Основы теории графов

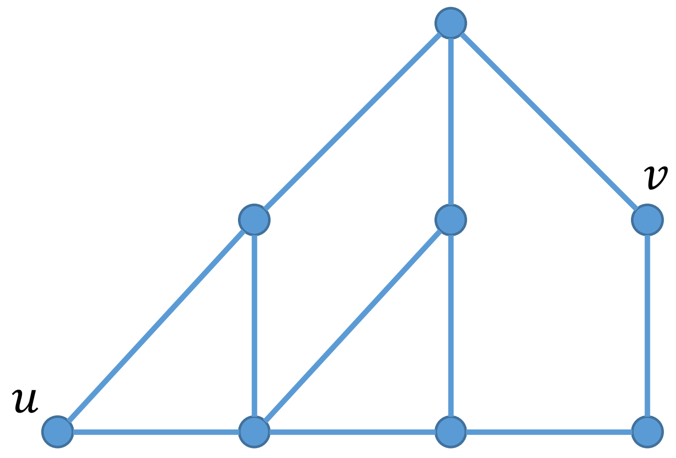
осень 2013

Александр Дайняк

www.dainiak.com

Расстояния в графе

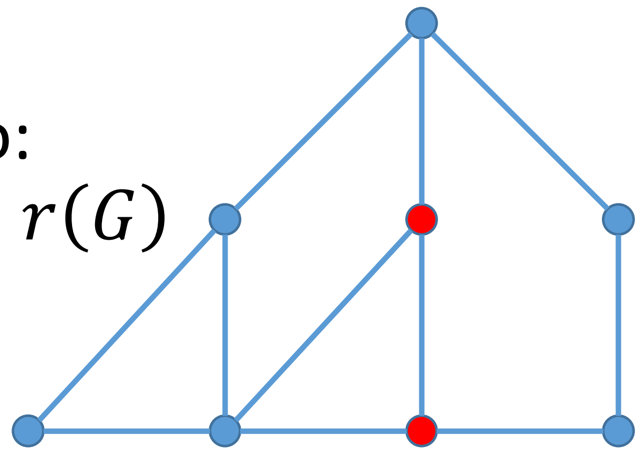
- *Расстояние между парой вершин* u, v — это длина кратчайшей цепи, соединяющей эти вершины. Обозначение: $d(u, v)$
- Диаметр графа — это максимальное из расстояний между парами вершин. Обозначение:
$$\operatorname{diam}(G) = \max_{u,v} d(u, v)$$
- *Диаметральная цепь* — это цепь, соединяющая пару вершин, расстояние между которыми равно диаметру



$d(G) = d(u, v) = 3$

Расстояния в графе

- *Эксцентриситет* вершины v — это величина
$$\max_u d(u, v)$$
- *Центр* графа — это вершина, имеющая минимальный эксцентриситет (центров у графа может быть много)
- *Радиус* графа — это эксцентриситет центра
Обозначение: $r(G)$
- Для любого связного графа G выполнено:
$$r(G) \leq \operatorname{diam}(G) \leq 2 \cdot r(G)$$



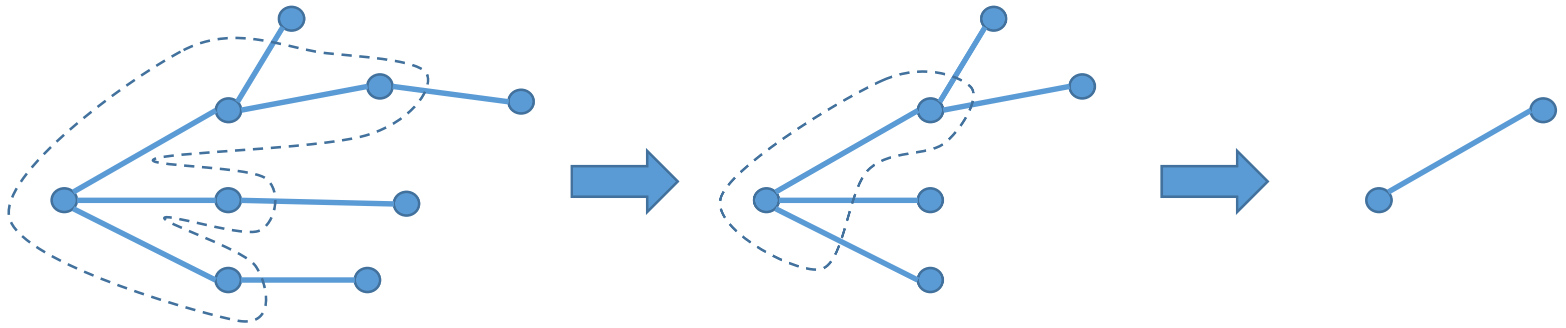
Расстояния в деревьях

- В дереве любой конец диаметральной цепи является листом
- В дереве ровно один центр, если диаметр дерева кратен 2, и ровно два центра иначе. Причём, если центра два, то они соседи.
- В любом дереве T

$$r(T) = \left\lceil \frac{1}{2} \cdot \operatorname{diam}(T) \right\rceil$$

Расстояния в деревьях

- Докажем, что в дереве один или два центра.
- Пусть дано произвольное T.
- «Обстрижём» T: удалим из T все листья.
- Получим новое дерево T', в котором центры останутся теми же, что и в T

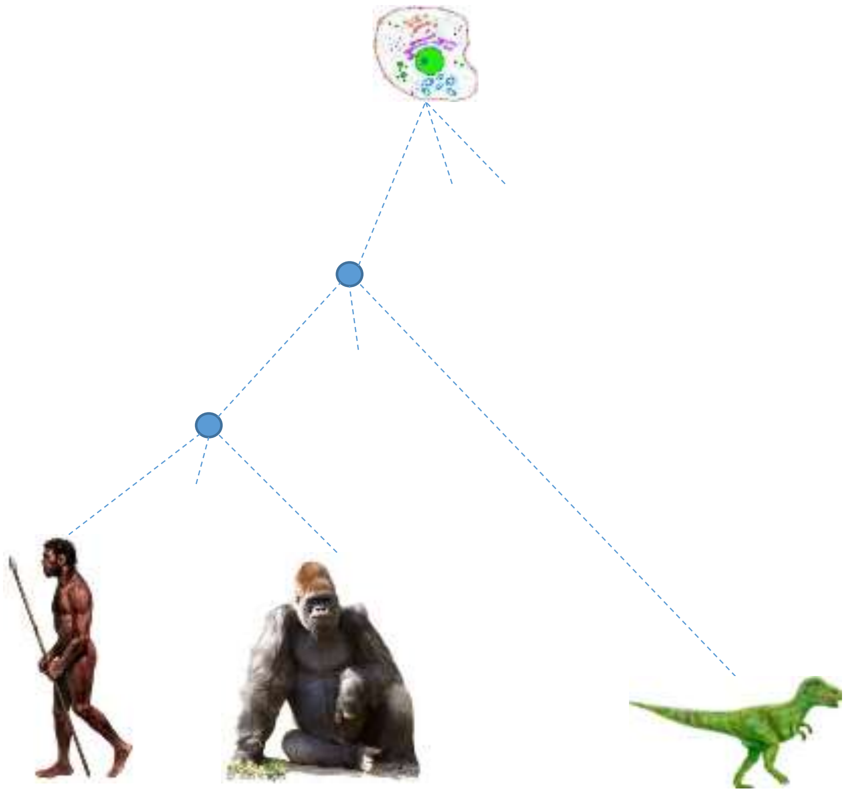


Расстояния в деревьях

- *Прямая задача* о расстояниях: в заданном графе найти расстояние между некоторыми/всеми парами вершин.
- *Обратная задача*: по заданным расстояниям попытаться восстановить граф.

Расстояния в деревьях

Обратная задача о расстояниях в деревьях актуальна, например, при построении *деревьев классификации*



Расстояния в деревьях

Расстояния между вершинами могут быть заданы матрицей расстояний:

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{pmatrix}$$

В матрице расстояний $d_{ij} \coloneqq d(v_i, v_j)$.

В общем случае расстояния могут быть заданы необязательно для всех пар вершин.

Расстояния в деревьях

Вопросы:

- При каких условиях на матрицу D существует реализующее её дерево?
- Если дерево существует, является ли оно единственным?
- Как это дерево построить?

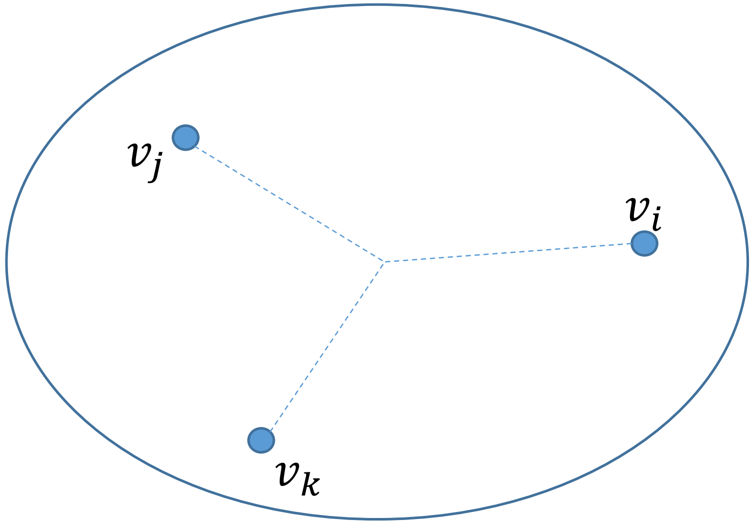
Расстояния в деревьях

Рассмотрим необходимые условия существования дерева:

1. Для любых i, j выполнены соотношения
$d_{ii} = 0, \;\; d_{ij} = d_{ji} > 0$

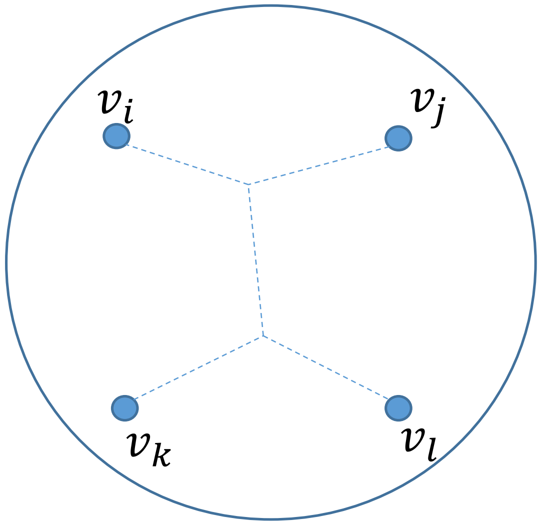
Расстояния в деревьях

2. Для любых i, j, k величина $d_{ij} + d_{jk} - d_{ik}$ является чётным числом.

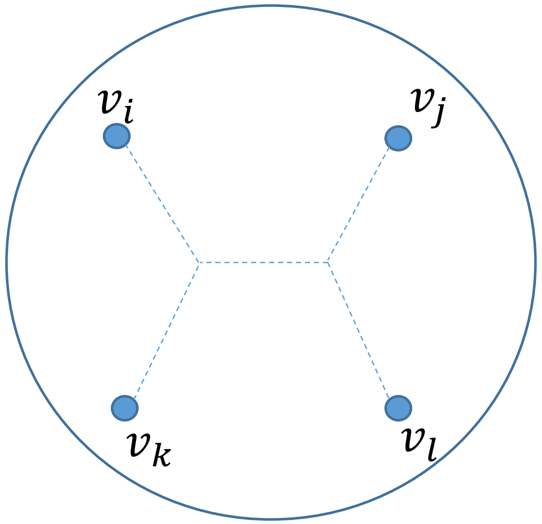


Расстояния в деревьях

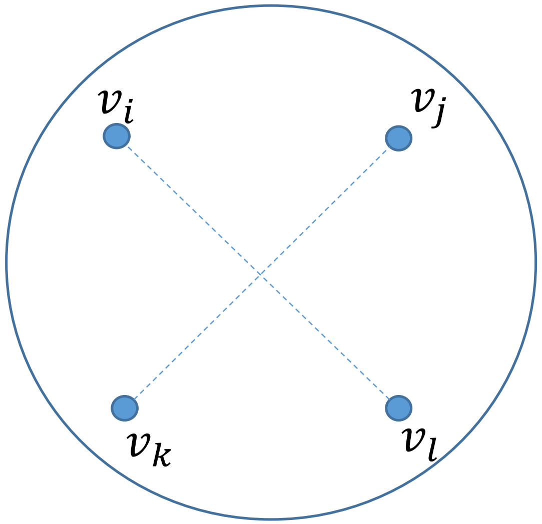
3. Для любых i, j, k, l из чисел $d_{ij} + d_{kl}$, $d_{ik} + d_{jl}$, $d_{il} + d_{jk}$ два числа равны между собой и не меньше третьего.



$$d_{ik} + d_{jl} = d_{il} + d_{jk} > \\ > d_{ij} + d_{kl}$$



$$d_{ij} + d_{kl} = d_{il} + d_{jk} > \\ > d_{ik} + d_{jl}$$



$$d_{ik} + d_{jl} = d_{il} + d_{jk} = \\ = d_{ij} + d_{kl}$$

Расстояния в деревьях

Теорема. Следующие три условия являются необходимыми и достаточными для существования дерева:

1. Для любых i, j выполнены соотношения $d_{ii} = 0, \quad d_{ij} = d_{ji} > 0$

2. Для любых i, j, k величина $d_{ij} + d_{jk} - d_{ik}$ является чётным неотрицательным числом

3. Для любых i, j, k, l из чисел $d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}$ два числа равны между собой и не меньше третьего

При этом если в D учтены все листья, то такое дерево будет единственным.

Расстояния в деревьях

План доказательства достаточности:

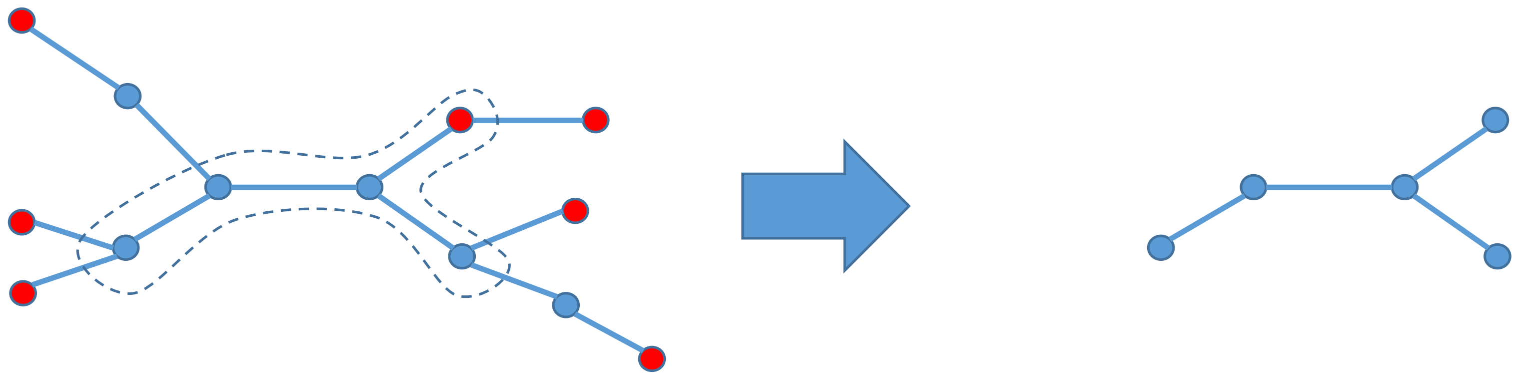
- Индукция по размеру матрицы
- По матрице D строим некоторую матрицу D' меньшей размерности
- Доказываем, что если для D выполнены условия 1—3, то они выполнены и для D'
- Указываем способ построения по дереву T', соответствующего матрице D', дерева T, соответствующего матрице D

Расстояния в деревьях

Идея перехода к новой матрице:

Допустим, у нас уже есть дерево T, соответствующее матрице D, и все листья дерева участвуют в D.

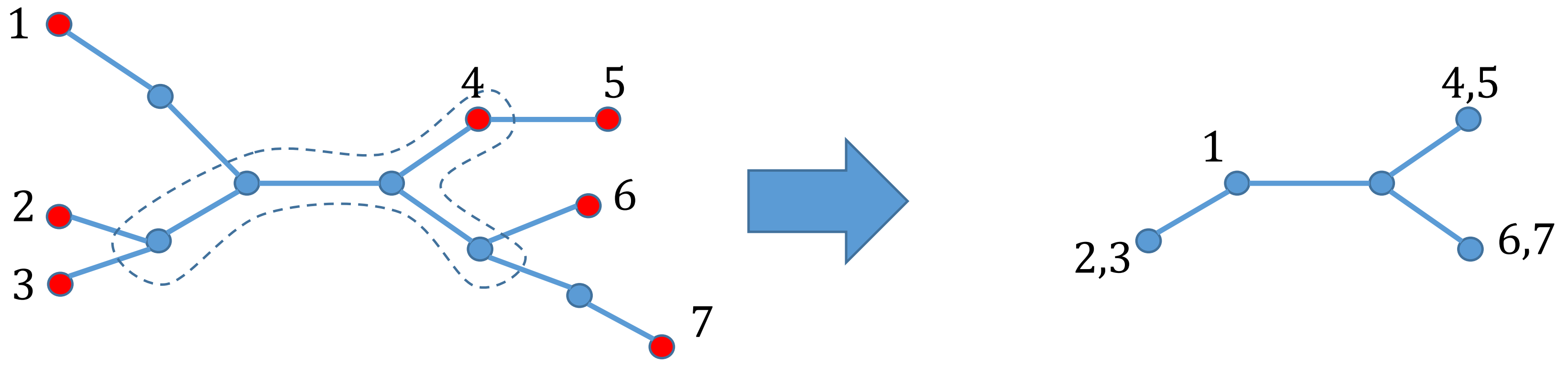
Дерево T можно «сильно остричь», удалив не листья, а сразу целые «висячие цепочки». Висячая цепочка — от листа дерева до вершины степени 3 или первой нелистовой вершины, участвующей в D.



Расстояния в деревьях

Надо понять, как меняются длины путей при переходе к «сильно обстриженному» дереву, и куда переходят вершины, по которым строилась исходная матрица D.

Для этого надо понять, как по матрице D определить длины цепочек, которыми висячие вершины присоединены к дереву.



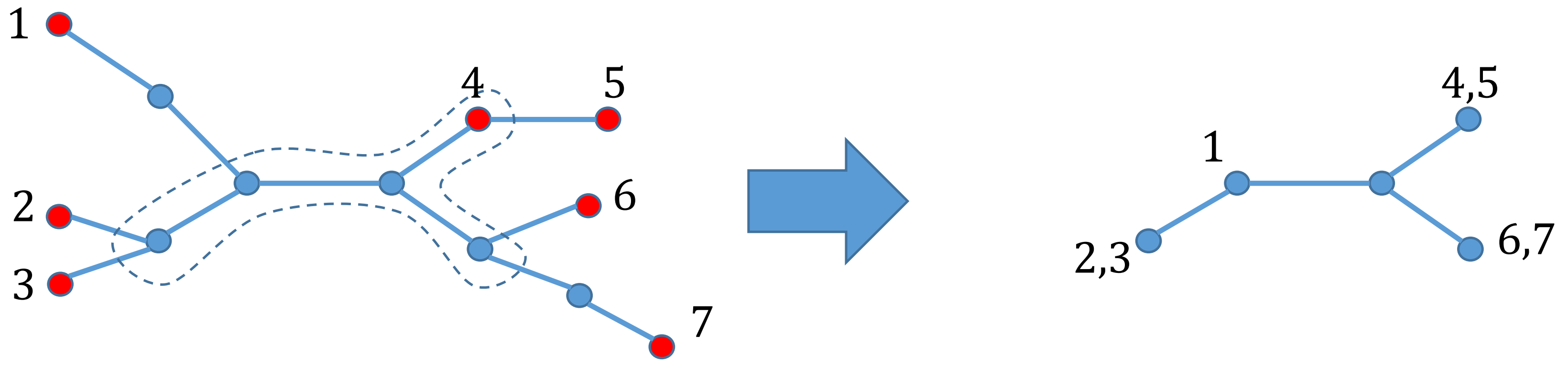
Расстояния в деревьях

Вспомним про величины $d_{ij} + d_{jk} - d_{ik}$.

Длина цепочки, которой v_j присоединяется к дереву, равна

$$\frac{1}{2} \cdot \min_{p,q \neq j}\left(d_{pj} + d_{jq} - d_{pq}\right)$$

(Для невисячих вершин длина такой цепочки принимается равной нулю)

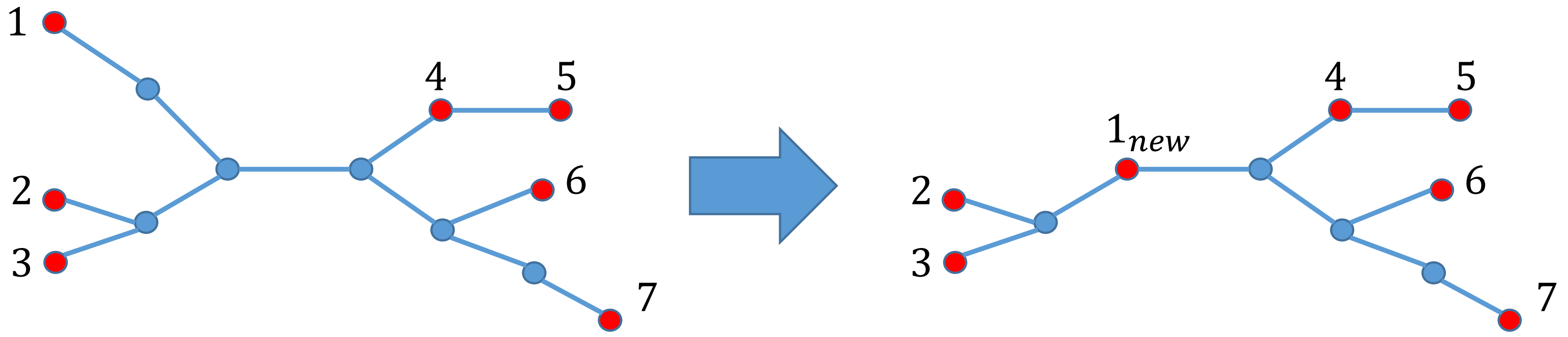


Расстояния в деревьях

Обозначим $c_j \coloneqq \frac{1}{2} \cdot \min_{p,q \neq j} \left(d_{pj} + d_{jq} - d_{pq} \right)$

Если мы обстригаем одну из цепочек, оканчивающуюся на листе j, при этом переводя j в другой конец этой цепочки, то расстояния от j_{new} до остальных вершин сократятся на длину цепочки:

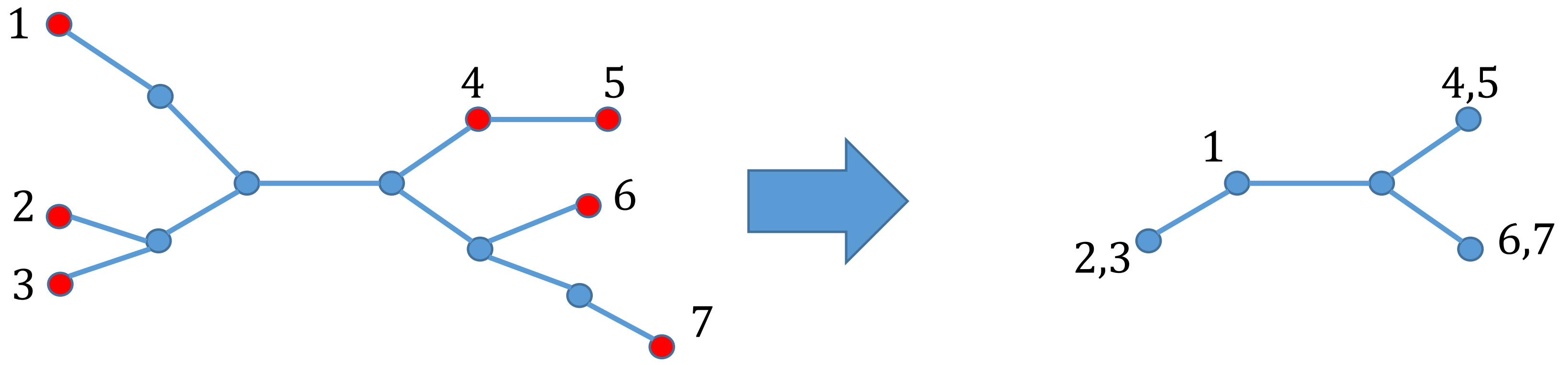
$$d_{ij}^{new} = d_{ij}^{old} - c_j$$



Расстояния в деревьях

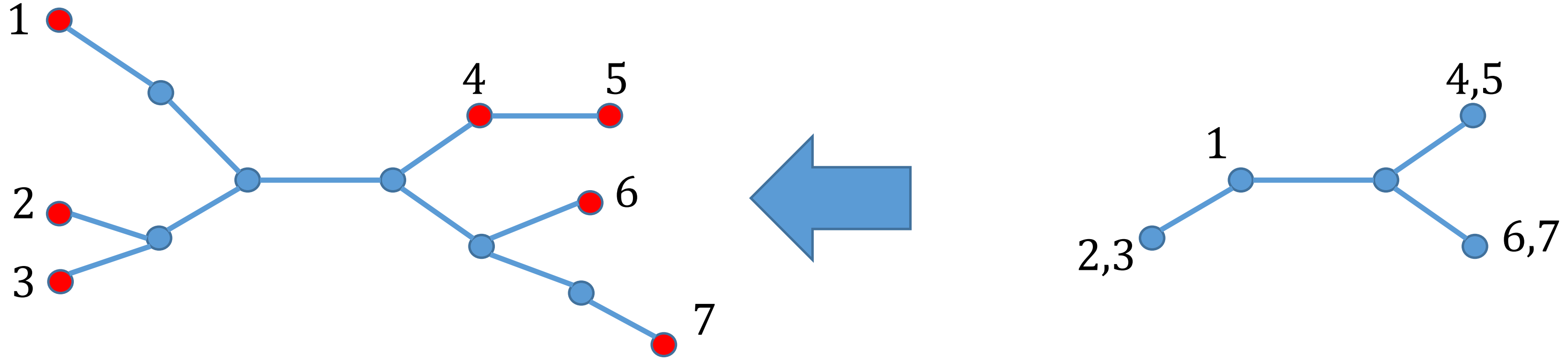
Обозначим $c_j := \frac{1}{2} \cdot \min_{p,q \neq j} \left(d_{pj} + d_{jq} - d_{pq} \right)$

Если мы «сильно обстригаем» дерево, смещая листья соответствующим образом, то расстояния в новом дереве вычисляются по формулам $d_{ij}^{new} = d_{ij}^{old} - c_i - c_j$



Расстояния в деревьях

Обратно, если у нас есть дерево T', реализующее матрицу D_{new}, то дерево, реализующее матрицу D, можно построить, присоединив к соответствующим вершинам дерева T', цепочки длины c_i



Расстояния в деревьях

Всё, сказанное до сих пор — *идея* доказательства. Для формального обоснования и индукции сделаем следующее:

- Рассмотрим матрицу $\widetilde{D}' = \{d'_{ij}\}$, в которой $d'_{ij} \coloneqq (d_{ij} - c_i - c_j)$ при $i \neq j$, и $d'_{ii} \coloneqq 0$
- Докажем, что если исходный набор чисел d_{ij} удовлетворяет условиям 1—3, то и набор чисел вида d'_{ij} тоже удовлетворяет этим условиям, кроме условия положительности всех d'_{ij} при $i \neq j$
- Докажем, что в матрице $\widetilde{D}'$ есть совпадающие строки/столбцы, так что можно, отождествив их, получить матрицу D' меньшей размерности, в которой уже все недиагональные элементы положительны

Расстояния в деревьях

Проверяем выполнение условия 3 для d'_{ij}:

$$d'_{ij} + d'_{kl} = d_{ij} - c_i - c_j + d_{kl} - c_k - c_l =$$
$$= d_{ij} + d_{kl} - (c_i + c_j + c_k + c_l)$$

Аналогично

$$d'_{ik} + d'_{jl} = d_{ik} + d_{jl} - (c_i + c_j + c_k + c_l)$$

и

$$d'_{il} + d'_{jk} = d_{il} + d_{jk} - (c_i + c_j + c_k + c_l).$$

Отсюда следует, что и для d'_{ij} условие 3 выполняется, раз оно выполнено для d_{ij}.

Расстояния в деревьях

Проверяем выполнение условия 2 для d'_{ij}:

$$d'_{ij} + d'_{jk} - d'_{ik} = d_{ij} - c_i - c_j + d_{jk} - c_j - c_k -$$
$$-d_{ik} + c_i + c_k = d_{ij} + d_{jk} - d_{ik} - 2c_j =$$
$$= d_{ij} + d_{jk} - d_{ik} - \min_{p,q \neq j}\left(d_{pj} + d_{jq} - d_{pq}\right) \geq 0$$

Очевидно также, что числа $\left(d'_{ij} + d'_{jk} - d'_{ik}\right)$ чётны, так что условие 2 выполняется.

Расстояния в деревьях

Проверяем неотрицательность d'_{ij}:

$$d'_{ij} = d_{ij} - c_i - c_j =$$
$$= d_{ij} - \frac{1}{2}\min_{p,q\neq i}\left(d_{pi} + d_{iq} - d_{pq}\right) - \frac{1}{2}\min_{r,s\neq j}\left(d_{rj} + d_{js} - d_{rs}\right) \geq$$
$$\geq d_{ij} - \frac{1}{2}\left(d_{ji} + d_{ik} - d_{jk}\right) - \frac{1}{2}\left(d_{ij} + d_{jk} - d_{ik}\right) = 0$$

Расстояния в деревьях

Теперь покажем, что если $d'_{ij} = 0$, то для любого k выполнено $d'_{ik} = d'_{jk}$.

Это означает, что если в $\widetilde{D}'$ есть нулевой недиагональный элемент, то соответствующие строки и столбцы совпадают, а значит в D' уже не будет нулевых недиагональных элементов.

Расстояния в деревьях

Пусть $d'_{ij} = 0$. Тогда и $d'_{ji} = 0$.

Для любого k имеем

$$d'_{ij} + d'_{jk} - d'_{ik} \geq 0 \quad \Rightarrow \quad d'_{jk} \geq d'_{ik}$$

Аналогично

$$d'_{ji} + d'_{ik} - d'_{jk} \geq 0 \quad \Rightarrow \quad d'_{ik} \geq d'_{jk}$$

Отсюда $d'_{ik} = d'_{jk}$.

Расстояния в деревьях

Осталось доказать теперь, что найдутся $i \neq j$, такие, что $d'_{ij} = 0$.

Это будет означать, что у матрицы D' размерность меньше, чем у D, а значит можно будет провести индукцию по размерности матрицы.

Расстояния в деревьях

Лемма. Для любых $i \neq j$ найдётся $k \notin \{i, j\}$, такое, что

$$c_j = \frac{1}{2}\left(d_{ij} + d_{jk} - d_{ik}\right).$$

Доказательство.

По определению c_j, существуют k, l, для которых $2c_j = d_{lj} + d_{jk} - d_{lk}$.

Так как $c_j = \frac{1}{2}\min_{p,q \neq j}\left(d_{pj} + d_{jq} - d_{pq}\right)$, то $2c_j \leq d_{lj} + d_{ji} - d_{li}$,

отсюда

$$d_{lj} + d_{jk} - d_{lk} \leq d_{lj} + d_{ji} - d_{li}$$

Расстояния в деревьях

Итак,

$$d_{lj} + d_{jk} - d_{lk} \leq d_{lj} + d_{ji} - d_{li},$$

поэтому $d_{jk} + d_{li} \leq d_{ji} + d_{lk}$. Отсюда и из условия 3 получаем $d_{ji} + d_{lk} = d_{jl} + d_{ik}$.

Имеем

$$\begin{aligned} 2c_j = d_{lj} + d_{jk} - d_{lk} = d_{jk} + \left(d_{lj} - d_{lk}\right) = \\ = d_{jk} + \left(d_{ji} - d_{ik}\right) = d_{ij} + d_{jk} - d_{ik}, \end{aligned}$$

что и требовалось.

Расстояния в деревьях

Теперь доказываем, что найдутся $i \neq j$, такие, что $d'_{ij} = 0$.

Возьмём произвольные $i \neq j_1$. По лемме, найдётся $j_2 \neq j_1$, такой, что $c_{j_1} = \frac{1}{2}\left(d_{i j_1} + d_{j_1 j_2} - d_{i j_2}\right)$

Аналогично, найдётся $j_3 \neq j_2$, для которого

$$c_{j_2} = \frac{1}{2}\left(d_{i j_2} + d_{j_2 j_3} - d_{i j_3}\right)$$

И так далее. Получаем последовательность $\{j_m\}$.

Пусть $j_\alpha = j_\beta$ — первое повторение в этой последовательности ($\beta \geq \alpha + 2$).

Расстояния в деревьях

Пусть $j_\alpha = j_\beta, \alpha < \beta$. Преобразуем $\sum c_{j_m}$:

$$2 \cdot \sum_{m=\alpha}^{\beta-1} c_{j_m} = \sum_{m=\alpha}^{\beta-1} \left(d_{i\,j_m} + d_{j_m j_{m+1}} - d_{i\,j_{m+1}}\right) =$$

$$= \sum_{m=\alpha}^{\beta-1} \left(d_{i\,j_m} - d_{i\,j_{m+1}}\right) + \sum_{m=\alpha}^{\beta-1} d_{j_m j_{m+1}} =$$

$$= d_{i\,j_\alpha} - d_{i\,j_{(\beta-1)+1}} + \sum_{m=\alpha}^{\beta-1} d_{j_m j_{m+1}} = \sum_{m=\alpha}^{\beta-1} d_{j_m j_{m+1}}$$

Расстояния в деревьях

Ещё немного преобразований:

$$0 = \sum_{m=\alpha}^{\beta-1} d_{j_m j_{m+1}} - 2 \cdot \sum_{m=\alpha}^{\beta-1} c_{j_m} =$$

$$= \sum_{m=\alpha}^{\beta-1} d_{j_m j_{m+1}} - \sum_{m=\alpha}^{\beta-1} c_{j_m} - \sum_{m=\alpha}^{\beta-1} c_{j_{m+1}} - c_{j_\alpha} + c_{j_\beta} =$$

$$= \sum_{m=\alpha}^{\beta-1} \left(d_{j_m j_{m+1}} - c_{j_m} - c_{j_{m+1}}\right) = \sum_{m=\alpha}^{\beta-1} d'_{j_m j_{m+1}}$$

Расстояния в деревьях

В итоге получаем:

$$\sum_{m=\alpha}^{\beta-1} d'_{j_m j_{m+1}} = 0$$

Отсюда $d'_{j_m j_{m+1}} = 0$ при всех $\alpha \leq m < \beta$.

Таким образом, в матрице $\widetilde{D}'$ есть ненулевой недиагональный элемент (например, $d_{j_\alpha j_{\alpha+1}}$).

Значит, размерность матрицы D' строго меньше размерности $\widetilde{D}'$, а значит, и размерности D.

Расстояния в деревьях

- Размерность матрицы D' строго меньше размерности $\widetilde{D}'$, а значит и размерности D.
- Условия 1—3 для D' выполнены.
- Следовательно, можно доказать теорему индукцией по размерности матрицы.
- Базис индукции: $D \in \mathbb{R}^{1\times1}$ или $D \in \mathbb{R}^{2\times2}$.
  В случае $D \in \mathbb{R}^{1\times1}$, дерево T — одна вершина.
  В случае $D \in \mathbb{R}^{2\times2}$ в качестве T берётся цепь длины d_{12}, концы этой цепи и будут вершинами, соответствующими строкам/столбцам D.

Расстояния в деревьях

Индуктивный переход:

- Пусть дана D, и для матриц меньших размерностей соответствующие деревья существуют.
- Строим по D матрицу D', как описано ранее. При этом некоторые строки/столбцы D' соответствуют группам строк/столбцов D.
  Для D', по предположению, есть дерево T'.

Расстояния в деревьях

Индуктивный переход:

- Строим дерево T для исходной матрицы D.
  Пусть группе строк $j_1, \dots, j_s$ матрицы D соответствует строка j матрицы D', а этой строке соответствует вершина v_j дерева T'.
  Тогда в дереве T к вершине v_j протягиваем цепочки длин $c_{j_1}, \dots, c_{j_s}$, а концы этих цепочек объявляем вершинами $v_{j_1}, \dots, v_{j_s}$.