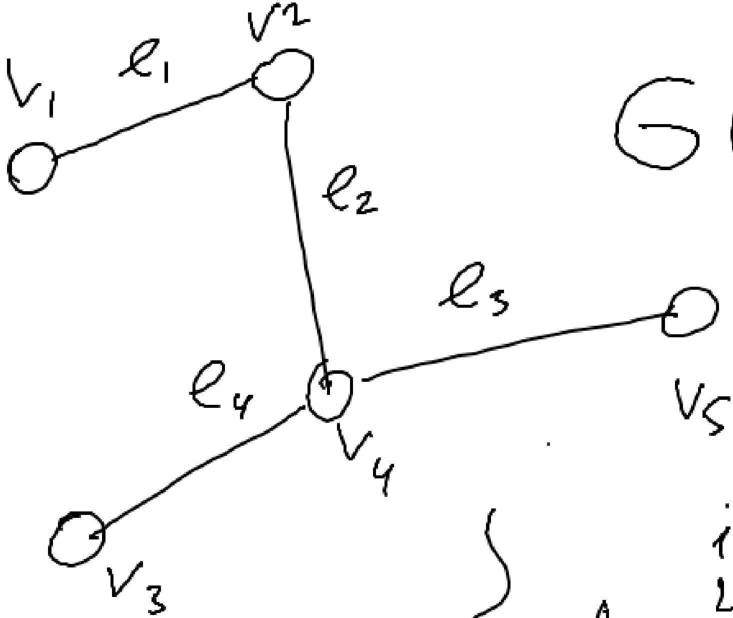
Introduction to Pagerank Algorithm



$$G(V, e)$$

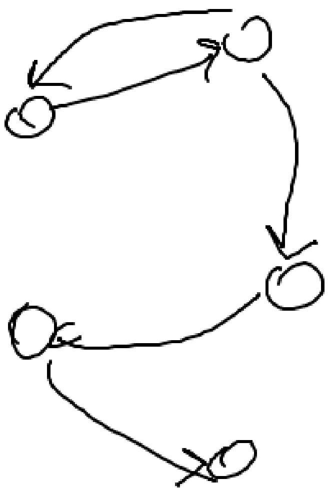
$$\underbrace{v_1, \dots, v_n}_{\text{вершины}}$$

$$\Downarrow$$

$$A \in \mathbb{R}^{n\times n}, \quad A_{ij} = \begin{cases} 1, \text{ если } v_i \text{ и } v_j \text{ связаны} \\ 0, \text{ если не связаны ребром} \end{cases} \quad (*)$$

$$A = \begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 1 & 0 \\ 4 & 0 & 1 & 1 & 0 & 1 \\ 5 & 0 & 0 & 0 & 1 & 0 \end{array}$$

матрица связности графа G

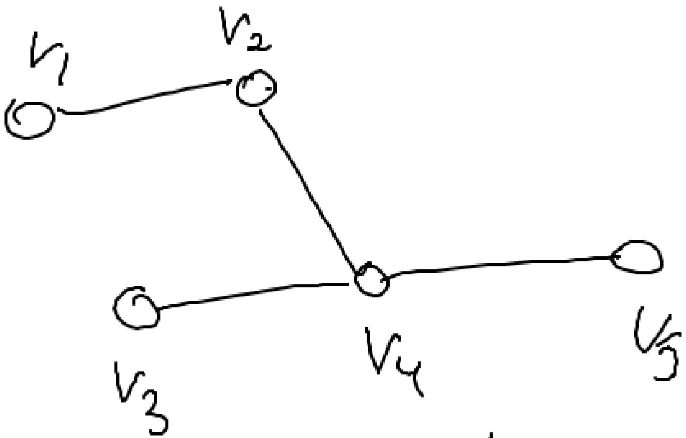


Если граф ненапр., то

$$A = A^T$$

иначе

$$A \neq A^T$$

① 

$G(V, e) \sim G(V', e')$

переименование вершин

A на уровне матриц

P - матрица перест.

получается при перестановке столбцов I, $P^T = P^{-1}$

$PP^T = I$ ← ортог.

преобр. подобие

$\underbrace{P^T A_G P}_{\text{матр. связности}} = A_{G'}$ $G \sim G'$

для ненаправ. графа

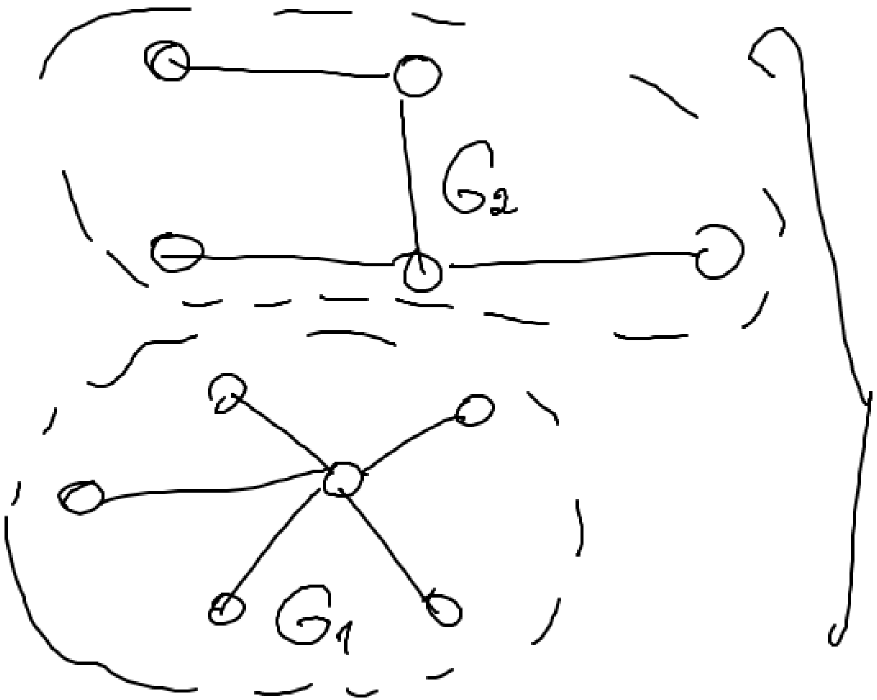
$A = A^T$, то $\lambda_i(A) \in \mathbb{R}$

Т. Шура (Shur th.) (см. лин. алг.)

Инварианты к подобию

① $\sum_{i=1} a_{ii} = tr(A_G) = tr(A_{G'})$

② $\lambda_i(A_G) = \lambda_i(A_{G'})$ - спектр



G

Если в графе неск. комп. связности, то анализ проводить надо раздельно

reducible (редуцируемое)

$$A = \begin{bmatrix} A_{G_1} & \Theta \\ \Theta & A_{G_2} \end{bmatrix}$$

нули

! Далее предполагаем, что G-связный

Если матр. соотв. связному G, то $\nexists P : P^T A P = \begin{bmatrix} A_{G_1} & \Theta \\ \Theta & A_{G_2} \end{bmatrix}$

1) $A^T = A$ $\in \mathbb{R}^{n \times n}$ - irreducible

2) $\Downarrow$ $A_{ij} \geq 0$ по построению

3) $\lambda_i(A) \in \mathbb{R}$ - упорядочим по убыв.

4) v_i - собств. векторы - ортонорм. $\in \mathbb{R}^n$

Т. Перрона - Фробениуса

Если A: $A_{ij} \geq 0$, то

Следствие!

$\exists \lambda_{max}$! $\lambda_{max} > \lambda_2 \ldots > \lambda_n$

$\lambda_{max} > 0$

$\lambda_1 = \lambda_{max}$ имеет кратность алг., равную 1

$\Downarrow$

Ему соотв. единств. с.з. V_1

$V_1 \in \mathbb{R} \Rightarrow V_{1_i} \geq 0$!

Как A_B привести к такому виду?

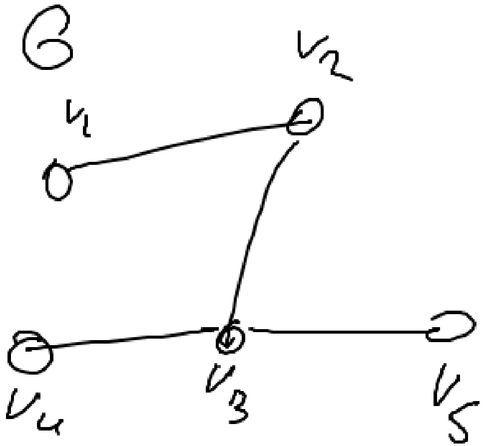
Идея!

Если бы

$\sum_{j=1}^{n} A_{ij} = 1, \forall i$, то $\lambda_{max} = 1$

$$\min_{i=1..n} \sum_{j=1}^{n} A_{ij} \leq \lambda_1 \leq \max_{i..n} \sum_{j=1}^{n} A_{ij} \; !$$

$$A^T = A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \rightsquigarrow \tilde{A} = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 0 & 1 & 1 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \end{bmatrix}$$



$$\lambda_{max}(\tilde{A}) = 1 > \lambda_i$$

$i = 2, \dots n$

простое

будет делать итерации

Цель : найти v_1

$$\det(A - \lambda I) = 0$$

$$(A - \lambda I)x = 0 \qquad \lambda = 1$$

Алг. Гаусса ($O(n^3)$)

$(A - I)x = 0$ – решить лин. сист.

$A = A^T \Rightarrow \exists$ ортонорм. базис из собств. векторов, то

$\Downarrow$

$\forall x = \sum_{i=1}^{n} \alpha_i v_i$ — веса

$\theta \neq x \in \mathbb{R}^n$

$A v_i = \lambda_i v_i$

$v_i \perp v_j, \ \|v_i\|_2 = 1$

$\boxed{\lambda_1, v_1}$

$\lambda_1 > \lambda_2, \lambda_3, \dots, \lambda_n$

$\Downarrow \ \lambda_1 = 1$

$\boxed{\lambda_i < 1, \ i = 2, n}$

$\alpha_i \cdot \lambda_i < \alpha_i$

$\wedge$

1

$$Ax = A\left(\sum_{i=1}^{n} \alpha_i v_i\right) =$$

$$= \sum \alpha_i A v_i = \sum \alpha_i \lambda_i v_i = \alpha_1 v_1 + \sum_{i=2}^{n} \alpha_i \lambda_i v_i =$$

$$= \underline{\alpha_1 v_1} + \sum \beta_i v_i = \underline{x^{(k+1)}}, \quad \frac{x^{(k+1)}}{\|x^{(k+1)}\|_2} = y^{(k+1)} \vee$$

Схема (метод простой итерации) power method

$x^{(0)}$

$y^{(k+1)} = A x^{(k)}$ ← $O(n^2)$

Для разр. матр. гораздо дешевле!

$x^{(k+1)} = \dfrac{y^{(k+1)}}{\|y^{(k+1)}\|}$ ← нормировка

$\|x^{(k+1)} - x^{(k)}\|_2 < \varepsilon$?

$x^{(k)} = x^{(k+1)}$

⇓ Вернуть $x^{(k+1)}$

⇓

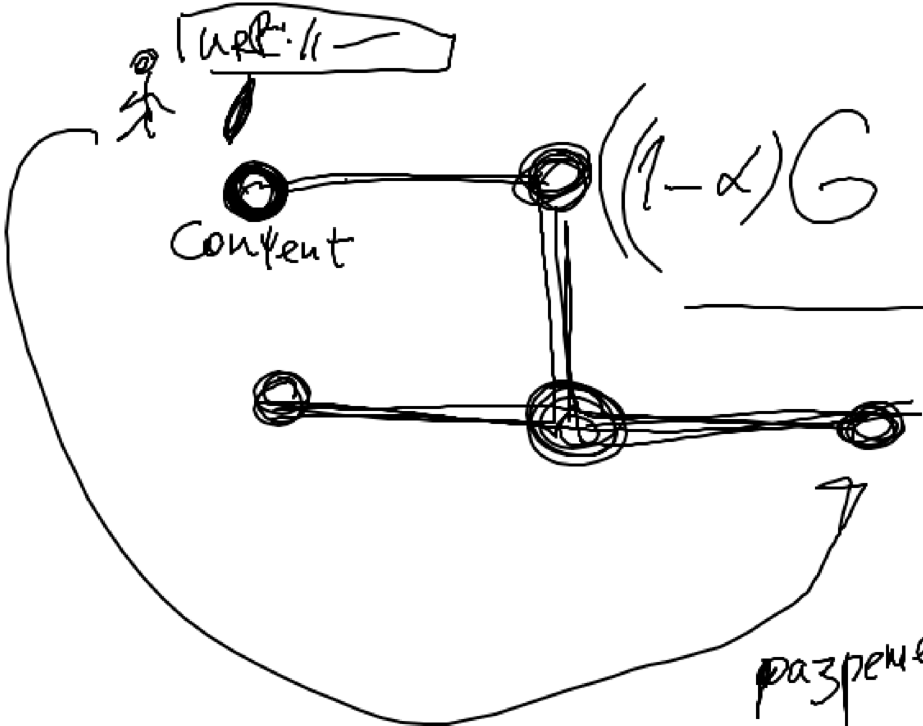
$\lim\limits_{k\to\infty} x^{(k)} = v_1$ $\quad v_{1i} > 0$

на практике

100 – 1000

$n \sim 10^6 - 10^8$

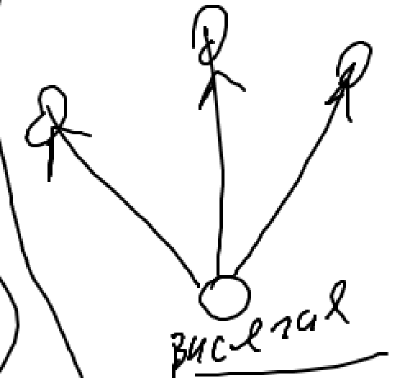
$\dfrac{v_{1i}}{\sum v_{1i}} = pr(A_G)$ ← вектор pagerank



$$\left((1-\alpha)G + \alpha K_n\right) \quad 1 \gg \alpha > 0$$

клика

(полносвязный граф)



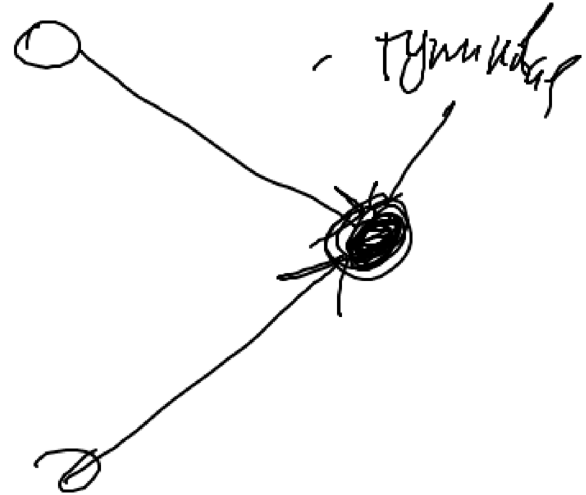
разреженная

D

плотная, но малого ранга!

$$\hat{A} = (1-\alpha)A + \alpha \frac{1}{n} \begin{bmatrix} 1 & \dots & 1 \\ 1 & \dots & 1 \\ \vdots & & \\ 1 & \dots & 1 \end{bmatrix}$$

damping factor



$\operatorname{rank} D = 1$

$$D = \begin{bmatrix} 1 & \cdots & 1 \\ 1 & & \vdots \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \underbrace{\underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{n\times 1} \times \underbrace{\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}}_{1\times n}}_{n\times n}$$

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \left(\underbrace{\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{1} \right)$$

$$\overbrace{1\times n \times n\times 1}^{1}$$

$$\underset{n\times 1}{n\times 1} \times \underbrace{}_{1}$$

базисный

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$Dx = \left(\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \left(\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} x \right) \right)$$

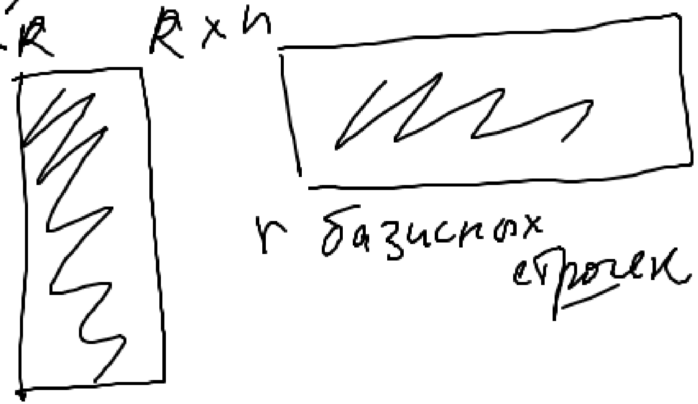
за $O(n)$!

Расширение этих идей:

rank $B = R$, то

$\mathbb{R}^{n\times n} \ni B = UV$

$n \times R$ $\quad R \times n$



$n^2 \to 2nR$ ячеек

r базисных столбцов

r базисных строчек

nR

$U(Vx)$, $\quad Vx = y$

nR

В результате Bx стоит $O(nR)$ вместо $O(n^2)$

numpy.linalg.svd

$B = \underbrace{U\Sigma V^*}$

! сингулярное разложение

$\sigma_i > 0$

$U \in \mathbb{R}^{n\times R}$ $\quad \Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_R \end{pmatrix}$

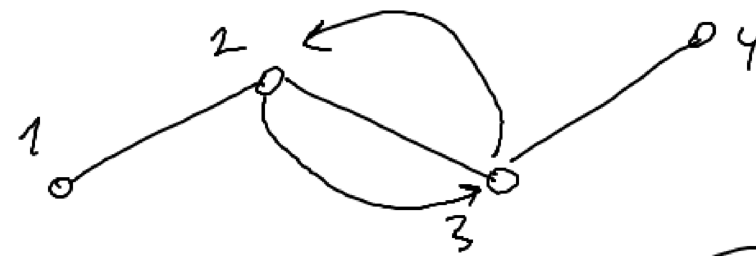
$V \in \mathbb{R}^{R\times n}$

$m=2$

$2\ 3\ 2$

A – матрица связности для G

G – ненапр. граф



$A = A^T$, то

$(A^m)_{ij}$ – дают нам количество маршрутов длины m из i в j

за $O(\log m)$ matmul

$A^4 = (A^2)(A^2)$

m – ?

A^m

Далее можно угадывать, что диаметр. графа, и его связность.