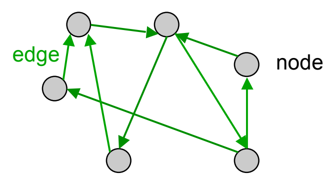
# PageRank

## Goal

- Given a directed graph with *n* nodes
- Assign each node a score that represents its importance in structure
  - Call score PageRank: pr(node)




1

# Conferring importance

Core ideas:

- A node should confer some of its importance to the nodes to which it points
  - If a node is important, the nodes it links to should be important
- A node should not transfer more importance than it has

2

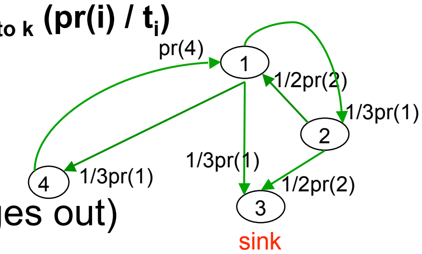
# Attempt 1

Refer to nodes by numbers 1, … , *n* (arbitrary numbering)
Let $t_i$ denote the number of edges out of *node i* (outdegree)

Define

$$\mathbf{pr_{new}(k) = \Sigma_{i\ with\ edge\ from\ i\ to\ k}\ (pr(i) / t_i)}$$

Iterate until converges




Problems
- Sinks (nodes with no edges out)
- Cyclic behavior

3

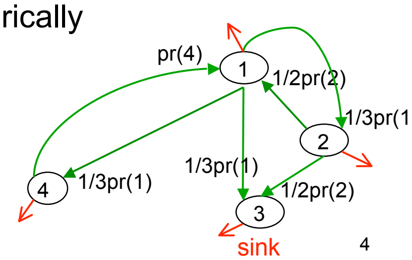
# Attempt 2

Random walk model
- Attempt 1 gives movement from node to linked neighbor with probability 1/outdegree
- Add random jump to any node

$$\mathbf{pr_{new}(k) = \alpha/n + (1-\alpha)\Sigma_{i\ with\ edge\ from\ i\ to\ k}\ (pr(i) / t_i)}$$

– $\alpha$ parameter chosen empirically

- Break cycles
- Escape from sinks




4

# Normalized?

- Would like $\boldsymbol{\Sigma_{1\le k\le n}}$ **(pr (k)) = 1**
- Consider $\boldsymbol{\Sigma_{1\le k\le n}}$ **($pr_{new}$(k))**

$$= \Sigma_{1\le k\le n}\big( \alpha/n + (1-\alpha)\Sigma_{\text{i with edge from i to k}}(pr(i) / t_i) \big)$$

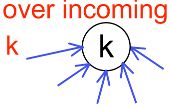
$$= \Sigma_{1\le k\le n}( \alpha/n)+\Sigma_{1\le k\le n}((1-\alpha)\Sigma_{\text{i with edge from i to k}}(pr(i) / t_i)) \;*$$

$$= \quad \alpha \quad + (1-\alpha)\Sigma_{1\le k\le n} \Sigma_{\text{i with edge from i to k}}(pr(i) / t_i)$$

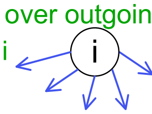
$$= \quad \alpha \quad + (1-\alpha)\Sigma_{1\le i\le n}\Sigma_{\text{k with edge from i to k}}(pr(i) / t_i) \;*$$

$$= \quad \alpha \quad + (1-\alpha)\Sigma_{\text{i with edge from i}}\; pr(i))$$

*inner sum $\Sigma_i$ over incoming edges for one k




*inner sum $\Sigma_k$ over outgoing edges for one i

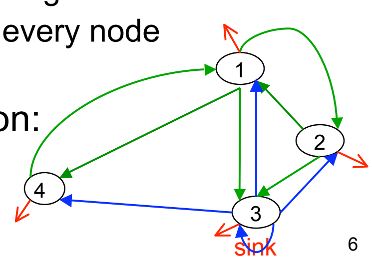





# Problem for desired normalization

- Have
  $\boldsymbol{\Sigma_{1\le k\le n}}$ **($pr_{new}$(k)) = $\alpha$ + (1-$\alpha$)$\Sigma_{\text{i with edge from i}}$ pr(i))**
- Missing **pr(i)** for nodes with no edges from them
  - sinks!
- Solution: add n edges out of every sink
  - Edge to every node including self
  - Gives 1/n contribution to every node

Gives desired normalization:
If $\Sigma_{1\le k\le n}$ ($pr_{initial}$(k)) = 1
then $\Sigma_{1\le k\le n}$ (pr(k)) = 1



# Matrix formulation

- Let E be the n by n adjacency matrix
  E(i,k) = 1 if there is an edge from node i to node k
  = 0 otherwise
- Define new matrix L:
  For each row i of E ($1 \le i \le n$)
  If row i contains $t_i > 0$ ones, $L(i,k) = (1/t_i)\, E(i,k)$, $1 \le k \le n$
  If row i contains 0 ones, $L(i,k) = 1/n$, $1 \le k \le n$
- Vector ***pr*** of PageRank values defined by
  $$\boldsymbol{pr} = (\alpha/n, \alpha/n, \ldots \alpha/n)^T + (1-\alpha)\, L^T \boldsymbol{pr}$$
- has a solution representing the steady-state values pr(k)

7

# Calculation

- Choose $\alpha$
  - No single best value
  - Page and Brin originally used $\alpha = .15$
- Simple iterative calculation
  - Initialize $pr_{initial}(k) = 1/n$ for each node k so $\Sigma_{1 \le k \le n} (pr_{initial}(k)) = 1$
  - $pr_{new}(k) = \alpha/n + (1-\alpha) \Sigma_{1 \le i \le n} L(i,k) pr(i)$
- Converges
  - Has necessary mathematical properties
  - In practice, choose convergence criterion
    - Stops iteration

8