binom{n}{k}^2 \cdot k! \cdot p^k (1-p^2)^{\binom{k}{2}}$$

**Лемма.** Если  $k = o(\sqrt{n})$ , то  $\binom{n}{k} \sim \frac{n^k}{k!}$ , к тому же, если  $k \rightarrow \infty$  при  $n \rightarrow \infty$ , то  $\binom{n}{k} \sim n^k k^{-k-\frac{1}{2}} e^k$ .

**Доказательство.** Применим неравенство  $\ln(1-x) < -x$

$$\begin{aligned} \binom{n}{k} &= \frac{n^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) = \\ &= \frac{n^k}{k!} e^{\ln\left(1-\frac{1}{n}\right) + \ln\left(1-\frac{2}{n}\right) + \cdots + \ln\left(1-\frac{k-1}{n}\right)} < \frac{n^k}{k!} e^{-\frac{1}{n} - \frac{2}{n} - \cdots - \frac{k-1}{n}} = \end{aligned}$$

$$= \frac{n^k}{k!} e^{-\frac{k(k-1)}{2n}} = \frac{n^k}{k!} e^{-O\left(\frac{k^2}{n}\right)}$$

Если использовать неравенство  $\ln(1-x) > -x - \frac{1}{2}x^2$ , то получаем

$$\begin{aligned} \binom{n}{k} &> \frac{n^k}{k!} e^{-\frac{1}{n} - \frac{1}{2} \cdot \frac{1}{n^2} - \frac{2}{n} - \frac{1}{2} \cdot \frac{2^2}{n^2} - \dots - \frac{1}{n} - \frac{k-1}{2} \cdot \frac{(k-1)^2}{n^2}} = \\ &= \frac{n^k}{k!} e^{-\frac{k(k-1)}{2n} - \frac{1}{2n} \sum_{i < k} i^2} > \frac{n^k}{k!} e^{-\frac{k^2}{2n} - O\left(\frac{k^3}{n^2}\right)} \end{aligned}$$

В итоге мы получили, что

$$\frac{n^k}{k!} e^{-\frac{k^2}{2n} - O\left(\frac{k^3}{n^2}\right)} < \binom{n}{k} < \frac{n^k}{k!} e^{-O\left(\frac{k^2}{n}\right)}$$

При  $k = o(\sqrt{n})$  мы получаем  $\binom{n}{k} \sim \frac{n^k}{k!}$ . Применяя формулу Стирлинга к  $k!$ , получаем второе утверждение леммы.  $\square$

Эта лемма позволяет найти точный порядок величины  $\mathbb{E}[f_k(G)]$  в предположении, что  $p = n^\alpha$ .

$$\begin{aligned} \mathbb{E}[I_k(G)] &= p^k (1-p^2)^{\binom{k}{2}} = n^{-\alpha k} e^{\binom{k}{2} \cdot \ln(1-\frac{1}{n^{2\alpha}})} = \\ &= n^{-\alpha k} e^{-\binom{k}{2} \cdot \frac{1}{n^{2\alpha}} + \binom{k}{2} \cdot \frac{1}{2n^{4\alpha}} + O\left(\frac{k^2}{n^{6\alpha}}\right)} = n^{-\alpha k} e^{-\frac{k^2}{2n^{2\alpha}} + \frac{k^2}{4n^{4\alpha}} + O\left(\frac{k^2}{n^{2\alpha}}\right)} \end{aligned}$$

Если предположить, что  $k = n^{2\alpha+\varepsilon}$ , то

$$\begin{aligned} \mathbb{E}[f_k(G)] &\sim n^{2n^{2\alpha+\varepsilon} - (2\alpha+\varepsilon)(n^{2\alpha+\varepsilon} + \frac{1}{2})} \cdot e^{n^{2\alpha+\varepsilon}} \cdot n^{-\alpha n^{2\alpha+\varepsilon}} \cdot e^{-\frac{1}{2}n^{2\alpha+2\varepsilon} + O(n^{2\varepsilon})} = \\ &= n^{n^{2\alpha+\varepsilon}(2-3\alpha-\varepsilon) - \alpha - \frac{\varepsilon}{2}} \cdot e^{-\frac{1}{2}n^{2\alpha+2\varepsilon} + O(n^{2\alpha+\varepsilon})} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

А если  $k = n^{2\alpha-\varepsilon}$ , то

$$\begin{aligned} \mathbb{E}[f_k(G)] &\sim n^{2n^{2\alpha-\varepsilon} - (2\alpha-\varepsilon)(n^{2\alpha-\varepsilon} + \frac{1}{2})} \cdot e^{n^{2\alpha-\varepsilon}} \cdot n^{-\alpha n^{2\alpha-\varepsilon}} \cdot e^{-\frac{1}{2}n^{2\alpha-2\varepsilon} + O(n^{-\varepsilon})} = \\ &= n^{n^{2\alpha-\varepsilon}(2-3\alpha+\varepsilon) - \alpha + \frac{\varepsilon}{2}} \cdot e^{n^{2\alpha-\varepsilon} + O(n^{2\alpha-2\varepsilon})} \xrightarrow{n \rightarrow +\infty} +\infty \end{aligned}$$

В итоге мы доказали следующую теорему:

**Теорема.** Пусть  $p = n^{-\alpha}$ , тогда

$$\bullet \quad k = n^{2\alpha-\varepsilon} \implies \mathbb{E}[f_k(G)] \xrightarrow{n \rightarrow +\infty} +\infty$$

$$\bullet k = n^{2\alpha+\varepsilon} \implies \mathbb{E}[f_k(G)] \xrightarrow[n \rightarrow +\infty]{} 0$$

## Список литературы

- [1] Kushilevitz Eyal, Nisan Noam. Communication Complexity. Cambridge University press, 2006.
- [2] Razborov Alexander. Communication Complexity. In: An Invitation to Mathematics: from Competitions to Research. Springer, 2011.
- [3] Gruber H., Holzer M. Finding lower bounds for nondeterministic state complexity is hard. Springer, 2006.
- [4] D. Caen D. A. Gregory, Pullman N. J. The Boolean rank of zero-one matrices. Department of Mathematics, University of the West Indies, 1981.
- [5] Jukna S., Kulikov A. S. On covering graphs by complete bipartite subgraphs. Discrete Math, 2009.
- [6] Kaced Tarik, Romashchenko A. E., Vereshchagin N. K. Conditional Information Inequalities and Combinatorial Applications. CoRR, 2015.
- [7] Оре Ойстин. Теория графов. Наука, 1968.