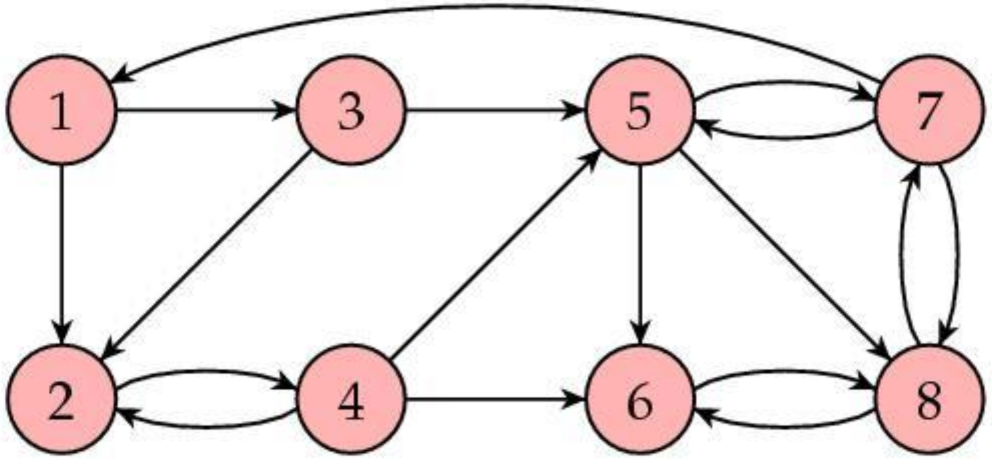
PageRank and Circularity

$P(d|q) \propto P(q|d) \cdot p(d)$

$$\boxed{I(P_i) = \sum_{P_j \in B_i} \boxed{\frac{I(P_j)}{l_j}}}$$

A page is important if it is pointed to by several important pages.



Example

$$\boxed{I = HI}$$

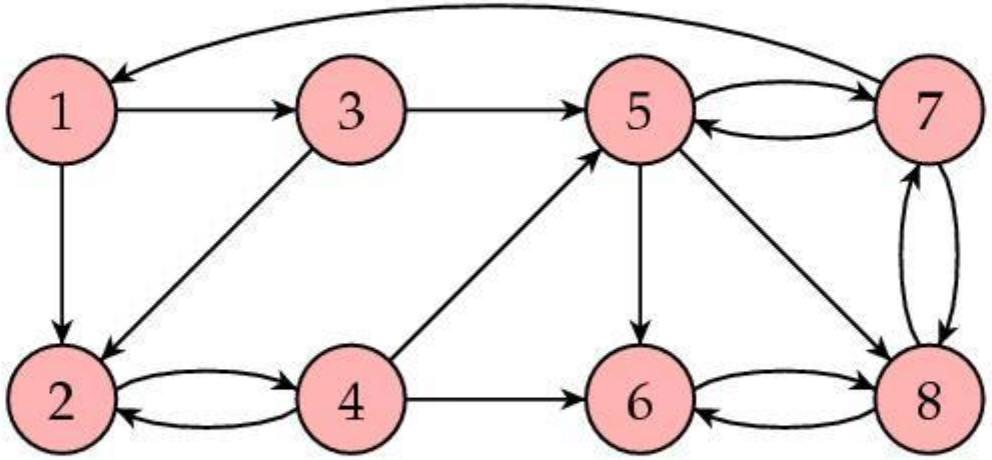
$$HI = \alpha I$$

$$\alpha = 1$$

$$I(P_1) = \frac{1}{3} I(P_7)$$

$$I(P_2) = \frac{1}{2} I(P_1) + \frac{1}{2} I(P_3) + \frac{1}{3} I(P_4)$$

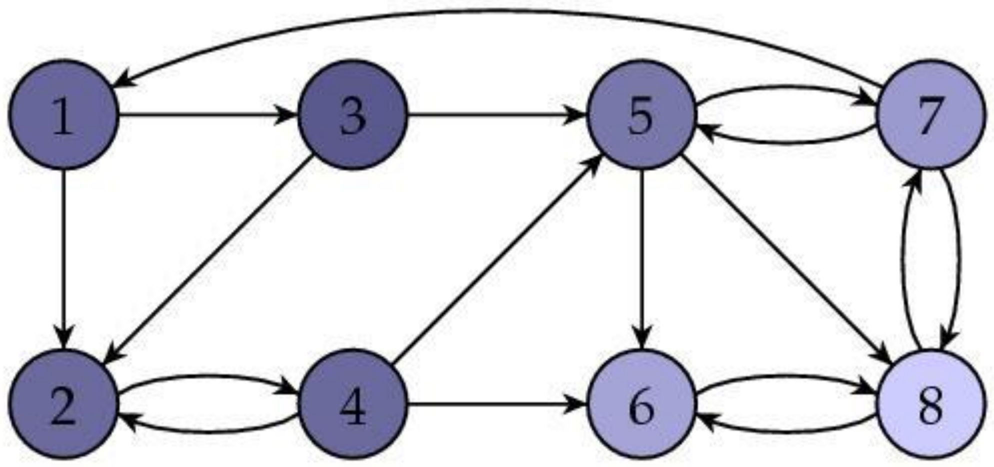
$$I(P_8) = \ldots$$



$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1/3} & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix} \qquad I = \begin{bmatrix} 0.0600 \\ 0.0675 \\ 0.0300 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.1800 \\ 0.2950 \end{bmatrix}$$

$$\boxed{\text{Stochastic}}$$

Example



$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix} \qquad I = \begin{bmatrix} 0.0600 \\ 0.0675 \\ 0.0300 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.1800 \\ 0.2950 \end{bmatrix}$$

The Power Method

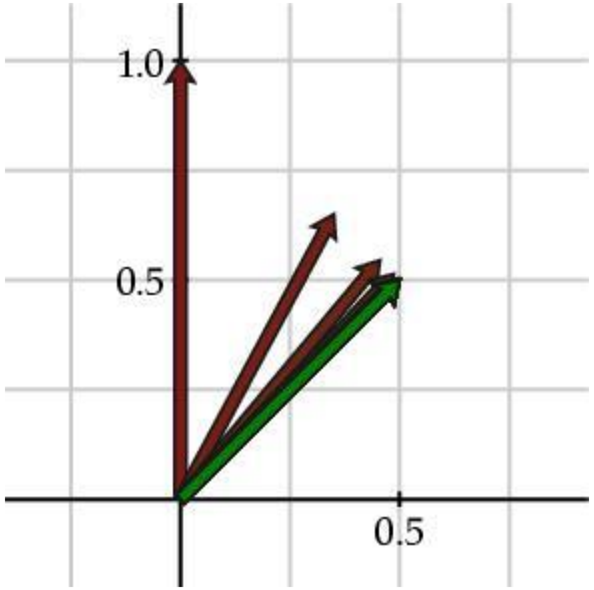
$I = HI$

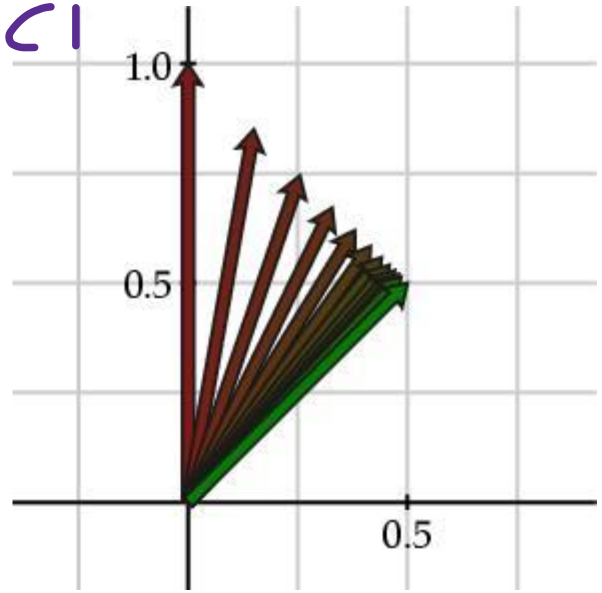
S

Stationary

$$
\begin{aligned}
I^0 &= c_1 v_1 + c_2 v_2 + \ldots + c_n v_n \\
I^1 = \mathbf{S} I^0 &= c_1 v_1 + c_2 \lambda_2 v_2 + \ldots + c_n \lambda_n v_n \\
I^2 = \mathbf{S} I^1 &= c_1 v_1 + c_2 \lambda_2^2 v_2 + \ldots + c_n \lambda_n^2 v_n \\
\vdots \quad & \quad \vdots \\
I^k = \mathbf{S} I^{k-1} &= c_1 v_1 + c_2 \lambda_2^k v_2 + \ldots + c_n \lambda_n^k v_n
\end{aligned}
$$

$\lambda_2 < 1$





Computing the eigenvector

$$I^{k+1} = \mathbf{H}I^k$$

General principle: *The sequence I^k will converge to the stationary vector I.*

I^0	I^1	I^2	I^3	I^4	...	I^{60}	I^{61}
1	0	0	0	0.0278	...	0.06	0.06
0	0.5	0.25	0.1667	0.0833	...	0.0675	0.0675
0	0.5	0	0	0	...	0.03	0.03
0	0	0.5	0.25	0.1667	...	0.0675	0.0675
0	0	0.25	0.1667	0.1111	...	0.0975	0.0975
0	0	0	0.25	0.1806	...	0.2025	0.2025
0	0	0	0.0833	0.0972	...	0.18	0.18
0	0	0	0.0833	0.3333	...	0.295	0.295

Three Questions

- Does the sequence I^k always converge?
- Is the vector to which it converges independent of the initial vector I^0?
- Do the importance rankings contain the information that we want?

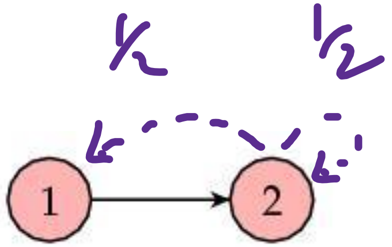
Problem 1

$$\mathbf{H} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

I^0	I^1	I^2	$I^3 = I$
1	0	0	0
0	1	0	0

Solution

$$\begin{bmatrix} \frac{1}{10^6} \\ \\ \end{bmatrix}$$



$$H = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

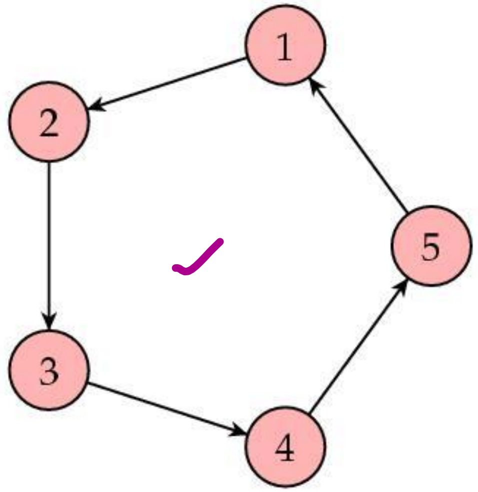
$$A = \begin{bmatrix} 0 & 1/2 \\ 0 & 1/2 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 0 & 1/2 \\ 1 & 1/2 \end{bmatrix} \qquad I = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix}$$

$$S = (H + A)$$

Problem 2

primitive



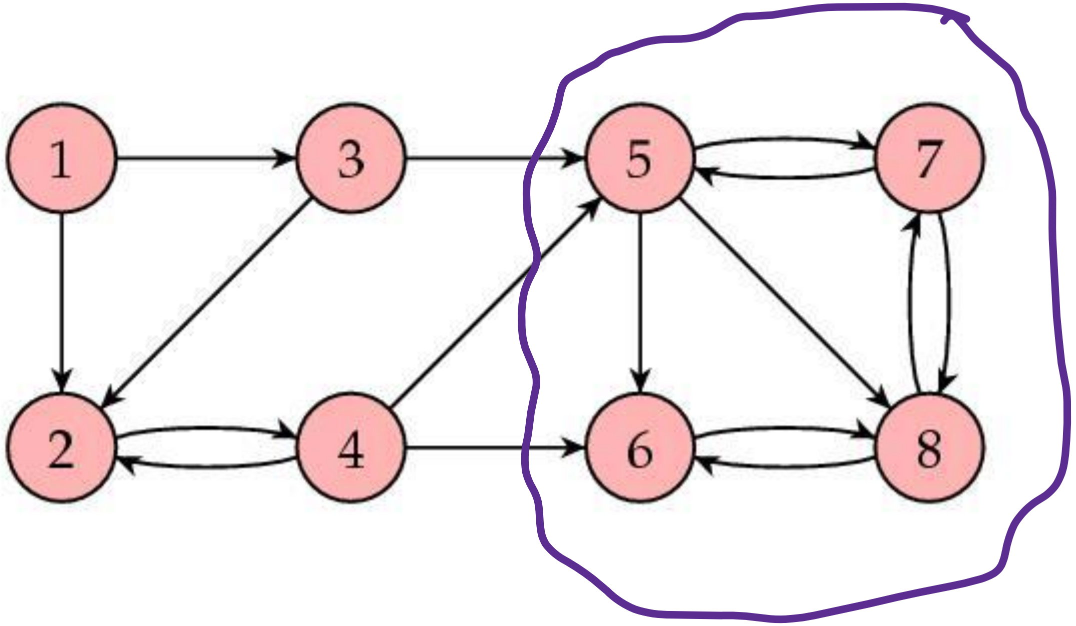
1 . . . .

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

I^0	I^1	I^2	I^3	I^4	I^5
1	0	0	0	0	1
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0

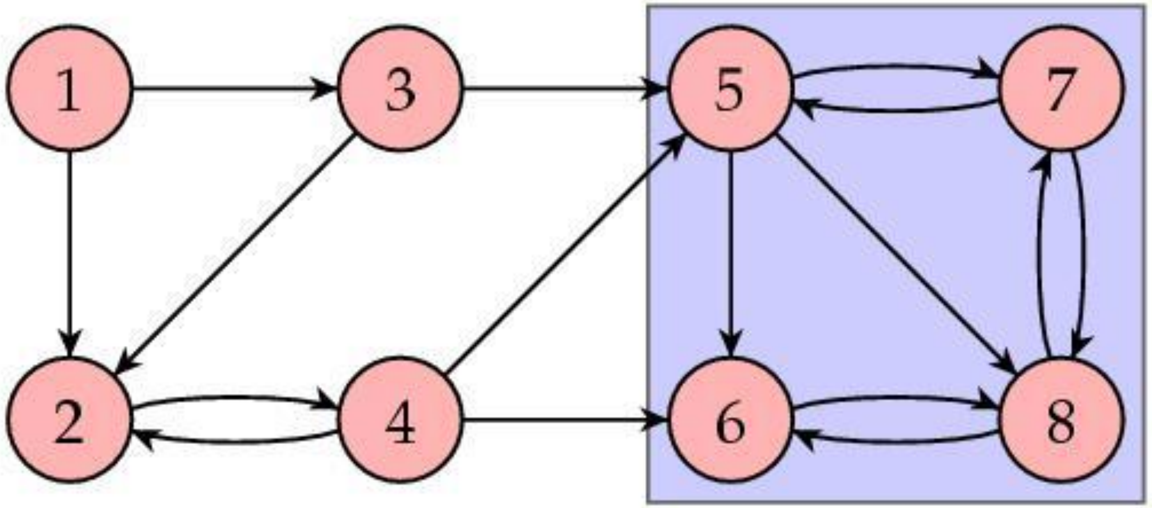
Problem 3

Irreducible



$$
\mathbf{S} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/2 & 0 \\
0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 1/3 & 1 & 1/2 & 0
\end{bmatrix}
\qquad
I = \begin{bmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0.12 \\ 0.24 \\ 0.24 \\ 0.4
\end{bmatrix}
$$

Observation



Google Equations

stochastic ✓

irreducible

primitive

$$\begin{bmatrix} 1/n & \cdot & \cdots & \cdot \\ 1/n & & & \cdot \\ & & & \vdots \\ 1/n & & & 1/n \end{bmatrix}$$

$$\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)\frac{1}{n}\mathbf{1}$$

$$\boxed{\mathbf{S} = \mathbf{H} + \mathbf{A}}$$

$$\mathbf{G} = \alpha\mathbf{H} + \alpha\mathbf{A} + \frac{1-\alpha}{n}\mathbf{1}$$

$$\mathbf{G}I^k = \alpha\mathbf{H}I^k + \alpha\mathbf{A}I^k + \frac{1-\alpha}{n}\mathbf{1}I^k$$

Seed value

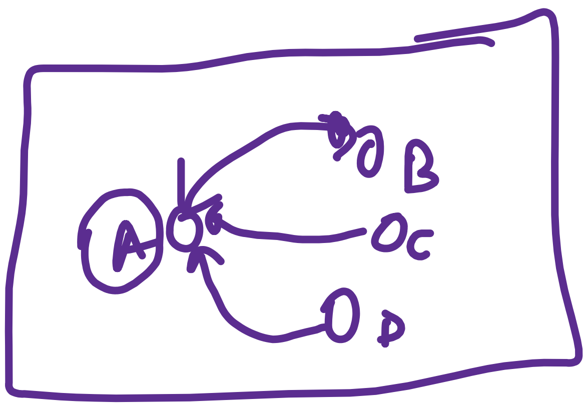
$[\alpha]$

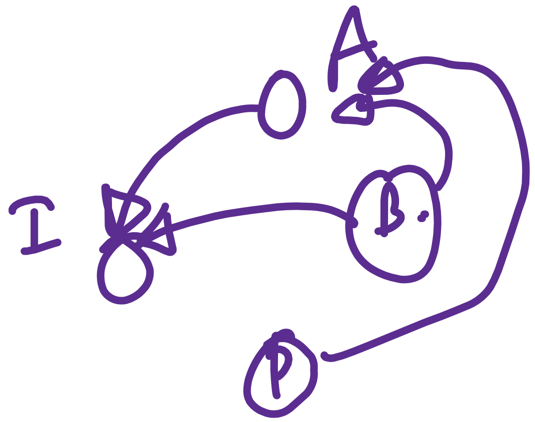
$\alpha S + (1-\alpha)\frac{1}{n}$

$[0,1]$

Applications

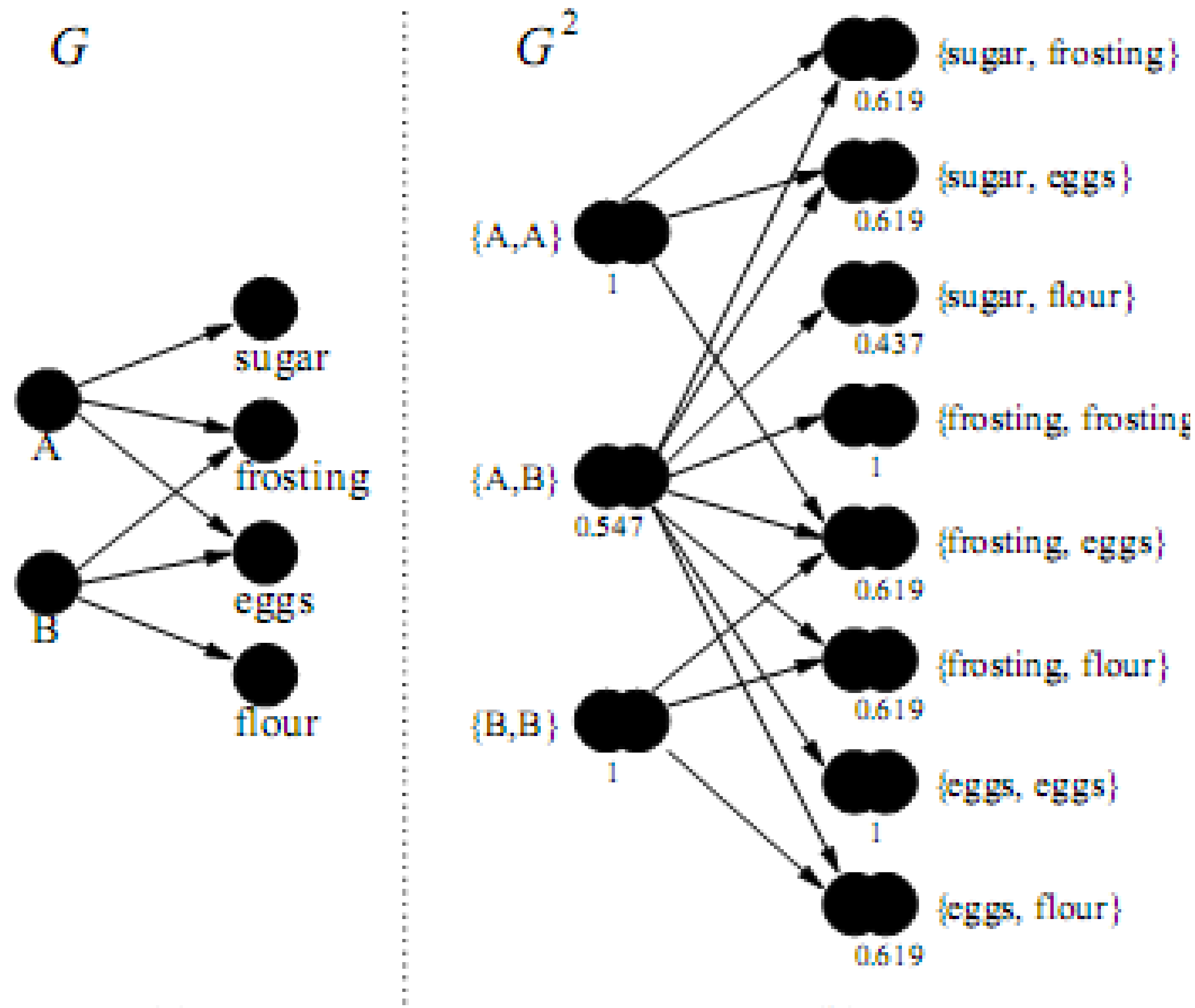
- Web Search
- Word Meanings
- Word Sense Disambiguation
- SentiWordNet
- SimRank





Problem Setting where PageRank is useful : Global ranking to be mined from pairwise preferences

SimRank



(a) (b)

$$s(A,B) = \frac{C_1}{|O(A)||O(B)|} \sum_{i=1}^{|O(A)|} \sum_{j=1}^{|O(B)|} s(O_i(A), O_j(B))$$

$$s(c,d) = \frac{C_2}{|I(c)||I(d)|} \sum_{i=1}^{|I(c)|} \sum_{j=1}^{|I(d)|} s(I_i(c), I_j(d))$$