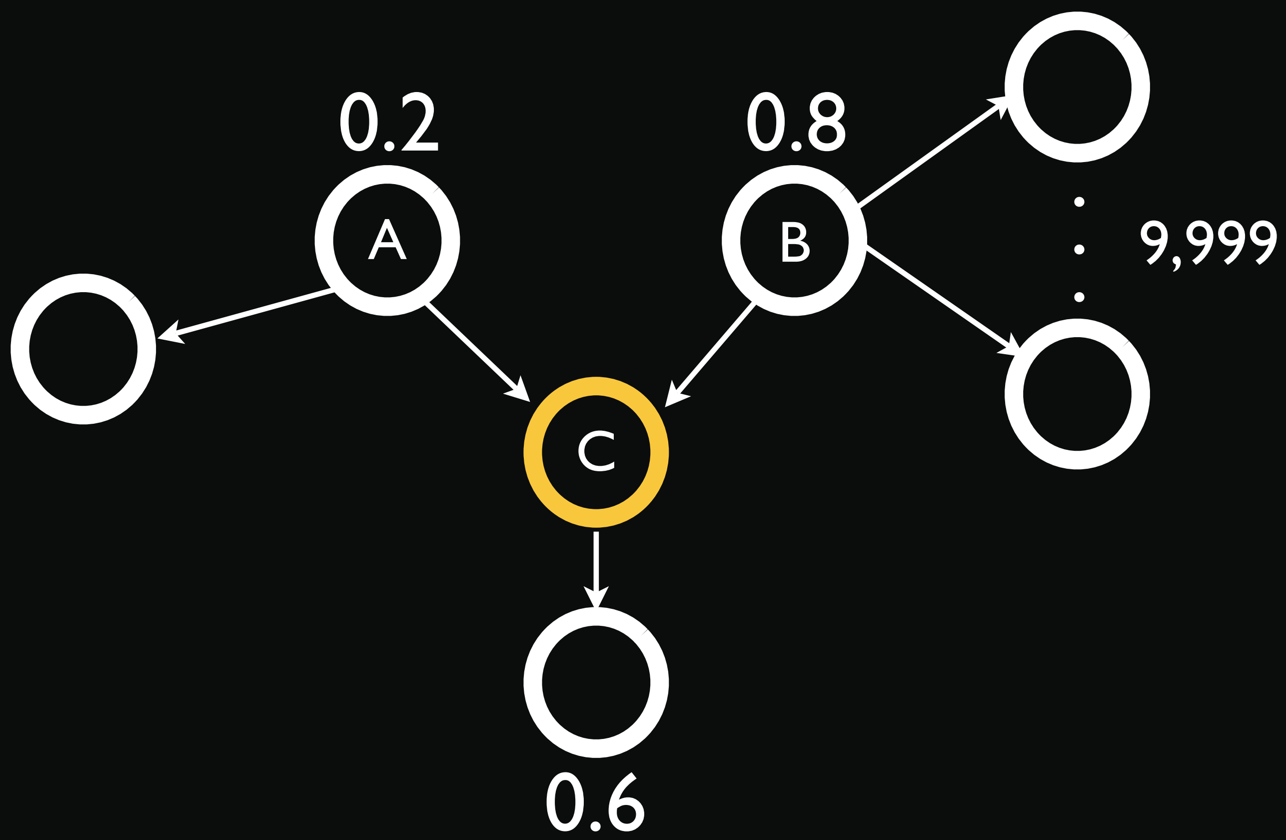
Importance

$$r^{(t+1)}(P_i) = \sum_{j \in E(i)} r^{(t)}(P_j) \,/\, l(P_j)$$



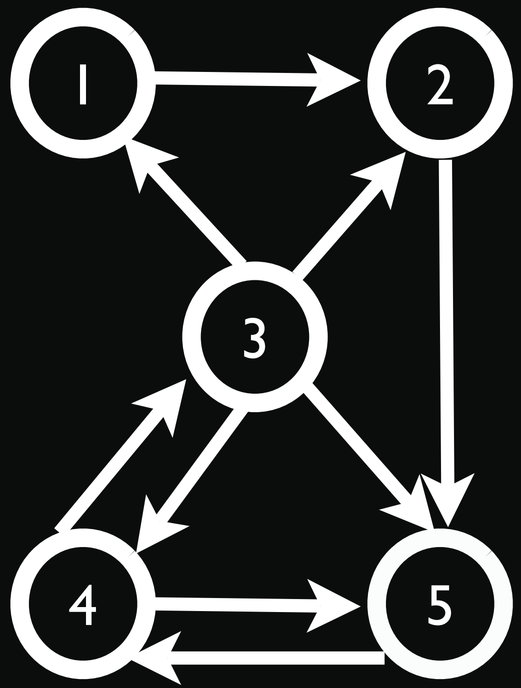
The Algorithm

Given a web graph with n nodes, where the nodes are pages and edges are hyperlinks

- Assign each node an initial page rank
- repeat until convergence

    calculate the page rank of each node (using the equation in the previous slide)

Example



	Iteration 0	Iteration 1	Iteration 2	Page Rank
P_1	1/5	1/20	1/40	5
P_2	1/5	5/20	3/40	4
P_3	1/5	1/10	5/40	3
P_4	1/5	5/20	15/40	2
P_5	1/5	7/20	16/40	1

$$r_1(P_5)=1/5 + 1/5\times 1/4 + 1/5 \times 1/2 = 7/20$$

Matrix representation

$$[1/20 \quad 5/20 \quad 1/10 \quad 5/20 \quad 7/20] = [1/5 \quad 1/5 \quad 1/5 \quad 1/5 \quad 1/5] \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$r_{(t+1)}^T \qquad = \qquad r_{(t)}^T \qquad H$$

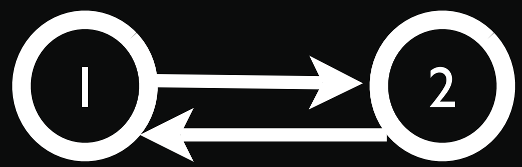
Three Questions

$$r_{(t+1)}^T = r_{(t)}^T H$$

Also known as the power method

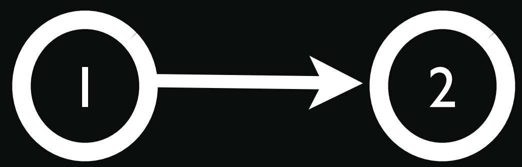
- Does this converge?
- Does it converge to what we want?
- Are the results reasonable?

Does it converge?



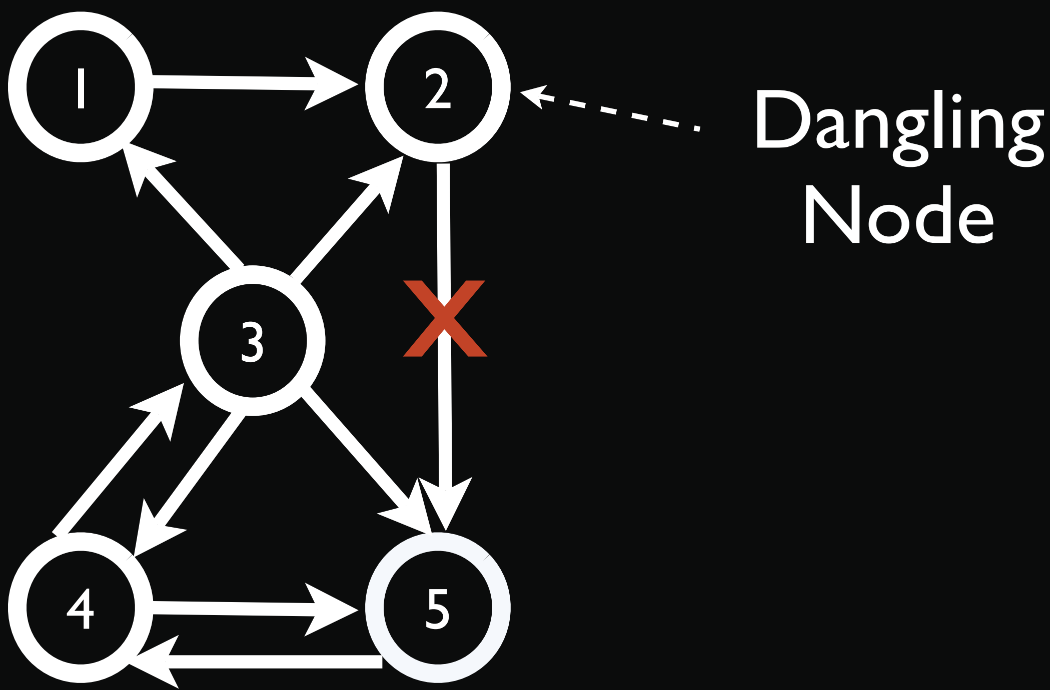
	Iteration 0	Iteration 1	Iteration 2	Iteration 3
P_1	1	0	1	0
P_2	0	1	0	1

Does it converge to what we want?



	Iteration 0	Iteration 1	Iteration 2	Iteration 3
P_1	1	0	0	0
P_2	0	1	0	0

Does it converge to what we want?



$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Page ranks to converge to 0.

Looks a lot like ...

$$r_{(t+1)}^T = r_{(t)}^T H$$

Markov Chains

- Set of states X
- Transition matrix P where $P_{ij} = P(X_t=j \mid X_{t-1}=i)$
- π specifying the probability of being at each state $x \in X$

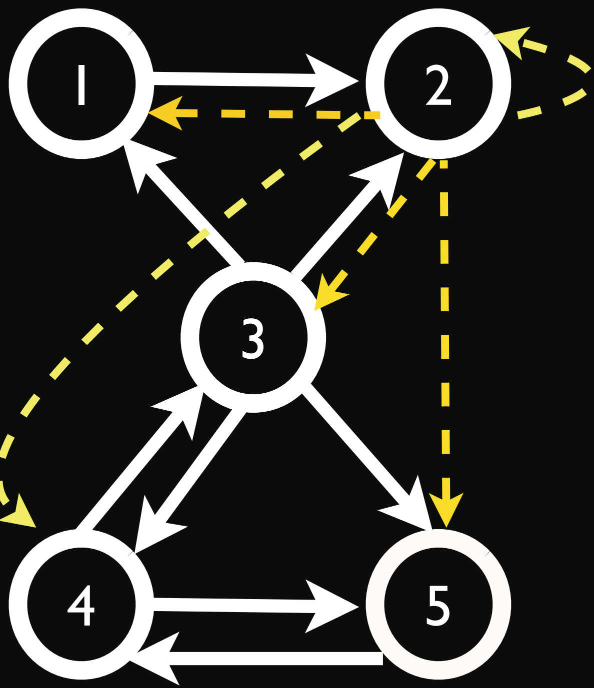
Goal is to find π such that $\pi^T = \pi^T P$

Why is this analogy useful?

There exists a theory about Markov chains that says that for any start vector, the power method applied to a Markov transition matrix P will converge to a unique positive stationary vector as long as P is stochastic, irreducible and aperiodic.

Make H stochastic

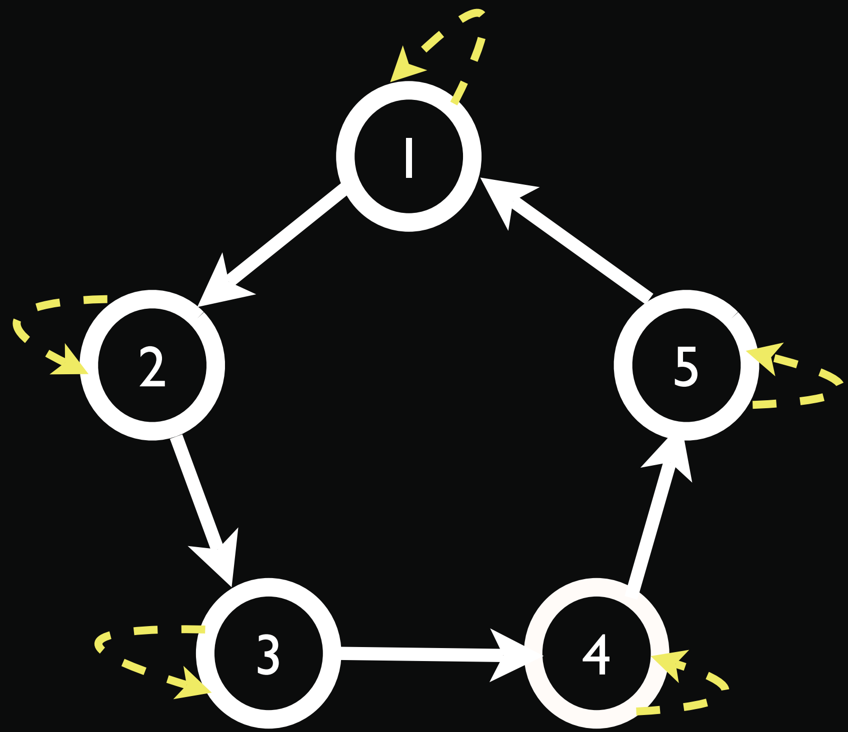
$$S = H + a(1/n\ e^T)$$



$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

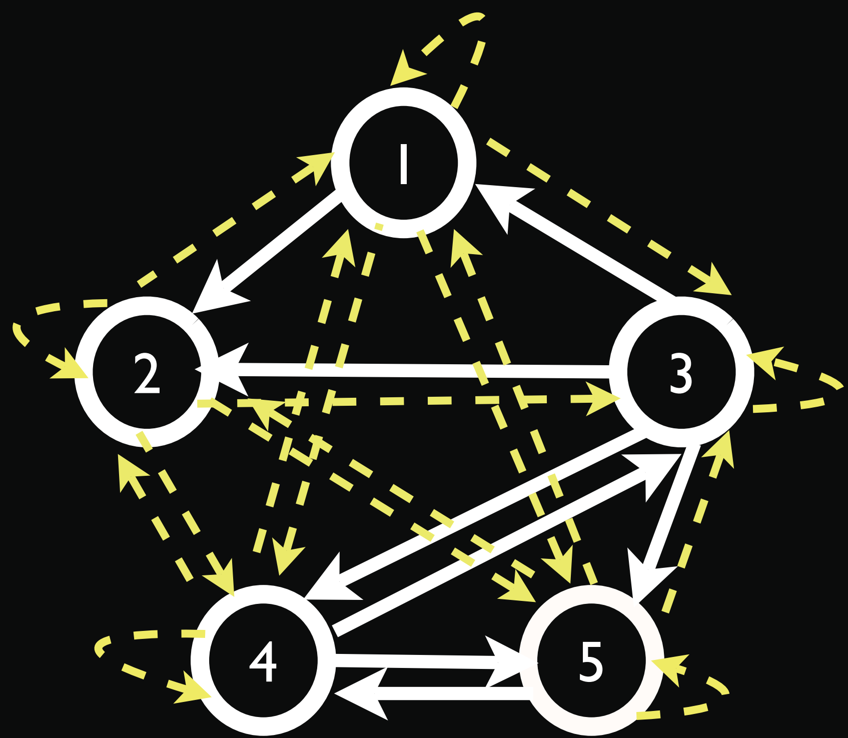
Make H aperiodic

A chain is periodic if there exists k > 1 such that the interval between two visits to some state s is always a multiple of k.



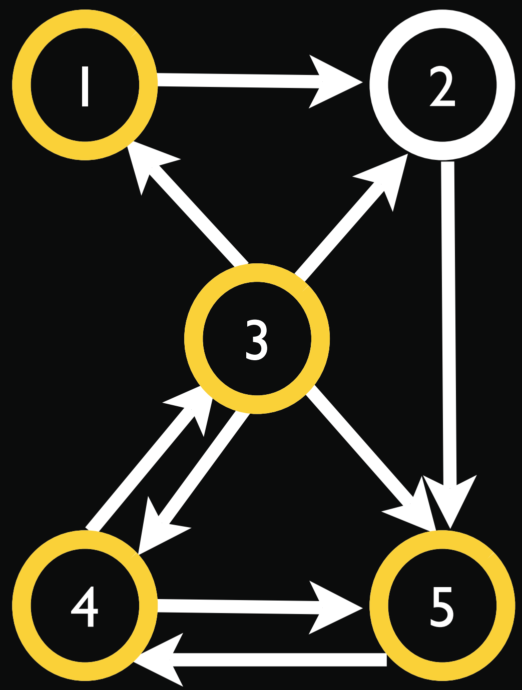
Make H irreducible

From any state, there is a non-zero probability of going from one state to another.



The Google Matrix

$$G = \alpha S + (1-\alpha)\ 1/n\ ee^T$$



The Random Surfer Model: for each page, time spent $\propto$ importance.

Are the results reasonable?

$$G = \alpha S + (1-\alpha)\ 1/n\ ee^T$$

G is stochastic, aperiodic and irreducible.

$$r_{(t+1)}{}^T = r_{(t)}{}^T\, G$$

G is dense but computable using the sparse matrix H.

$$\begin{aligned} G &= \alpha S + (1-\alpha)\ 1/n\ ee^T \\ &= \alpha(H + 1/nae^T) + (1-\alpha)\ 1/n\ ee^T \\ &= \alpha H + (\alpha a + (1-\alpha)e)\ 1/n\ e^T \end{aligned}$$