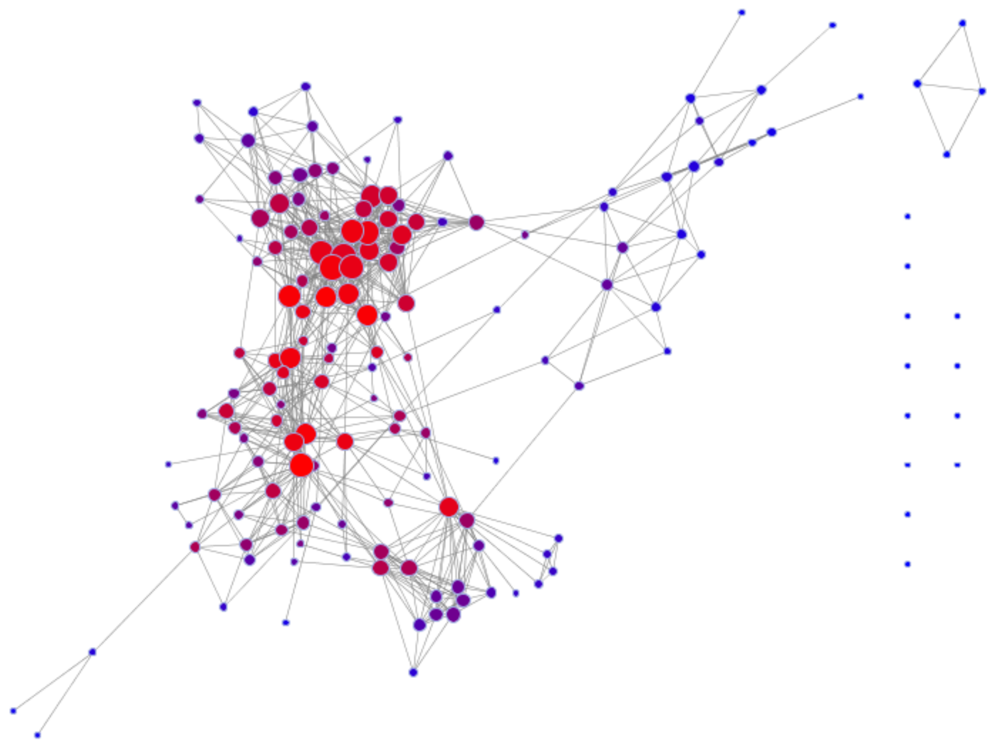
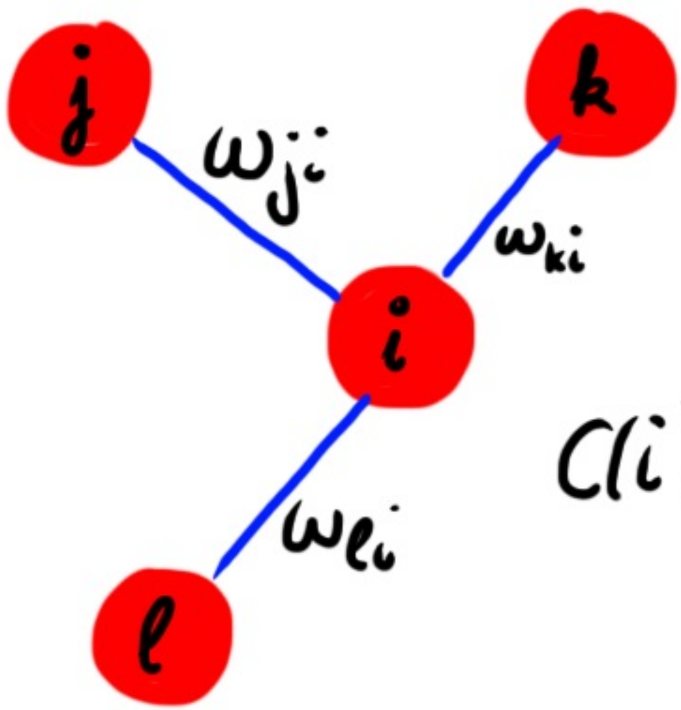
Closeness and Lada's fb network



- **degree** (number of connections) denoted by size
- **closeness** (length of shortest path to all others) denoted by color

Eigenvector centrality

- How central you are depends on how central your neighbors are



$$C(i) = \omega_{ji} \cdot C(j) + \omega_{ki} \cdot C(k) + \omega_{\ell i} \cdot C(\ell)$$

Bonacich eigenvector centrality

$$c_i(\beta) = \sum_j (\alpha + \beta c_j) A_{ji}$$

$$c(\beta) = \alpha (I - \beta A)^{-1} A\mathbf{1}$$

- α is a normalization constant
- β determines how important the centrality of your neighbors is
- $\mathbf{A}$ is the adjacency matrix (can be weighted)
- $\mathbf{I}$ is the identity matrix (1s down the diagonal, 0 off-diagonal)
- $\mathbf{1}$ is a matrix of all ones.

Bonacich Power Centrality: attenuation factor β

small β ➔ high attenuation
only your immediate friends matter, and their importance is factored in only a bit

high β ➔ low attenuation
global network structure matters (your friends, your friends' of friends etc.)

= 0 yields simple degree centrality

$$c_i(\beta) = \sum_j (\alpha \quad) A_{ji}$$

Bonacich Power Centrality: attenuation factor β

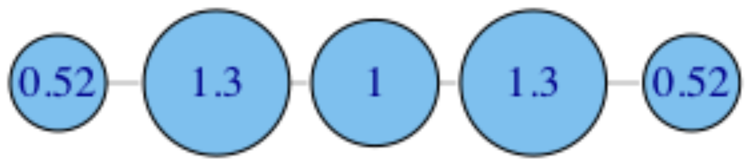
If $\beta > 0$, nodes have higher centrality when they have edges to other central nodes.

If $\beta < 0$, nodes have higher centrality when they have edges to less central nodes.

Bonacich Power Centrality: examples

β=.25

β=-.25



Why does the middle node have lower centrality than its neighbors when β is negative?

Centrality in directed networks

- WWW
- food webs
- population dynamics
- influence
- hereditary
- citation
- transcription regulation networks
- neural networks

Betweenness centrality in directed networks

- We now consider the fraction of all directed paths between any two vertices that pass through a node

betweenness of vertex i

paths between j and k that pass through i

$$C_B(i) = \sum_{j,k} g_{jk}(i) / g_{jk}$$

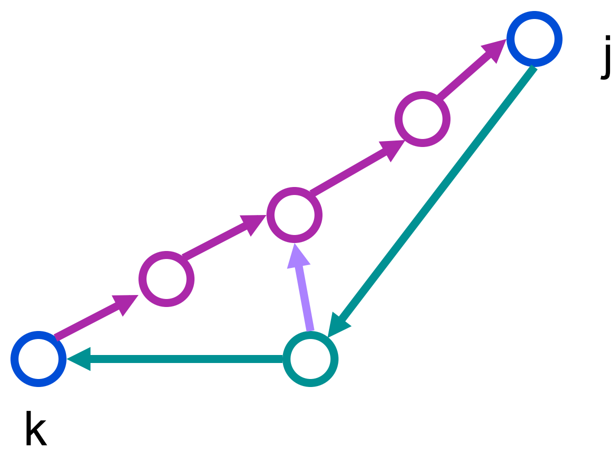
all paths between j and k

- Only modification: when normalizing, we have (N-1)*(N-2) instead of (N-1)*(N-2)/2, because we have twice as many ordered pairs as unordered pairs

$$C'_B(i) = C_B(i)/[(N-1)(N-2)]$$

Directed geodesics

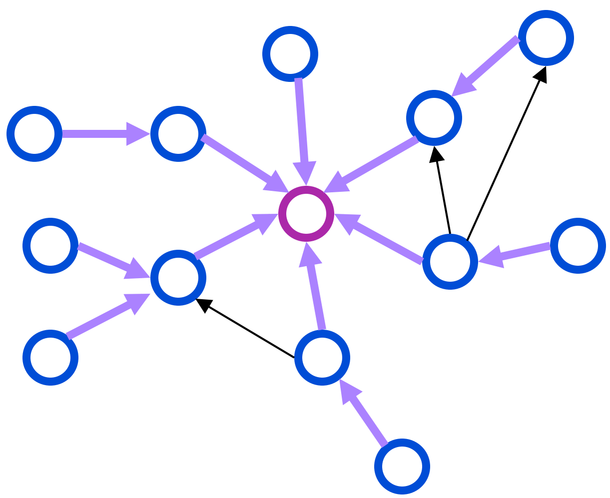
- A node does not necessarily lie on a geodesic from *j* to *k* if it lies on a geodesic from *k* to *j*



Directed closeness centrality

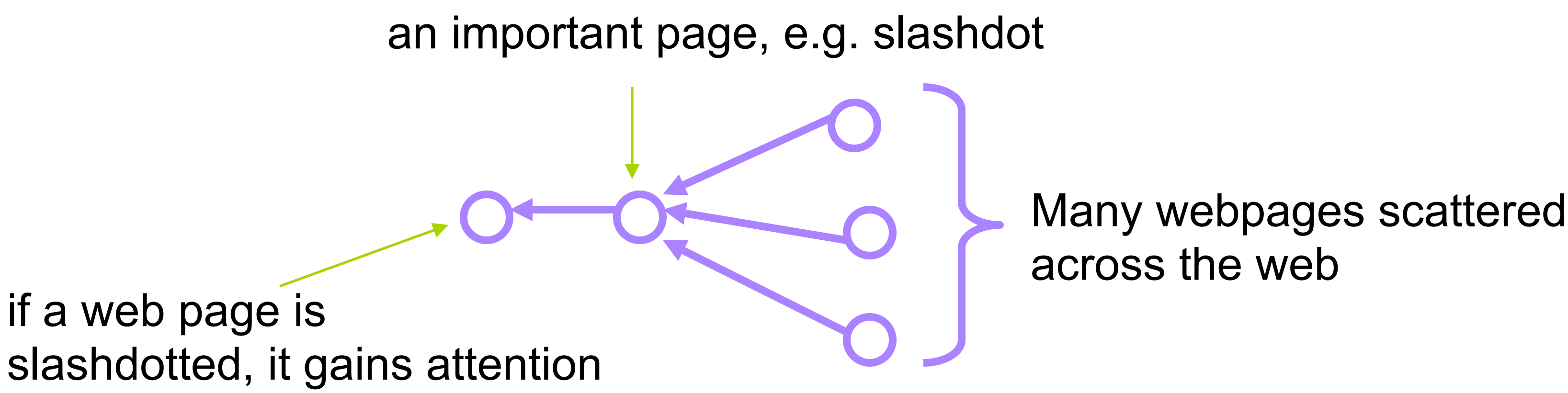
- choose a direction
  - in-closeness (e.g. prestige in citation networks)
  - out-closeness

- usually consider only vertices from which the node *i* in question can be reached



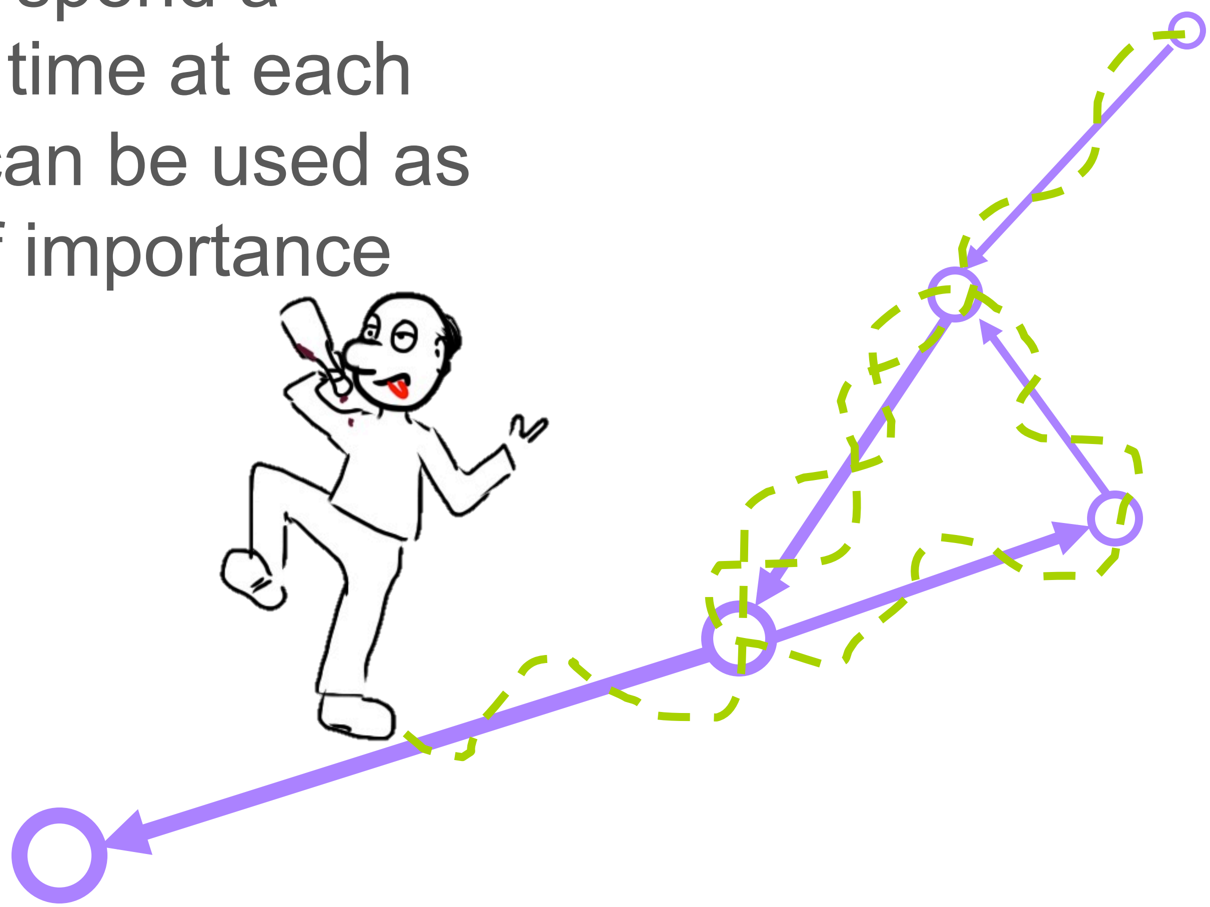
Eigenvector centrality in directed networks

- PageRank brings order to the Web:
  - it's not just the pages that point to you, but how many pages point to those pages, etc.
  - more difficult to artificially inflate centrality with a recursive definition



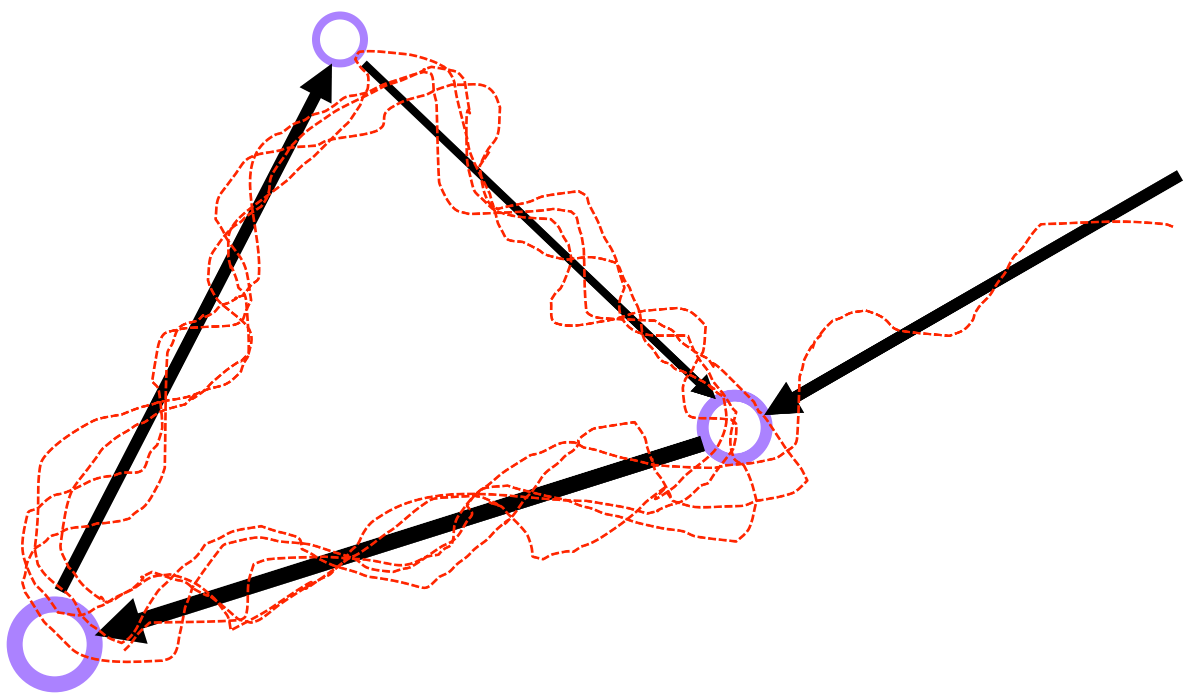
Ranking pages by tracking a drunk

- A random walker following edges in a network for a very long time will spend a proportion of time at each node which can be used as a measure of importance



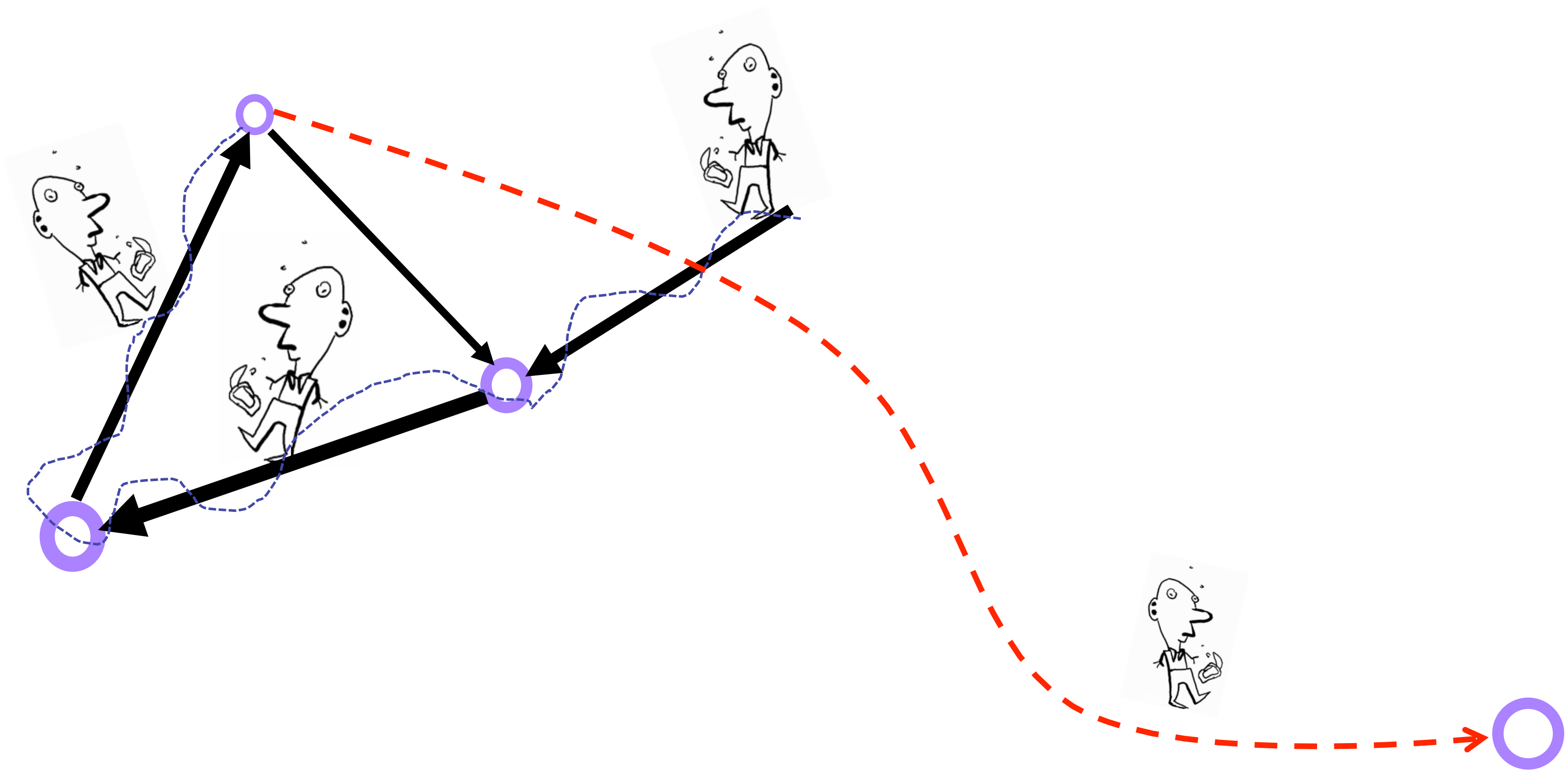
Trapping a drunk

- Problem with pure random walk metric:
  - Drunk can be “trapped” and end up going in circles



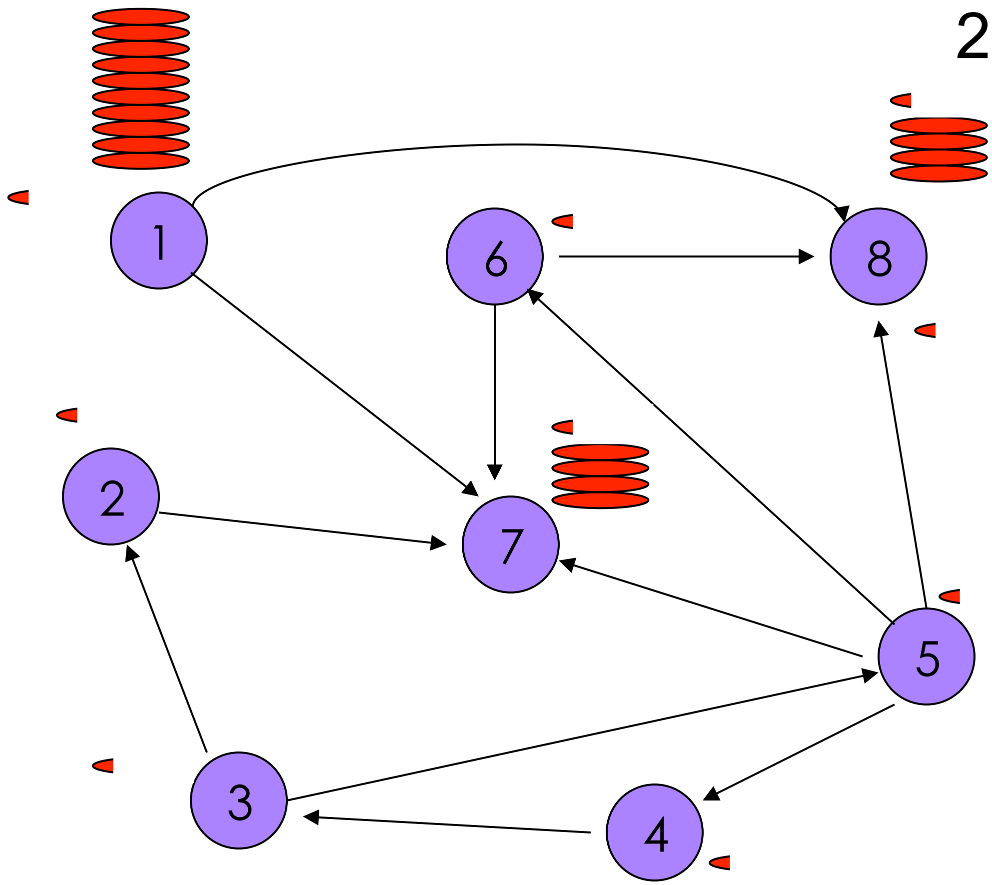
Ingenuity of the PageRank algorithm

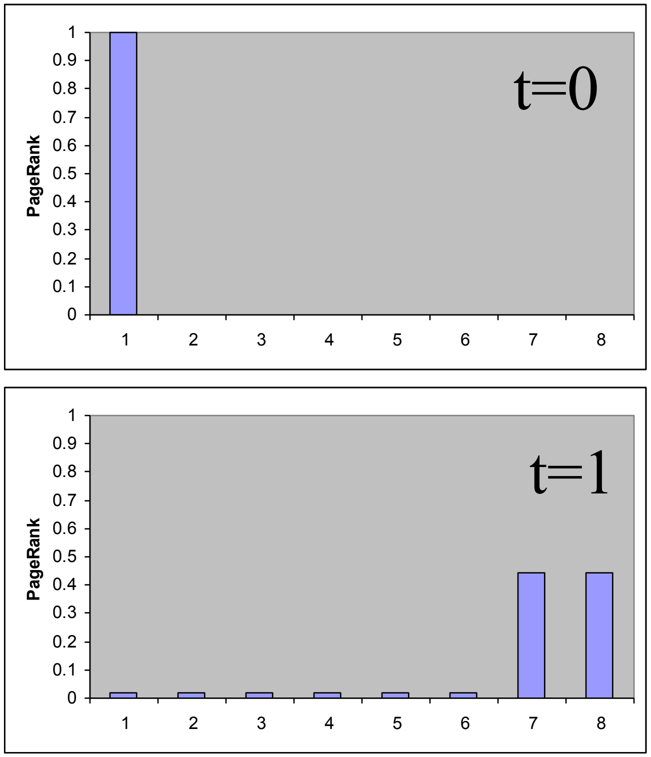
- Allow drunk to teleport with some probability
  - e.g. random websurfer follows links for a while, but with some probability teleports to a "random" page (bookmarked page or uses a search engine to start anew)



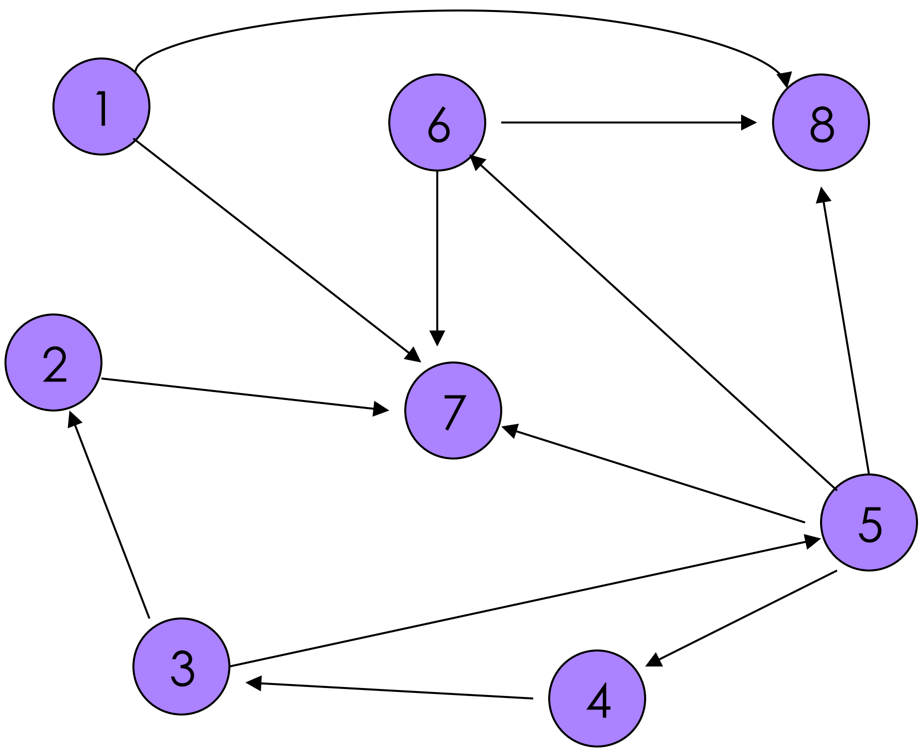
example: probable location of random walker after 1 step

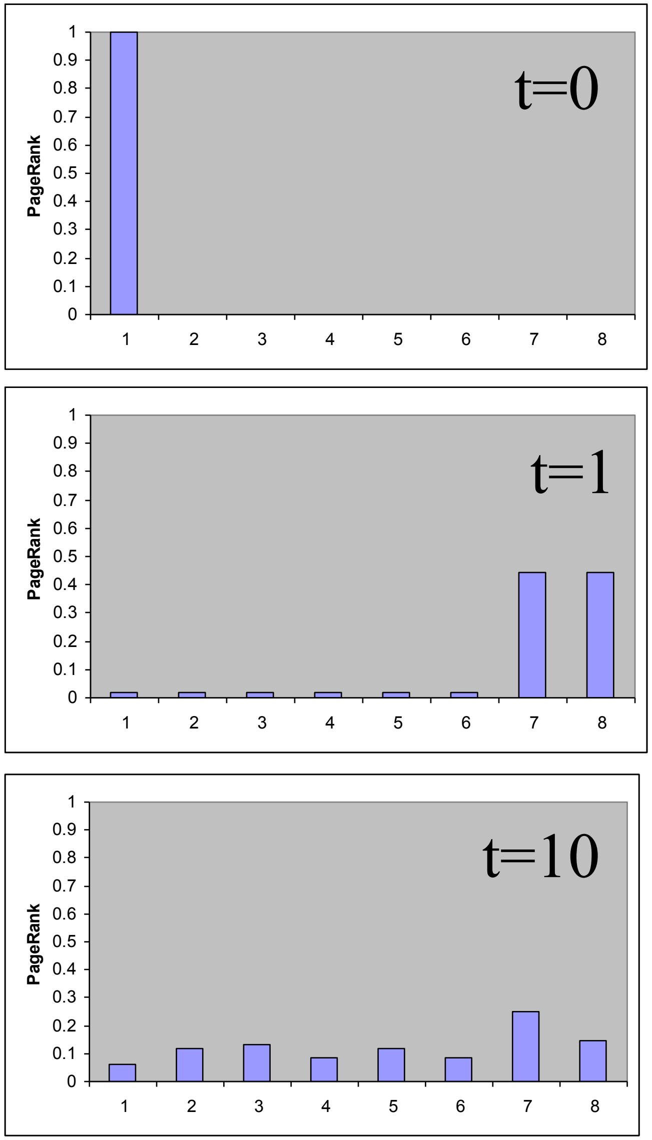
20% teleportation probability





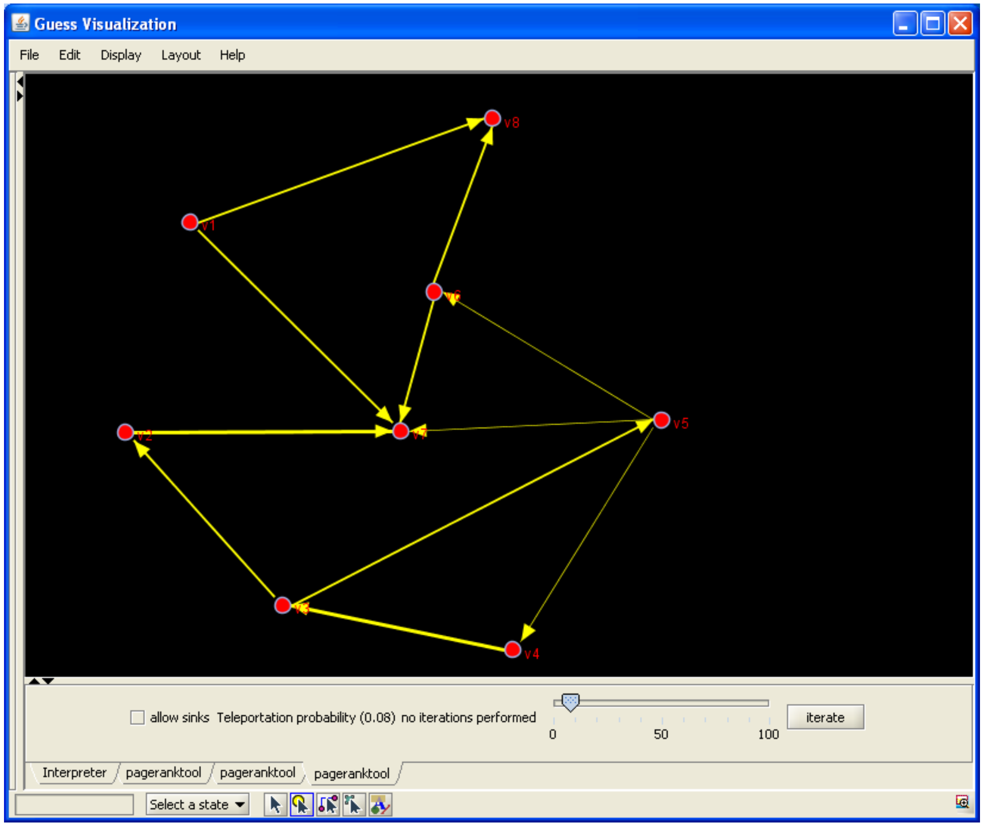
example: probable location of random walker after 10 steps





slide from: Dragomir Radev

Quiz Q:



- What happens to the relative PageRank scores of the nodes as you increase the teleportation probability?
  - they equalize
  - they diverge
  - they are unchanged

http://www.ladamic.com/netlearn/GUESS/pagerank.html

wrap up

- Centrality
  - many measures: degree, betweenness, closeness, eigenvector
  - may be unevenly distributed
    - measure via distributions and centralization
  - in directed networks
    - indegree, outdegree, PageRank
  - consequences:
    - benefits & risks (Baker & Faulkner)
    - information flow & productivity (Aral & Van Alstyne)