Perfect Matching

$x_1 \ldots x_n$

$h: x_i \to i$

"Статическая хеш-таблица".

m ячеек

n ключей

Каждый ключ попадает в случ. ячейку.

Утв 1. Пусть $m = n^2$.

Хешируем каждый ключ ф-ней $h \in \mathcal{H}$, где $\mathcal{H}$ – универс. сем-во.

$\Rightarrow$ вероятность коллизии $\le 1/2$

Д-во! E[Количество пар коллизий] $=$

$$= \sum_{1 \le i < j \le n} P[h(i) = h(j)] \le$$

$$\le \sum_{1 \le i < j \le n} \frac{1}{m} = \binom{n}{2} \frac{1}{m} =$$

$$= \frac{n \cdot (n-1)}{2} \cdot \frac{1}{n^2} < \frac{1}{2}$$

P[хотя бы одна пара поколлизилась] $\le$

$\le E$[числа коллизий].

<u>Следствие:</u> $\exists h \in \mathcal{H}$: хеширование без коллизий.

<u>Второе сл-вие</u>: Такую h можно эффективно построить рандом. алгоритм.

Недостаток: нужно n^2 ячеек на n ключей.

<u>Утв 2.</u> Пусть $m=n$.

Каждый ключ k положим в ячейку $h(k)$.

где $h \in \mathcal{H}$ выбрана случайно.

Пусть X_i — кол-во ключей в ячейке i

Тогда $E[\sum X_i^2] \leq 2n-1$

<u>Д-во:</u> $E[\sum X_i^2] = E[\sum(X_i^2 - X_i)] +$

$$+ \underbrace{E[\sum X_i]}_{n} =$$

ключ #i

$$E[\sum X_i(X_i-1)] \overset{?}{=} 2 \sum_{1 \leq i < j \leq n} P[h(i)=h(j)]$$

(*)

$$\leq 2\cdot\binom{n}{2}\cdot\frac{1}{n} = n-1$$

(*): И левая сторона и правая ст. это мат. ожид. кол-во пар (i, j) где $i\neq j$ и i и j в одной ячейке

$$E\left[\sum x_i^2\right] = E\left[\sum x_i(x_i-1)\right] + E\left[\sum x_i\right] \leq 2n-1$$

$\square$.

Следствие 1: $\exists h: \sum x_i^2 \leq 2n-1$.

Следствие 2: Сущ. эффективный алгоритм по поиску $h\in\mathcal{H}$, что $\sum x_i^2 \leq 3n$.

$$P_h\left[\sum x_i^2 \leq 3n\right] = 1 - \underbrace{P\left[\sum x_i^2 > 3n\right]}_{\leq 2/3 \text{ по нер-ву Маркова}} \geq 1/3.$$

<u>Нер-во Маркова</u>: Пусть $X \geq 0$ — случ. величина. Тогда $P[X\geq \lambda E[X]] \leq \frac{1}{\lambda}$.

Д-во:

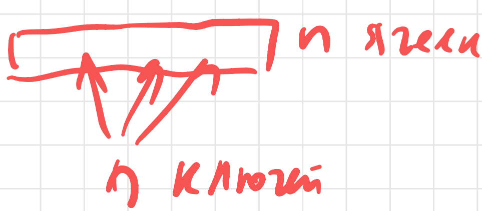
$$E[X] = P[X \geq \lambda E[X]] \cdot E[X \mid X \geq \lambda E[X]] + P[X < \lambda E[X]] \cdot E[X \mid X < \lambda E[X]]$$

$$\geq P[X \geq \lambda E[X]] \cdot E[X \mid X \geq \lambda E[X]]$$

$$\geq P[X \geq \lambda E[X]] \cdot \lambda E[X].$$

$$P[X \geq \lambda E[X]] \leq \frac{1}{\lambda}$$

Следствие 2: Сущ. эффективный алгоритм по поиску $h \in \mathcal{H}$, что $\sum x_i^2 \leq 3n$.



Двух-уровневая схема.

1) n ключей распределим по ячейкам

запомнить хорошую $h \in \mathcal{H}$.

2	3	0

2) в каждой ячейке построим хеш-таблицу размера x_i^2.

используя утв 1.

Запомнить хорошую h_i для хеширования ключей в ячейке i.

Итоговое потребл. памяти $= \Theta(n)$

т.к. $\sum x_i^2 \leqslant 3n$.

(см следствие 2)

Парадокс Д.Р.

P[коллизии] = ? (функция от n и m).

Утв. 3 $n = m^{0.5}$ $P[\text{коллизии}] \leq \frac{1}{2}$

(Д-во в рамках утв. 1)

Утв. 4 $n = m^{0.5} + 1$ $P[\text{коллизии}] \geq 0.39$.

<u>Утв 5:</u> $n = m^{\alpha}$ $\alpha > 1/2$

то $P[\text{коллизии}] \xrightarrow[m\to\infty]{} 1$

<u>Утв 6:</u> $n = m^{\alpha}$ $\alpha < 1/2$

$P[\text{коллизии}] \to 0.$

Док-во (утв 4):

$$P[\text{нет коллизии}] = \frac{m}{m} \cdot \frac{m-1}{m} \cdot \ldots \frac{m-n+1}{m}$$

$$= \prod_{i=0}^{n-1} \left(1 - \frac{i}{m}\right) \boxed{\le}$$

Лм: $1 - x \le e^{-x}$; $e^{-x} = 1 - x + \frac{x^2}{2} + \ldots$

$$\le \prod_{i=0}^{n-1} \exp\left(-\frac{i}{m}\right) =$$

$$= \exp\left(\sum_{i=0}^{n-1} -\frac{i}{m}\right) = \exp\left(-\frac{n(n-1)}{2m}\right)$$

$$P[\text{коллизия}] = 1 - P[\text{нет коллизий}]$$

$$\geq 1 - \exp\left(-\frac{n(n-1)}{2m}\right)$$

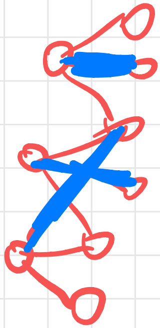
$n = \sqrt{m} + 1$:

$$P[\text{коллизия}] \geq 1 - \exp\left(-\frac{m + \sqrt{m}}{2m}\right)$$

$$= 1 - \exp\left(-\frac{1}{2} - \frac{1}{2\sqrt{m}}\right)$$

$$\geq 1 - \exp\left(-\frac{1}{2}\right) \geq 0.39$$

Паросочетание



$G = (V, E)$

M - паросочетание, если

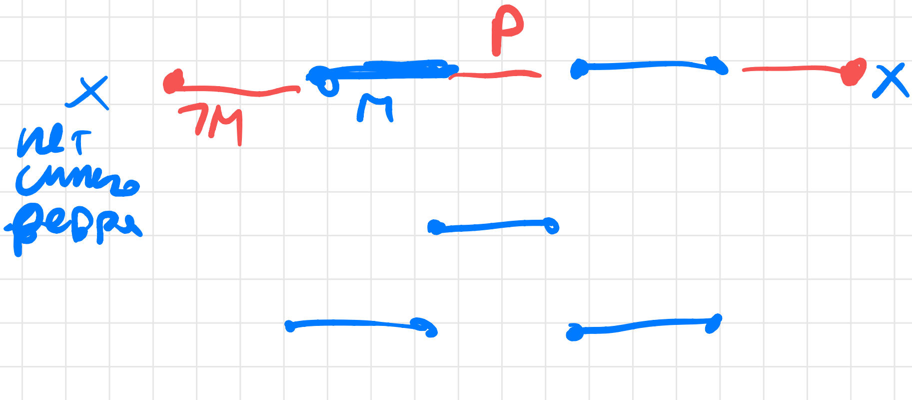
каждая вершина покрыта M

не более 1 раза

Задача: Найти наибольшее паросоч.

(Двудольный граф)

Def: Удлиняющая цепочка:



обе крайние вершины в P не покрыты M.

Наблюдение: $M \Delta P$ — паросочетание и $|M \Delta P| = |M| + 1$

Lem: M — максимально $\Leftrightarrow$ $\nexists$ удл. цепочки для M

Д-во: "$\Rightarrow$" - очевидно

"$\Leftarrow$"

Пусть от противного M не макс.

т.е. $\exists$ паросочетание M': $|M'| > |M|$.

в $M' \Delta M$ степень каждой в-ны ≤ 2.

P: 

т.к. $|M'| > |M|$, то $\exists$ путь P в $M' \Delta M$, такой что в нём M'-рёбер больше чем M.

P – удл. цепочка!

противоречие с предп. что удл. цепочек нет $\square$

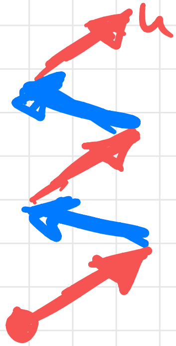
"Алгоритм" поиска макс паросоч.

$M = \varnothing$

while $\exists P$ – удл. цепочка:

$M = M \Delta P$

вопрос: как искать удлиняющую цепочку



— : не паросочетание

— паросочетание

v — непокрытая в-на левой доли

и u — непокрытая в-на правой дол.

$\exists$ увл. цепочка $(\Leftrightarrow)$

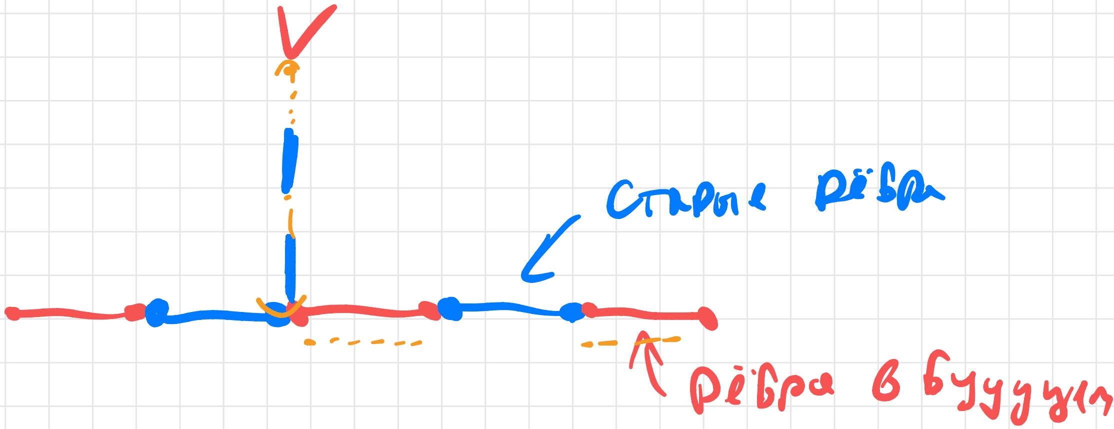
в ориент. графе $(\rightarrow, \leftarrow)$ есть путь из $v \in L$ в $u \in R$, v и u не покрыты.

(DFS)

```
M = ∅.
while modified:
    modified = false
    for v ∈ L
        if v не покрыта M
            вычислить по M ориент. рёбер
            dfs(v)
            if посетили u ∈ R, u — непокрыта.
                восстановить путь P: v → u
                M = M △ P.
                modified = true.
```

Утв: Внешний цикл не нужен.

(Если уже один раз поискали удл путь из $v \in L$, то в будущем нет смысла искать ещё раз)



for $v \in L$

вычислить по M ориент. рёбер

$dfs(v)$

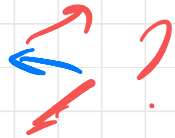
if посетили $u \in R$, u — непокрыта.

Восстановить путь $P: v \to u$

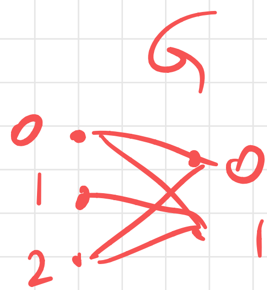
$M = M \triangle P$

Алгоритм "Куна":


```
match = [-1 for v in 0..R-1]

ans = 0

for v = 0..L-1:
```




```
    used = [false for u in 0..L-1]

    if dfs(v, used):
        ans += 1
```



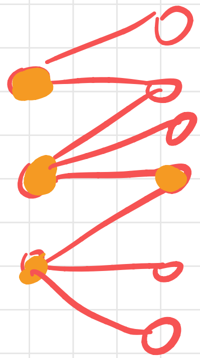
— blank —

adj[0]: [0,1]
[1] = [1]
[2]: [0,1]

вершина левой доли

```
dfs(v, used):
    if used[v]:
        return false
    used[v] = true

    for u ∈ adj[v]:
        if match[u] == -1:
            match[u] = v
            return true

        if dfs(match[u], used):
            match[u] = v
            return true

return false
```

Теорема König's



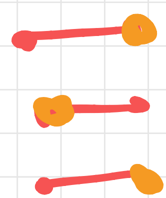
$C \subseteq V(G)$ — это vertex cover, если

C покрывает каждое ребро

$\forall e \in E(G) \quad e \cap C \neq \emptyset$.

- $\forall C$ – VC. $\forall M$ – паросочетание

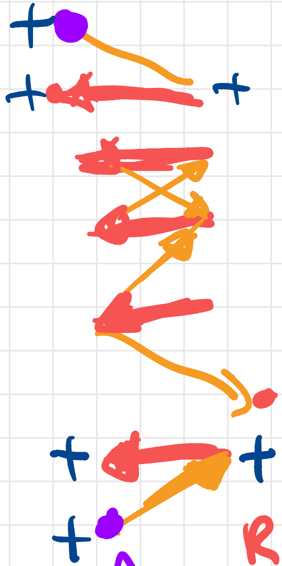
$$|C| \geq |M|$$



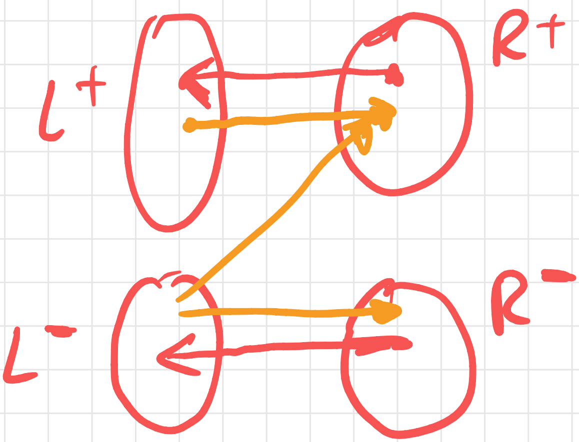
$$\min VC \geq \max M$$

Теорема (König's, 1931):

$$\min VC = \max M$$



L dfs из вершин L, не покрытых M



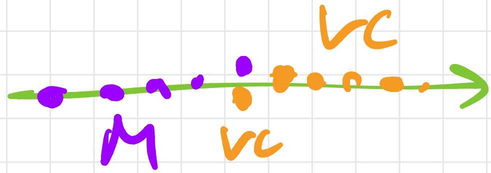
Lm 1. $L^- \cup R^+$ – vertex cover

Lm 2. $|L^- \cup R^+| = |M|$

Следств: $L^- \cup R^+$ — мин. vc

$\forall\, vc: \quad |vc| \geq |M|$

$\forall c$



Lm 3: $\forall\, C$ - vc $\quad \overline{C}$ — нез. мн-во

$\Rightarrow \quad \overline{L^- \cup R^+} = L^+ \cup R^-$ — макс. нез. мн-во