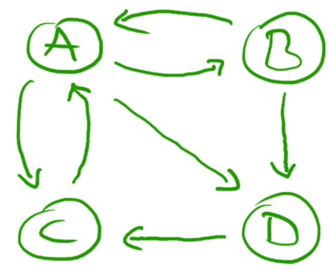
PageRank algorithm

1998: Google founded by Larry Page + Sergey Brin

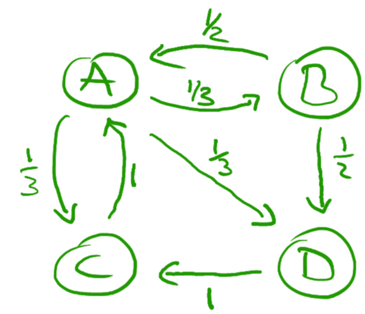
EG rank webpages A, B, C, D by computing the PageRank vector



idea surfer randomly clicking on links

⇝ rank according to how frequently surfer is on each page

a_t : probability surfer is on page A at time t
likewise b_t, c_t, d_t for B, C, D



transition from one click to next :

$$\begin{bmatrix} a_{t+1} \\ b_{t+1} \\ c_{t+1} \\ d_{t+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} b_t + 1\, c_t \\ \frac{1}{3} a_t \\ \frac{1}{3} a_t + 1\, d_t \\ \frac{1}{3} a_t + \frac{1}{2} b_t \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} a_t \\ b_t \\ c_t \\ d_t \end{bmatrix}$$

Markov chain
cols sum to 1

transition matrix

PageRank vector
= equilibrium state is a 1-eigenvector of transition matrix

$$= \frac{3}{16} \begin{bmatrix} 2 \\ 2/3 \\ 5/3 \\ 1 \end{bmatrix} \approx \begin{bmatrix} 0.375 \\ 0.125 \\ 0.313 \\ 0.188 \end{bmatrix}$$

$$\text{null}\left(\begin{bmatrix} -1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & -1 & 0 & 0 \\ \frac{1}{3} & 0 & -1 & 1 \\ \frac{1}{3} & \frac{1}{2} & 0 & -1 \end{bmatrix} \right) \quad \text{basis: } \begin{bmatrix} 2 \\ 2/3 \\ 5/3 \\ 1 \end{bmatrix}$$

corresponding ranking:
A > C > D > B

$$2 + \frac{2}{3} + \frac{5}{3} + 1 = \frac{16}{3}$$

to think about : webpages without links, avoiding Gaussian elimination, ...