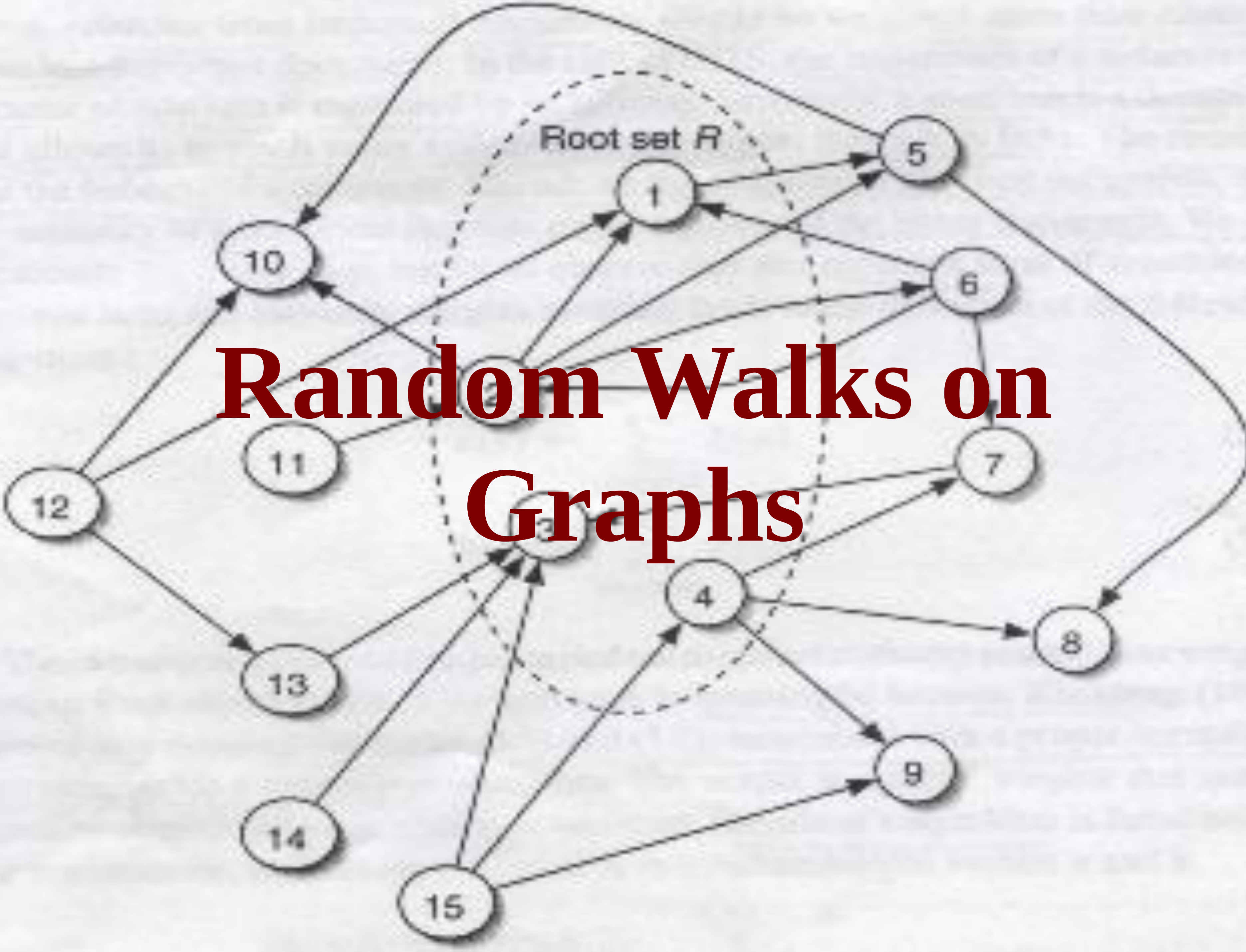
Random Walks on Graphs



What is a Random Walk

- Given a graph and a starting point (node), we select a neighbor of it at random, and move to this neighbor;
- Then we select a neighbor of this node and move to it, and so on;
- The (random) sequence of nodes selected in this way is a *random walk* on the graph

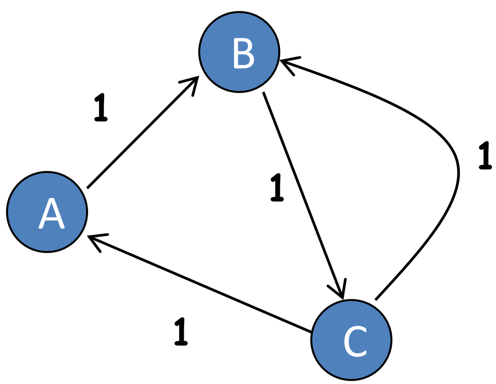
An example

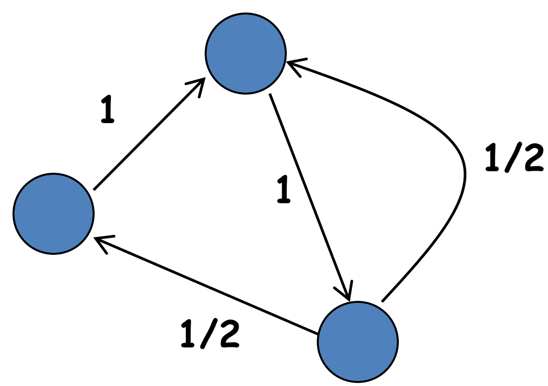
0	1	0
0	0	1
1	1	0

Adjacency matrix W

0	1	0
0	0	1
1/2	1/2	0

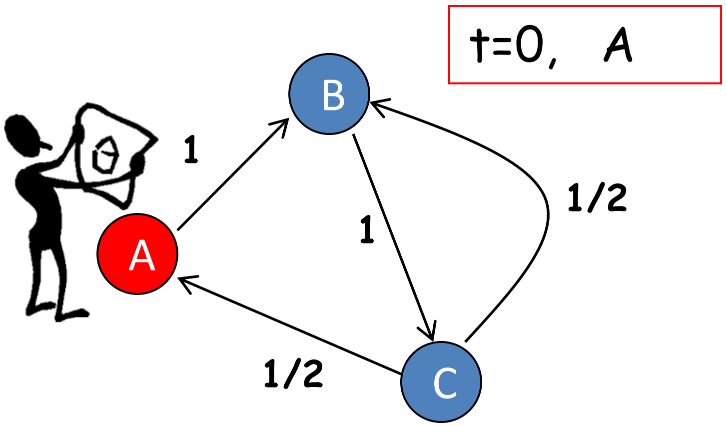
Transition matrix Q





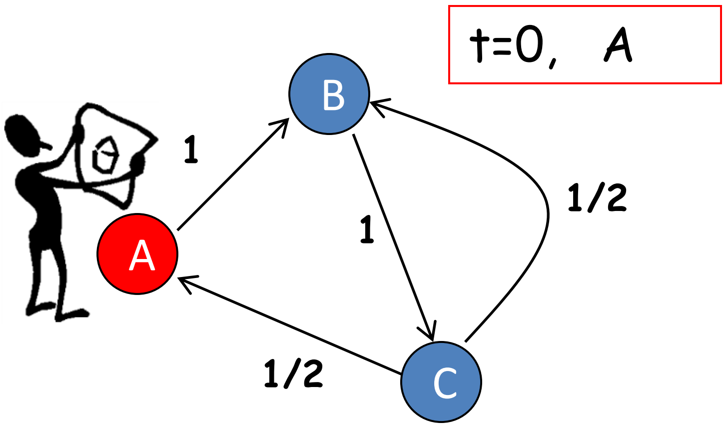
Slide from Purnamitra Sarkar, *Random Walks on Graphs: An Overview*

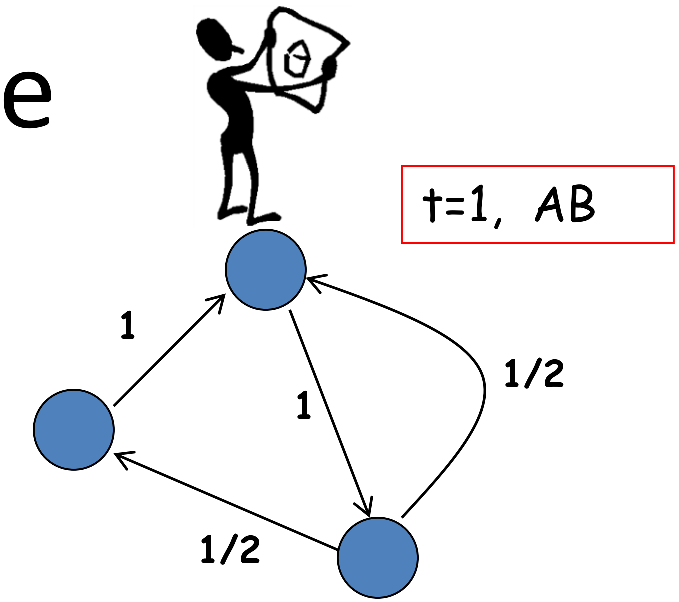
An example



Slide from Purnamitra Sarkar, *Random Walks on Graphs: An Overview*

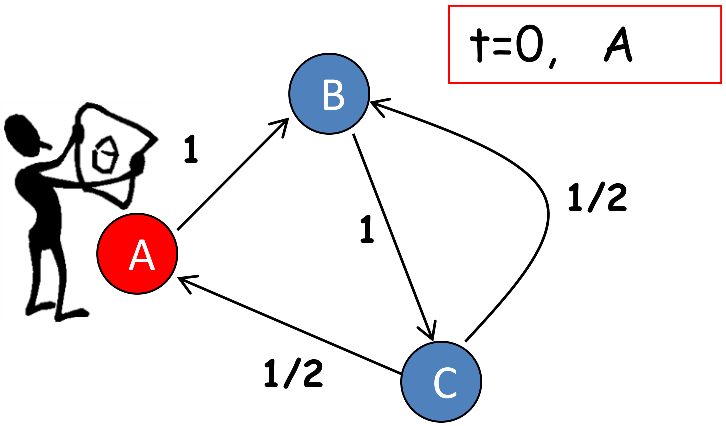
An example

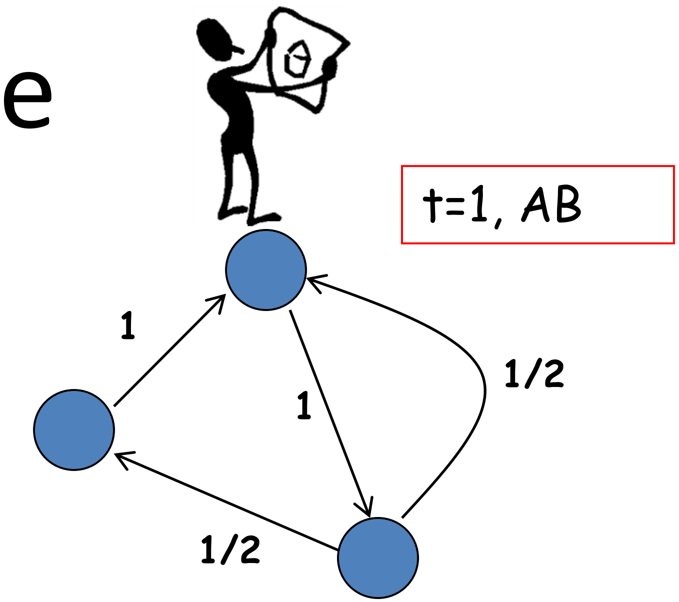


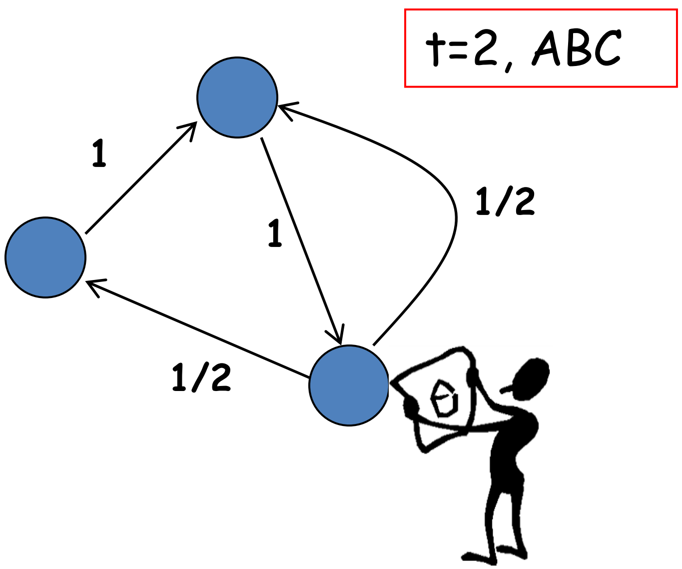


Slide from Purnamitra Sarkar, *Random Walks on Graphs: An Overview*

An example

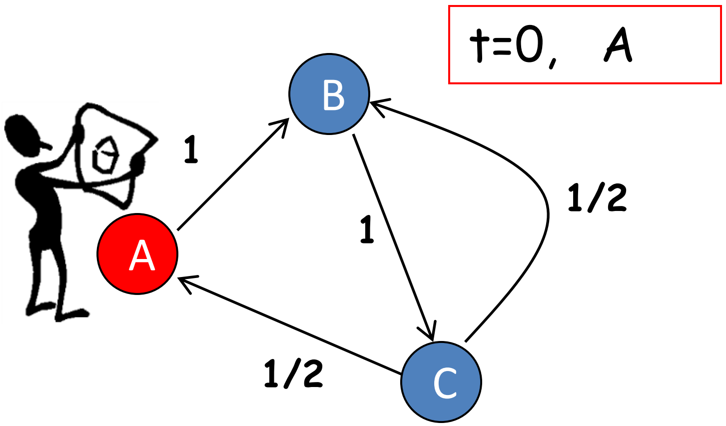


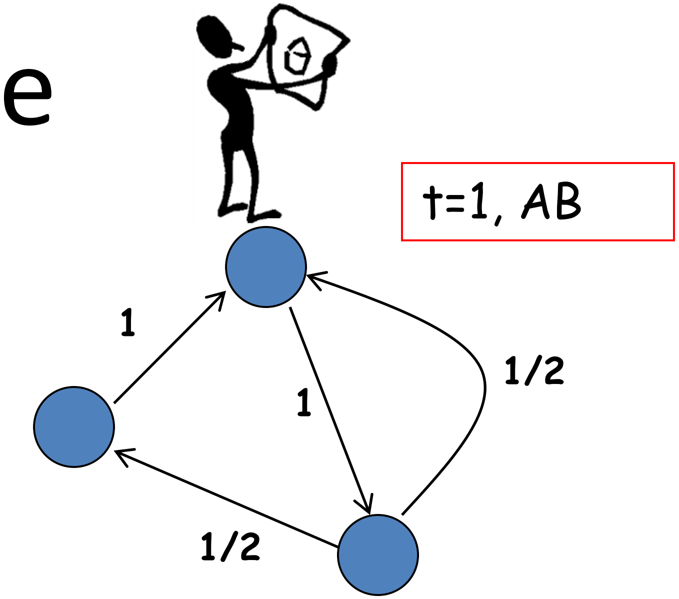


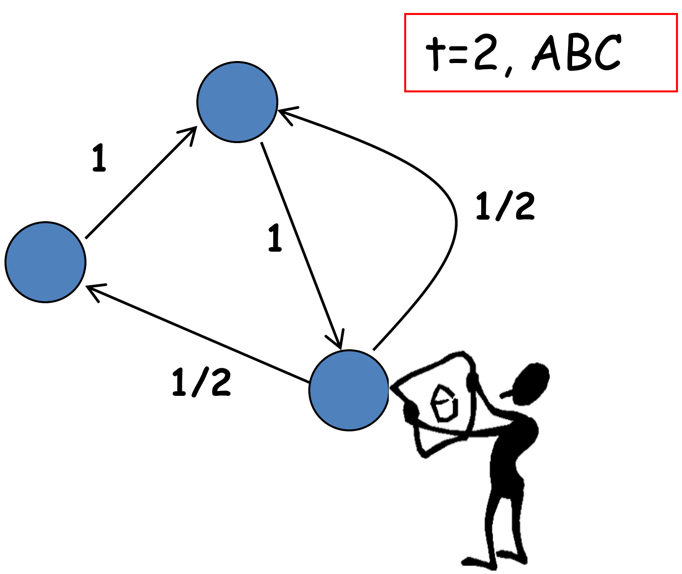


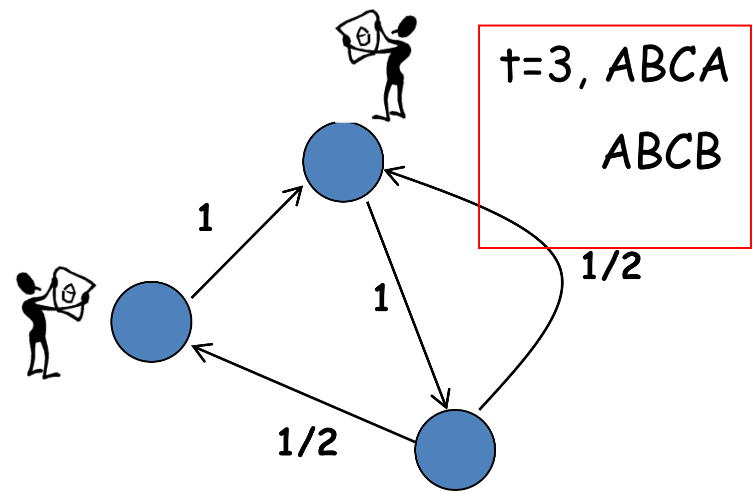
Slide from Purnamitra Sarkar, *Random Walks on Graphs: An Overview*

An example









Slide from Purnamitra Sarkar, *Random Walks on Graphs: An Overview*

Random walks and Markov chains

- A *Markov chain* describes a stochastic process over a set of states according to a transition probability matrix
- Markov chains are *memoryless*
- *Random walks correspond to Markov chains*:
  - The set of states is the set of nodes in the graph
  - The elements of the transition probability matrix are the probabilities to follow and edge from one node to another

Random Walk algorithm

Input:
- the adjacency matrix $\boldsymbol{W}$ of a graph $G=<V,E>$
- A subset of nodes V_C having property C

• Initialization of nodes:
if $v \in V_c$ then $p_0(v) = 1 / |V_C|$ else $p_0(v)=0$

• Set transition matrix: $\boldsymbol{Q} = \boldsymbol{D}^{-1}\boldsymbol{W}$
where $\boldsymbol{D}$ is a diagonal matrix with

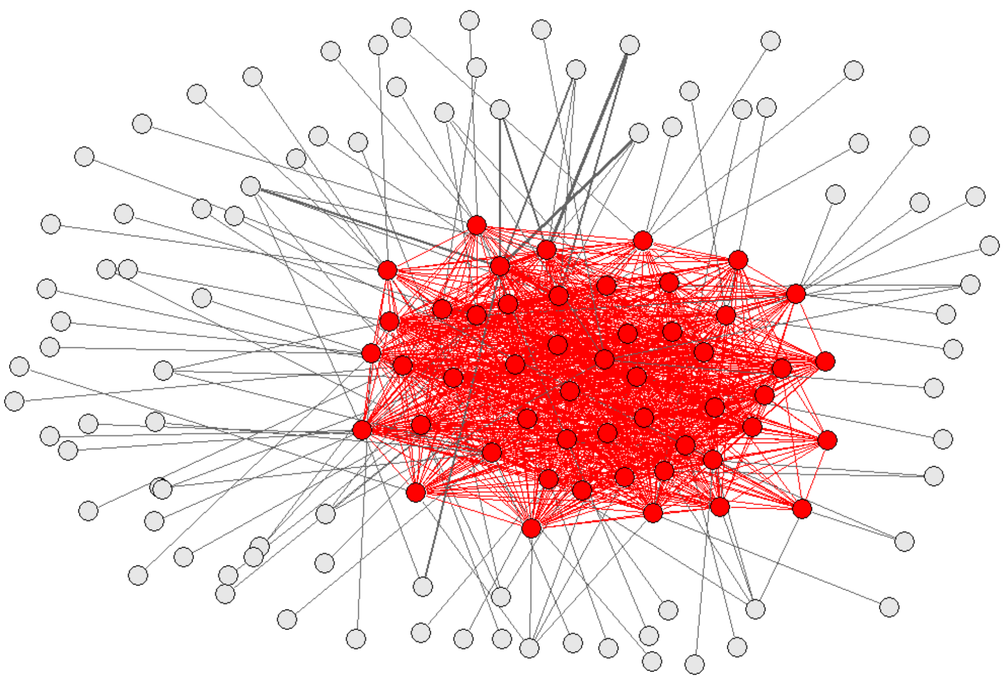
$$d_{ii} = \sum_j w_{ij}$$

• Iteratively update until convergence or until $t=k$
$\boldsymbol{p}_t = \boldsymbol{Q}^T \boldsymbol{p}_{t-1}$

Output: $\boldsymbol{p}_t$

Random walking algorithm to rank genes w.r.t to a given "property" C

- A subset V_C of a set of genes V have "a priori" known property C
- Can we rank the other genes in the set $V \setminus V_C$ w.r.t their likelihood to belong to V_C ?



Random walk algorithm

C can be e.g. a *disease* (gene disease prioritization) or a GO term (gene function prediction)