Applications of Mathematics in Computer Science (MACS)

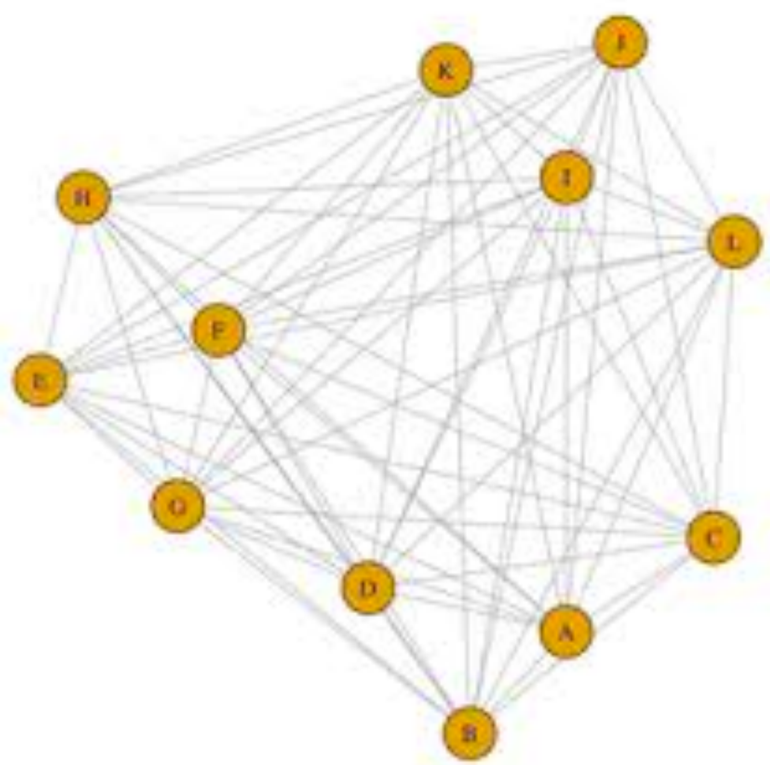
LECTURE #3: GOOGLE PAGERANK

The CS Problem:

- Searching in Google for "google pagerank" has 10,800,000 results
- Google has to present them in some order

How do we rank the page results?

The (abstract) CS Problem:

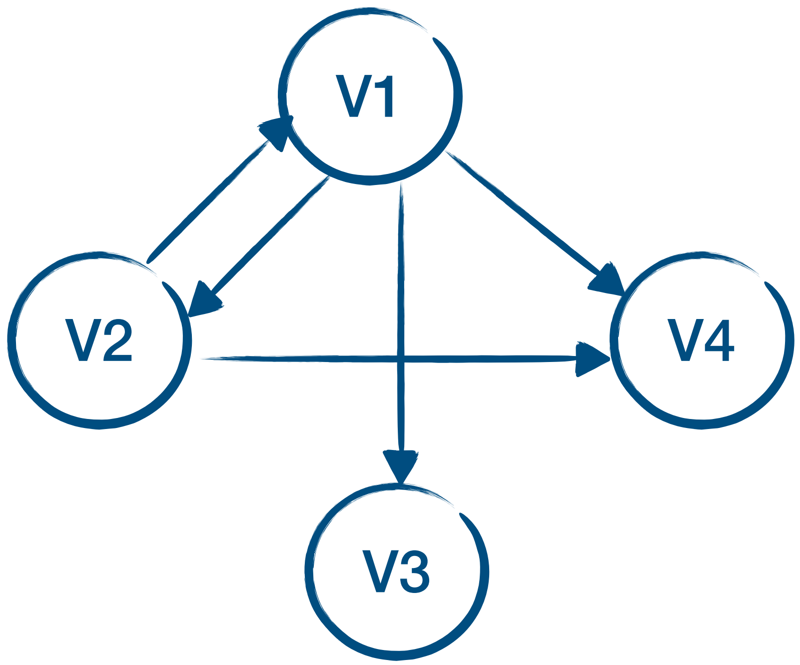


Given a graph

webpages / scientific papers / people relations / …

How do we rank nodes in the graph?

How to represent graphs?



$$V = \{v_1, v_2, v_3, v_4\}$$

$$E = \{e_{1,2}, e_{2,1}, e_{1,3}, e_{1,4}, e_{2,4}\}$$

Graph adjacency matrix

$$a_{ij} \neq 0$$

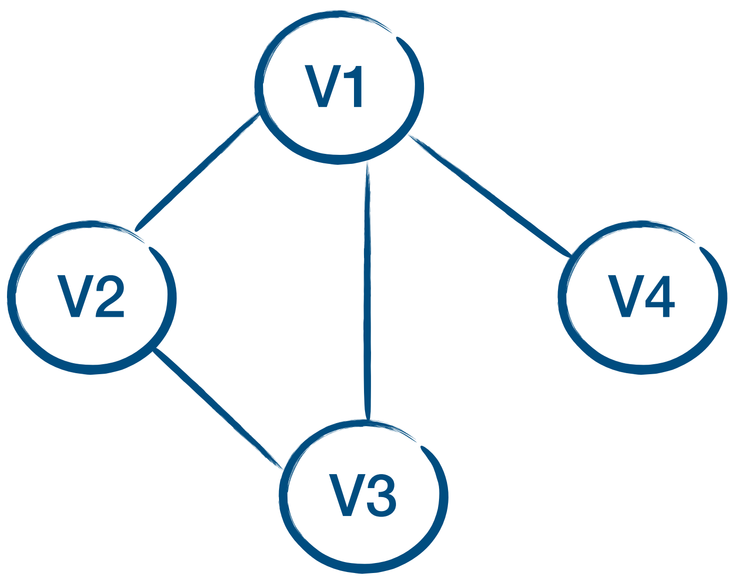
$\Updownarrow$

There is an edge from v_i to v_j

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Variations

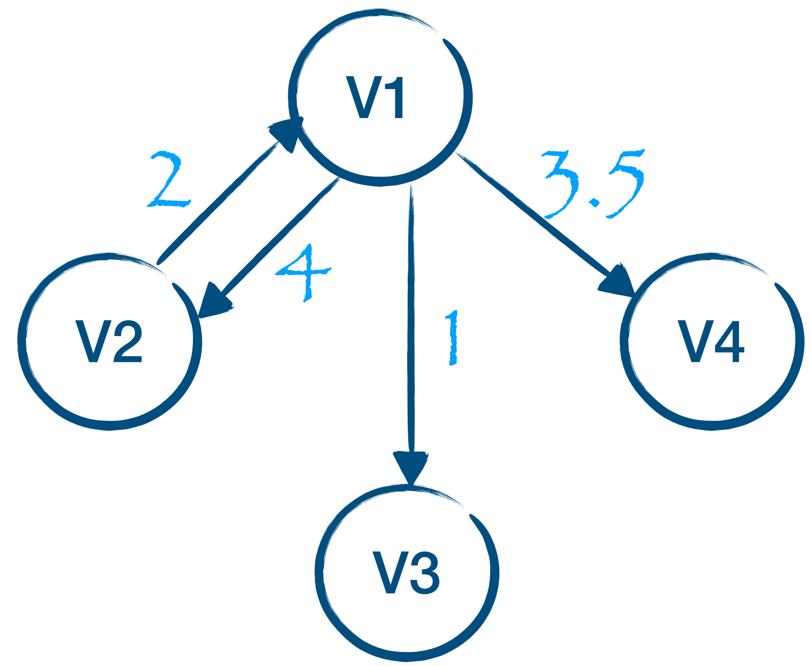
Undirected graph



$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

A is symmetric

Graph with weights



$$A = \begin{pmatrix} 0 & 4 & 1 & 3.5 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The (concrete) Problem:

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix}$$

Given an adjacency matrix with weights

How do we rank nodes?

The math technique:

Matrices

(Linear Algebra)

Vectors and matrices

Vector (1-dimensional)

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

Matrix (2-dimensional)

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix}$$

Scalar

$$\lambda = (\lambda)$$

a vector (or a matrix) with 1 element

Matrix multiplication

$$C = A \cdot B \quad \longleftrightarrow \quad c_{ij} = \sum_k a_{ik} b_{kj}$$

$$(2)\,(3) = (6)$$

$$\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ -3 & 0 \end{pmatrix}$$

Matrix inverse

$$A \cdot A^{-1} = I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 0.4 & -0.2 \\ 0.2 & 0.4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Eigenvalues and eigenvectors

$$Av = \lambda v$$

Eigenvector

Eigenvalue

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Eigenvector

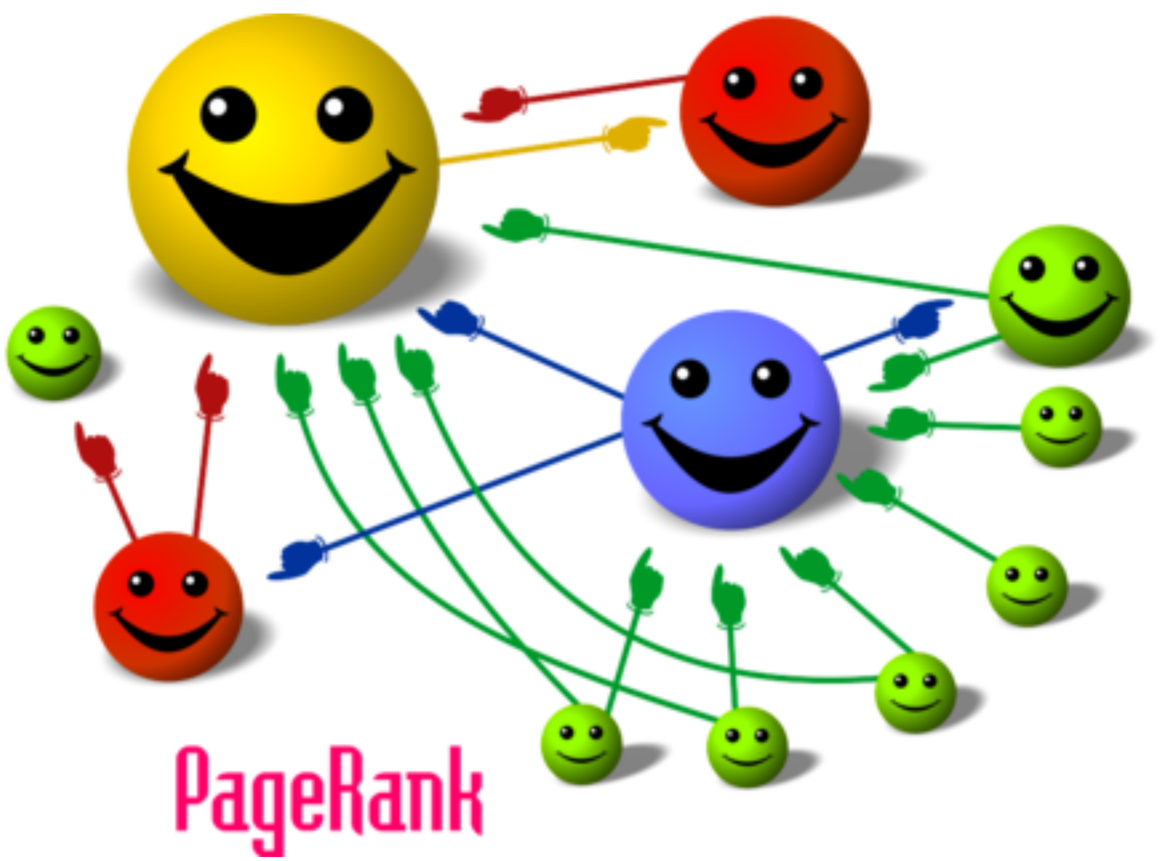
$\lambda = 0$

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$$

Eigenvector

$\lambda = 2$

Google PageRank



The CS Problem:

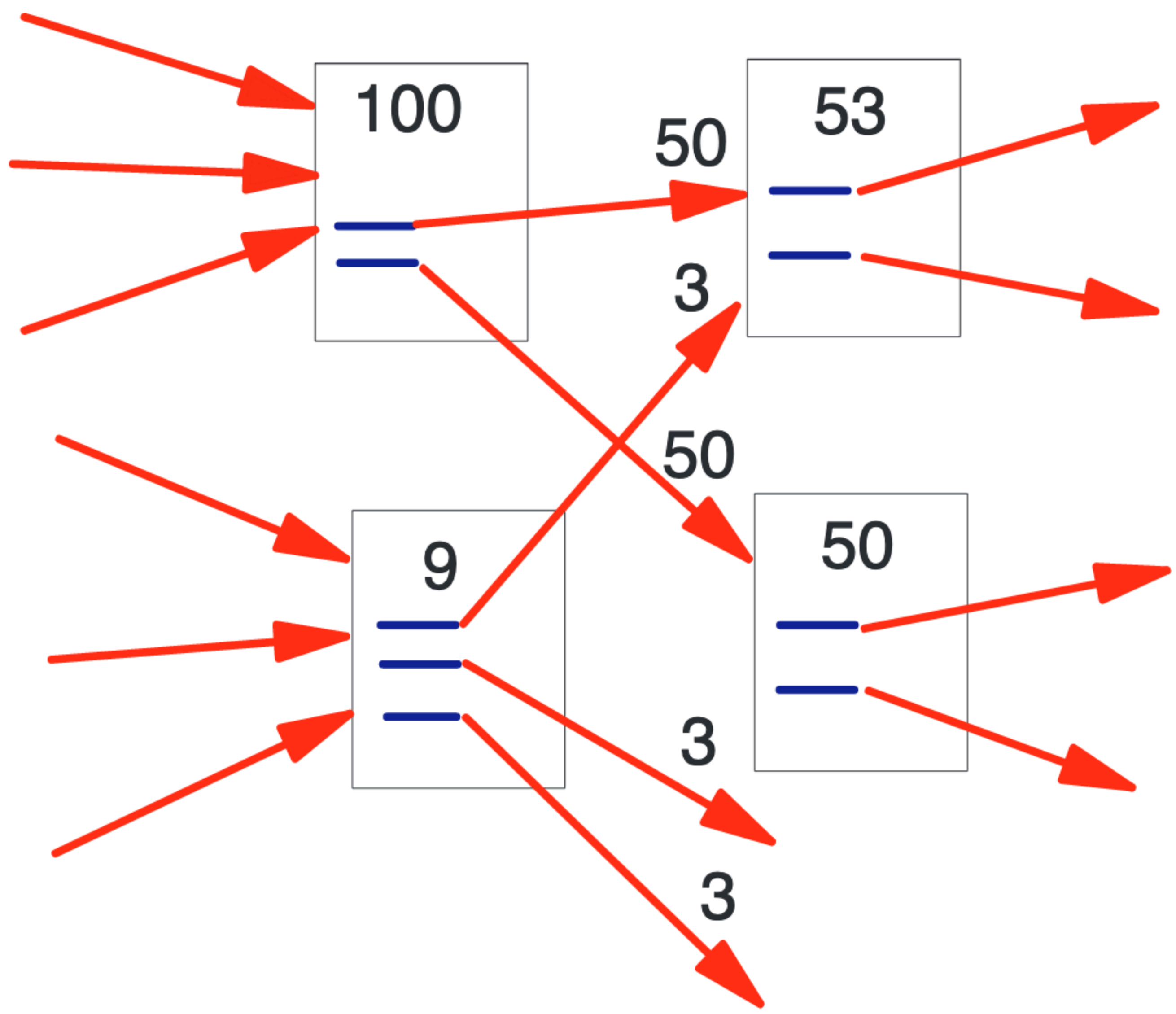
- Searching in Google for "google pagerank" has 10,800,000 results
- Google has to present them in some order

How to assign page importance?

PageRank (R)

Measures centrality

Simplified PageRank



$$R(u) = \sum_{\forall(v,u)} \frac{R(v)}{Nv}$$

Simplified PageRank

$$R = dAR \quad , \quad d \leq 1$$

Dumping factor

$$a_{ij} = \begin{cases} \dfrac{1}{N_i} & \text{if edge (j,i) in the graph} \\ 0 & \text{otherwise} \end{cases}$$

$$AR = \frac{1}{d}R$$

$\boxed{R \text{ is an eigenvector with eigenvalue } \frac{1}{d}}$

PageRank

$$R = \frac{1-d}{N}\mathbf{1} + dAR \quad , \quad d \leq 1$$

Computing PageRank

Algebraic: $$R = (\mathbf{I} - \mathbf{dA})^{-1} \cdot \frac{1-d}{N}\mathbf{1}$$

Iterative:

1. Set $\forall i \,.\, R_i = \dfrac{1}{N}$
2. At each step $R' = \dfrac{1-d}{N}\mathbf{1} + dAR$
3. Stop when $|\, R' - R \,| < \varepsilon$
4. $R := R'$

Let's try…