Social Network Analysis

CS 185 Lecture 14
May 16th 2018

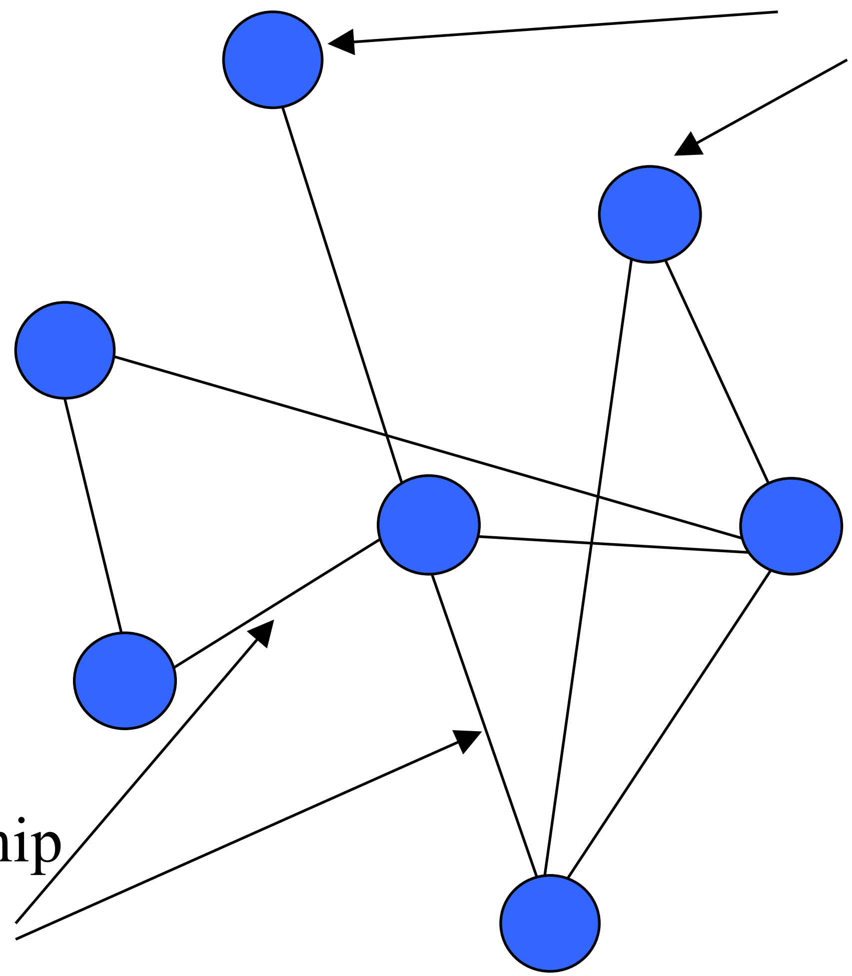
Thanks to Prof. Dr. Daning Hu (Original Author)
Department of Informatics
University of Zurich

Outline

- Node Level Analysis
- Link and Group Level Analysis
- Network Level Analysis
- Network Topological Analysis
- Ref Book: Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)
  - http://www.amazon.com/Social-Network-Analysis-Applications-Structural/dp/0521387078

What is a Network?

Node: Any entity in a network (person, system, group, organization)



Tie/Link: Relationship or interaction between two nodes.

Network Topological Analysis Vs. Dynamic Network Analysis

- Topology (from the Greek τόπος, “place”, and λόγος, “study”) is a major area of mathematics concerned with the most basic properties of space, such as connectedness.
  - as a field of study out of geometry and set theory, through analysis of such concepts as space, dimension, and transformation.

- Network topology is the arrangement of the various elements (links, nodes, etc). Essentially, it is the topological structure of a network.
  - **Physical topology** refers to the placement of the network's various components, including device location and cable installation;
  - while **logical topology** shows how data flows within a network, regardless of its physical design.

Node Level Analysis: Node Centrality

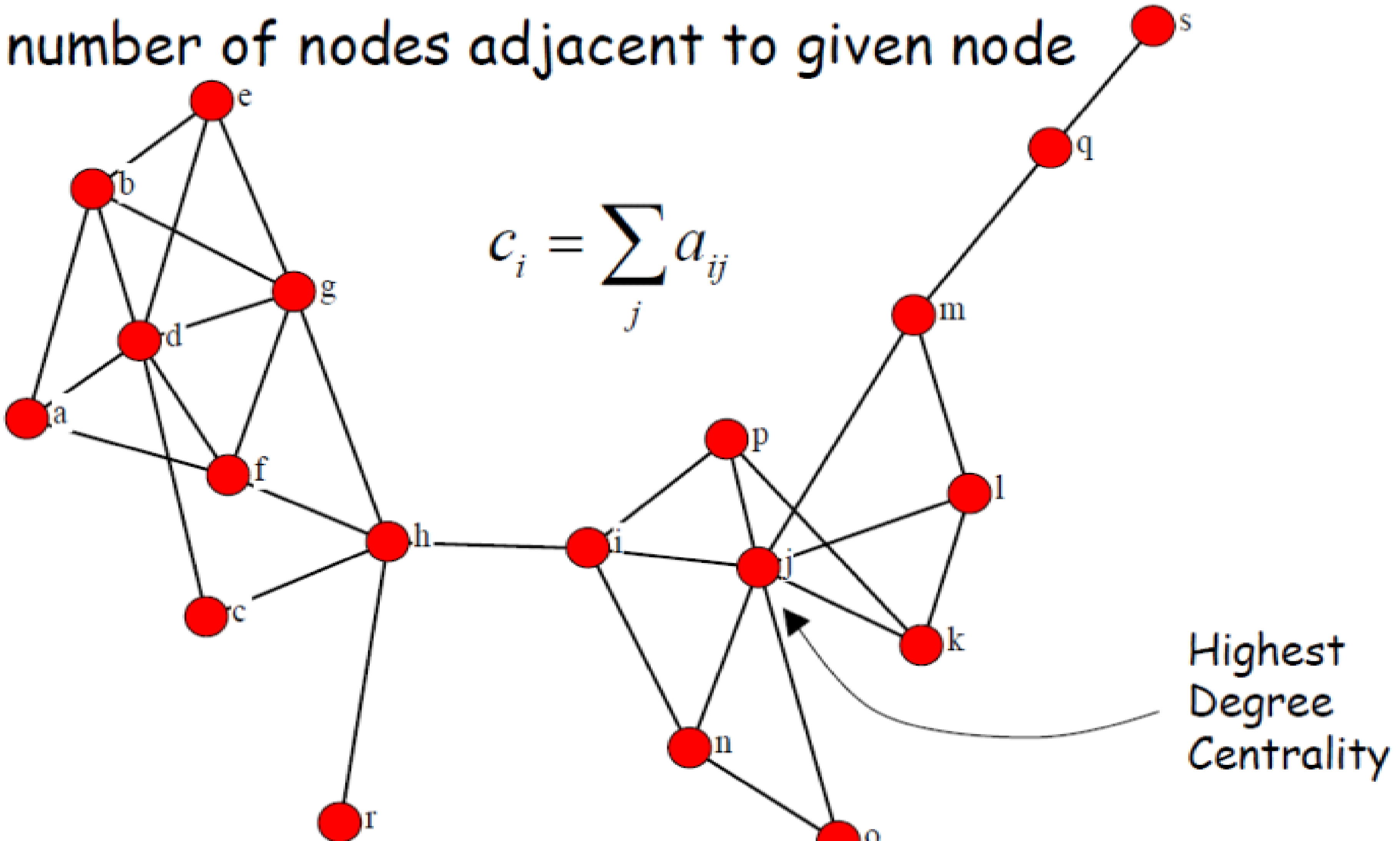
- Node Centrality can be viewed as a measure of influence or importance of nodes in a network.

- Degree
  - the number of links that a node possesses in a network. In a **directed** network, one must differentiate between in-links and out-links by calculating in-degree and out-degree.
- Betweeness
  - the number of shortest paths in a network that traverse through that node.
- Closeness
  - the average distance that each node is from all other nodes in the network

Node Level Analysis: Degree Centrality

- The number of nodes adjacent to given node

$$c_i = \sum_j a_{ij}$$

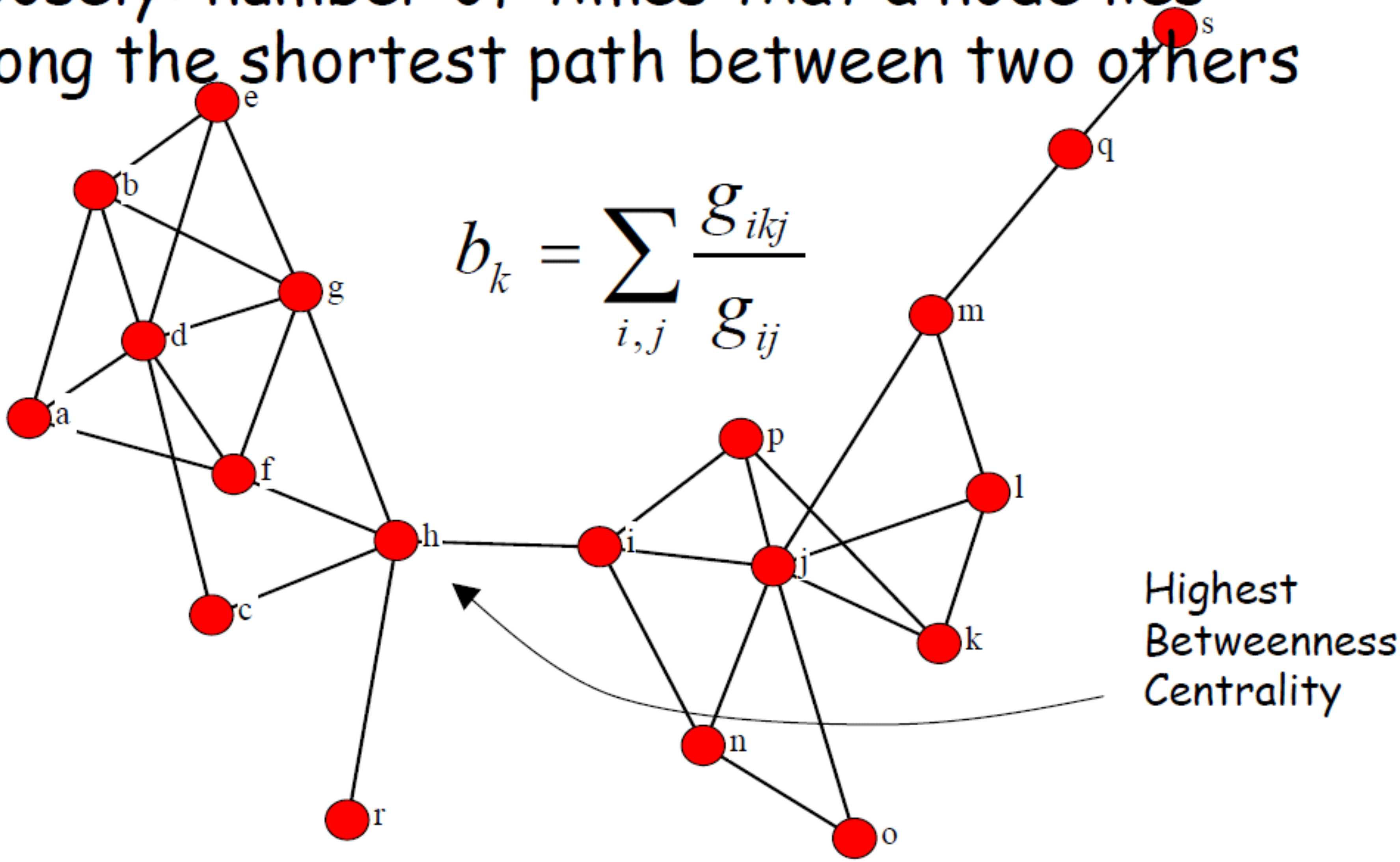


From Steve Borgatti

Node Level Analysis: Betweenness Centrality

- Loosely: number of times that a node lies along the shortest path between two others

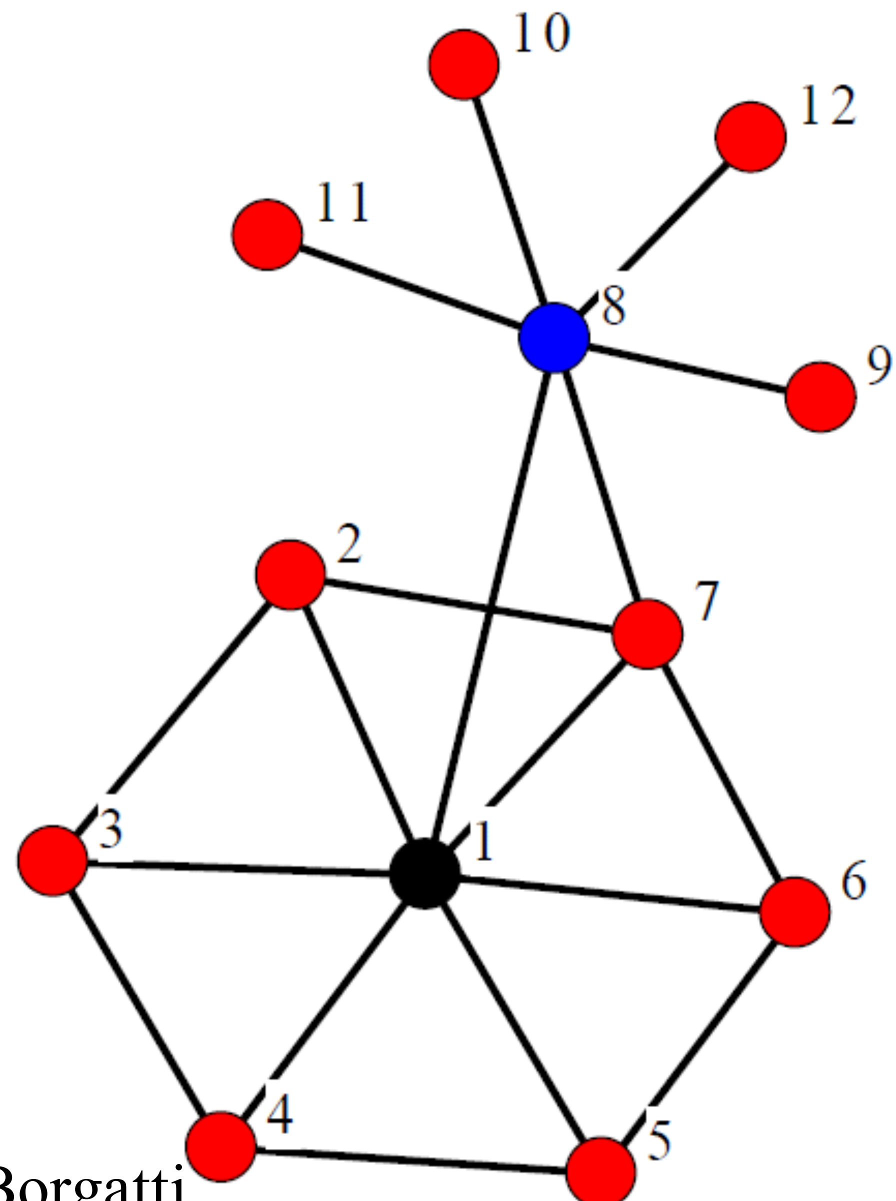
$$b_k = \sum_{i,j} \frac{g_{ikj}}{g_{ij}}$$



From Steve Borgatti

Link Level Analysis: Length and Distance

- Length of a path is the number of the links
- Distance between two nodes is the length of shortest path (i.e., geodesic)

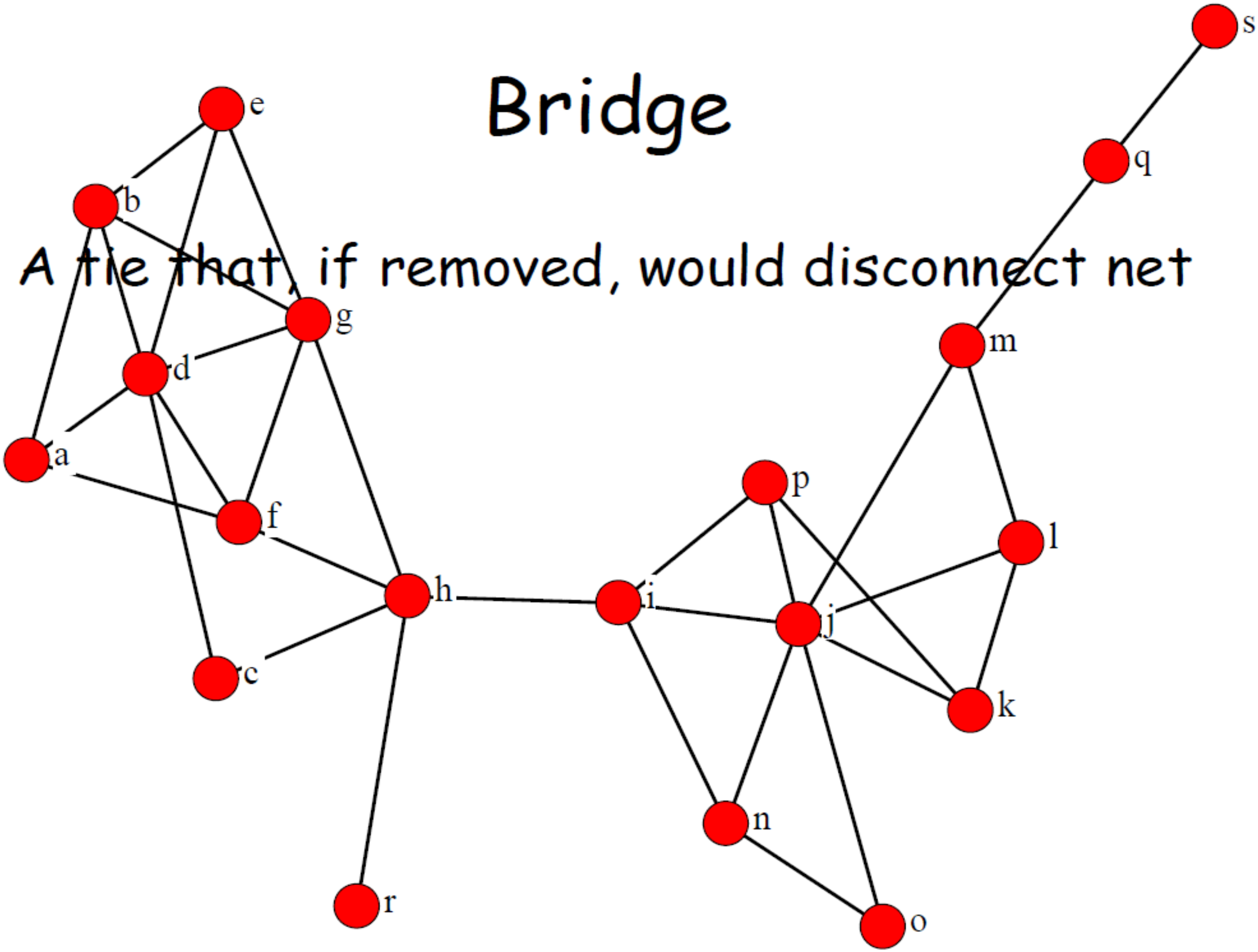


From Steve Borgatti

Group Level Analysis: Cutpoints and Bridge

Bridge

- A tie that, if removed, would disconnect net



From Steve Borgatti

The Strength of Weak Tie (Granovetter 1973)

- Strong ties create transitivity
  - Two nodes linked by a strong tie will have mutual acquaitances

- Ties that are part of transitive triples cannot be bridges

- Therefore, only weak ties can be bridges
  - the value of weak ties!!

- Strong ties embeded in tight homophilous clusters, while weak ties connect to diversity
  - Weak ties is a major source of novel information

Network Level Analysis: Cohesion

- Network Topology Analysis takes a macro perspective to study the physical properties of network structures. Network topological measures include:
  - **Size**, i.e., number of nodes and links
  - **Network Cohesion**
  - **Average Degree**, **Distance**
  - **Average Path Length**: on average, the number of steps it takes to get from one member of the network to another.
  - **Diameter**
  - **Clustering Coefficient**: a measure of an "all-my-friends-know-each-other" property; small-world feature

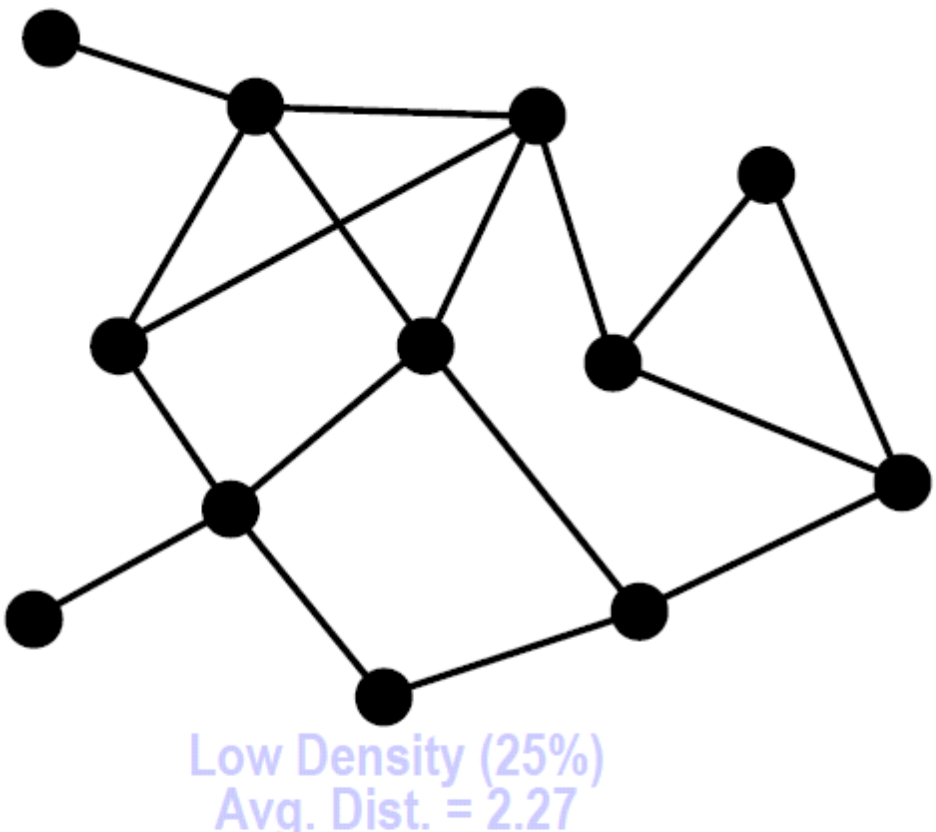
Network Level Analysis: Cohesion

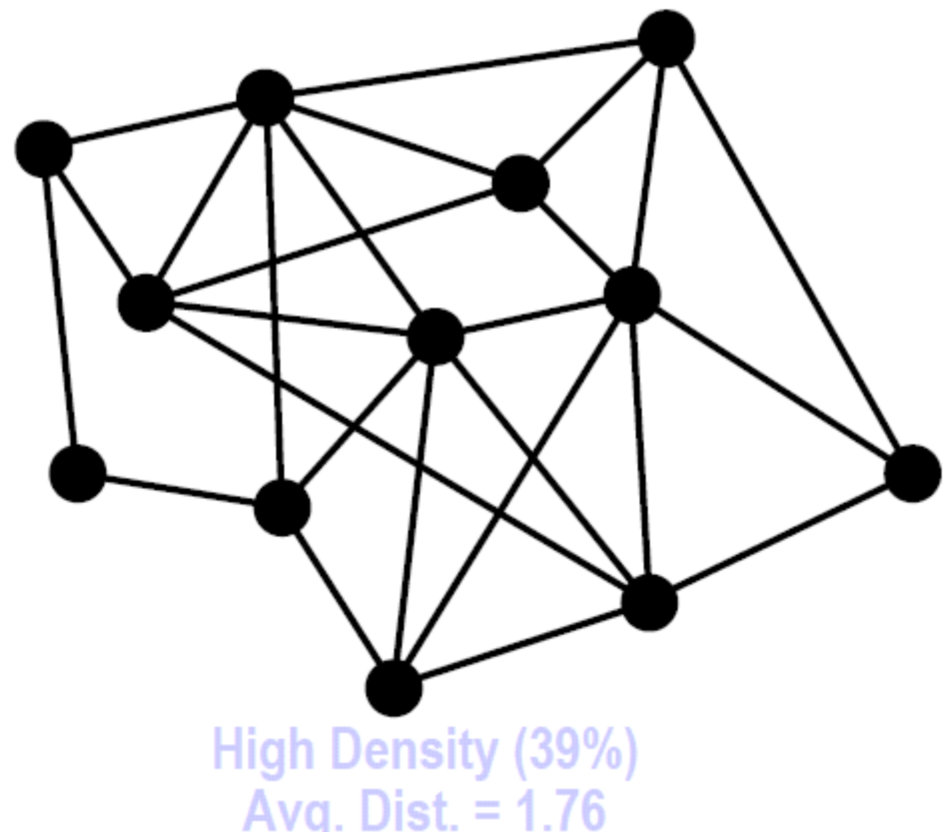
- Fragmentation: Percentage of pairs of nodes that are unreachable from each other.
  - Calculated as

$$F = 1 - \frac{\sum_k s_k (s_k - 1)}{n(n-1)}$$

Network Level Analysis: Cohesion

- Density: the percentage of the number of links over all possible pairs of links.

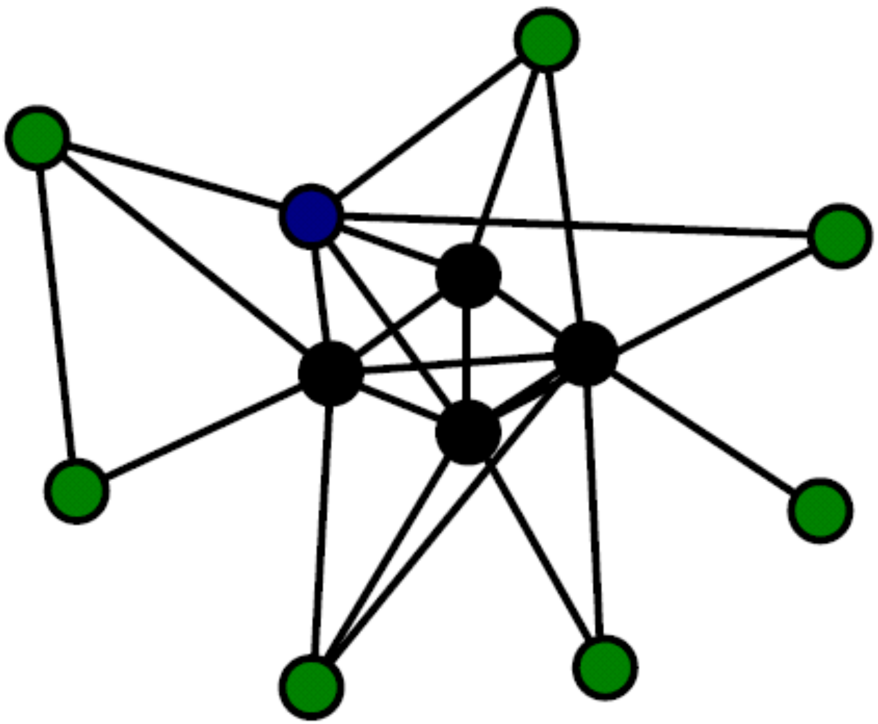


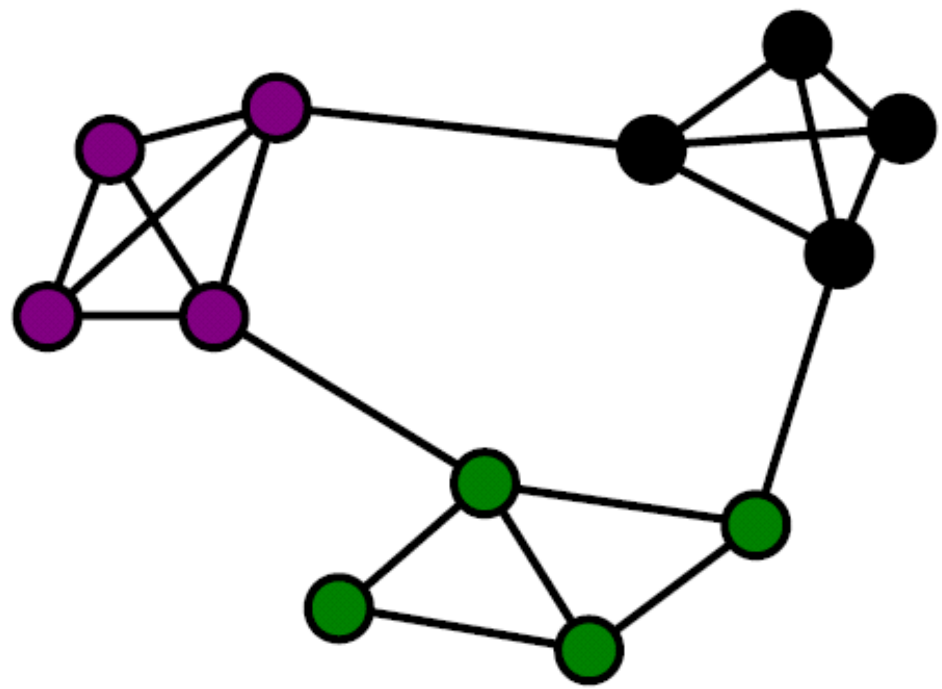


From Steve Borgatti

Network Level Analysis: Cohesion

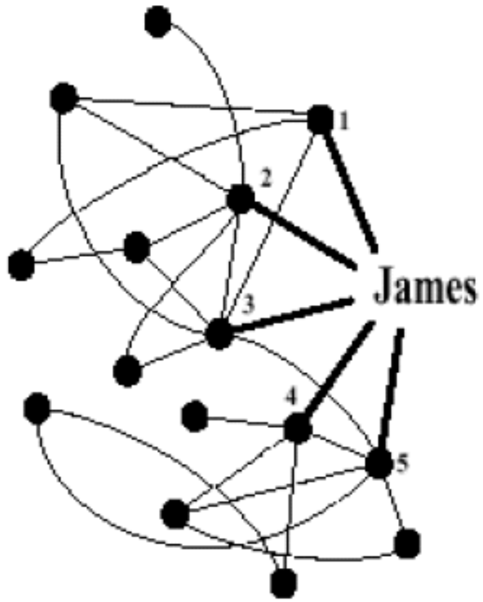
- Average distance: average distance between all pairs of nodes.





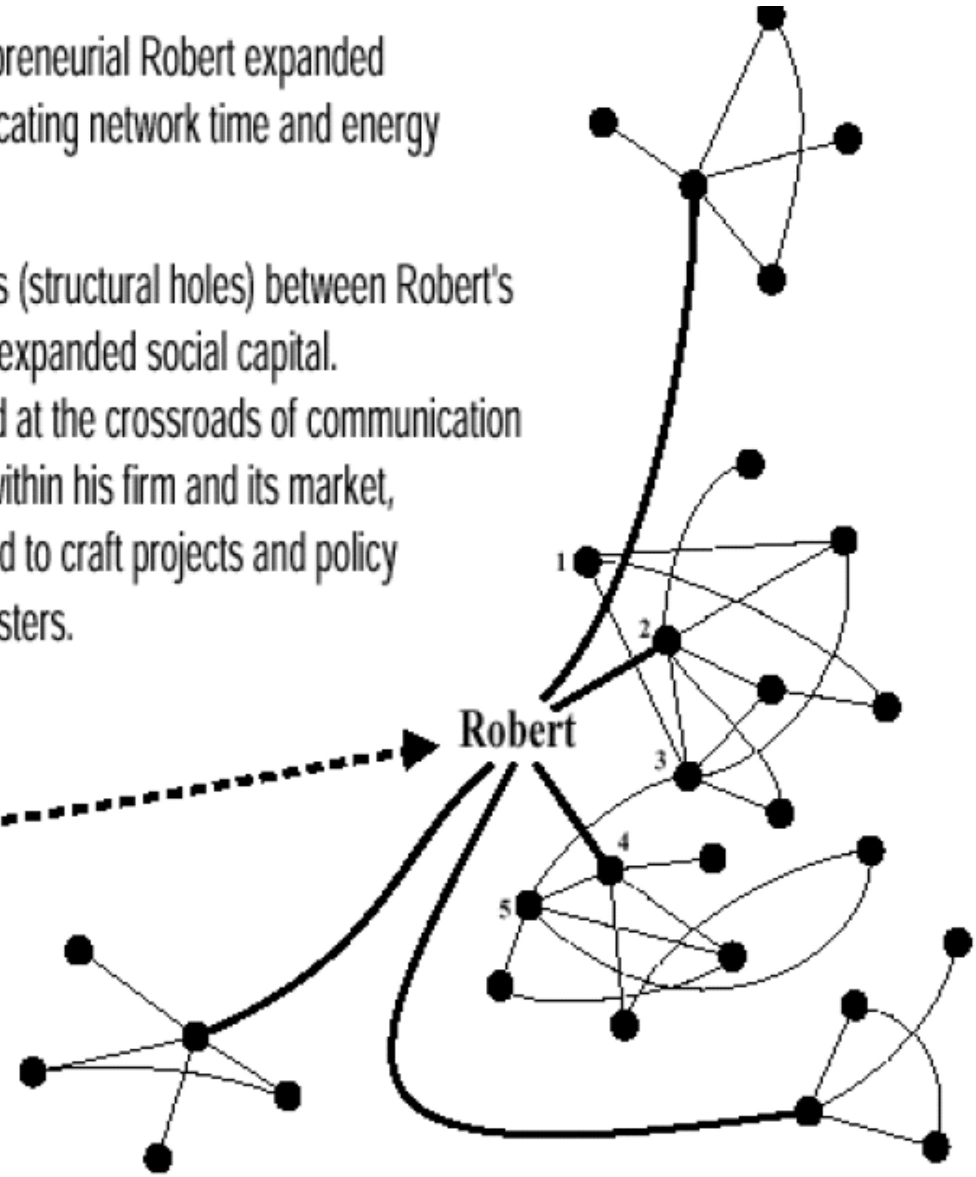
From Steve Borgatti

Network Level Analysis: Structural Holes



Robert took over James' job. Entrepreneurial Robert expanded the social capital of the job by reallocating network time and energy to more diverse contacts.

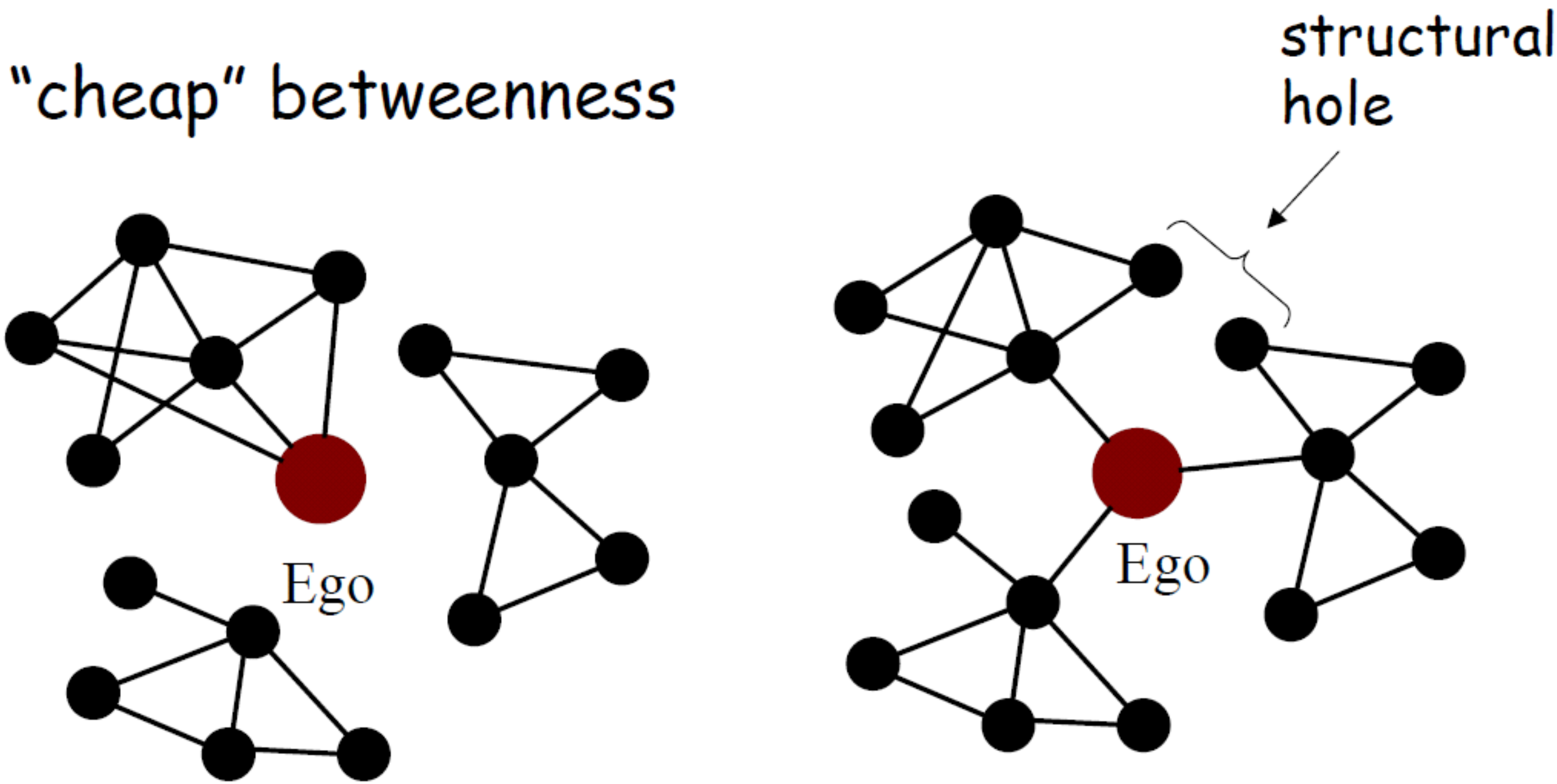
It is the weak connections (structural holes) between Robert's contacts that provide his expanded social capital.
Robert is more positioned at the crossroads of communication between social clusters within his firm and its market, and so is better positioned to craft projects and policy that add value across clusters.



Research shows that people like Robert, better positioned for entrepreneurial opportunity, are the key to integrating across functions and across the people of increasingly diverse backgrounds in today's flatter organizations. In research comparisons between managers like James and Robert, it is the people like Robert who get promoted faster, earn higher compensation, receive better performance evaluations, and perform more successfully on teams.

Network Level Analysis: Structural Holes

- "cheap" betweenness



Few structural holes

Many structural holes:
- power, info, freedom

Network Topological Analysis

- Network topology is the arrangement of the various elements (links, nodes, etc). Essentially, it is the topological structure of a network.

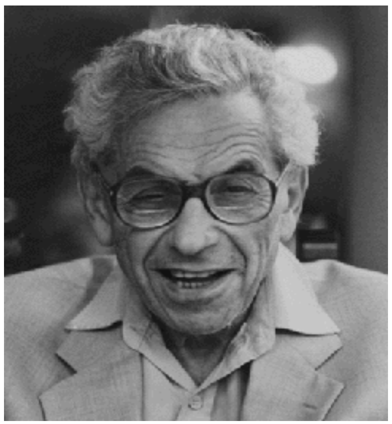
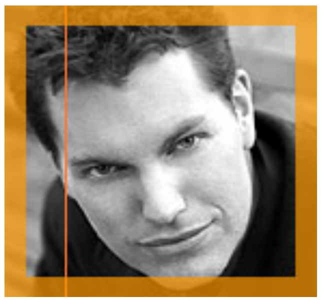
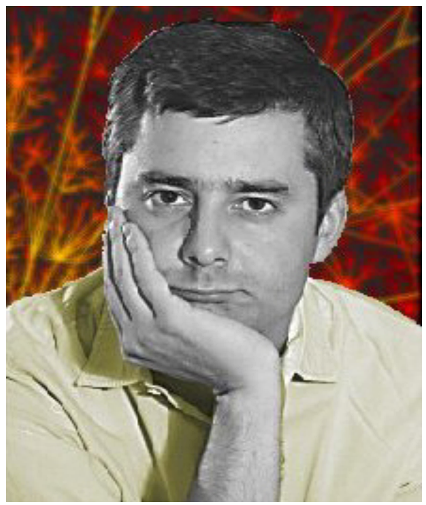
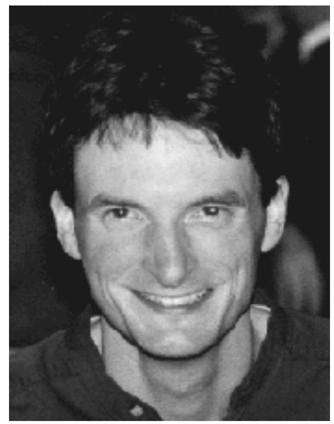
- How to model the **topology** of large-scale networks?

- What are the **organizing principles** underlying their topology?

- How does the topology of a network affect its **robustness** against errors and attacks?

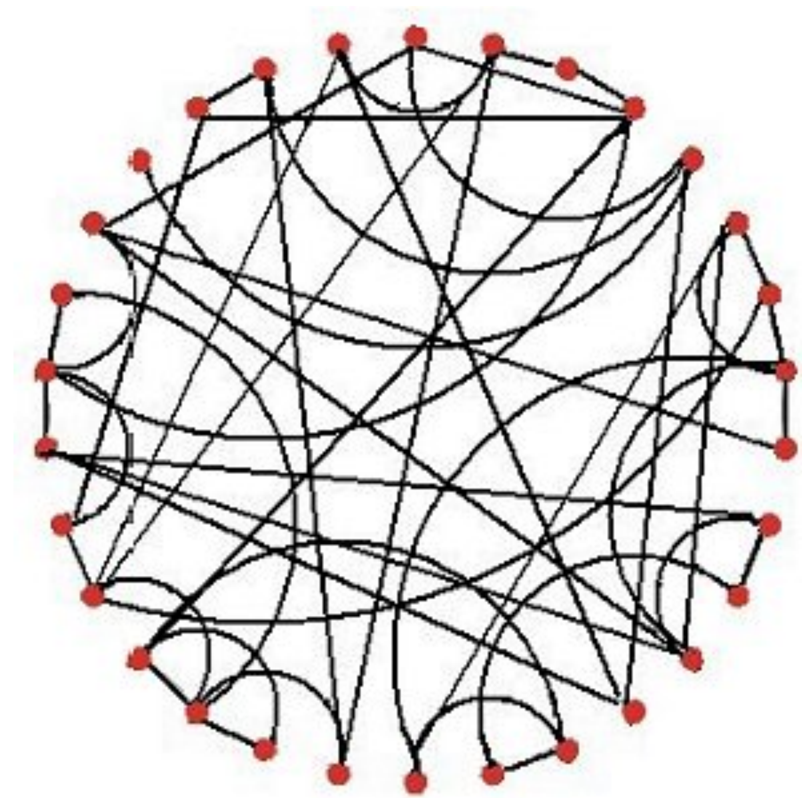
Network Models

- Random graph model (Erdős & Rényi, 1959)
- Small-world model (Watts & Strogatz, 1998)
- Scale-free model (Barabasi & Alert, 1999)

Random Networks

- Erdős–Rényi Random Graph model is used for generating random networks in which links are set between nodes with equal probabilities
  - Starting with n isolated nodes and connecting each pair of nodes with probability p
  - As a result, all nodes have roughly the same number of links (i.e., ***average degree, <k>***).

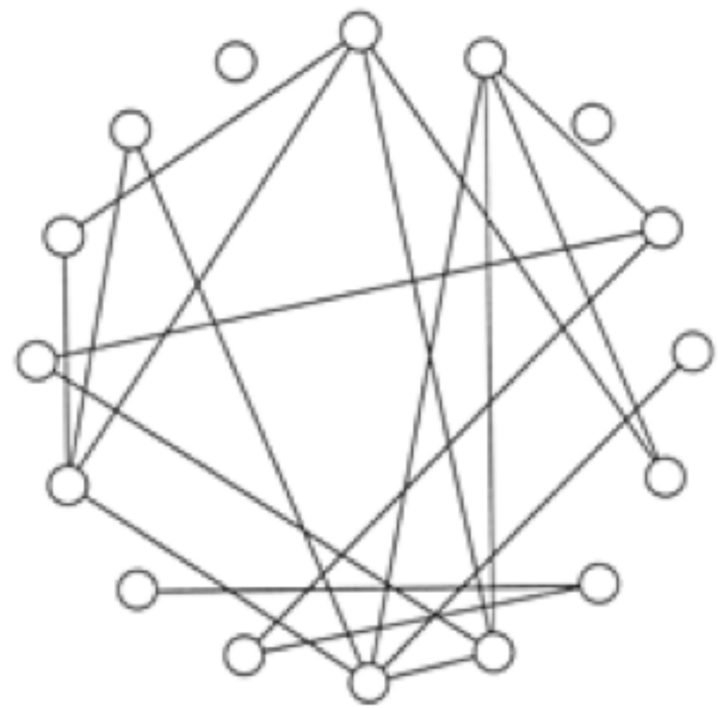


Random Networks

- In a **random network,** each pair of nodes *i, j* has a connecting link with an independent probability of *p*

- This graph has 16 nodes, 120 possible connections, and 19 actual connections—about a 1/7 probability than any two nodes will be connected to each other.

- In a random graph, the presence of a connection between A and B as well as a connection between B and C will not influence the probability of a connection between A and C.

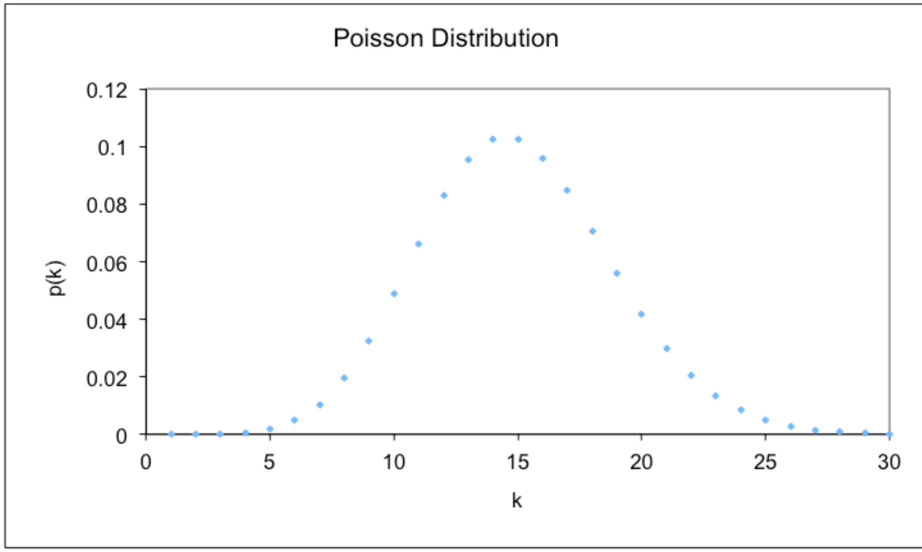


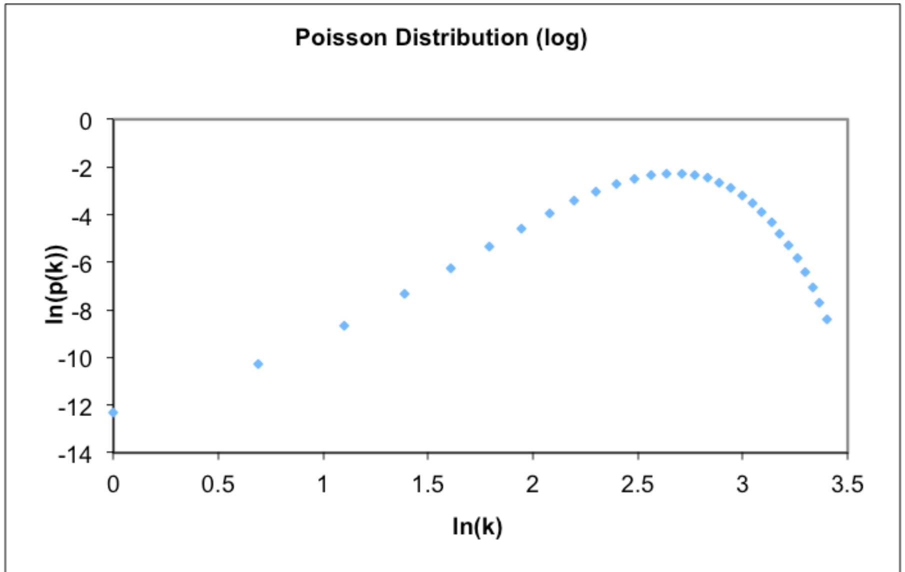
Random Graphs (Cont' d)

- Average path length: $L \sim \dfrac{\ln(n)}{\ln(\langle k \rangle)}$
- Clustering coefficient: $C = p = \dfrac{\langle k \rangle}{n}$
- Degree distribution
  - Binomial distribution for small *n* and Poisson distribution for large *n*
  - Probability mass function (PMF)

$$p(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

However, real networks are not random!



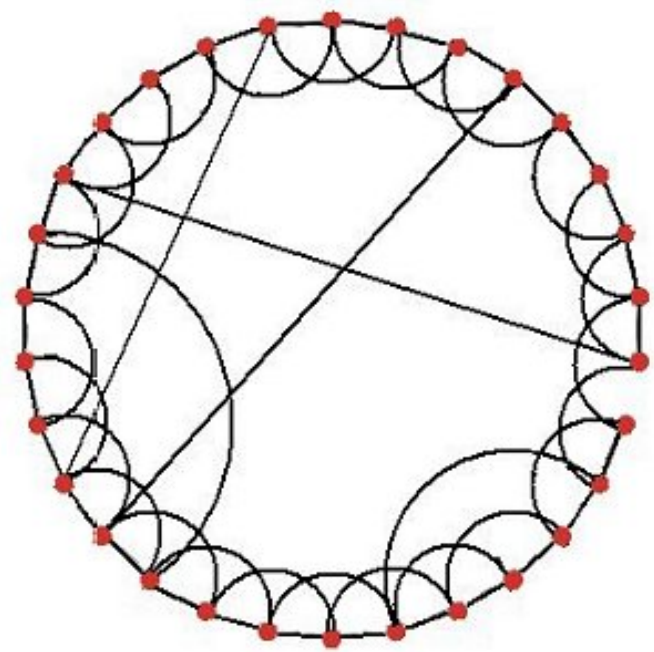


Small-World Network

- Social networks usually are small world networks in which a group of people are closely related, while a few people have far-reaching connections with people out side of the group

- Starting with a ring lattice of *n* nodes, each connected to its neighbors out to form a ring <*k*>. Shortcut links are added between random pairs of nodes, with probability ϕ (Watts & Strogatz, 1998)

- Watts-Strogatz Small World model
  - large clustering coefficient
  - high average path length



Small-World Networks

- A small-world network is defined to be a network where the typical distance L between two randomly chosen nodes (the number of steps required) grows proportionally to the logarithm of the number of nodes N in the network, that is: $L \propto \log N$
- and L_{sw} ⊠ L_{rand}

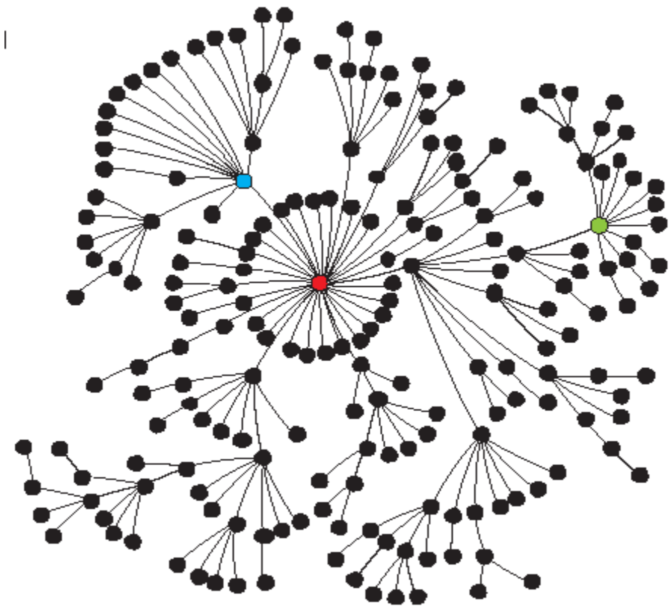
- Clustering coefficient:
  - $C_{sw} >> C_{random}$

Thus, small-world networks are characterized by **large clustering coefficient**, **small path length** relative to n.

- Degree distribution
  - Similar to that of random networks

Scale-Free (SF) Networks: Barabási–Albert (BA) Model

- "Scale free" means there is no single characterizing degree in the network



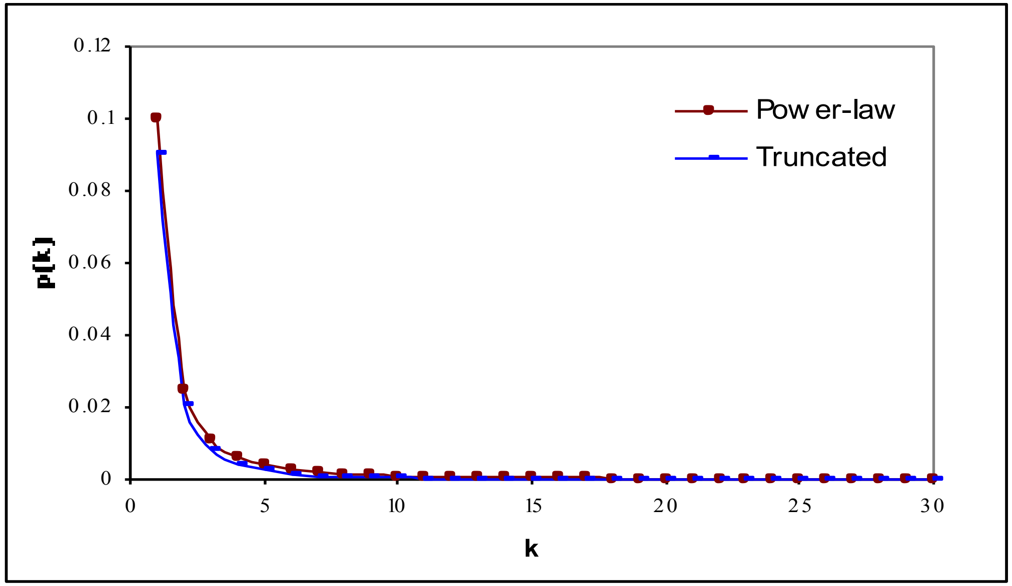
- **Growth**:
  - starting with a small number (n_0) of nodes, at every time step, we add a new node with $m(<=n_0)$ links that connect the new node to m different nodes already present in the system

- **Preferential attachment**:
  - When choosing the nodes to which the new node will be connected to node i depends on its degree k_i

Scale-Free Networks (Cont’d)

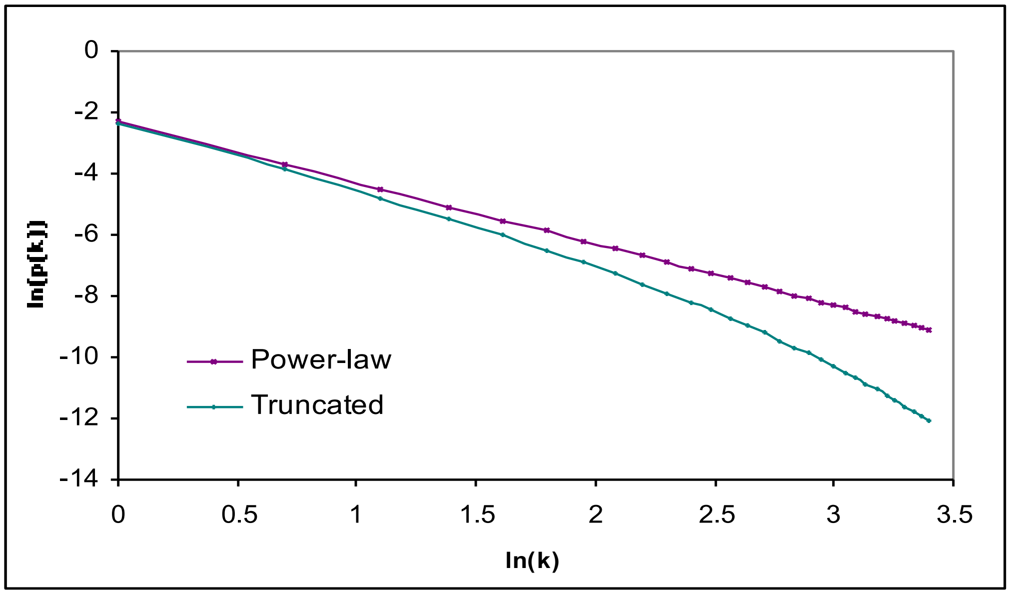
- The degree of scale-free networks follows **power-law distribution** with a flat tail for large *k*

$$p(k) \sim k^{-\gamma}$$



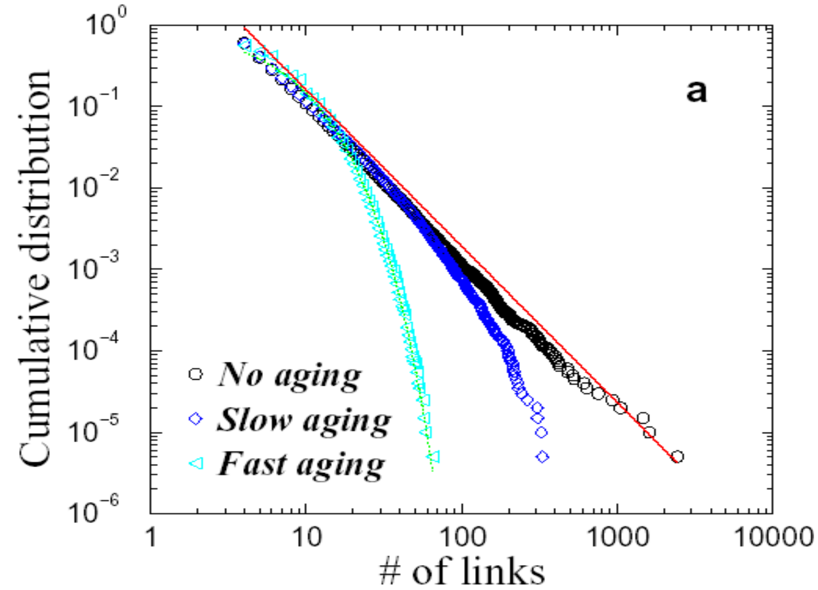
- Truncated power-law distribution deviates at the tail

$$p(k) \sim k^{-\gamma} e^{-\frac{k}{\kappa}}$$

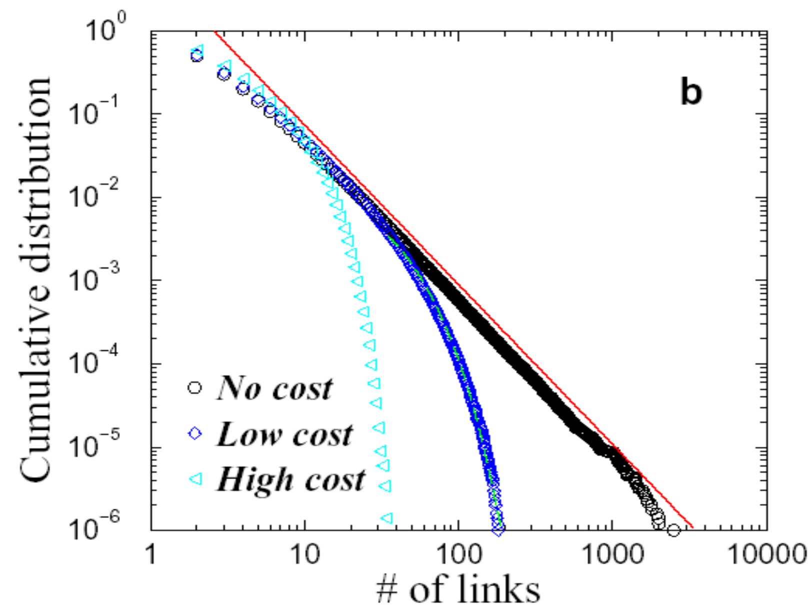


Evolution of SF Networks

- The emergence of scale-free network is due to
  - Growth effect: new nodes are added to the network
  - Preferential attachment effect (Rich-get-richer effect): new nodes prefer to attach to "popular" nodes



- The emergence of truncated SF network is caused by some constraints on the maximum number of links a node can have such as (Amaral, Scala et al. 2000)
  - Aging effect: some old nodes may stop receiving links over time
  - Cost effect: as maintaining links induces costs, nodes cannot receive an unlimited number of links



Network Analysis: Topology Analysis

Topology	Average Path Length (L)	Clustering Coefficient (CC)	Degree Distribution ($P(k)$)
Random Graph	$L_{rand} \sim \dfrac{\ln N}{\ln\langle k\rangle}$	$CC_{rand} = \dfrac{\langle k\rangle}{N}$	Poisson Dist.: $P(k) \approx e^{-\langle k\rangle}\dfrac{\langle k\rangle^{k}}{k!}$
Small World (Watts & Strogatz, 1998)	***L_{sw} ☒ L_{rand}***	***CC_{sw} AAA CC_{rand}***	Similar to random graph
Scale-Free network	*L_{SF} ☒ L_{rand}*		**Power-law Distribution: $P(k) \sim k^{-■}$**

$\langle k\rangle$: Average degree