COM
CITY UNIVERSITY OF HONG KONG
DEPARTMENT OF MEDIA AND COMMUNICATION (COM),
COLLEGE OF LIBERAL ARTS AND SOCIAL SCIENCES (CLASS)
CityU

GE2234 Social Networks

for Media, Business and Technological Applications

Lecture Note 6: Cohesive Sub-groups

By Dr. Wang Xiaohui, Vincent

- Triadic closure
- Local bridge (low embeddedness)
- The strength of weak ties
- Structural holes and good ideas

Course Outline

- Cohesive subgroups
  - Cliques
  - K-core
  - Communities
  - Homophily

What makes a community (cohesive subgroup):

A cohesive subgroup is a subgroup in which a high proportion of the nodes within the subgroup share strong, direct, mutual, frequent or positive ties.

- **Mutuality of ties.** Everyone in the group has ties (edges) to one another.
- **Compactness.** Closeness or reachability of group members in small number of steps, not necessarily adjacency.
- **Density of edges.** High frequency of ties within the group.
- **Separation.** Higher frequency of ties among group members compared to non-members.

Wasserman & Faust, 1994

Why study communities (cohesive subgroups):

Greater homogeneity among persons who have relatively frequent face to face contact or who are connected through intermediaries, and less homogeneity among person who have less frequent contact (Friedkin, 1984).

"the more tightly that individuals are tied into a network, the more they are affected by group standards…"

--- Collins (1988, page 416)

Component

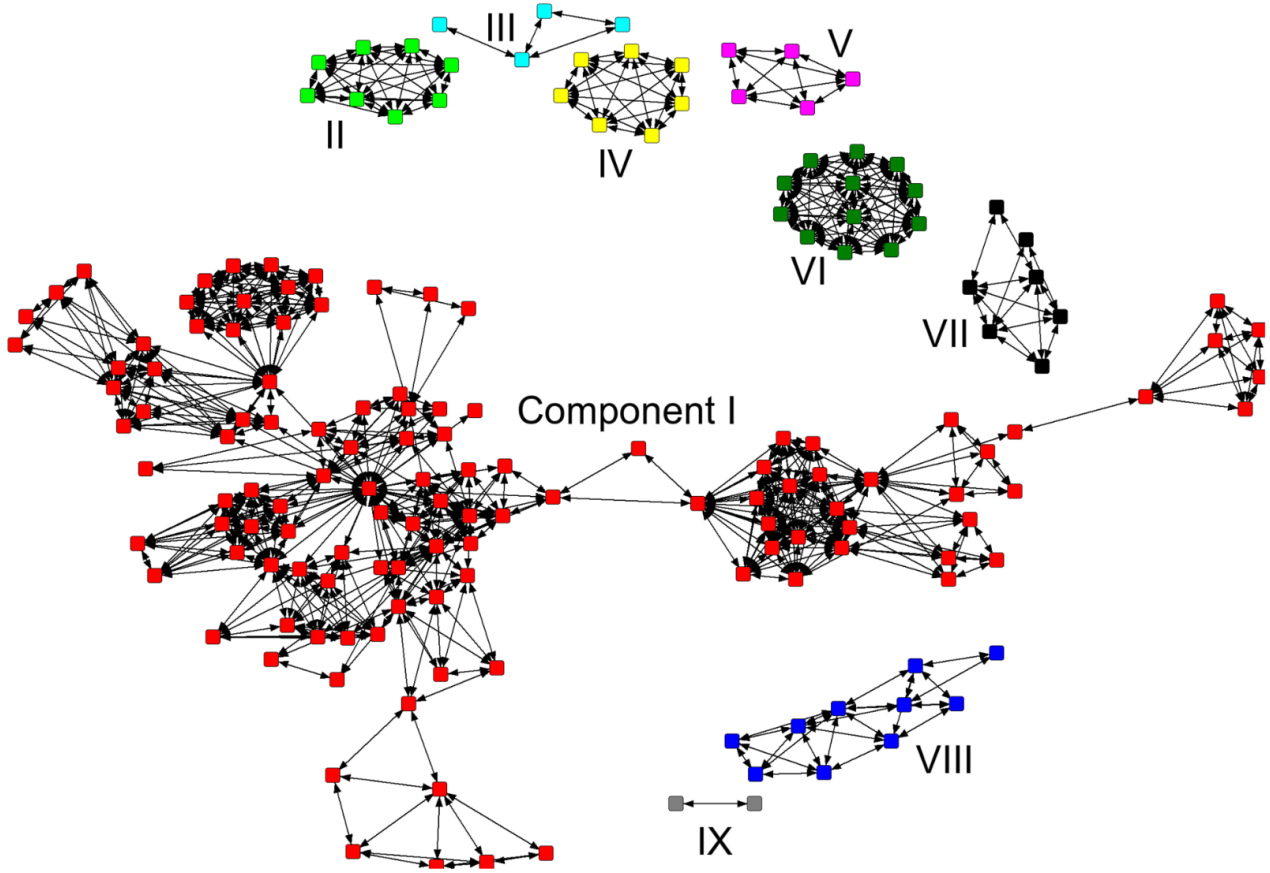
- A component consists of a subgroup of individuals, whereby all the individuals are connected to one another by at least one path (Connectivity).

Weak/Strong Component, and Isolates

- A weak component refers to nodes who are connected to one another, regardless of the direction of the edges
  - Ignore the direction of the edges
- A strong component refers to nodes being connected to one another via direct or indirect edges
  - Pay special attention to the directionality of the edges
- An isolate can be considered as a component, although it has not any edges to anyone else

Structural Cohesion and Components

- A component is generally considered the most minimum requirement for a cohesive subgroup (Moody & White, 2003)

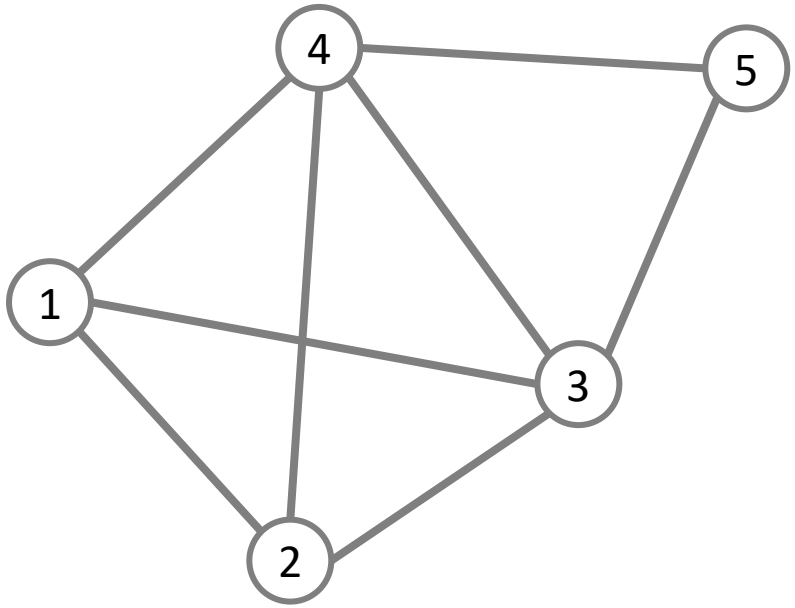


Components in Dengue Research Co-authorship Network
Source: http://www.plosntd.org/article/info:doi/10.1371/journal.pntd.0000501

Cliques

- A clique is a complete (fully connected) subgraph, i.e. a set of vertices where every pair of vertices is directly connected.

- Clique is a maximal complete subgraph (Luceand Perry, 1949)
  - Complete means that every node in the clique is adjacent to every other
  - Maximal means that we cannot increase its size and still have it to be complete

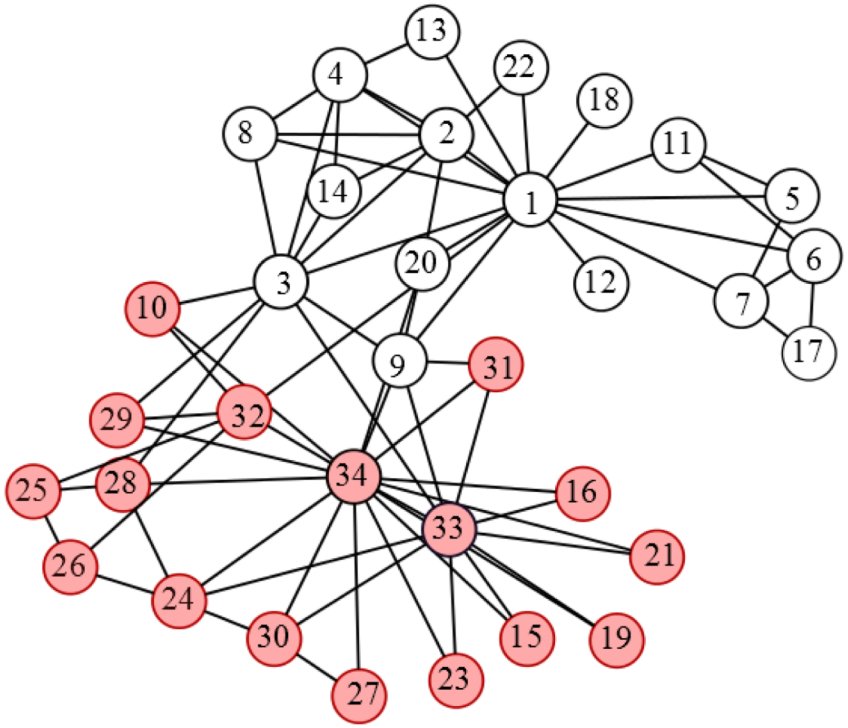


Cliques in the network:
{1, 2, 3, 4}
{3, 4, 5}

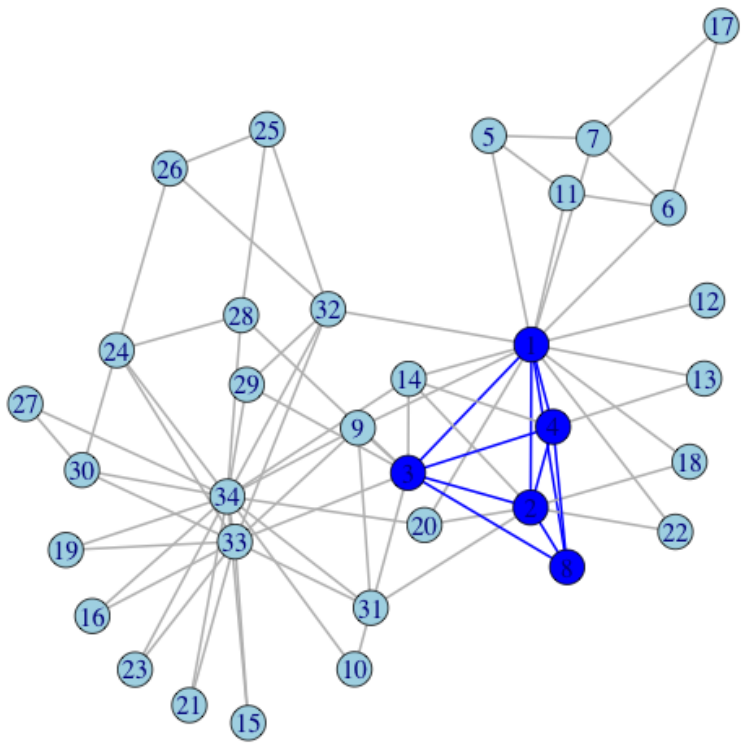
Cliques

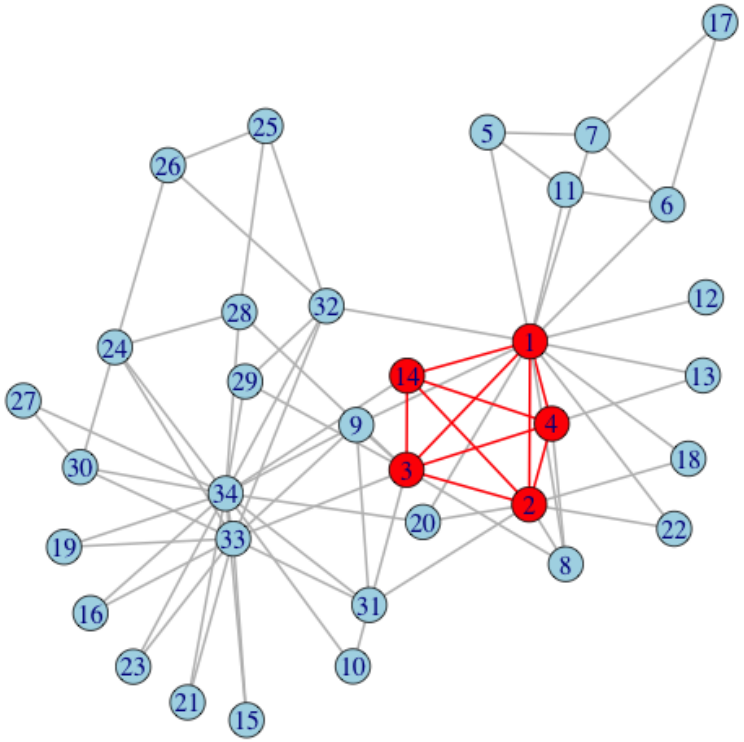
- Clique refers to informal groupings of people in which feelings of intimacy exist, and where the presence of particular group norms and sub-culture exist (Warner & Lunt, 1941).

Example: Zachary's Karate Club



Clique size	3	4	5
Number of cliques	21	2	2

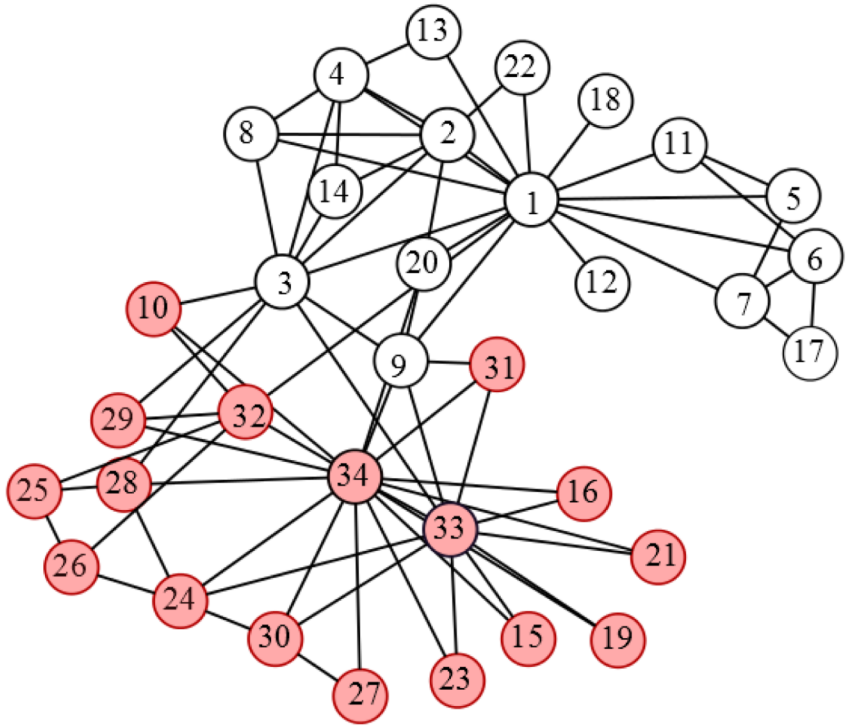


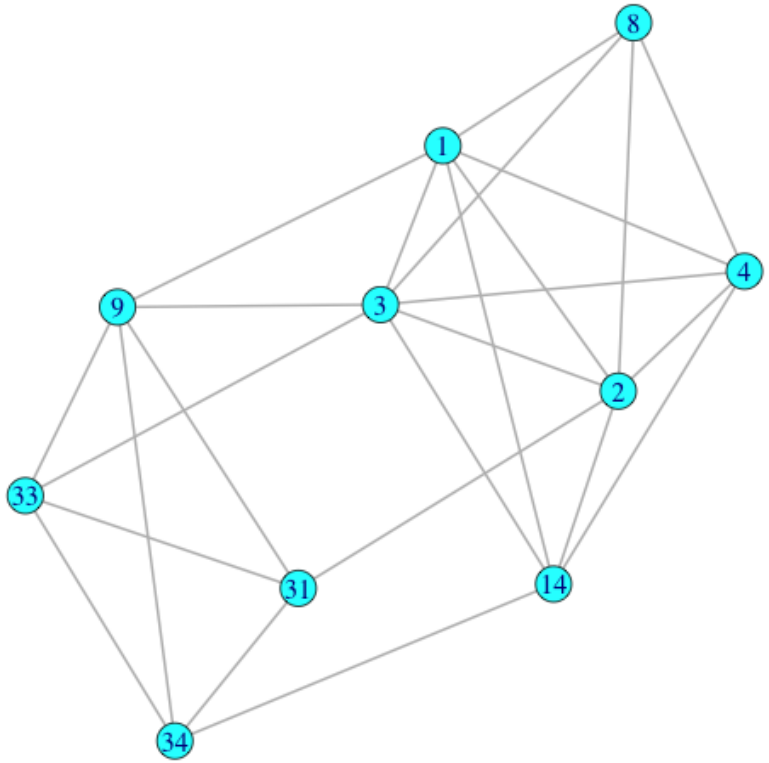


Maximum cliques

Limitation of clique

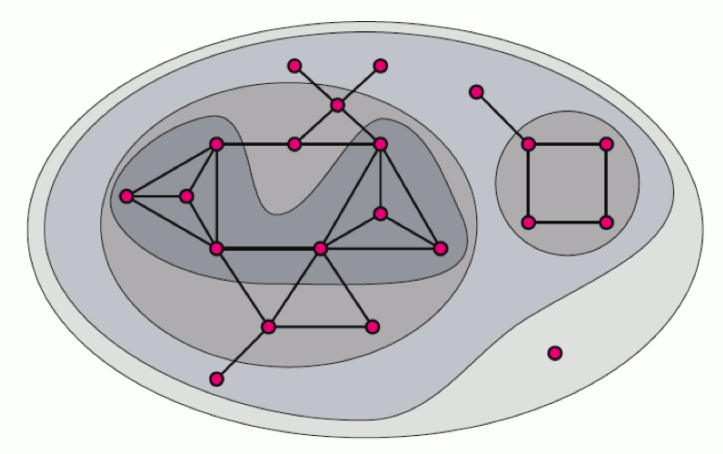
- However, cliques seldom are useful in analysis of actual data because the definition is too strict. In addition, cliques that do occur are often quite small, and overlapping one another.

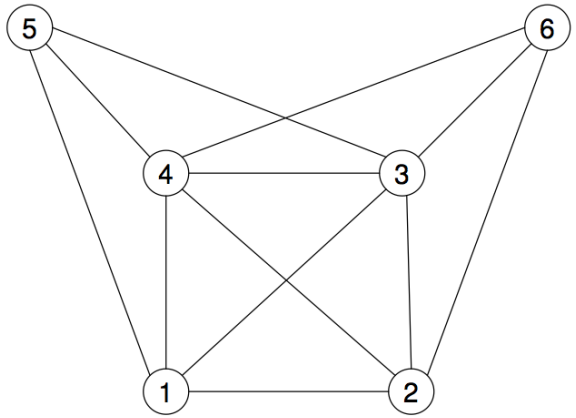




Relaxation of a clique

- k-core is a maximal subset of vertices such that each is connected to at least k others in the subset
  - degree of every vertex in k-core $C_D(i) >= k$.
  - (k+1)-core is always a subgraph of k–core.

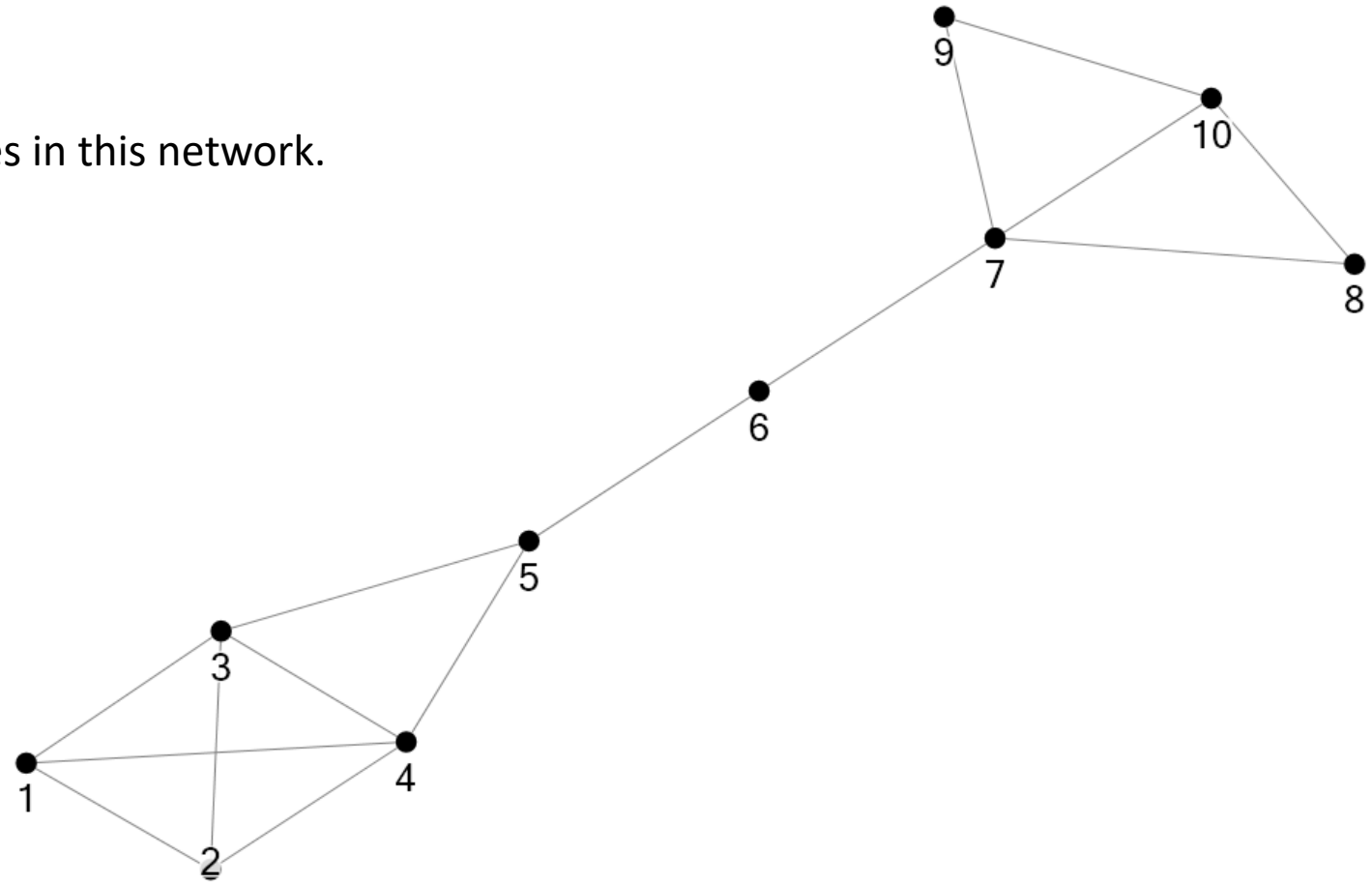




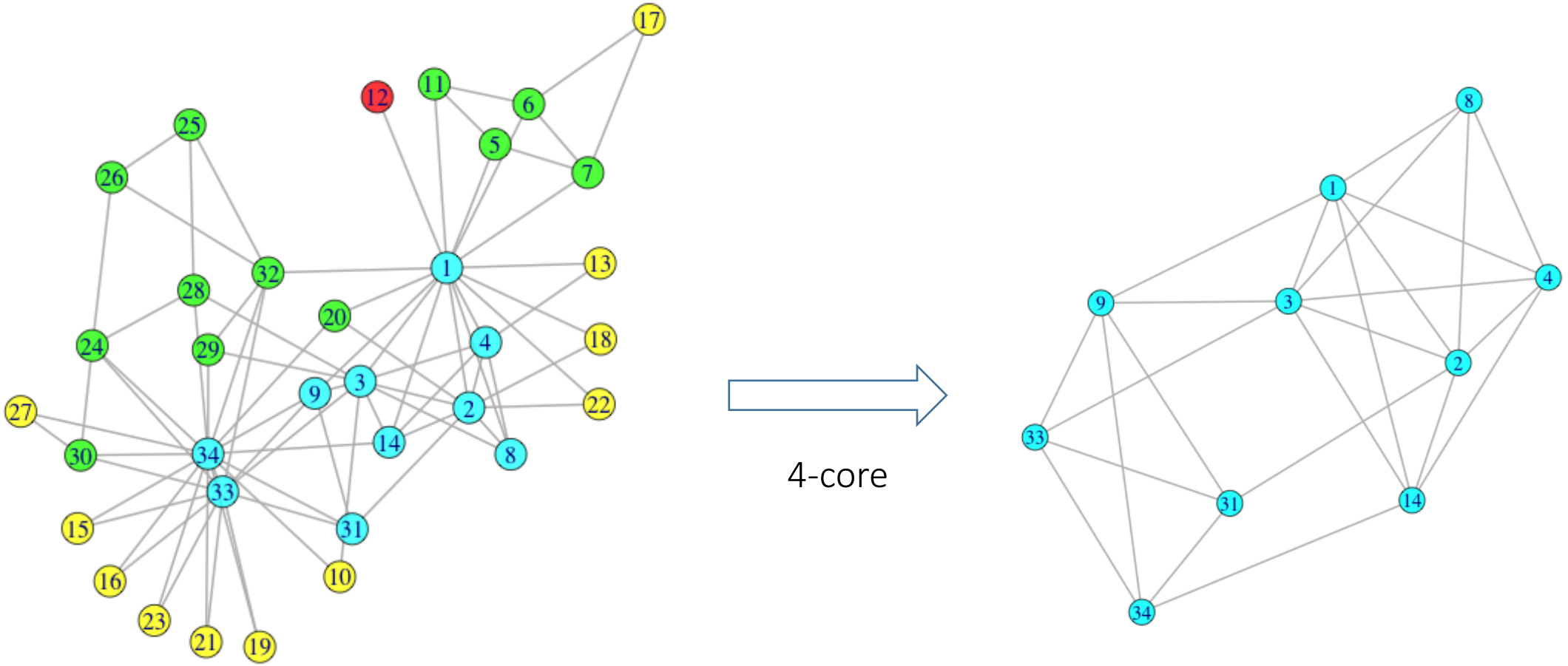
3-core

An Illustration

Find k-cores in this network.

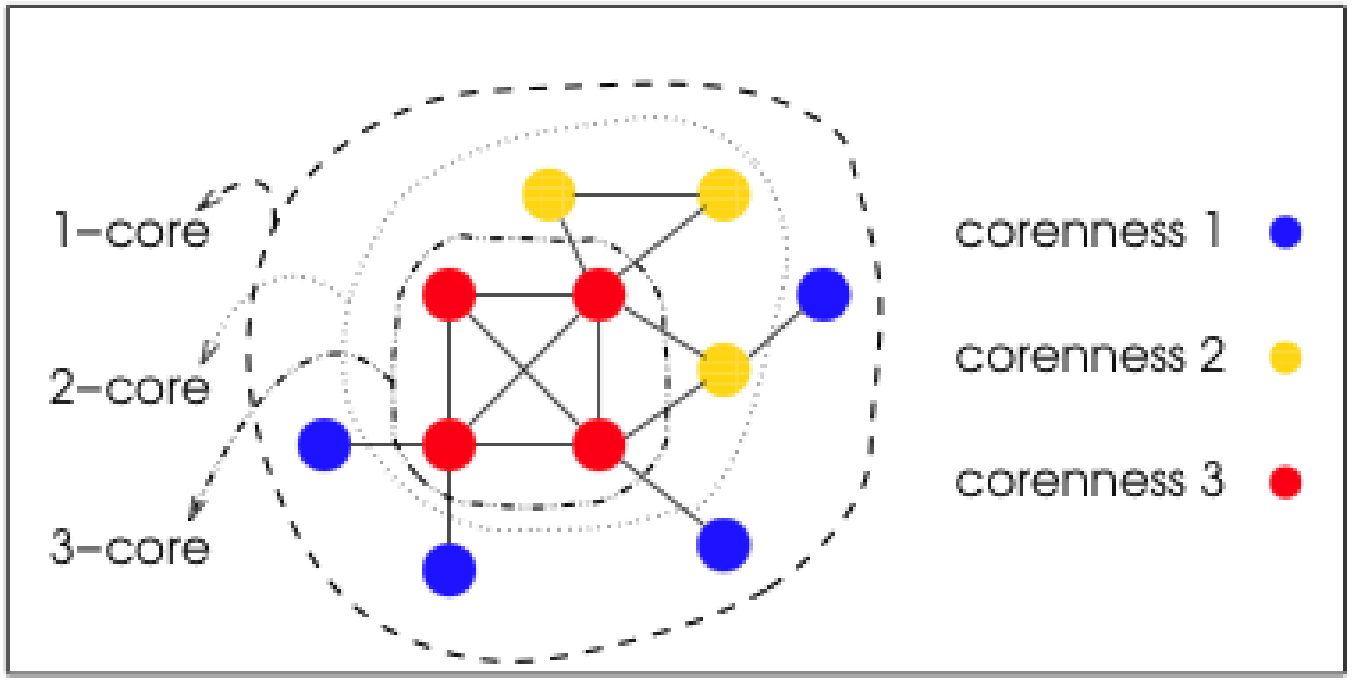


Example: Zachary's Karate Club



k-core

Corenness: the core number of a vertex, equal to the highest order of a core that contains this vertex



Limitation of clique

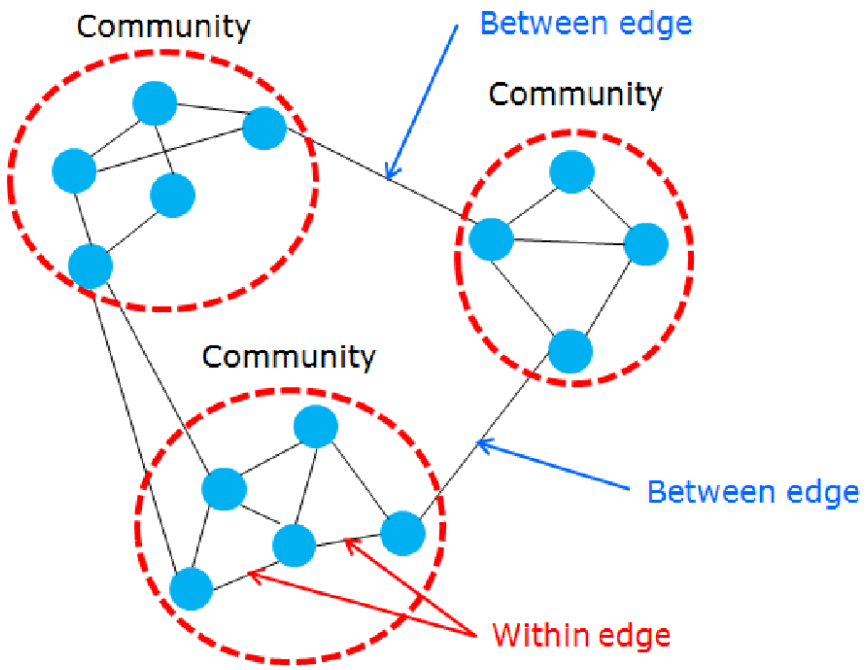
- One problem with previous group identification methods, such as clique, n-core, and k-plex, is that they return lists of groups in which individuals are members of multiple groups.
- Can be difficult to conduct analysis comparing group membership to behavior because people are not classified in mutually exclusive groups.

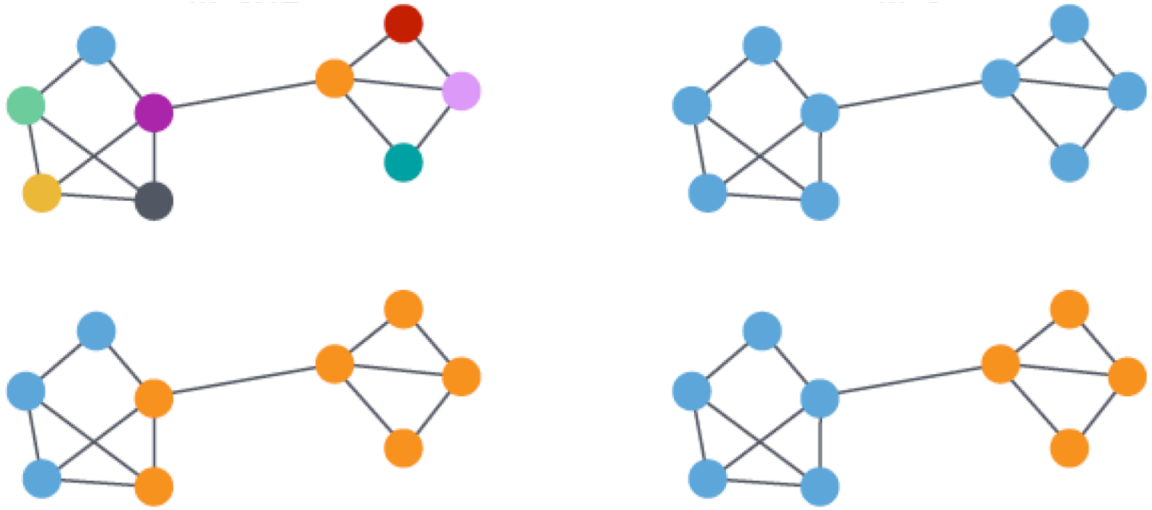
Community Detection in Social Network Analysis

- Community detection is an assignment of vertices to communities.
- Non-overlapping communities (every vertex belongs to a single group)

Network communities

- Network communities are groups of vertices such that vertices inside the group connected with many more edges than between groups.
- Basic Idea: relatively more links within communities.

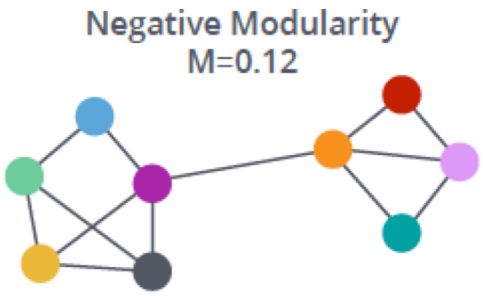


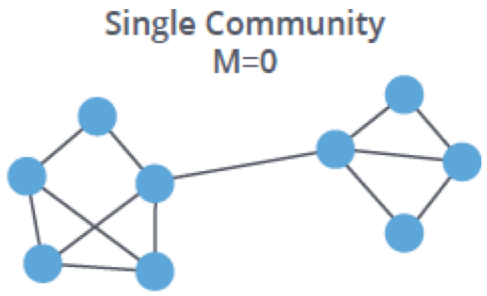


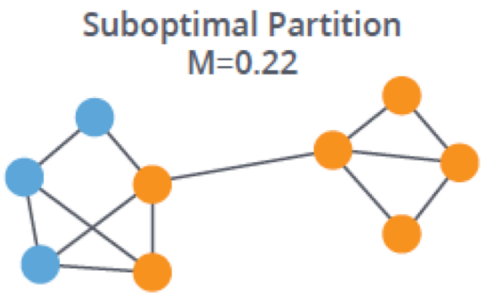
How to decide and measure the quality of community detection?

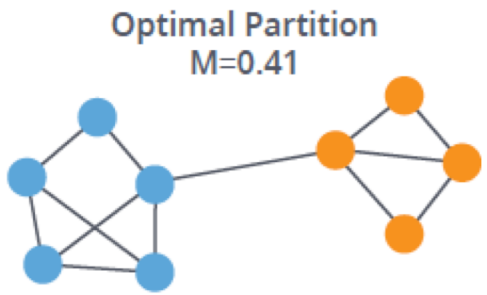
Modularity - used as a measure for quality of community detection

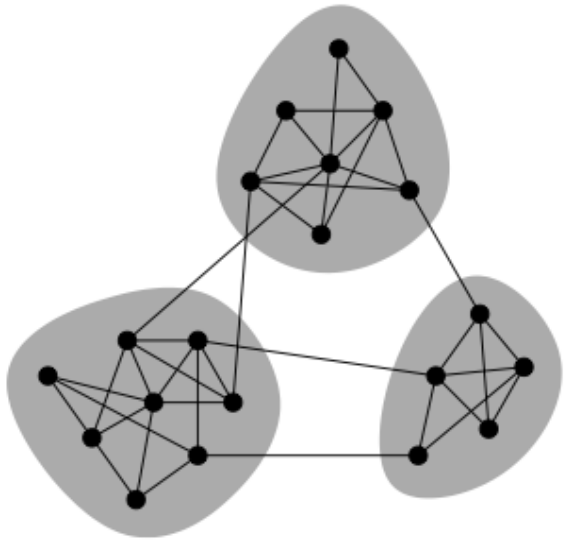
- **Modularity** measures the strength of division of a network into communities. Networks with high modularity have dense connections between the nodes within communities but sparse connections between nodes in different modules.
- Modularity is a score between 0 and 1.
- **The higher the modularity score - the better quality of community detection.**



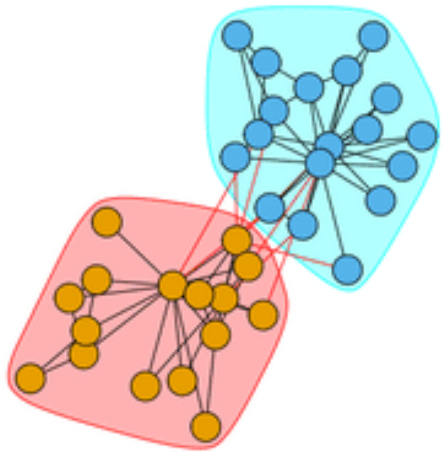






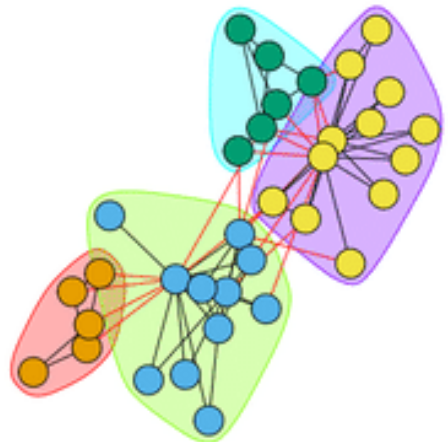


a) Ground Truth



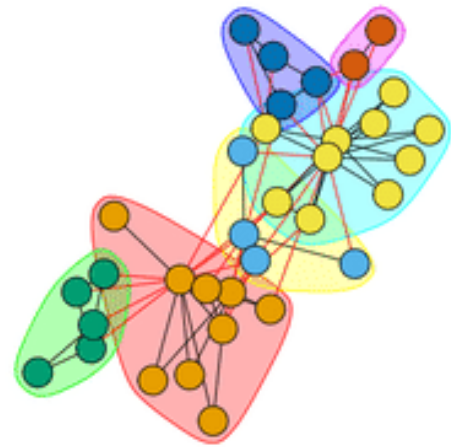
Modularity: 0.37

b) Louvain



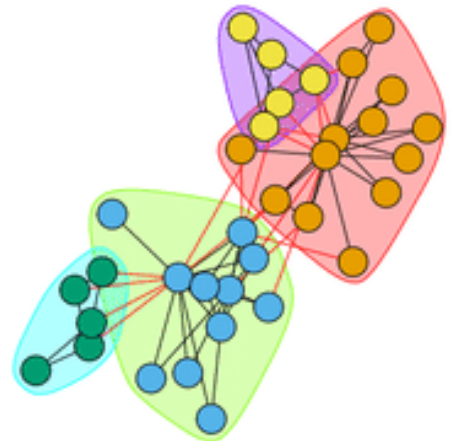
Modularity: 0.45

c) Girvan-Newman



Modularity: 0.34

d) Walktrap



Modularity: 0.44

The Karate Club

Cohesion - interpretation

- Cohesion is perceived as a field of forces that keep people in a group (or network) together (Lewin, 1951).
    - Psychology-oriented view on cohesion emphasizes people's feelings of belonging or identification with a group
    - Relational view on cohesion emphasizes the ties linking nodes together

- Usually, feelings or belonging are seen as going hand in hand with relational belonging, and groups that are tied together relationally are also seen as composed of individuals who share similar beliefs and values (Friedkin, 1984; Collins, 1988; Moody & White, 2003)

- When conceptualized as a form of network structure, cohesion is seen as influencing the behavior and/or values of individuals and the collective.
- Individuals within a cohesive subgroup tend to share norms and often have common goals and ideals.

Homophily and Social Influence

Networks in their surrounding contexts

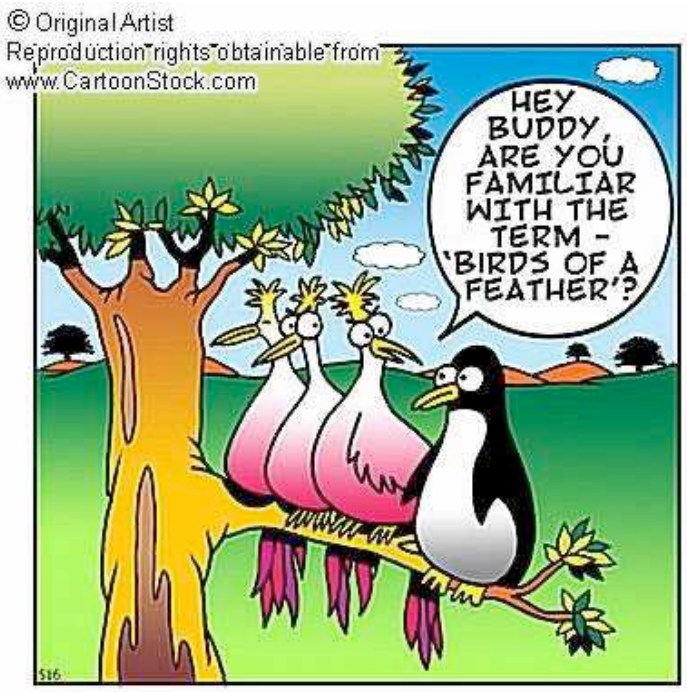
- The contexts in which a social network is embedded will generally have significant effects on its structures
- Each individual in a social network has a distinctive set of personal characteristics, and similarities and compatibilities between two people's characteristics can strongly influence whether a link forms between them
- In addition to endogenous mechanism (e.g., triadic closure), how factors existing outside the nodes and edges of a network affect how the network's structure evolves.

Homophily

- We tend to have similar characteristics with our friends
  - Age, gender, race/ethnicity, education, music
- “We are a moving average of our associates.” (Miller McPherson)

Birds of a Feather Flock Together

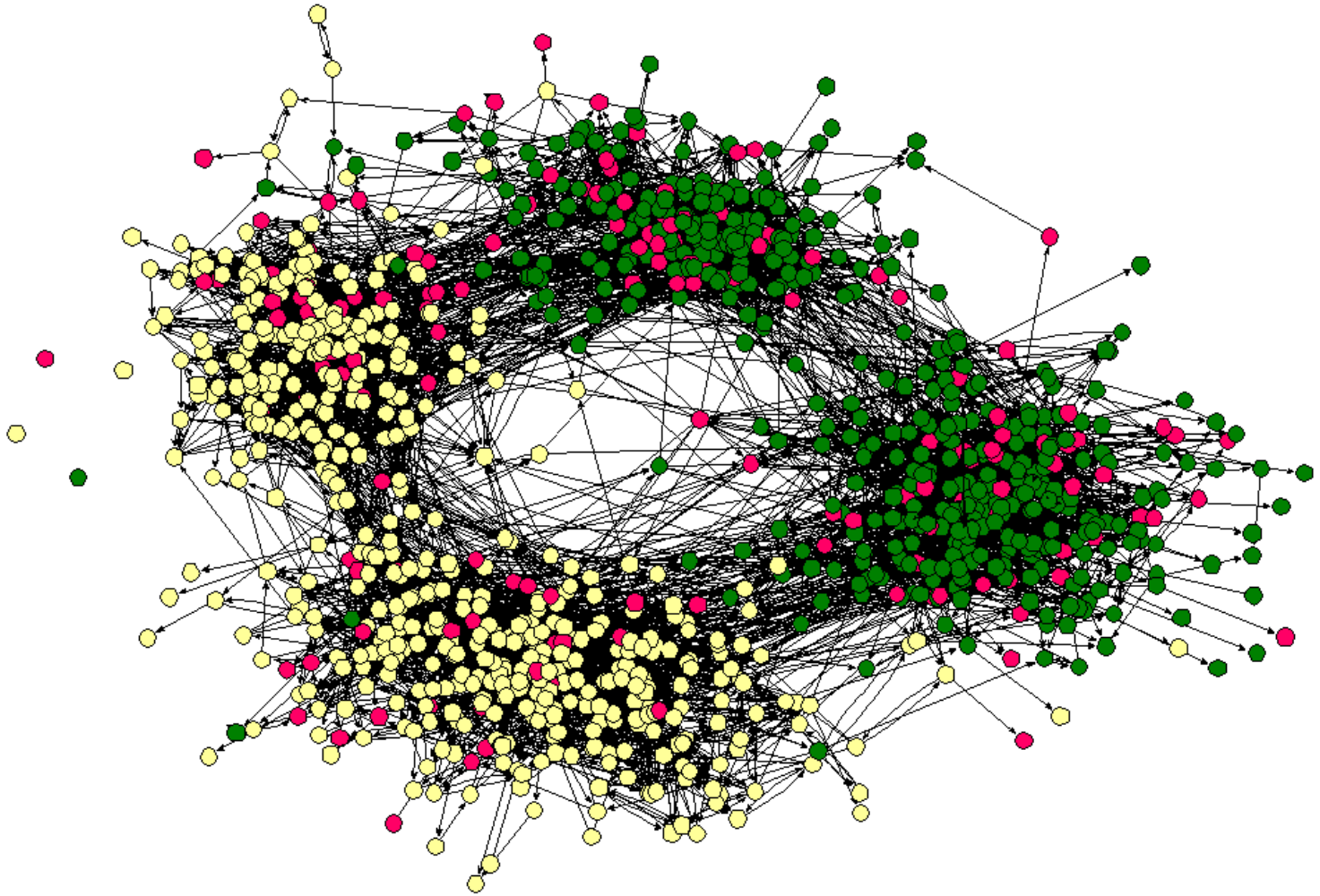
- People of similar character, background, or taste tend to congregate or interact with one another (like likes like)
- Expression appears in the 16th century, a literal translation of Plato's Republic
- “Similarity begets friendship” (Plato)
- People “love those who are like themselves” (Aristotle)



Homophily

- Links in a social network tend to connect people who are similar to one another
- Basic notions governing the structure of social networks
- Its role in modern sociology by influential work in the 1950s (Lazarsfeld and Merton)

Friendship in student



- Social network from a town's middle school and high school (students of different races drawn as differently colored circles)
- 2 divisions: one based on race and the other based on friendships in the middle and high schools

Gang Violence Network

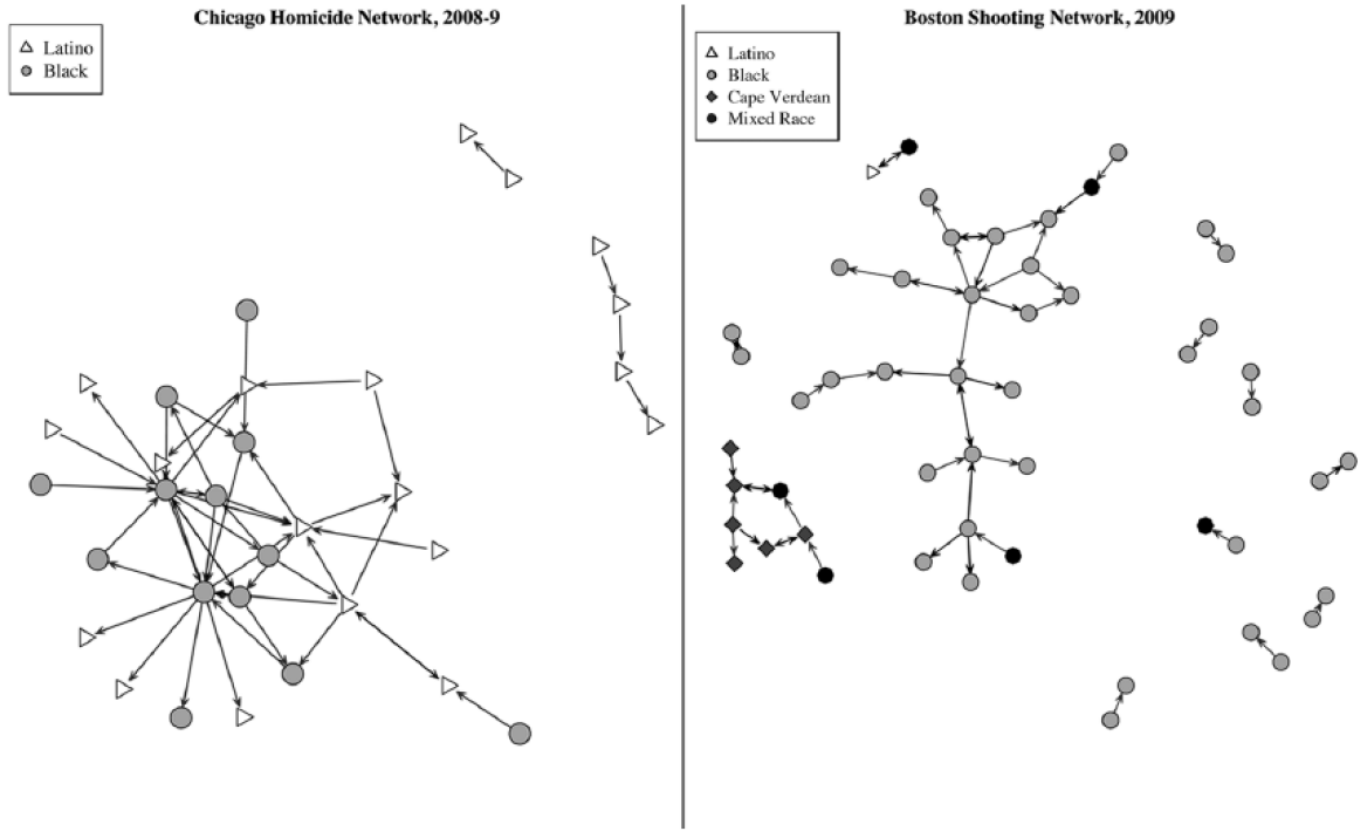
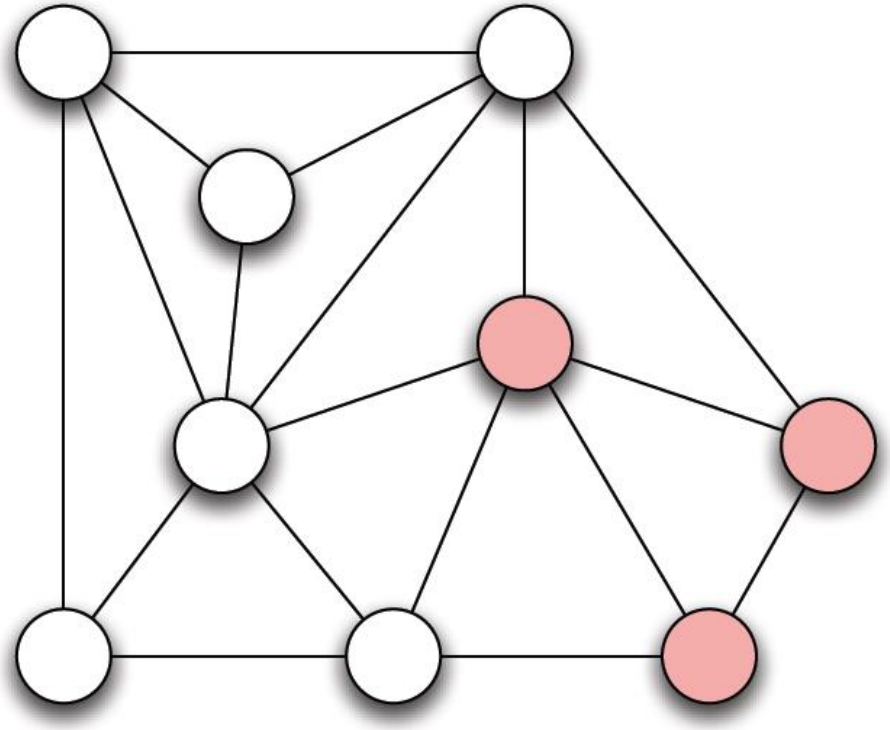


Figure 2. Gang Violence Networks in Chicago (2008 to 2009) and Boston (2009)

Source: Papachristos et al., 2013

How to measure Homophily

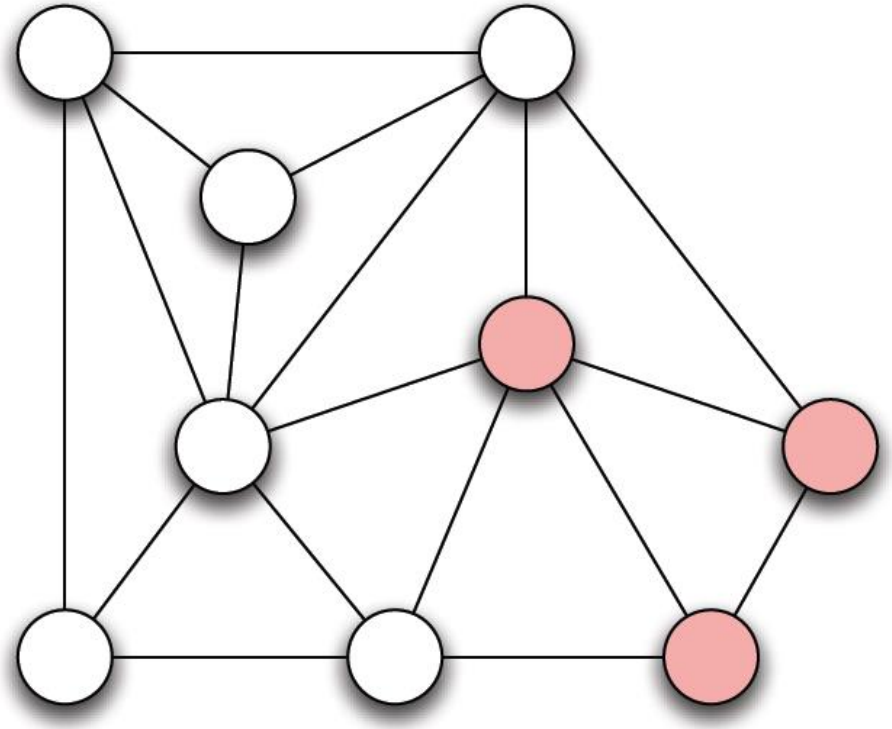
- Given a characteristic (e.g., race or education), how to test if a network exhibits homophily according to it?
- Example: friendship network
  - Exhibits homophily by gender?
  - Boys tend to be friends with boys, and girls tend to be friends with girls
  - Cross-gender edges exist



friendship network of a (hypothetical) classroom: shaded nodes are girls and the six unshaded nodes are boys

How to measure Homophily

- p the probability of males
- q = 1-p the probability of females
- For a given edge
  - Homophily
    - Prob (both ends male) = p * p
    - Prob (both ends female) = q * q
  - Cross gender
    - Prob (ends male and female) = 2 * p * q
- Homophily Test: if the probability of cross-gender edges is significantly less than 2pq, then there is evidence for homophily



Mechanisms Underlying Homophily

Homophily can be explained by two mechanisms:

- Selection: select friends with similar characteristics
  - Individual characteristics drive the formation of links
  - Involves immutable characteristics (determined at birth)
- Social influence: modify behavior close to behaviors of friends
  - The reverse of selection
  - Involves mutable characteristics (behaviors, activities, interests, beliefs, and opinions)

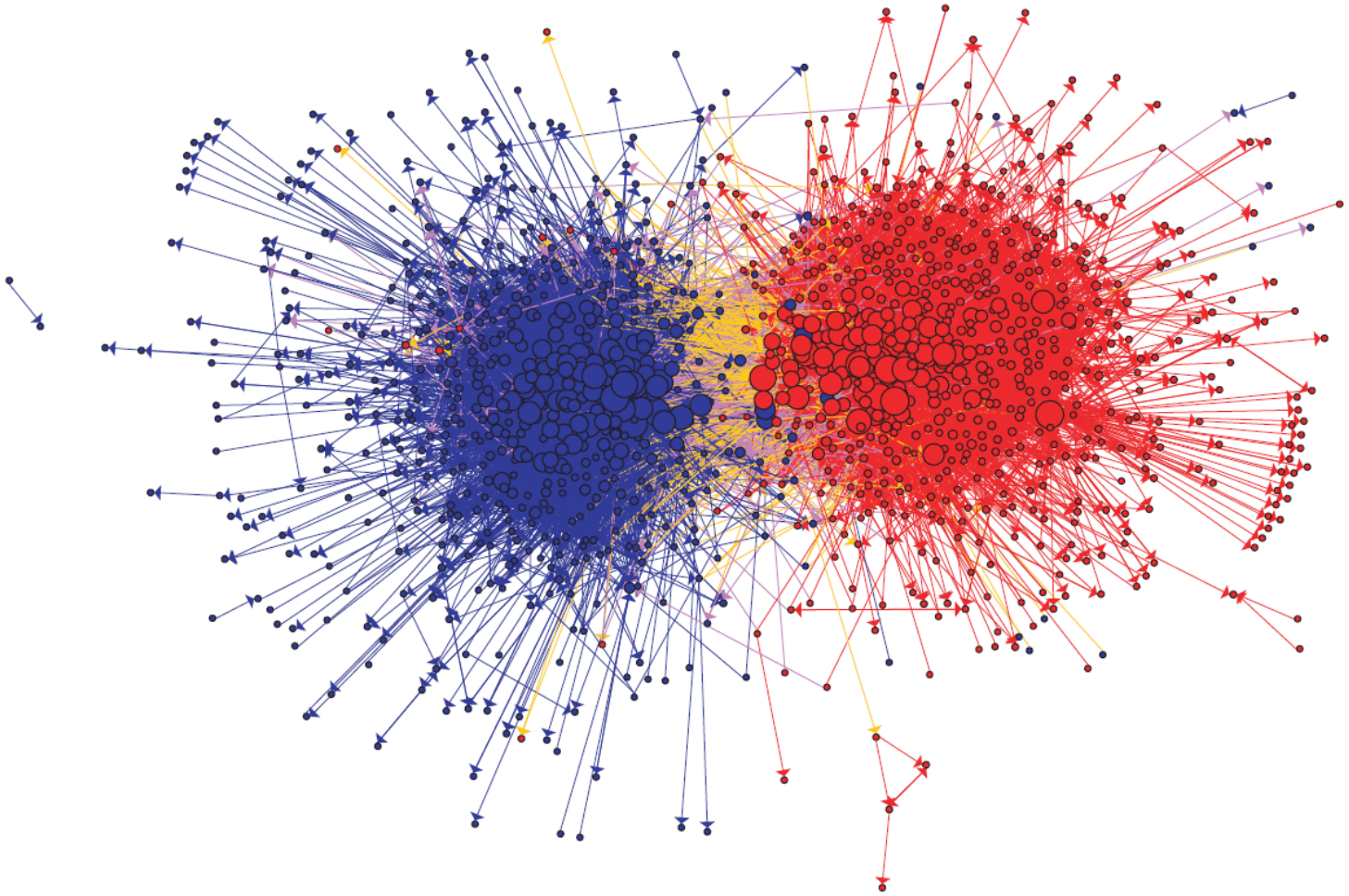
The Interplay between Selection and Social Influence

When homophily is observed, is it a result of selection or social influence or both?

- Have people adapted their behaviors to become more like their friends, or have they selected friends who were already like them?

- Most of the time, both mechanisms apply and interact with each other

- Studies show that teenage friends are similar to each other in their behaviors, and both selection and social influence apply:
  - Teenagers seek social circles of people like them (Social Selection)
  - Peer pressure causes conform to behavioral patterns within these circles (Social Influence)



red-conservative blogs, blue -liberal, orange links from liberal to conservative, purple from conservative to liberal (echo chamber)

image from Adamic & Glance, 2005

Homophily vs. Triadic Closure for Link Formation

With triadic closure:

- A new link is added for reasons that are intrinsic to the network (no need to look beyond the network)
- Example: a friendship that forms because two people are introduced through a common friend

With homophily:

- A new link is added for reasons that are beyond the network (in surrounding contexts)
- Example: a friendship that forms because two people attend the same school or work for the same company

- Strong interactions between intrinsic and contextual effects
- Both operating concurrently
- Most links arise from a combination of several mechanisms
- difficult to attribute any individual link to a single mechanism