Application 3: Social Network Analysis

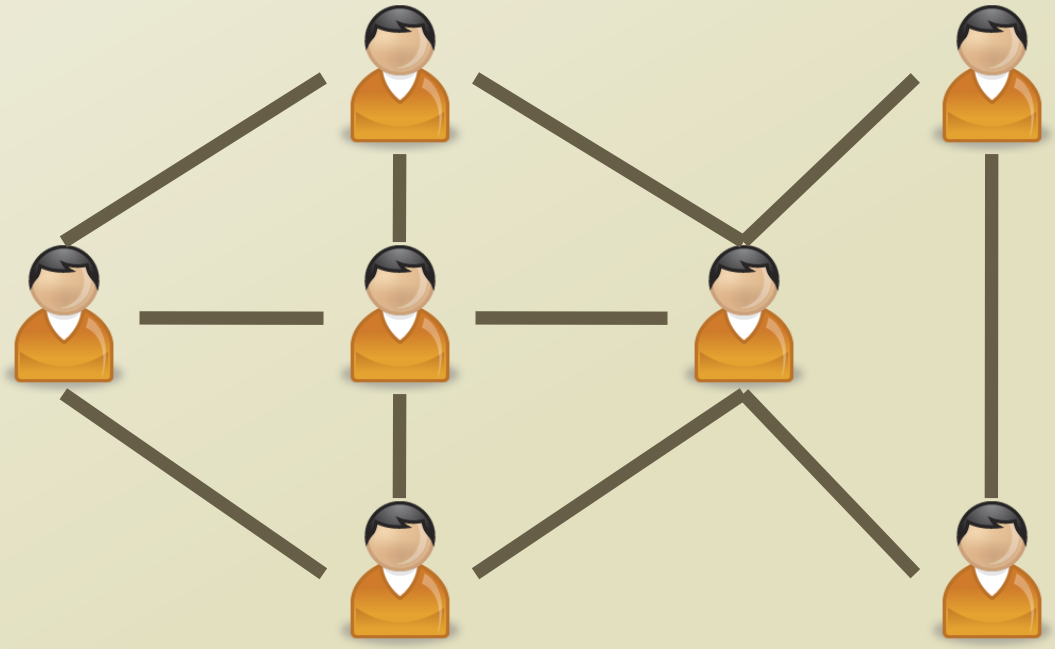
Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- This week, we study social networks and society structures.
  - Understand the basics in social networks and society structures.
    - Standard analytical techniques
      - Measures and algorithms
  - Understand the basic topologies of social networks and society structures

What is a social network?

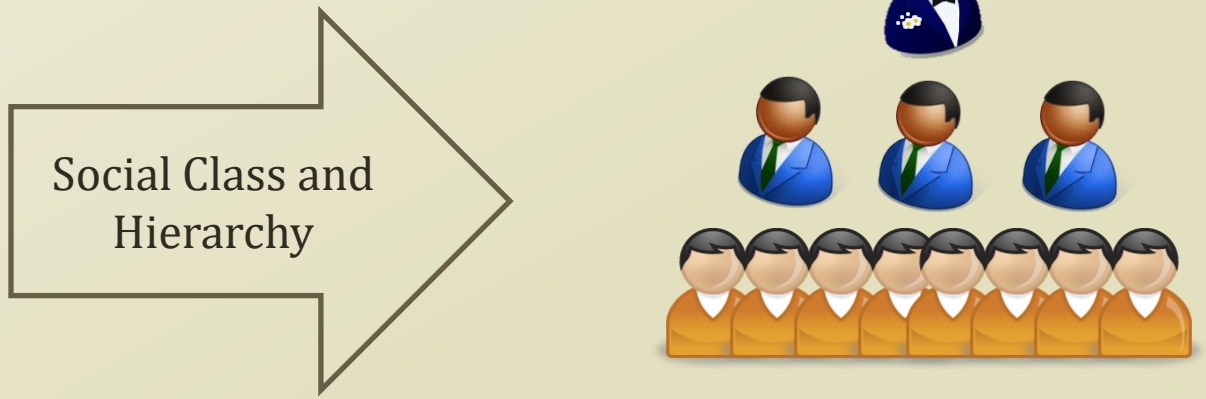
- Relations among people
  - Nodes are people, and links are relations.
- Difference between network and graph
  - Graph: binary matrix
  - Network: weighted matrix



Organization and Society

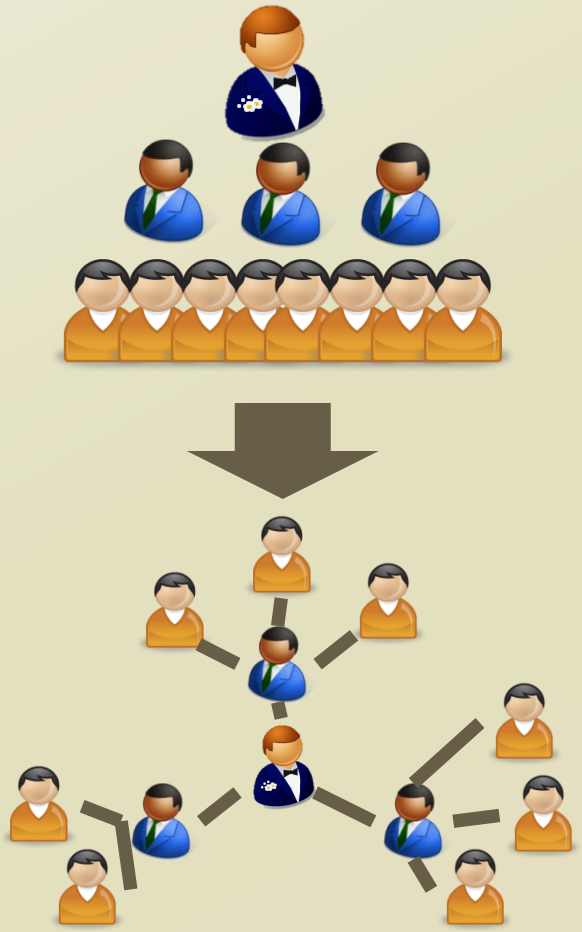
Max Weber
One of the three principal architects of modern social science

- Let's go back to the father of modern sociology!
  - Max Weber (1864-1920) think
    - An institution to control individuals in the interest of the organization leaders' goals (Weber, 1978)
  - Leader? Control? Individuals?
  - Institution=System=Tool?
  - Okay, then where are the links?



Modern View on Organizational Structure

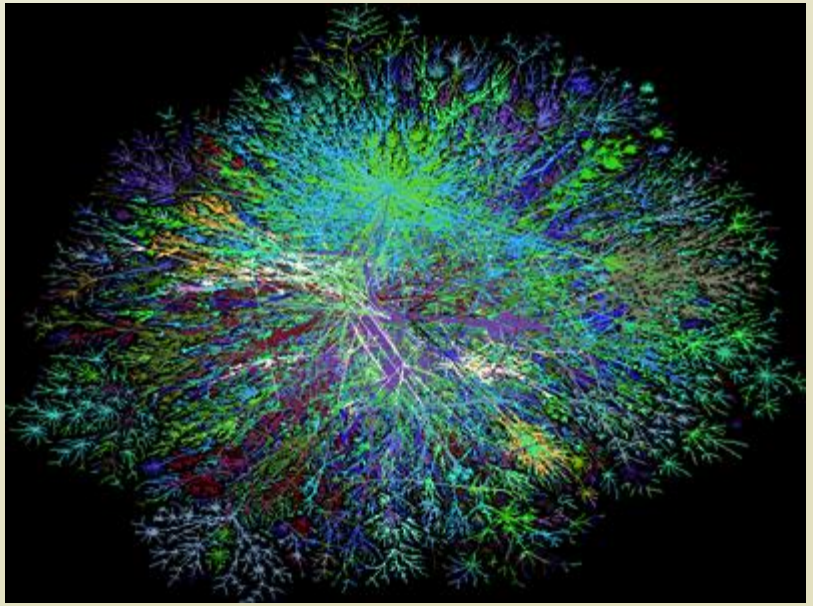
- Social networks as a metaphor of a social system
  - Leader at the center. Minions as pendants
- Barnes, 1954
  - Started focusing on the patterns of ties
    - Pattern between bounded groups and social categories
- March and Simon, 1958
  - Organization analysis
    - From the social structure viewpoint
    - They started to see social networks as social structures.

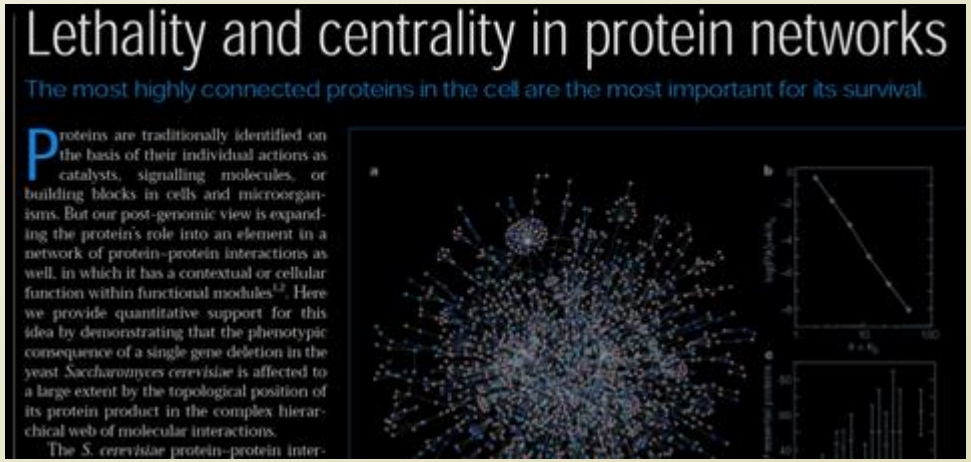


New Breeds in SNA

- Internet and world wide web
  - Computer Scientists
- Protein structures and material structures
  - Biologists and Physicists
- Their common interests?

Albert, Jeong, and Barabasi (1999)





Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival

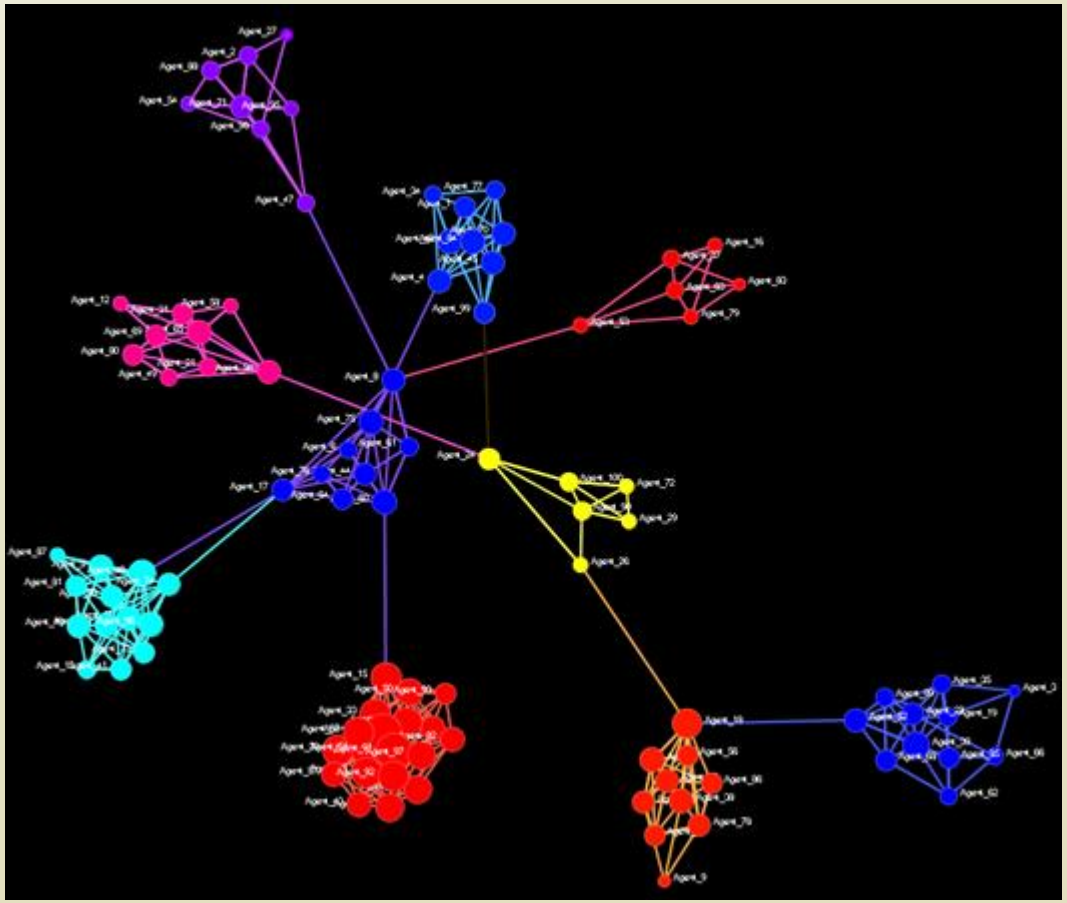
Proteins are traditionally identified on the basis of their individual actions as catalysts, signalling molecules, or building blocks in cells and microorganisms. But our post-genomic view is expanding the protein's role into an element in a network of protein–protein interactions as well, in which it has a contextual or cellular function within functional modules[1,2]. Here we provide quantitative support for this idea by demonstrating that the phenotypic consequence of a single gene deletion in the yeast *Saccharomyces cerevisiae* is affected to a large extent by the topological position of its protein product in the complex hierarchical web of molecular interactions.

The *S. cerevisiae* protein–protein inter-

Jeong, Mason, Barabasi, and Oltvai (2001)

Key Techniques in Social Network Analysis

- So, they started to see social networks as social structures.
- Then, how to find the leaders, the organizations, and the
  system?
- Measures
  - Required by sociologists looking for a key personnel
- Clusters
  - Required by sociologists looking for a sub-group
- Dynamics
  - Triadic closure and strength of weak ties



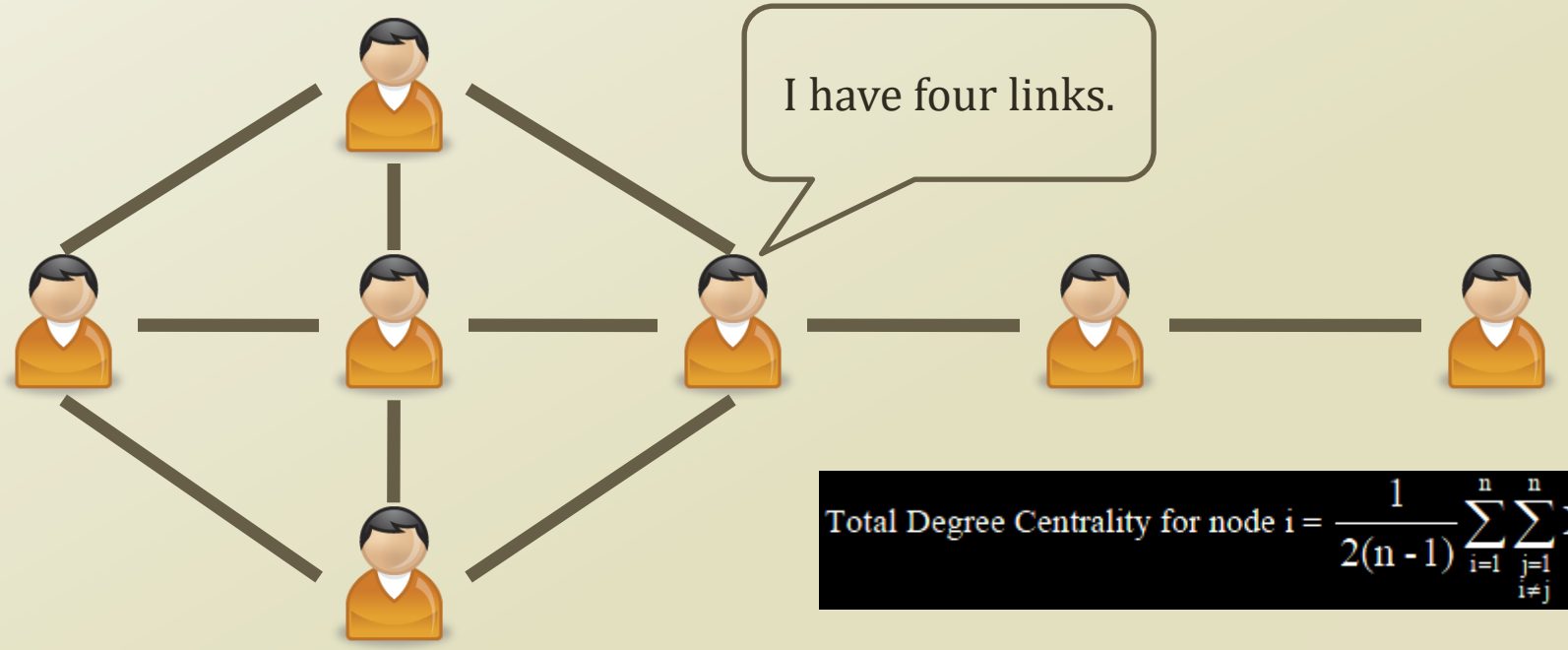
Centralities as Measures

- From a suggestive metaphor to an analytic approach
  - Metaphor: This person at the center must be the leader!
  - Analytics: This person with a high score must be the leader!
- We need Numbers, Scores!
- Mathematical sociologists searched correlations between numbers and roles
  - Some numbers found to be useful became metrics and named as centralities
  - Degree centrality
  - Betweenness centrality
  - Eigenvector centrality

  (Freeman, 1979; Bonacich, 1972)

Centralities: Degree Centrality

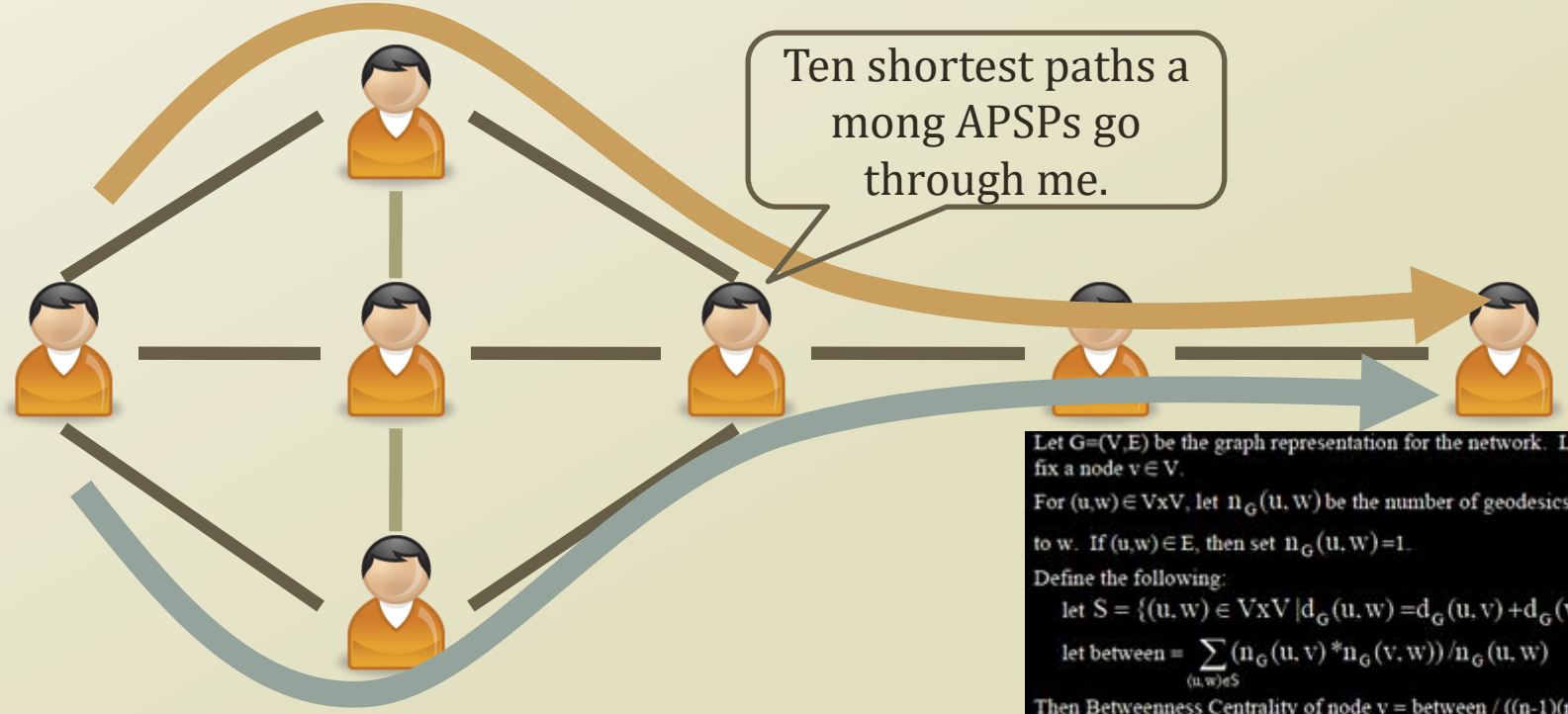
- Simple! Counting the number of links
- Local measure
  - Do not need the network-wise calculation
- Approximate hubs in a network



$$\text{Total Degree Centrality for node } i = \frac{1}{2(n-1)} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} X(i,j)$$

Centralities: Betweenness Centrality

- Simple! Counting the number of all-pair shortest paths going through the node
- Global measure
  - Need the whole network structure to calculate a value for a node
- Interesting implications
  - Has been suggested as a information/influence transfer indicator



Let G=(V,E) be the graph representation for the network. Let n=|V|, and fix a node $v \in V$.

For $(u,w) \in VxV$, let $n_G(u,w)$ be the number of geodesics in G from u to w. If $(u,w) \in E$, then set $n_G(u,w)=1$.

Define the following:

let $S = \{(u,w) \in VxV \mid d_G(u,w) = d_G(u,v) + d_G(v,w)\}$

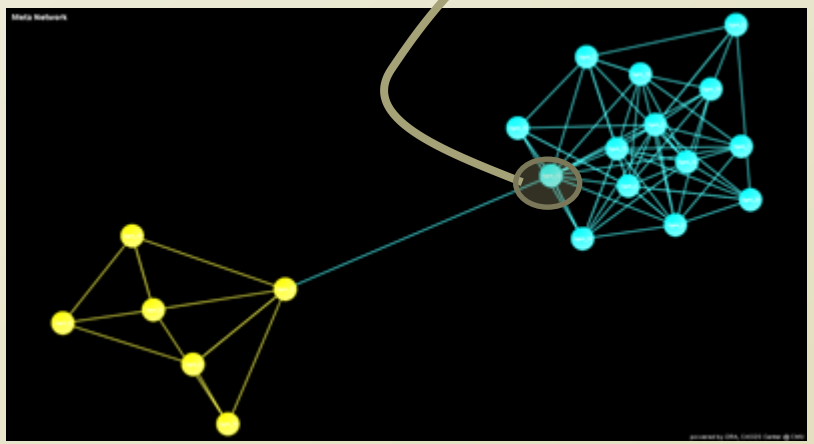
let between = $\sum_{(u,w) \in S} (n_G(u,v) * n_G(v,w)) / n_G(u,w)$

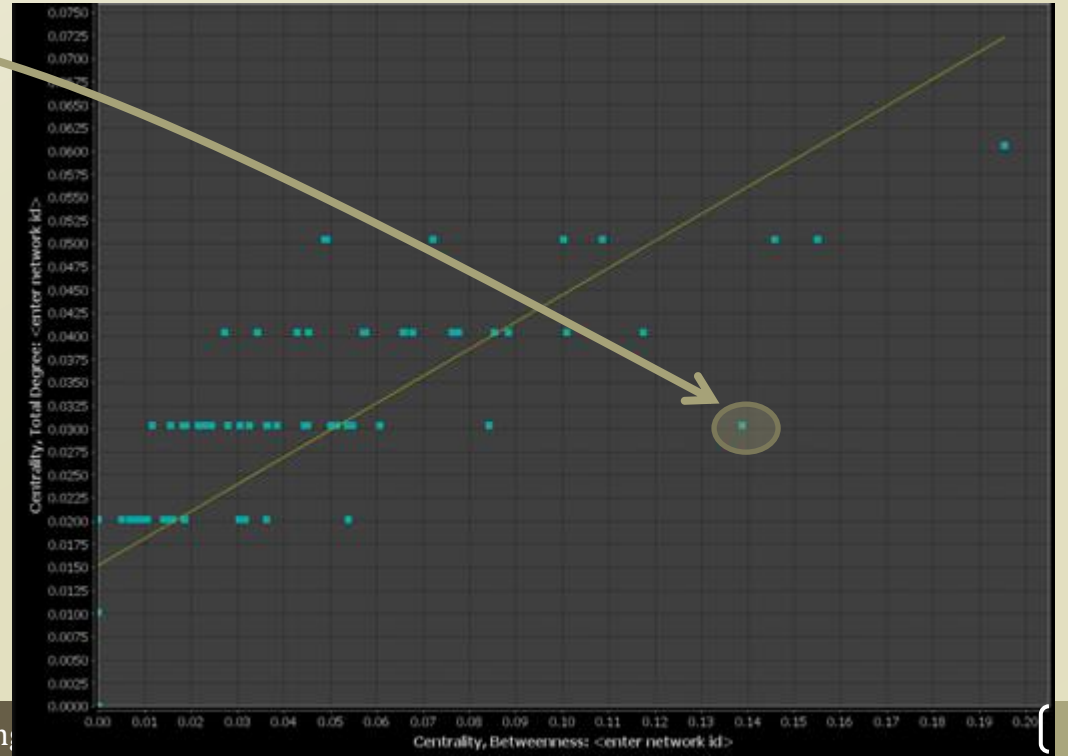
Then Betweenness Centrality of node v = between / ((n-1)(n-2)/2).

Note: if G is not symmetric, then between is normalized by (n-1)(n-2).

Low Degree and High Betweenness

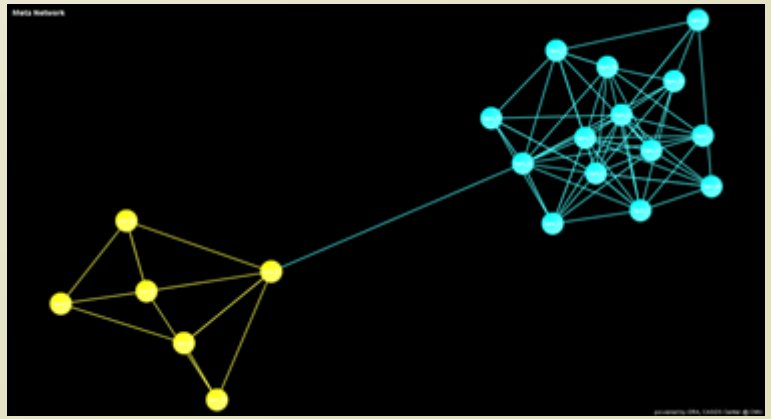
- High degree centrality people
  - Often, they are hubs. They are the elites.
- High betweenness centrality people
  - Often, they are information brokers.
- Then, low degree and high betweenness people?
  - They are not network elites, but they control information!
  - Bridging nodes!

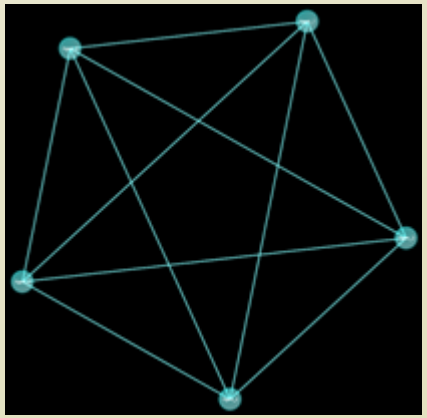


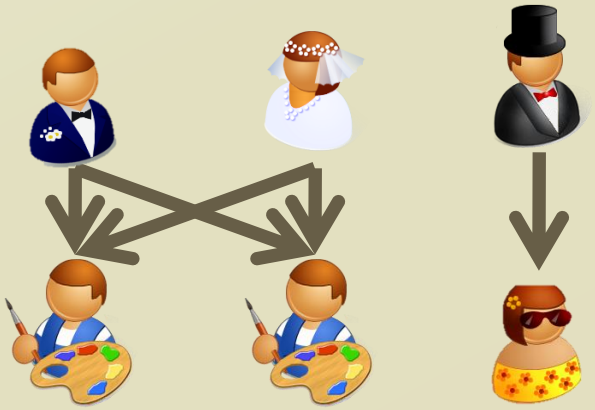


Clusters as Social Groups

- How to divide the social entities into sub groups?
- Graph theoretic groups
  - K-Cliques
- Cohesive groups
  - Form a tightly linked components



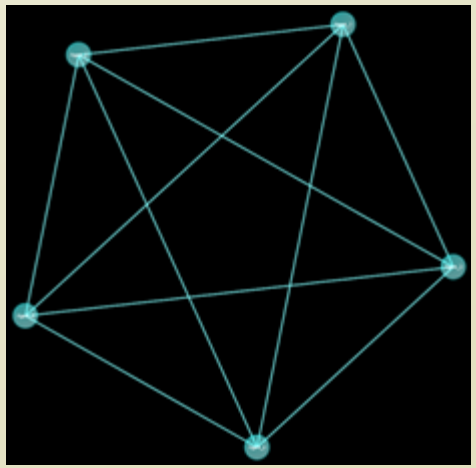




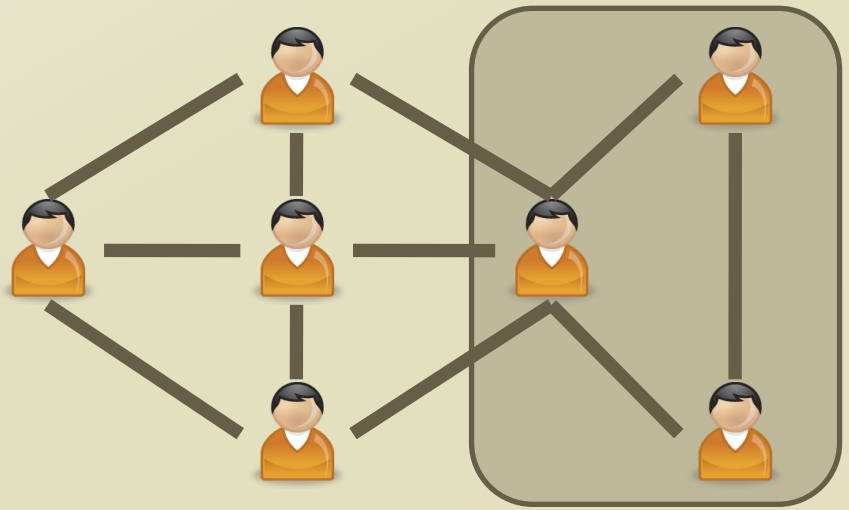
K-Clique

- Clique in an undirected graph G
  - A set of vertices V such that for every two vertices in V, there exists an edge connecting the two.
- Complete network
- NP-Complete problem, only works in a small network or very small K

5-Clique

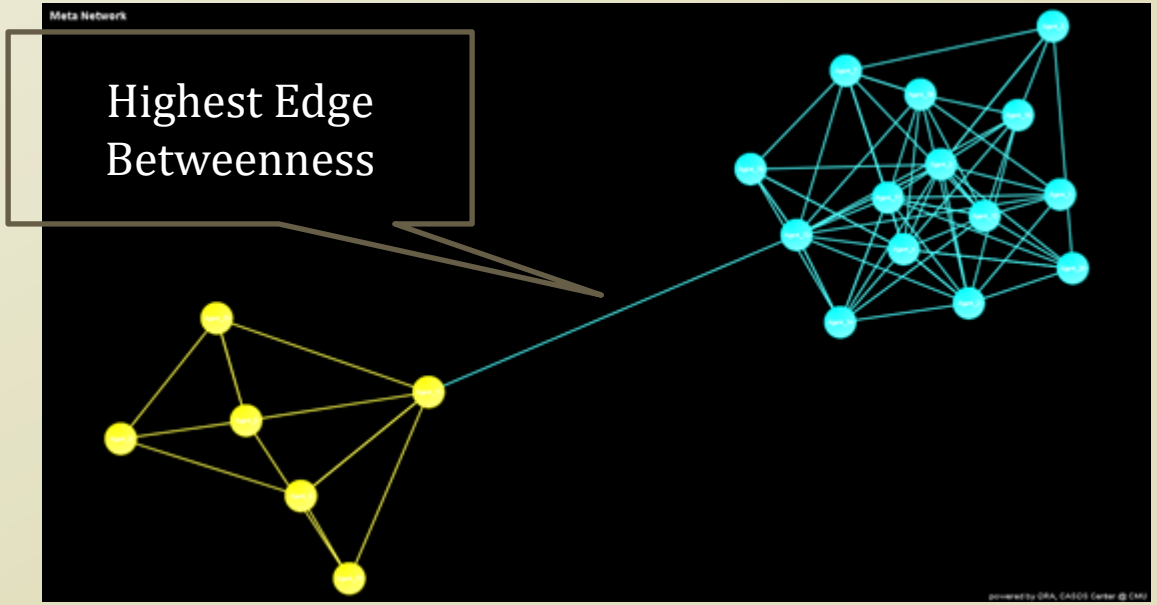


3-Clique



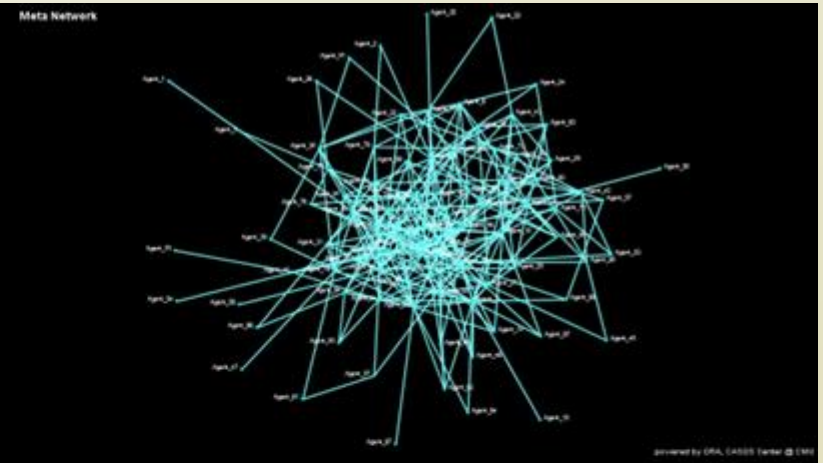
Newman Clustering

- Girvan-Newman Algorithm (Girvan and Newman, 2002)
  - The betweenness of all existing edges in the network is calculated first.
  - The edge with the highest betweenness is removed.
  - The betweenness of all edges affected by the removal is recalculated.
  - Steps 2 and 3 are repeated until no edges remain.
- Pretty nice tool to find a cohesive group

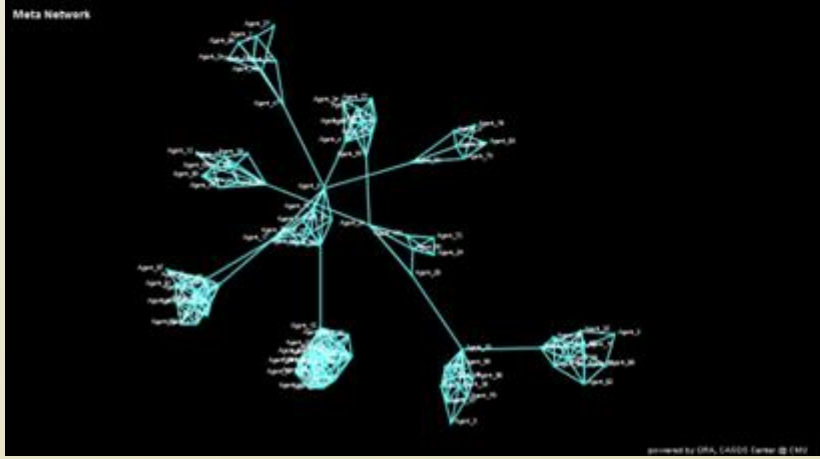


Network Visualization

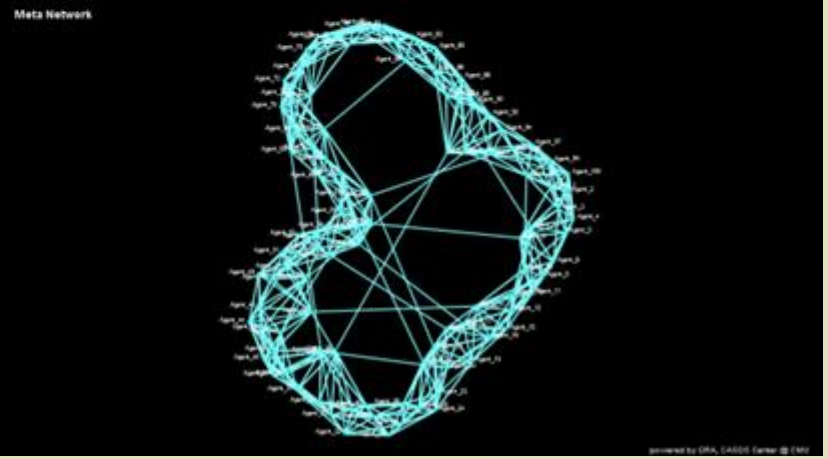
Core-Periphery Network



Scale Free Network

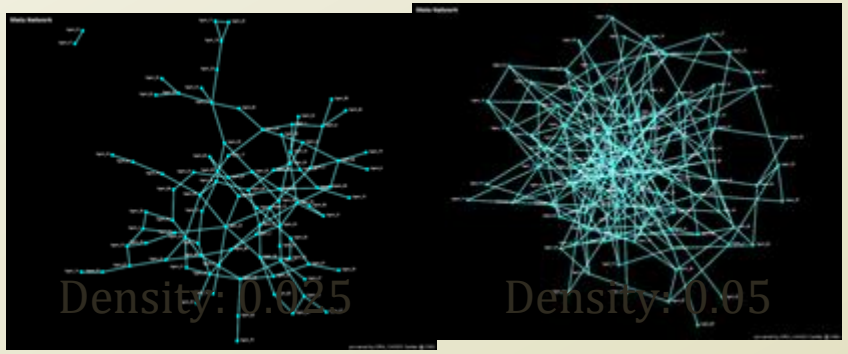


Cellular Network



Small World Network

Network Density



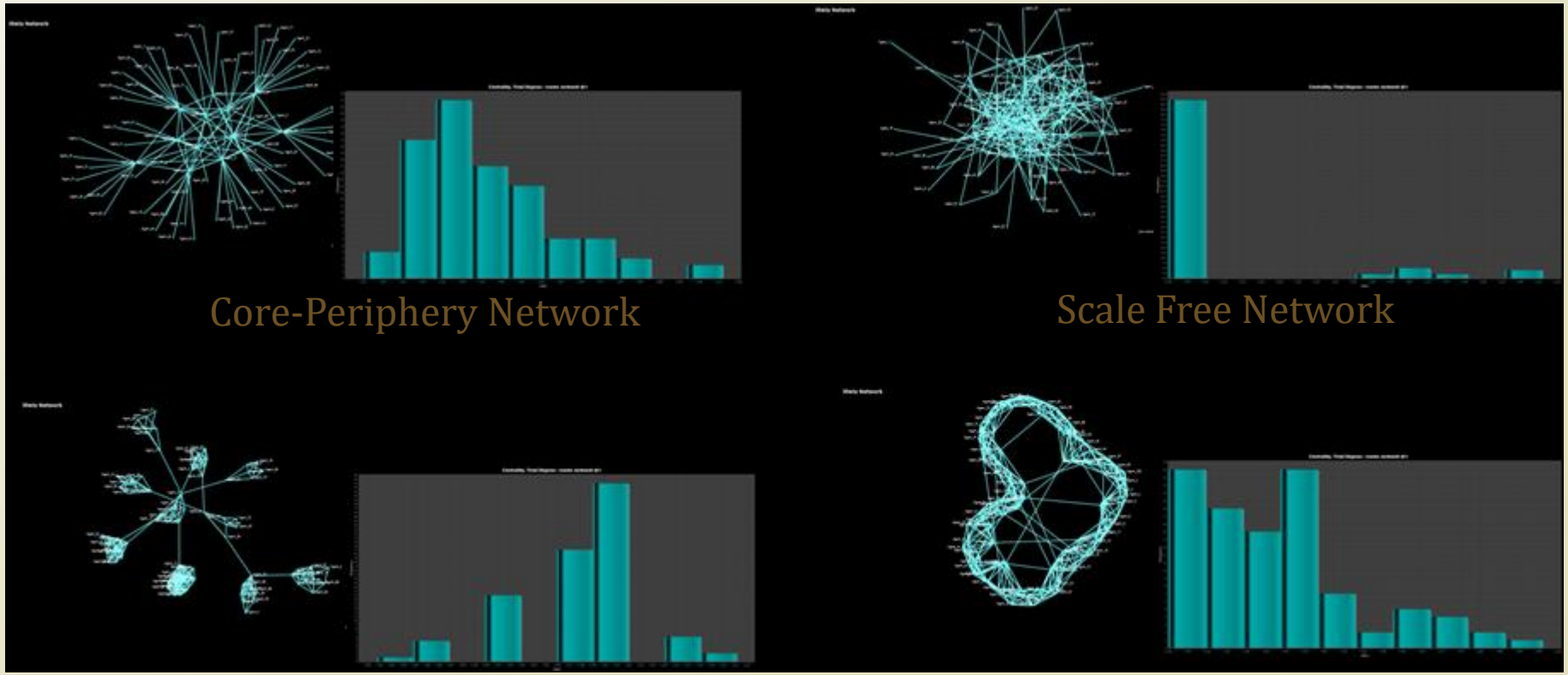
- One of the simplest metrics about network status
  - Things you need to report when you write a SNA paper
    - Number of nodes
    - Network Density
    - Sometimes
      - Network diameter
      - Pareto distribution parameter if the degree distribution follows the power-law

$$Density = \frac{sum(M)}{|M|(|M|-1)}$$

- One trend
  - Social network density is usually very low.
    - Why?
    - One pendant node increase will induce huge network density drop.
    - Remember the adjacency matrix will grow $O(n^2)$

Degree Distribution

- The overall shape of networks
- How to statistically recognize a network topology?



Core-Periphery Network

Scale Free Network

Cellular Network

Small World Network