

Конспекты по алгоритмам и структурам данных

Анатолий Коченюк, Георгий Каданцев, Константин Бац

2022 год, семестр 4

1 Паросочетания и потоки

Мы начнём с паросочетаний и потоков.

1.1 Паросочетания

Паросочетание графа — это набор его ребер, не имеющих общих вершин.

1.1.1 Паросочетание в двудольном графе

Паросочетания в двудольных графах ищутся гораздо проще, чем в произвольных. Алгоритмы решают эту задачу понемногу. Берется маленькое (пустое) паросочетание и расширяется с помощью дополняющих цепочек.

Дополняющая цепочка — это путь, который начинается и заканчивается в вершинах вне паросочетания, и ребра в нём чередуются (принадлежит - не принадлежит пар. сочетанию). Из такой цепочки можно получить паросочетание большего размера (на 1).

Алгоритм начинает строить цепочки и удаляет дополняющие цепочки такой заменой.

Теорема 1. В графе нет дополняющего пути тогда и только тогда, когда паросочетание максимально.

Это утверждение сформулировано для произвольных графов: для двудольных и не очень. А в чём проблема? Проблема в поиске дополняющих путей.

В двудольном графе это делается просто. Это алгоритм Куна.

Двудольный граф можно хранить не как нормальный граф. "Алгоритм КУНА. Это не связано с аниме никак!" "Код, на самом деле, за пять копеек." Асимптотика: $O(nm)$.

Алгоритм никогда не освобождает вершины — если вершина попала в пар. соч., она там останется, можно dfs из неё не запускать. Чуть сложнее: если dfs однажды не нашел доп. пути из одной вершины, он никогда его не найдёт — можно его не запускать.

Немного о полных сочетаниях.

Теорема 2. В двудольном графе $G_{n,n}$ существует полное паросочетание в том и только том случае, когда для любого подмножества A вершин одной доли выполнено $|N(A)| \geq |A|$.

Паросочетания позволяют решать кучу прикольных задач.

1.1.2 Вершинное покрытие

Вершинное покрытие в графе — это набор вершин, таких, что никакие две вершины не лежат на одном ребре. Это понятие двойственное понятию паросочетания. Мы ищем минимальное по мощности вершинное покрытие. Это NP -полная задача в произвольном графе. В двудольном, однако, всё очень мило.

Размер покрытия в графе всегда не превосходит размер вершинного покрытия: $M \leq S$.

1.1.3 Парасочетание в недвудольном графе

Так же понемногу будем искать дополняющие пути и достраивать наше паросочетание.

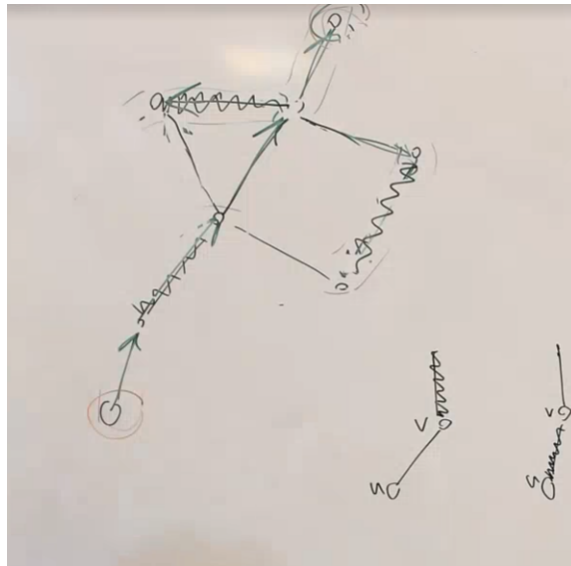
Как уже было доказано, M — макс парсоч \iff не. т дополняющих путей.

Как искать дополняющие пути? Может быть заюзать dfs?

```
1  dfs(v, state)
2      ...
3  for n
4      if state:
5          (u,v) ∈ M
6      else:
7          (u,v) ∈ M
8          dfs(n, !state)
```

Но так нельзя!

На таком графе не работает:

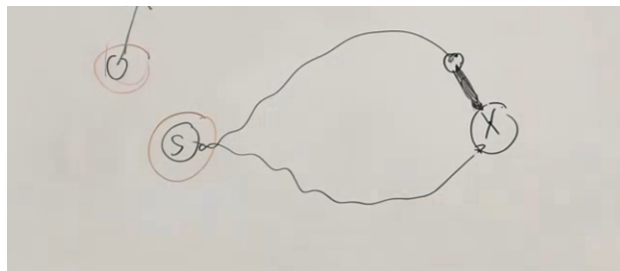


Проблема в том, что в какую-то вершину мы можем прийти с неправильной четностью.

Утверждение: если такого не случилось, то все будет ок.

Пусть есть стартовая вершина. Назовем вершину сомнительной, если до нее есть четный и нечетный пути.

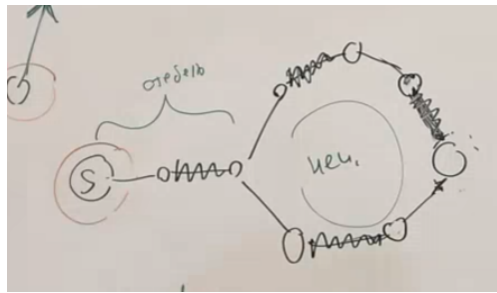
Если в графе нет сомнительных вершин, такой граф можно переделать в двудольный и легко в нем постить паросочетание.



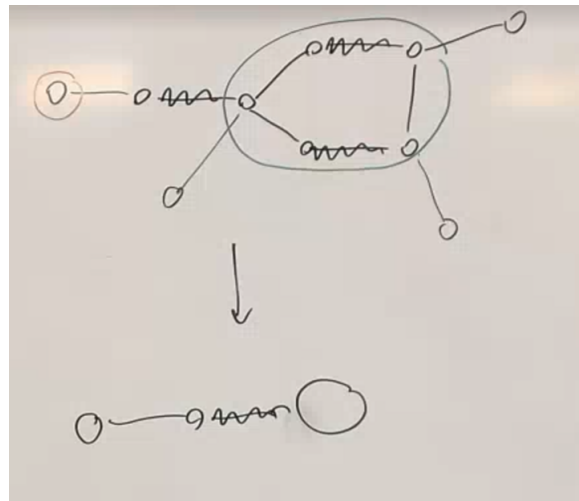
Что делать? Давайте запустим такой dfs. Может быть три случая:

1. Нашли дополняющий путь;
2. Не нашли дополняющий путь и нет сомнительных вершин \implies паросочетание максимальное;
3. Нашли сомнительную вершину.

Если случилось третье, будем веселиться. При помощи того же dfs-а найдем соцветие (blossom). Внутри соцветия цикл нечетной длины.



Алгоритм называется blossom cut. Возьмем все вершины соцветия и заменим на большую вершину.

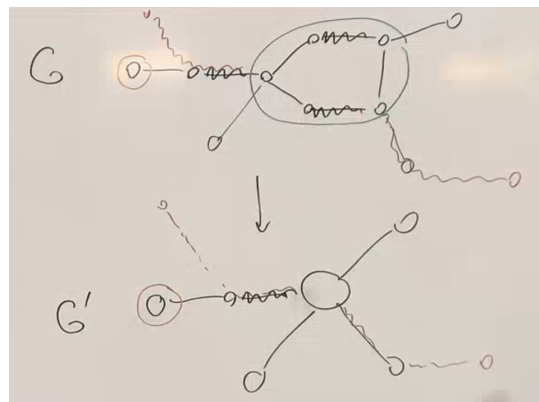


Утверждение: если в исходном графе G был дополняющий путь, то и в полученном графе G' .

Доказательство. Докажем в две стороны

(\Leftarrow) Либо дополняющий путь вообще не проходит через большую вершину соцветия, тогда вообще все просто, ничего точно не ломается.

Если дополняющий путь проходит через большую вершину соцветия, то выберем кусок цикла нужной четности и соединим кусочки дополняющего пути.

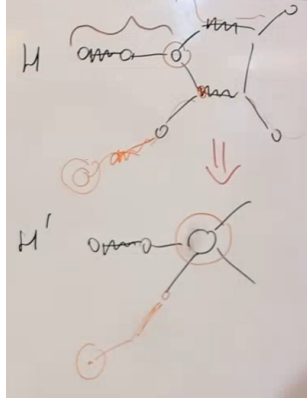


(\Rightarrow) Все сложно. Могут быть всякие сложные пути и после сжатия не понятно, что с ними делать. Конструктивно доказать сложно.

Докажем при помощи теоремы о дополняющем пути. От противного.

1. Возьмем граф G вместе с соцветием и инвертируем ему стебель, получим граф H с валидным парсочем.

2. Возьмем граф H вместе с соцветием и инвертируем ему стебель, получим граф H' с валидным парсоchem.
3. Если в G есть дополняющий путь, то и в G' тоже есть дополняющий путь (очевидно, так как и то и другое паросоветание не максимального размера).
 Если в H' есть дополняющий путь, то и в H тоже есть дополняющий путь (очевидно, так как и то и другое паросоветание не максимального размера).
 Если в G' есть дополняющий путь, то и в H' есть дополняющий путь (конструктивно).



■

```

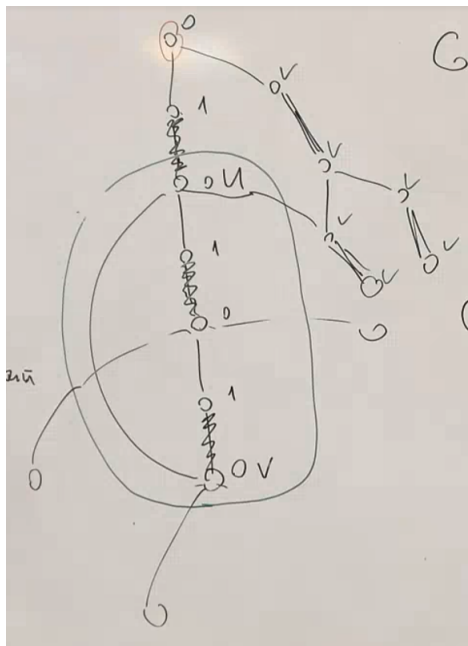
1  M = ∅
2  while True:
3      dfs
4      if нашли доп путь:
5          разжать соцветие
6          M++
7          continue
8      if нет додоп пути, не соцветий:
9          break
10     if соцветие:
11         сжать соцветие

```

Итого, время работы алгоритма — $\mathcal{O}(n^2m)$.

Как реализовать побыстрее? *Немного пострадать.*

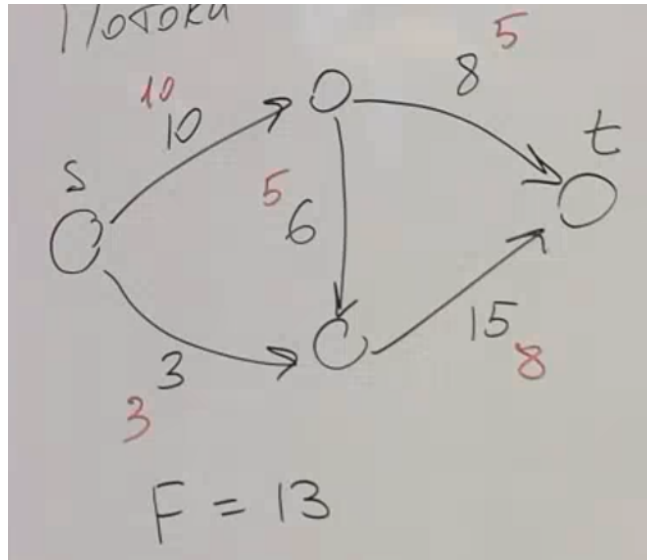
Пооптимизируем *DFS*. Прямо проходя в *DFS*—е можно сжимать вершины. Для этого можно быстро мерджить списки ребер.



В общем, если постараться, можно получить решение $\mathcal{O}(nma(m, n))$. А вообще, пацаны умеет делать за $\mathcal{O}(m\sqrt{n})$.

1.2 Потоки

Пусть у нас есть ориентированный граф. По этому графу течет некоторая *жизжа*. Есть исток и есть сток, для каждого ребра известна пропускная способность.



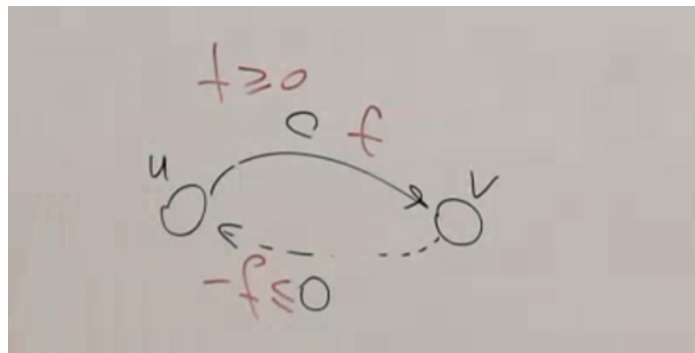
Обозначим C_{uv} — пропускная способность ребра, f_{uv} — величина потока.

$$f_{uv} \geq 0, f_{uv} \leq C_{uv}.$$

$$\forall v \neq s, v \neq t : \sum_u f_{uv} = \sum_w f_{vw}.$$

$$F = \sum_u f_{uv} - \sum_w f_{vw}$$

Решать такую задачу в такой формулировке не очень удобно. Удобнее мыслить в симметричной формулировке, хочется избавиться от двух сумм. Давайте считать, что для каждого ребра есть фиктивный обратный поток.



$$f_{vu} = -f_{uv}, c_{vu} = 0, f_{uv} \leq c_{uv}$$

$$\sum_v f_{vu} = 0, \forall v \text{ кроме } s \text{ и } t.$$

$$F = \sum_u f_{su} = \sum_w f_{wt}.$$

$$\sum_v \sum_u f_{vu} = \sum_w f_{su} + \sum_t f_{tu}.$$

На самом деле, не всегда можно просто так складывать потки $F = F_1 + F_2$. — не всегда валидный поток (может переполниться ребро)

Теорема 3. (Как бы утверждение) Любой поток можно декомпозировать на маленькие потоки

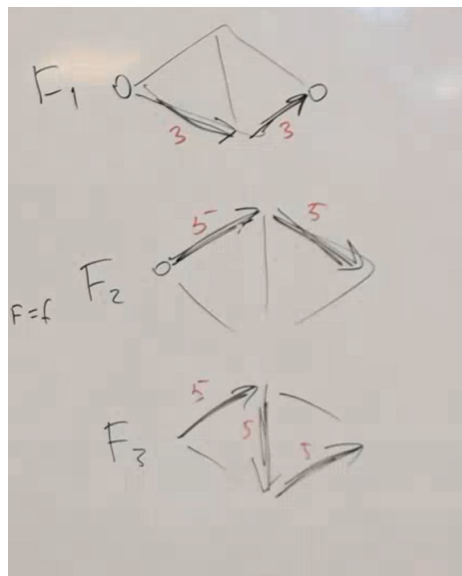
$$F = F_1 + F_2 + \dots + F_k$$

F_i — путь $S \rightarrow t$ или цикл

Доказательство. План: взять путь и вычесть его. Как найти какой-нибудь путь?

1. Пойдём из S
2. Возьмём любое ребро по которому течёт жижа.
3. Рано или поздно либо дойдём до t , либо заиклимся.

Хорошо, мы нашли путь. Давайте удалим его, в терминах потоков надо взять минимальную пропускную способность ребра и вычесть из всех ребер.



Может ситуация, что из S все рёбра нулевые, в t все рёбра нулевые, а где-то в середине что-то происходит. Это значит, что остались только циклы. Для их нахождения нужно взять любое ненулевое ребро и запускаться уже от него.

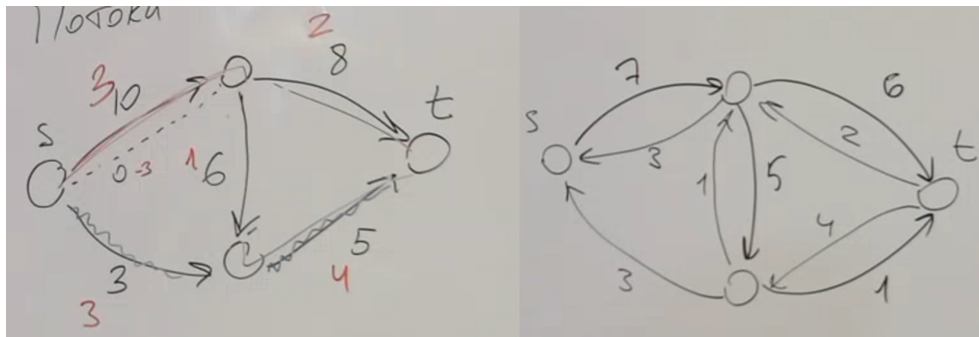
Всего будет $O(m)$ путей и циклов, т.к. мы каждый раз обнуляем хотя бы одно ребро. ■

Вообще, мы хотим найти максимальный поток. Мы сперва возьмем пустое разбиение, а потом будем его постепенно дополнять.

Определение 1. Ребро называется насыщенным, если вся его пропускная способность задействована.

Определение 2. Блокирующий поток — такой поток, что любой путь содержит насыщенное этим потоком ребро. Иными словами, в данной сети не найдётся такого пути из истока в сток, вдоль которого можно беспрепятственно увеличить поток.

Определение 3 (Остаточная сеть). Пусть у нас был валидный поток. Для каждого ребра посчитаем $c'_{uv} = c_{uv} - f_{uv} \geq 0$. Остаточная сеть — подграф исходного, где на прямых ребрах написаны c_{uv} , а на обратных ребрах остаточная пропускная способность c'_{uv} .



Идея алгоритма (схема Форда–Фалкерсона)

- Возьмём поток F , возьмём максимальный поток F_{\max} и посмотрим на их разность $\Delta F = F_{\max} - F$
- ΔF — поток в остаточной сети C'_F $\Delta f_{uv} = f_{\max uv} - f_{uv} \leq C_{uv} - f_{uv} = C'_{uv}$.
- Нет пути $S \rightarrow t$ в сети C'_F тогда и только тогда когда $F = F_{\max}$

```

1  int dfs(v, minΔ):
2      if mark(v):
3          return 0
4      mark[v] = True
5      for (u): Cvu - fuv > 0
6          Δ = dfs(u, min(minΔ, Cuv - fuv))
7          if Δ > 0:
8              fvu += Δ
9              fuv -= Δ
10             return Δ
11  return 0

```

Время работы алгоритма — $\mathcal{O}(F \cdot m)$.

1.2.1 Разрезы

Задача 1. Пусть есть граф. Мы хотим удалить некоторые ребра, чтобы не было пути из S в T . Определим $C = \sum c_{uv}$, для таких u, v , что нужно удалить ребро uv .

Надо найти C_{\min} .

На самом деле это двойственная задача к поиску максимального потока.

Как искать минимальный разрез? Найдём максимальный поток, рассмотрим его дополняющую сеть.
... ДОПИСАТЬ

1.2.2 Алгоритм Эдмондса–Карпа

Сделаем все то же самое, но dfs заменим на bfs.

Пусть $d[v]$ — расстояние (число ребер) от S до v в остаточной сети.

Теорема 4. $d[v]$ не убывает, если выбирать кратчайшие дополняющие пути.

Доказательство. Возьмем кратчайший путь $S \rightarrow T$. Ребра с минимальной дельтой на пути могли пропасть из остаточной сети. Какие могли добавиться? Те, у которых были обратные ребра.

Заметим, что $d[v]$ не убывает.

Ребро может пропасть из дополняющей сети $d[u] = d[u] + 1$.

Ребро может восстанавливаться в дополняющей сети $d'[u] = d'[v] + 1$.

Таким образом $d'[u] \geq d[u] + 2$.

Поэтому ребро может быть добавлено и удалено не больше $\mathcal{O}(n)$ раз, а поскольку ребер m , то всего таких запусков bfs может быть $\mathcal{O}(nm)$. ■

Итого, наш алгоритм будет работать за $\mathcal{O}(nm^2)$. На самом деле, время работы асимптотически не больше, чем $\mathcal{O}(m \cdot F)$.

Анонс на дальше (что умеют люди):

1. $\mathcal{O}(nm)$ — но страшно, такое смотреть небудем
2. Плавнo дойдём до $\mathcal{O}(nm \ln n)$
3. $\mathcal{O}(nm \ln C)$
4. $\mathcal{O}(n^3)$

Определение 4 (Масштабирование). Почему наш алгоритм работал долго? ПОтому что он искал плохие пути. Давайте запретим ему искать единичные пути. Запустим ту же штуку, но на фазе k заставим его брать пути длиной хотя бы 2^k

$k = \log C \dots 0$ искать пути $\Delta \geq 2^k$

Доказательство. $k + 1 \rightarrow k$:

$$F_{\max} - F \leq 2^{k+1} \cdot m$$

$$\text{Путей с } \Delta = 2^k \leq 2m$$

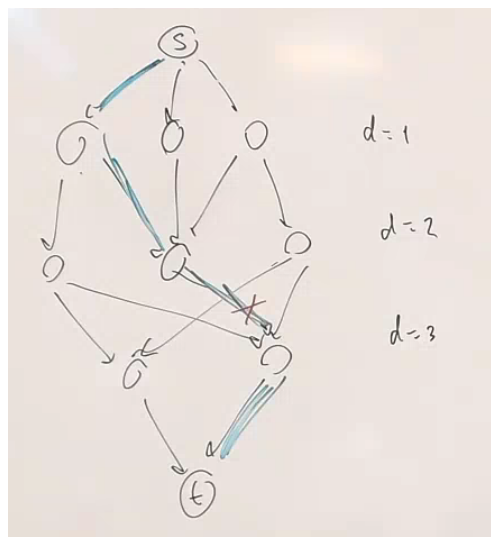
Финальная асимптотика: $\mathcal{O}(\ln C \cdot m^2)$ — слабополиномиальный алгоритм. ■

Алгоритм Диница

Определение 5. Алгоритм Диница.

Проведём bfs, оставим только те рёбра, которые лежат на кратчайших путях (с меньшего на больший слой). Уберём все тупики.

Найдём так кратчайший путь (идём по любому ребру), запустим на нём поток величины равной минимальной пропускной способности на пути. Таким образом хотя бы одно из рёбер уйдёт. Повторим. Остановимся когда s станет тупиком.



```

1  while есть путь  $s \rightarrow t$ :
2      bfs()
3      while dfs():
4           $\Delta = \min$ 
5           $f += \Delta$ 
6          удалить тупики
7  можно( удалять лениво:
8      если dfs не смог дойти до  $t$ ,
9      возвращаемся из рекурсии и
10     убираем за собой ребро)

```

Внешний *while* выполняется не больше n раз. Внутренний не больше m . Суммарная асимптотика: $\mathcal{O}(n^2m)$. n на каждый *dfs*. Удаление тупиков – суммарно n , не страшно.

Если добавить масштабирование, то асимптотика станет $\mathcal{O}(nm \ln C)$

Алгоритм Гольберга-Торьяна План: не терять все пути, которые мы нашли раньше.

1. Нашли какой-то кратчайший путь
2. Нужно найти на нём минимальное C' , только значение.
3. Прибавить этот C' ко потоку
4. Нужно найти минимальное ребро, чтобы его убрать.
5. Помним предыдущий путь. На нём есть ребро идущее в тупик. Уберём его. Будем делать так пока таких не будет, а тогда мы сможем пройти в t .
6. В целом будем идти до ближайшего осколка предыдущего пути.
7. Из этих путей будем дерево.

План закончился, теперь алгоритм:

1. У каждой вершины будет не более одного исходящего ребра – дерево.
2. s принадлежит дереву. Пойдём по помеченным рёбрам, пока не упёрлись. Дошли до корня дерева s .
3. Если дошли до t – радуемся и на этом пути делаем все действия (минимальная пропускная способность, ко всем на пути прибавляем Δ , убираем минимальные)
4. Если не дошли – идём куда-нибудь. Идём до корня уже той вершины. Заодно помечаем ребро p которому мы “куда-нибудь” перешли. Повторяем.

Псевдо-псевдо-кот:

```

1  while есть путь  $s \rightarrow t$ :
2      bfs()
3      while не удалили  $s$ :
4           $v = \text{root}(s)$ 
5          if  $v == t$ :
6               $\Delta, e = \min(s..v)$ 
7              add( $s..v$ ,  $-\Delta$ )
8              cut( $e$ )
9          else:
10             if  $v$  -- тупик:
11                 for all  $wv$ :
12                     cut( $wv$ )
13             else:
14                 ( $vu$ ) -- любое ребро ( $d(u) = d(v) + 1$ )
15                 link( $vu$ )

```

- Большой while – n раз
- bfs – m
- корень суммарн $m \ln n$. Каждый if тоже $m \ln n$

Время работы алгоритма – $\mathcal{O}(nm \ln n)$.

Алгоритм Малхотры – Кумара – Махешвари Идея: удалять за раз не ребро, а целую вершину. Какую вершину проще всего удалить? У которой минимальная пропускная способность...

$$c(v) = \min(\sum c_{uv}, \sum c_{uw})$$

1. Берём вершину. Течём из неё по одному ребру. Если не хватает, пишем сколько получится. Пишем в другие, чтобы в сумме было Δ .
2. Повторяем на вершинах ниже.
3. От s делаем по обратным рёбрам.

Время работы алгоритма – $\mathcal{O}(n \cdot (n \cdot n + m)) = \mathcal{O}(n^3)$ Рассмотрим такие графы, в которых все $c_{uv} = 1$

Можем находить все непересекающиеся пути между двумя вершинами. Если c_{uv} маленькие, например 1 или 2, и в нём нет параллельных потоков, можно разрешить параллельные рёбра.

Алгоритмы:

Таблица 1: Асимптотики

	$c_{uv} \in \mathbb{R}$	$c_{uv} = 1$	$c_{uv} = 1$ и нет параллельных
ФФ	$\mathcal{O}(Fm)$	$\mathcal{O}(m^2)$	$\mathcal{O}(nm)$
ЭК	$\mathcal{O}(m^2n)$	$\mathcal{O}(m^2)$	$\mathcal{O}(nm)$
Диниц	$\mathcal{O}(mn^2)$	$\mathcal{O}(m^{\frac{3}{2}})$	$\mathcal{O}(\min\{\sqrt{m}, n^{\frac{2}{3}}\}m)$

Утверждение 4.1. Диниц не делает n фаз. $c_{uv} = 1 \implies \text{Фаз} \leq \sqrt{m}$

Нашли пути длин $1, 2, 3, \dots, \sqrt{m}$

$F_{\max} - F$ разбивается на непересекающиеся пути $s \rightarrow t$

Так как всего рёбер m И каждое ребро участвует только в одном пути, то количество таких путей будет $\leq \sqrt{m}$

Утверждение 4.2. c_{uv} и нет параллельных рёбер, то $\mathcal{O}(n^{\frac{2}{3}})$

1.3 Вспомним паросочетания

Был двудольный граф. Мы в нём искали паросочетания, потом содили его к потоку.

Если просто натравить Диница, то уже получится лучше $\mathcal{O}(nm)$. Давайте ещё лучше оценим.

Утверждение 4.3. $\mathcal{O}(\sqrt{m})$ фаз для диница и паросочетаний.

$F_{\max} - F$ можно декомпозировать на пути.

Финальная асимптотика нахождения паросочетания с помощью Диница – $\mathcal{O}(\sqrt{nm})$

1.4 Push-relabel

Определение 6. Предпоток:

$$f_{uv} \leq c_{uv}, \text{ но баланс } b_u = \sum_v f_{uv} \geq 0$$

Слой: $l[v]$ – уровень вершины.

$l[s] = n$ $l[t] = 0$ Это константно. $l[v] = 0$ для всех остальных вершин, потом растёт.

Свойство 1. Если ребро очень сильно уходит вниз. $l[u] \leq l[v] - 2 \implies f_{uv} = c_{uv}$

Все рёбра, которые идут сильно вниз, они насыщены.

Изначально в алгоритме свойство не выполняется для рёбер из s . Насытим все эти рёбра

Две операции:

- push. $l[u] = l[v] - 1$ $f_{uv} + = \min(b_v, c'_{uv})$
- relabel. $c'_{uv} > 0$ $l[v] = \min(l[u]) + 1$

1. Все вершины с балансом $b_v = 0 \implies F_{\max}$

2. На любом пути от s до t будет хотя бы одно насыщенное ребро. Значит насыщенные рёбра образуют разрез. Значит поток максимален.

Давайте теперь время на всё это.

$O(n)$ на один relabel. $O(nm)$ на все.

Насыщающее проталкивание: Каждый раз поднимает уровень. Таких не больше n для каждого ребра. Все вместе они тоже дают $O(nm)$

Остались ненасыщающие, их может быть много, с ними нужно аккуратно.

Заведём баланс как сумму уровней вершин с ненулевым балансом. $\Phi = \sum_{b_u > 0} l(u)$

relabel: $n^2 \cdot n$. Насыщающий push $nm \cdot n$. Ненасыщающий -1 . Последних $\leq n^2 m$ соответственно. насыщающий push

Если взять хорошую очередь и в правильном порядке релаксировать, то будет $O(n^3)$ вместо $O(n^2 m)$

Ещё можно $O(n\sqrt{m})$

relabel не поднимает потенциал больше, чем на n , а проталкивание бесплатное.

1.5 Задача о назначениях

Есть работы и работники. Стоимость назначения работника i на работу j — c_{ij} . Нужно найти такое назначение, чтобы суммарное количество потраченных денег было минимальным.

$$i \rightarrow p(i) \quad \min \sum C_{i,p(i)}.$$

Другая формулировка: построим матрицу. В неё запишем стоимости, какой работник за какую работу хочет денег. Тогда задача: выбрать в каждом столбце и строке по одной ячейке и минимизировать их сумму.

Замечание. Простое действие: изменить все стоимости у одного работника на константу...

$$i : C_{i,j} + = \Delta \quad j = 1..n$$

Пример. Матрица работников

5	8	6	4
1	2	1	2
3	2	8	5
5	4	6	3

Утверждение 4.4. С помощью таких преобразований можно нормализовать матрицу и получить в некоторых ячейках 0.

Пример. Преобразуем её, вычтя из каждой строки минимум:

Матрица работников получше

1	4	2	0
0	1	0	1
1	0	6	3
2	1	3	0

$$c_{i,j} \geq 0.$$

Если по

Посмотрим на граф нулей. Он двудольный, поэтому нашу задачу можно свести к поиску максимального паросочетания.

```

1  for v ∈ L // O(n)
2      while True: // O(n)
3          if dfs(v): // O(n^2)
4              M++
5              break
6          else: // O(n^2)
7              Δ = min Cuv: u ∈ L+, v ∈ R-
8              Cuv -= Δ, u ∈ L+
9              Cuv += Δ, v ∈ R+

```

Алгоритмическая сложность: $\mathcal{O}(n^4)$.

Теорема 5. 1. После наших действий не появятся отрицательные элементы.

	$+\Delta: R^+$	R^-	$\Delta = \min.$
$-\Delta: L^+$	0	$-\Delta$	
L^-	$+\Delta$	0	

2. После нашей операции достижимых вершин стало не меньше и старые остались достижимы.

Давайте ускорять. Что мы делаем долго? Давайте не запускать dfs из нуля каждый раз. Мы находим те же вершины каждый раз вместе с новыми. Давайте хранить bfs и просто идти по новому ребру, когда оно появляется.

Операции, которые хочется:

1. $add_row(i, \Delta)$ $a[j] += \Delta$
2. $add_col(i, \Delta)$ $b[j] += \Delta$
3. $get(ij)$ $c[ij] + a[j] + b[j]$

Давайте дополнительно хранить потенциалы для столбцов и строку Пусть a — массив потенциалов для строк, b — массив потенциалов для столбцов.

Недостижимые вершины становятся достижимыми, но никогда не наоборот. Как будет находить минимум? Посчитаем минимум по нужным столбцам. Дальше будем считать минимум за n как минимум по столбцам. Если столбец уходит — за n пересчитается. Если строка добавляется, за n Обновляем минимумы во всех столбцах

```

1  for v ∈ L: // O(n)
2      v ∈ L+, остальные ∈ L-
3      R- = R
4      m[j] += c[v, j]
5      while True: // O(n)
6          Δ, (vu) = minj ∈ R- (m[j]) // O(n)
7          a[i] -= Δ i ∈ L+
8          b[j] += Δ j ∈ R+
9          m[j] -= Δ
10         R+ = R+ ∪ {u}
11         if p[u] = ∅:
12             инвертируем дополняющий путь // O(n)
13         else:
14             L+ = L+ ∪ { p(n) }
15             m[j] = min (m[j], c[p(n), j] + a[p(n)] + b[j]) // O(n)

```

Итого, алгоритмическая сложность — $\mathcal{O}(n^3)$.

Давайте дальше ускорять. Для нахождения минимального ребра можно использовать двоичную кучу. К этой куче нужно добавить операцию изменения весов всех элементов на константу.

Теперь $\mathcal{O}(nm \ln n)$.

Можно сделать фббоначиеву кучу. $\mathcal{O}(n \ln n + m)$