Complex Network Theory

Lecture 4

Network analysis

Instructor: S. Mehdi Vahidipour

(vahidipour@kashanu.ac.ir)

Spring 2017

Thanks A. Rezvanian

A. Barabasi, L.Adamic and J. Leskovec

Outline

- Overview of class topics
  - Homophily in social networks
  - Triadic closure
  - Clustering coefficient
  - Bridge
  - The strength of weak ties
  - Signed Networks
  - Structural balance
  - Cohesive subgroups

- Next class
  - Network Models

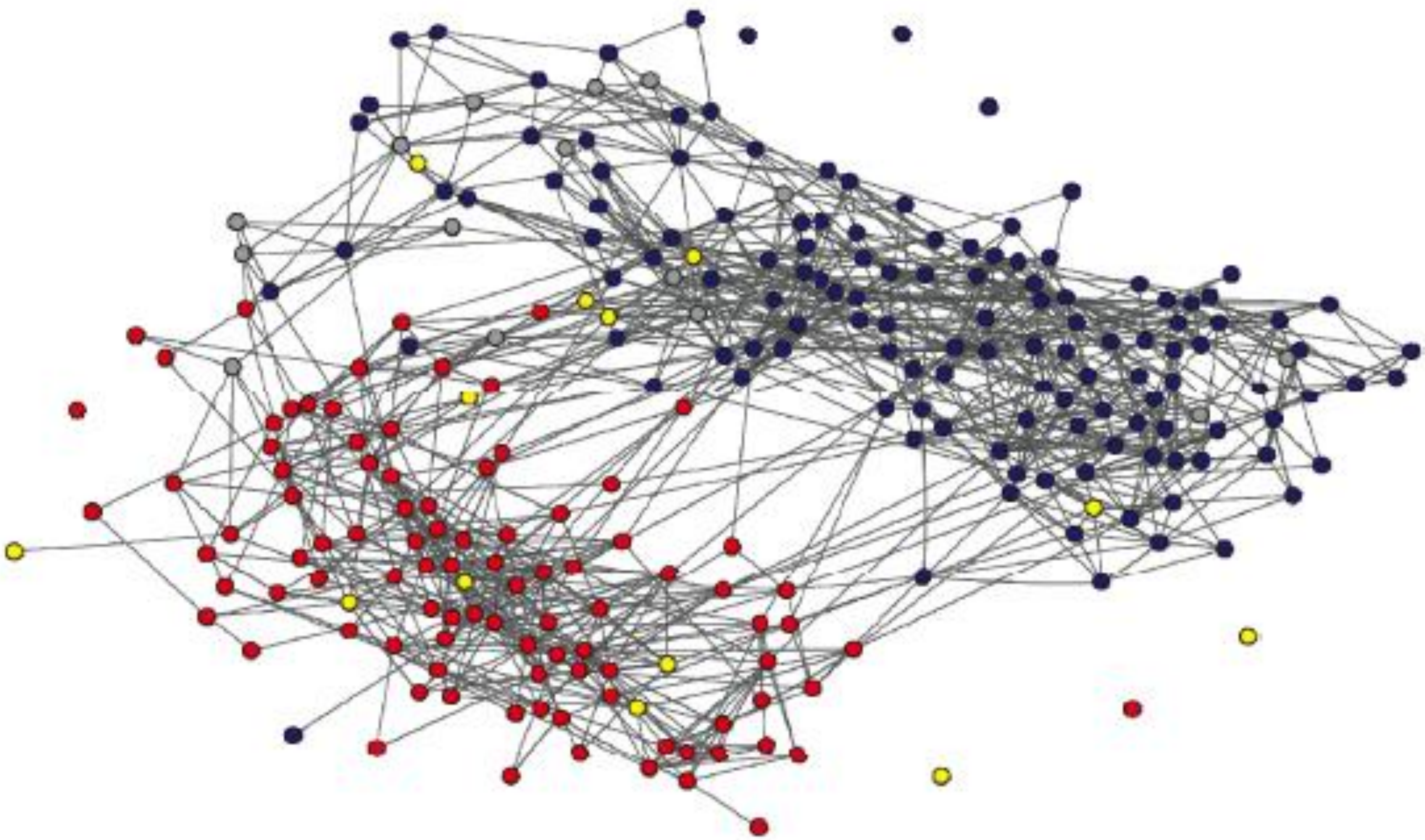
Homophily in social networks

- **Homophily**: tendency of people to connect to other people similar to themselves
- Certainly not a new observation: Aristoteles: “people love those who are like themselves”, Plato: “similarity begins friendship”
- Early studies: school friendships (1929). Homophily in play is observed in race, gender, age, intelligence, attitudes.
- Mid-century: strong interest in homophily driven by school segregation and peer effects on behavior.
- From ‘70s: application of statistical inference allows to study large networks.

Homophily in social networks

- Hypothesizing intrinsic mechanisms:
  - Individuals B and C have a common friend A
  - So, there are increased opportunities and sources of trust on which to base their interactions,
  - As a results, A will also have incentives to facilitate their friendship.
- Since we know that A-B and A-C friendships already exist, the principle of homophily suggests that B and C are each likely to be similar to A in a number of dimensions
- As a result, based purely on this similarity, there is an elevated chance that a B-C friendship will form; and this is true even if neither of them is aware that the other one knows A.

Example

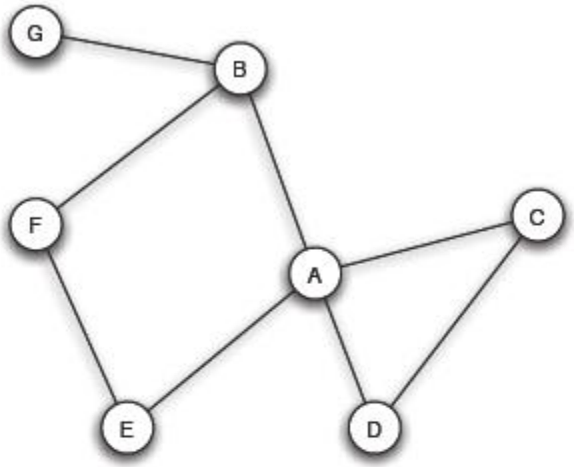


- U.S. Midwest Urban school. Red = Black, Blue = White, Yellow = Hispanic, Grey = Asian. A link means a nominated friendship.
- Source: Add Health Dataset and Currarini-Jackson-Pin (2009).

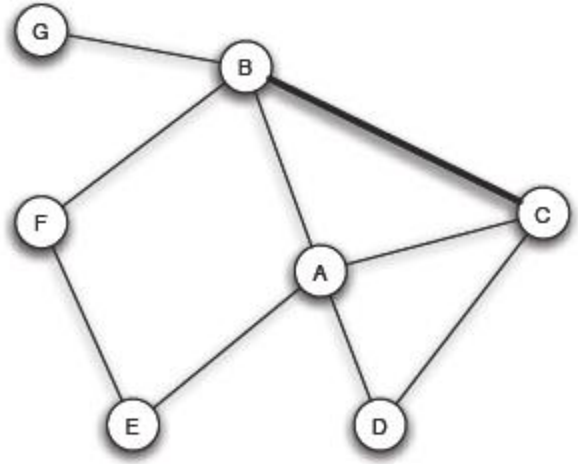
Triadic closure

- Triadic closure principle:
  - If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.
- If three nodes are all-to-all connected, they form a triangle.
- If we observe snapshots of a social network at two distinct points in time, then in the later snapshot, we generally find a significant number of new edges that have formed through this triangle-closing operation, between two people who had a common neighbor in the earlier snapshot.

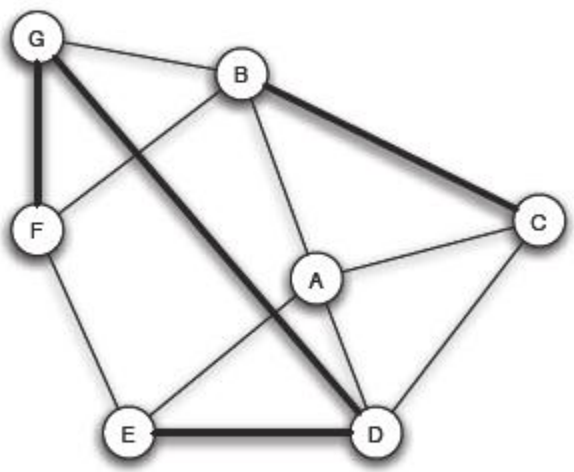
Triadic closure



(a)



(b)



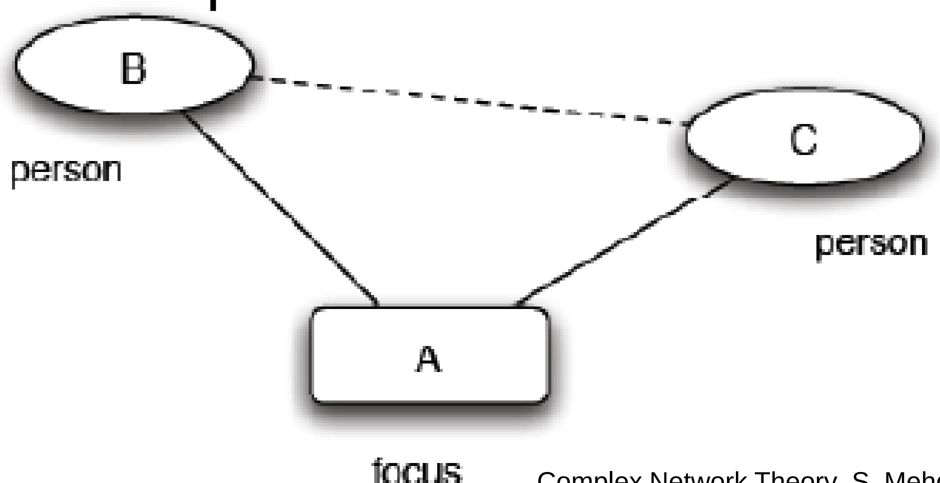
(c)

Reasons for Triadic Closure

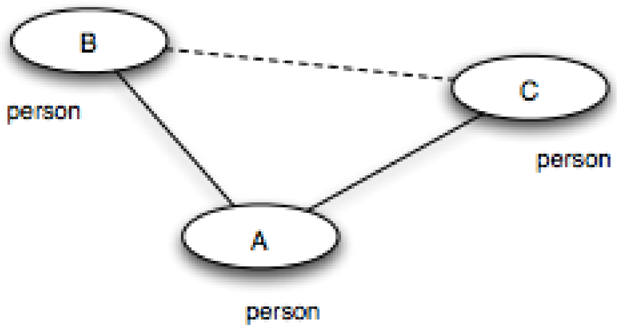
- **Opportunity**: If **A** spends time with both **B** and **C**, then there is an increased chance that they will end up knowing each other and potentially becoming friends.
- **Trust**: The fact that each of **B** and **C** is friends with **A** (provided they are mutually aware of this) gives them a basis for trusting each other that an arbitrary pair of unconnected people might lack.
- **Incentive**: If **A** is friends with **B** and **C**, then it becomes a source of latent stress in these relationships if **B** and **C** are not friends with each other.

Focal closure

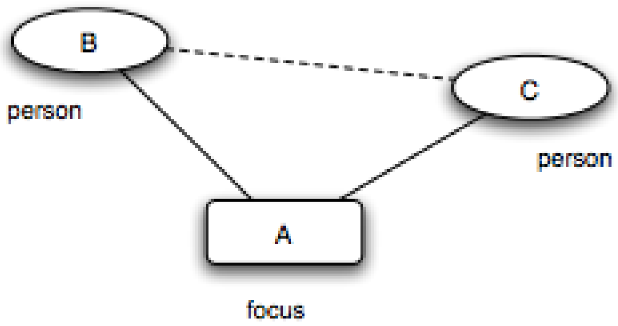
- B and C represent people, but A represents a focus
  - **Foci, or "focal points" of social interaction** – social, psychological, legal, or physical entities around which joint activities are organized (workplaces, hangouts, etc.)
- It is the tendency of two people to form a link when they have a focus in common.
- This is an aspect of the more general principle of selection, forming links to others who share characteristics with you.
- This process has been called focal-closure



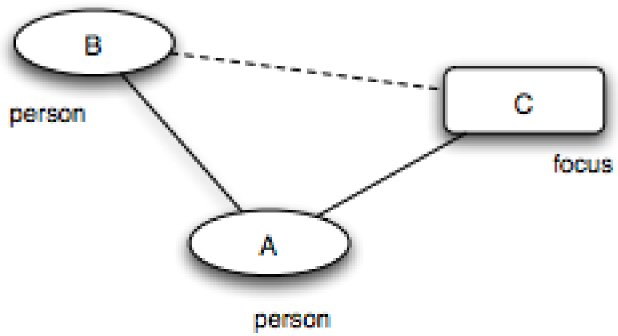
closures



(a) *Triadic closure*



(b) *Focal closure*



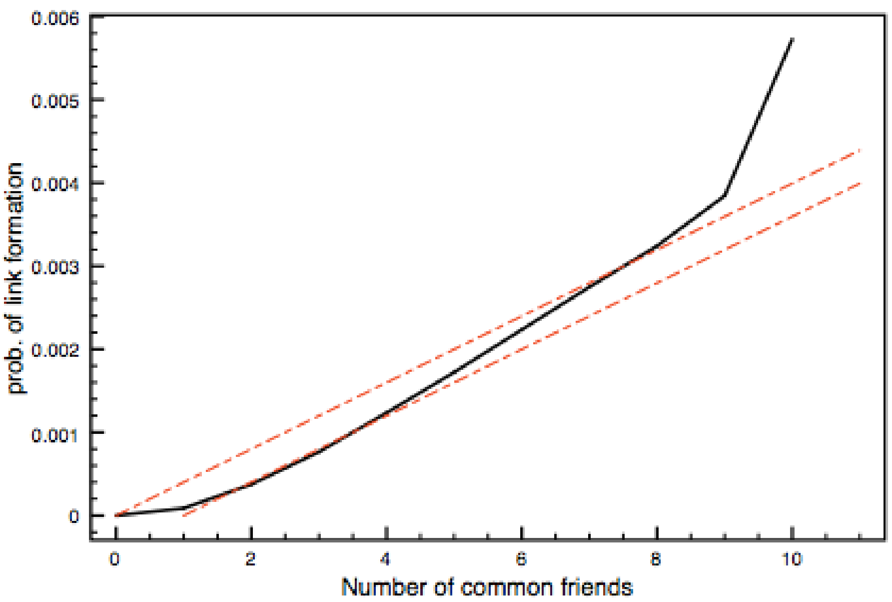
(c) *Membership closure*

Tracking Link Formation in Online Data

- How much more likely is a link to form between 2 people in a social network if they have a friend in common? Multiple (k) friends in common?

Answer empirically:

- Snapshots of network at different times
- For each k, find all pairs of nodes with exactly k friends in common in 1st snapshot, but not directly connected by edge.
- T(k) = fraction of these pairs that formed an edge by 2nd snapshot
- Plot T(k) as function of k to show effect of common friends



Quantifying the effects of triadic closure in an email dataset. The curve determined from the data is shown in the solid black line; the dotted curves show a comparison to probabilities computed according to two simple baseline models in which common friends provide independent probabilities of link formation

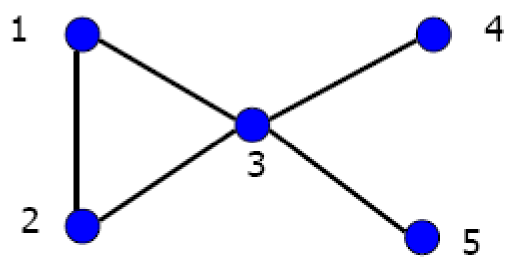
Clustering coefficient

- It is to measure the local connectivity in the network
- Shows somehow the local information exchange
- Originally comes from social sciences
- Shows to how much extent the friends (neighbors) of two connected nodes are connected themselves
- Measures the density of triangles (local clusters) in the networks
- Two different ways to measure it

Global Clustering coefficient

$$\begin{cases} C = \dfrac{3N_\Delta}{N_3} \\ N_\Delta = \sum\limits_{k>j>i} a_{ij}a_{ik}a_{jk} \\ N_3 = \sum\limits_{k>j>i} \left(a_{ij}a_{ik} + a_{ji}a_{jk} + a_{ki}a_{kj}\right) \end{cases}$$

- C is clustering coefficient and A= (a_{ij}) is the adjacency matrix
- N_Δ is the number of triangles (local clusters) in the network
- N_3 is the number of connected triples is the network

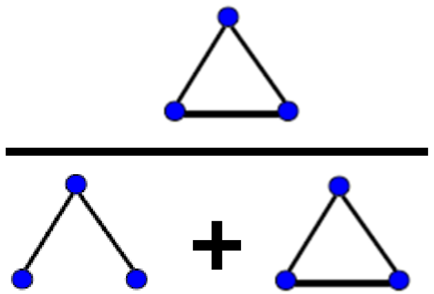


$$\begin{cases} N_\Delta = 1 \\ N_3 = 8 \\ C = \dfrac{3}{8} \end{cases}$$

Local Clustering coefficient

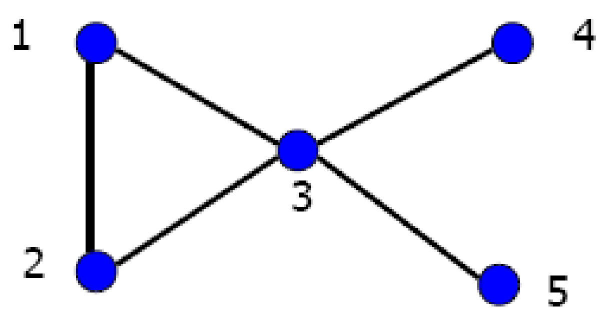
$$\begin{cases} C_i = \dfrac{N_\Delta(i)}{N_3(i)}; C = \dfrac{1}{N}\sum_i C_i \\ N_\Delta = \sum_{k>j} a_{ij}a_{ik}a_{jk} \\ N_3 = \sum_{k>j} a_{ij}a_{ik} \end{cases}$$

$$C_i = \frac{\text{triangles centered at node i}}{\text{triples centered at node i}}$$



$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i-1)}.$$ Undirected networks

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i-1)}.$$ Directed networks



$$\begin{cases} C_1 = 1 \\ C_2 = 1 \\ C_3 = \dfrac{1}{6} \Rightarrow C = \dfrac{1}{5}\left(1+1+\dfrac{1}{6}\right) = \dfrac{13}{30} \\ C_4 = 0 \\ C_5 = 0 \end{cases}$$

Clustering coefficient of weighted networks

$$\begin{cases} s_i = \sum_i w_{ij} \\ k_i = \sum_i a_{ij} \\ C_i^w = \dfrac{1}{s_i\left(k_i - 1\right)} \sum_{j,k} \dfrac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk} \\ C^w = \dfrac{1}{N} \sum_i C_i^w \end{cases}$$

- A= (a_{ij}) is the adjacency matrix and W=(w_{ij}) is the weight matrix
- k_i is the degree and s_i is the strength of node i

Clustering coefficient of real networks

Table 1: Clustering coefficients, C, for a number of different networks; n is the number of node, z is the mean degree. Taken from [146].

Network	n	z	C measured	C for random graph
Internet [153]	6,374	3.8	0.24	0.00060
World Wide Web (sites) [2]	153,127	35.2	0.11	0.00023
power grid [192]	4,941	2.7	0.080	0.00054
biology collaborations [140]	1,520,251	15.5	0.081	0.000010
mathematics collaborations [141]	253,339	3.9	0.15	0.000015
film actor collaborations [149]	449,913	113.4	0.20	0.00025
company directors [149]	7,673	14.4	0.59	0.0019
word co-occurrence [90]	460,902	70.1	0.44	0.00015
neural network [192]	282	14.0	0.28	0.049
metabolic network [69]	315	28.3	0.59	0.090
food web [138]	134	8.7	0.22	0.065

Signed Networks and Structural balance

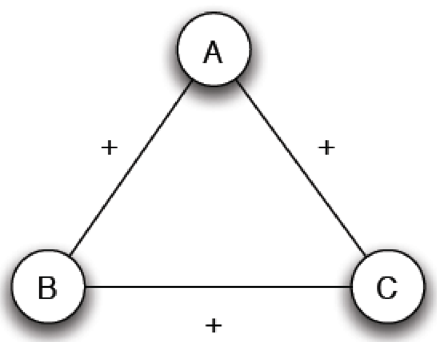
- A rich part of social network theory involves annotating edges with positive and negative signs representing friendship and antagonism.
- An important problem in social networks is to understand the tension between these two forces.
- Signed graphs: we can label graph edges with positive (+) and negative (-) signs
- + signs show positive aspects of relationships
  - Friend, like, trust, follow, ...
- - sigs show negative aspects of relationships
  - Antagonistic, dislike, distrust, …
- The way in which local effects can have global consequences

Assumptions

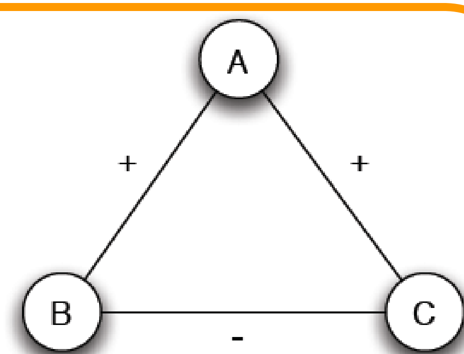
- Given:
  - A complete graph
  - Edges labeled with + Or – expressing
- Assumption:
  - all nodes know each other that each pair
  - Nodes are either friends or enemies
- The model makes sense for a group of people small enough to have this level of mutual awareness
  - E.g. a classroom, a small company, a sports team, a fraternity or sorority, international relations
- Patterns of relations
  - If we look at any two people in the group in isolation, the edge between them can be labeled + or -; that is, they are either friends or enemies.
  - But when we look at sets of three people at a time, certain configurations of +'s and -'s are socially and psychologically more plausible than others.

Structural Balance

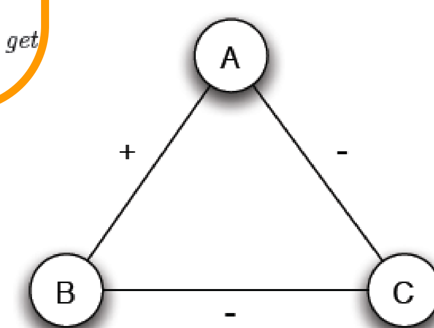
- Theories in social Psychology by Heider, Cartwright and Harary
- Balanced Triangles: we can classify triads in this graph into two categories, balanced & unbalanced



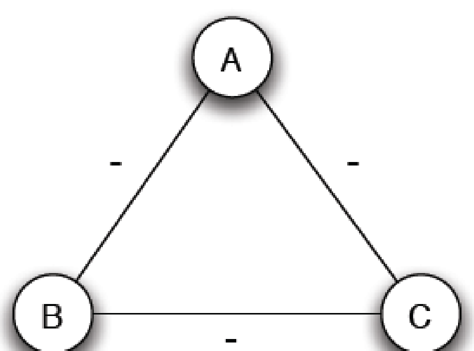
(a) *A, B, and C are mutual friends: balanced.*



(b) *A is friends with B and C, but they don't get along with each other: not balanced.*



(c) *A and B are friends with C as a mutual enemy: balanced.*



(d) *A, B, and C are mutual enemies: not balanced.*

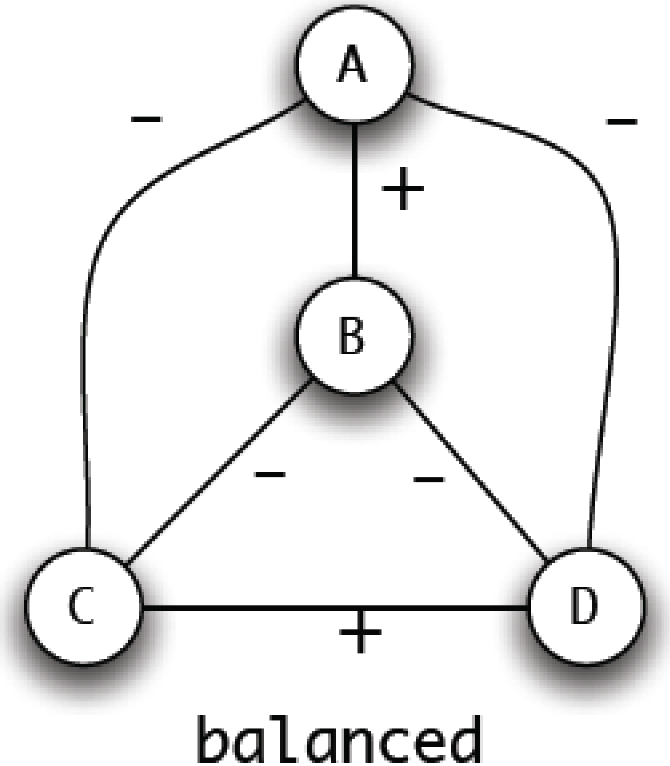
structural balance theorists

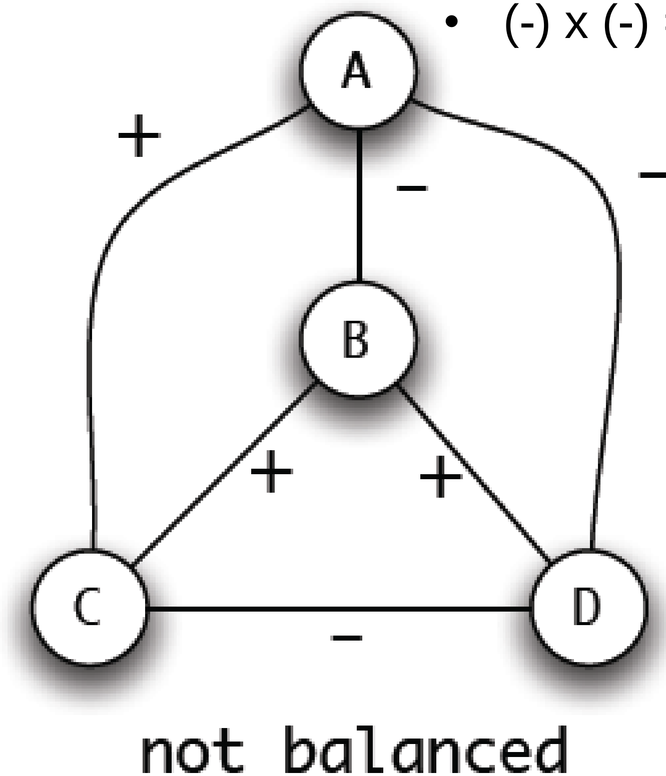
- Argument of structural balance theorists
  - Unbalanced triangles are sources of stress or psychological dissonance
    - People strive to minimize them in their personal relationships
    - They will be less abundant in real social settings than balanced triangles
- Structural Balance Property
  - We say that a signed complete graph is balanced if every one of its triangles is balanced
- For every set of three nodes, if we consider the three edges connecting them,
  - either all three of these edges are labeled +, or else
  - exactly one of them is labeled +

Balanced/Unbalanced Graphs

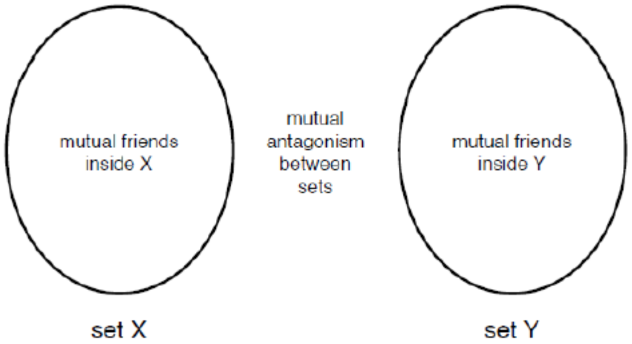
- Signed graphs are balanced iff all of its triads are balanced

  - (+) x (+) = (+)
  - (-) x (+) = (-)
  - (-) x (-) = (+)





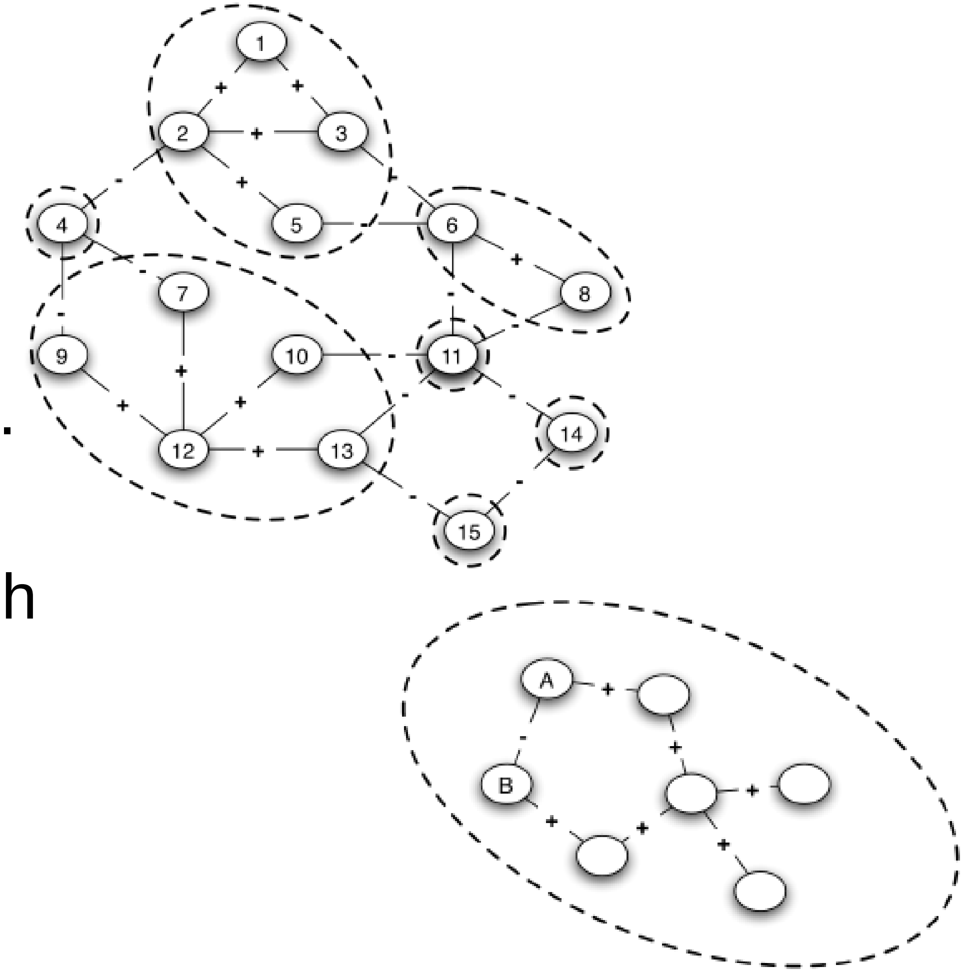
Balanced networks



- Suppose:
  - we have a signed complete graph
  - The nodes can be divided into two groups, X and Y
  - Every pair of people in X like each other
  - Every pair of people in Y like each other,
  - Everyone in X is the enemy of everyone in Y
- Such a network is balanced:
  - a triangle contained entirely in one group has three + labels,
  - and a triangle with two people in one group and one in the other has exactly one + label

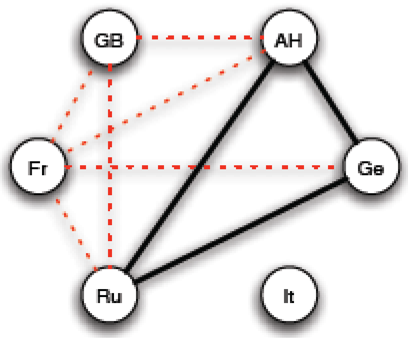
Structural Balance in Arbitrary Graphs

- Theorem: A signed graph is balanced if and only if it contains <u>no cycle</u> with an odd number of negative edges
- Algorithm
  1. Convert the graph to a reduced one in which there are only negative edges.
  2. Solve the problem on the reduced graph

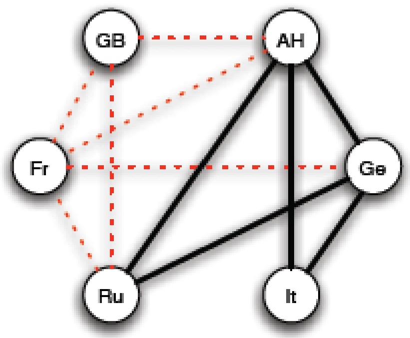


International Relations

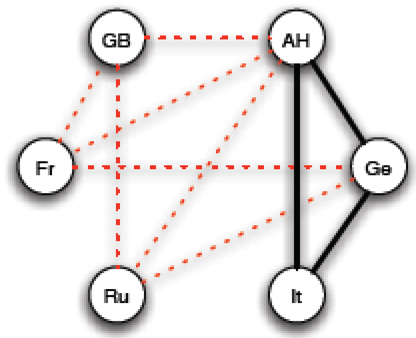
- The evolution of alliances in Europe, 1872-1907 (the nations GB, Fr, Ru, It, Ge, and AH are Great Britain, France, Russia, Italy, Germany, and Austria-Hungary respectively). Solid dark edges indicate friendship while dotted red edges indicate enmity. Note how the network slides into a balanced labeling - and into World War I



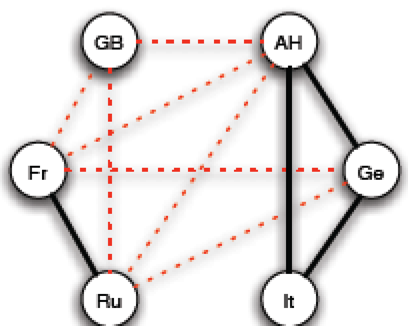
(a) *Three Emperors' League 1872–81*



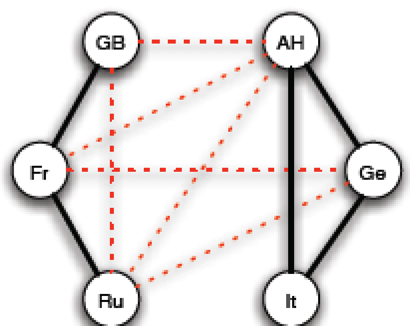
(b) *Triple Alliance 1882*



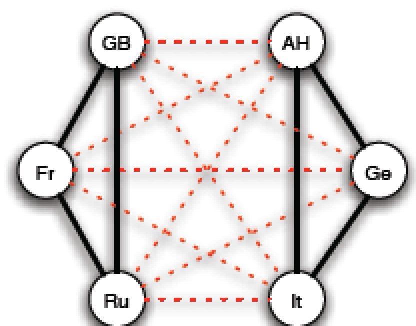
(c) *German-Russian Lapse 1890*



(d) *French-Russian Alliance 1891–94*



(e) *Entente Cordiale 1904*



(f) *British Russian Alliance 1907*

World War I

Cohesive subgroups

- Informal definition
- Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties
- There are many possible social network subgroup
  - One mode networks
  - Two mode networks e.g. affiliation networks based on joint membership
- Valuable for studying the emergence of consensus
- Homophily: more homogeneity among cohesive subgroups

Operating factors

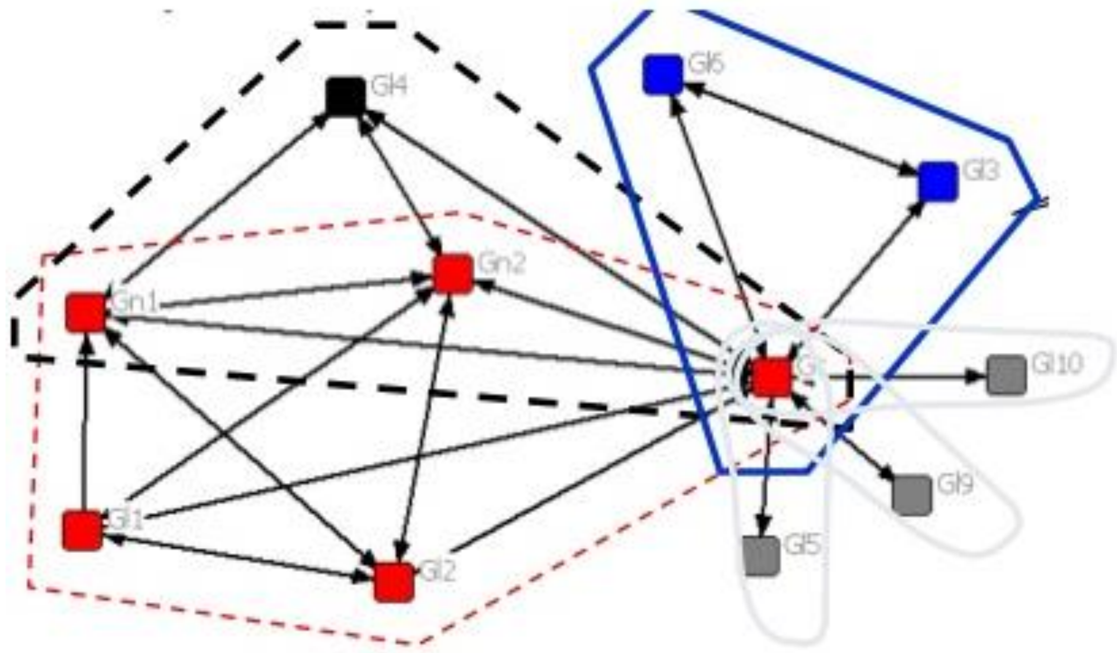
- Two operating factors:
  - How many links an individual has to the group
  - How many links an individual has to the outsiders
- Social forces: operating through
  - Direct contact among subgroup members,
  - Indirect contact transmitted via intermediaries
  - The relative cohesion within as compared to outside the subgroup.

Some Criteria for Cohesion

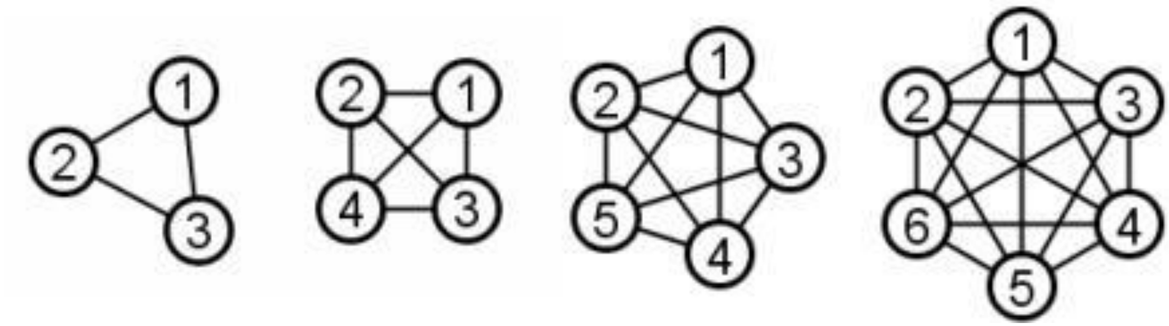
- Distance
  - The mutuality and completeness of edges
  - The closeness or reachability of subgroups
- Density
  - The density of edges among members
  - The relative density of edges among subgroup members compared to non-members

Subgroups based on Complete Mutuality

- **Clique**: A subset of nodes, all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the members of the clique
- Number of nodes: at least three
- Cliques in a graph may overlap



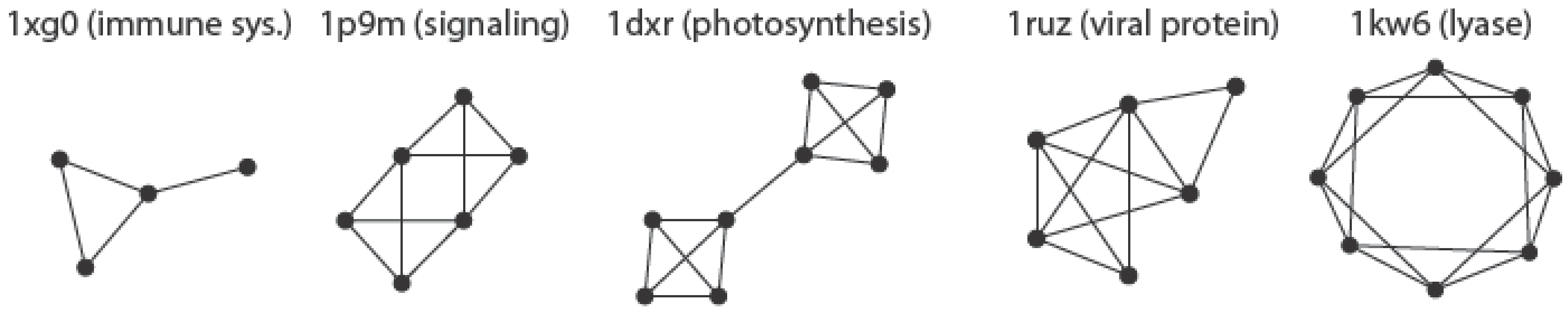
Usefulness of Cliques

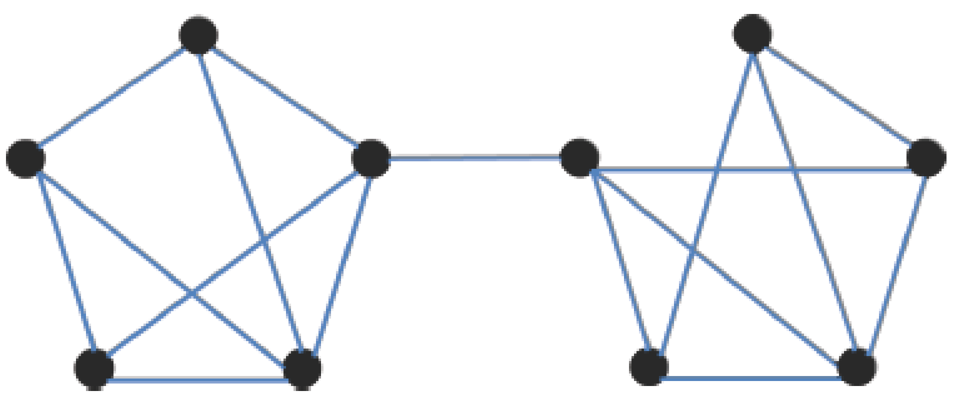


- A clique is a very strict definition of cohesive subgroup.
  - The absence of a single edge will prevent a subgraph from being a clique.
- The size of the Cliques will be limited by the degree of the nodes.
- There is no internal differentiation among nodes within a clique
  - No place for Internal structure or hierarchy

Some examples

- Perfect may mean impractical.



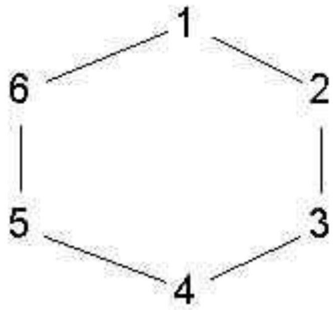


Subgroups based on Reachability and Diameter

- Relaxing the distance
- There should be relatively short paths of influence or communication between all members of the subgroup.
- Subgroup members might not be adjacent, but if they are not adjacent, then the paths connecting them should be relatively short.

n- clique (Luce 1950)

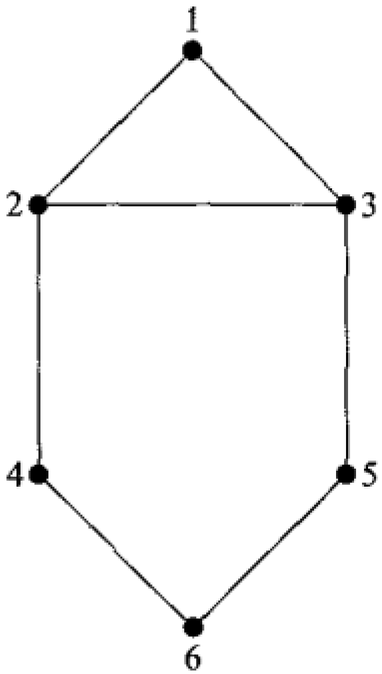
- A maximal set of nodes such that all pairs of nodes in the set are distance *n* or less
  - One problem with n-clique is that even a 2-clique is not very cohesive.
    - In the graph {1,3,5} is a 2-clique, but none are connected to each other.
  - From substantive point of view is weird that shortest path between two members required outside intermediary



n-clan

- An n-clique such that the induced sub-graph has diameter n or less

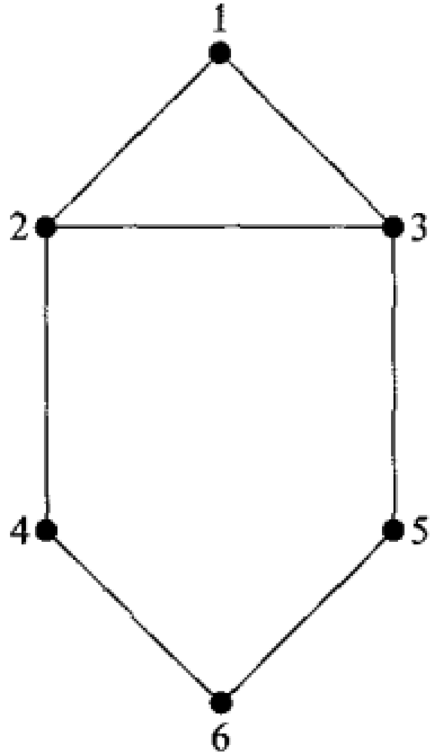
- 2-cliques:
  - {1,2,3,4,5},
  - {2,3,4,5,6};
- **2-clan:**
  - {2,3,4,5,6}



n-club (Mokken 1979)

- **n-club** is a maximal subgraph in which the distance between all nodes within the sub-graph is less than or equal to n (diameter at most *n*)

- **2-clubs:**
  - {1,2,3,4},
  - {1,2,3,5},
  - {2,3,4,5,6}



N-Clique & N-Club & N-Clan

- Every n-clan is an n-club
- Every n-clan is an n-clique
- But every n-club is not an n-clan or n-clique, although it is contained in them
  - (fail n-clique maximal condition)

Usefulness of subgroups based on distance

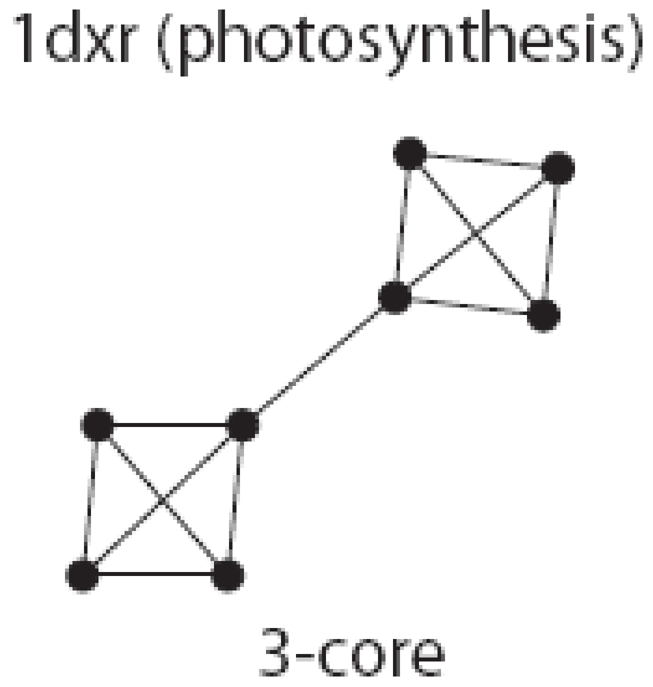
- Cohesive subgroups based on indirect connections of relatively short paths provide a reasonable approach for studying network processes such as information diffusion

Subgroups based on nodal degree

- Relaxing the density
- All subgroup members should be adjacent to some minimum number of other subgroup members.
- Useful when network processes require direct contact among nodes, and perhaps repeated, direct, contact to several nodes.
- Multiple redundant channels of communication increase the accuracy of information
- Robustness: The degree to which the structure is vulnerable to the removal of any given individual.

K-core

- Each node has degree at least **k** (Seidman 1983)
- N-1 core is complete graph

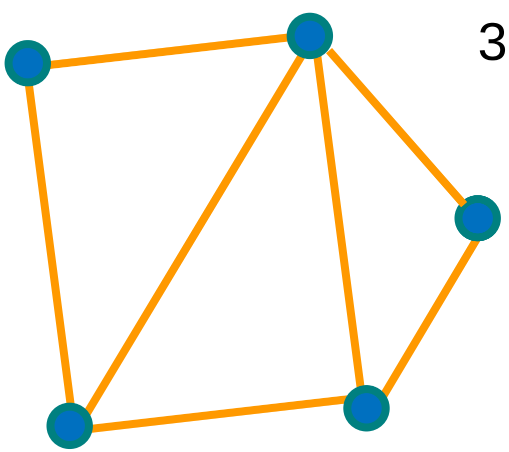


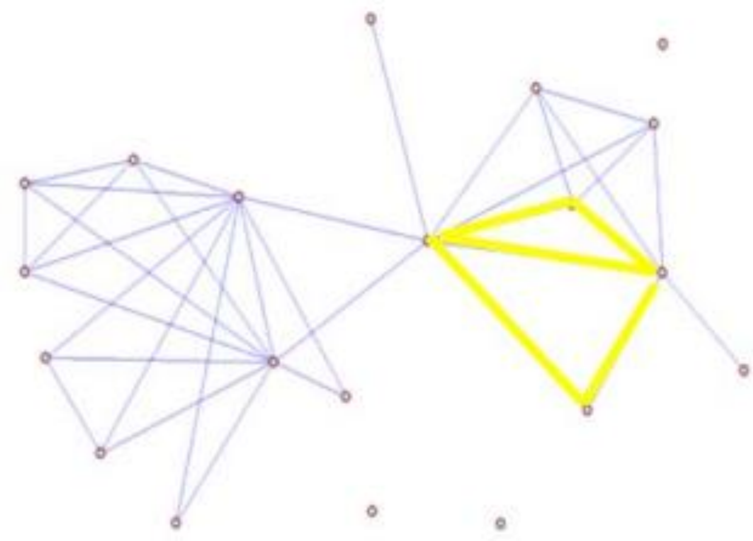
K-plex

- A **k**-plex of size **n** is a maximal sub-set of **n** nodes within a network such that each node is connected to at least **n-k** of the others. (Seidman & Foster 1978)
- A 1-plex is the same as a clique.
- Obviously, a k-core of *n* vertices is also an (n – k)-plex.

2-core

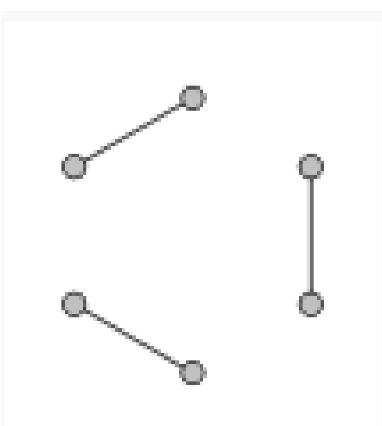
3-Plex



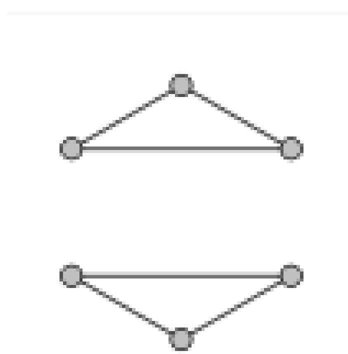


Regular graph

- A **regular graph** is a graph where each node has the same number of neighbors (i.e. every vertex has the same degree).
- A regular graph with vertices of degree *k* is called a **k-regular graph** or regular graph of degree *k*.



1-regular

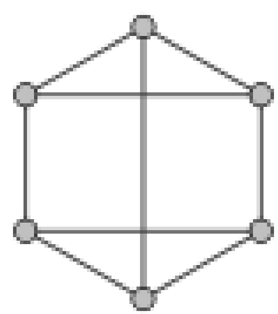


2-regular

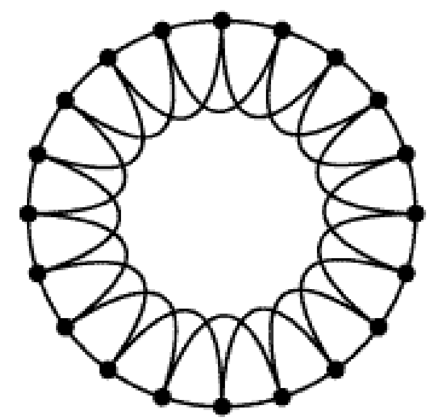
2-Club

3-Plex

3-Core



3-regular



4-regular

Readings

- Newman, Mark. ***Networks: an introduction***. Oxford University Press, 2010. (Ch. 7, 10)
- Easley, David, and Jon Kleinberg. **Networks, crowds, and markets: Reasoning about a highly connected world**. Cambridge University Press, 2010. (Ch. 3, 5)
- Van Steen, Maarten. "**Graph theory and complex networks**." ,2010. (Ch. 6)
- Pattillo, Jeffrey, Nataly Youssef, and Sergiy Butenko. "**On clique relaxation models in network analysis**." *European Journal of Operational Research* 226.1 (2013): 9-18.