SC435
Introduction to Complex Networks

Groups of nodes

sub-graphs $\Rightarrow$ local structure of networks.

Group of nodes

- **clique:** is a maximal subset of vertices in an undirected network such that every member of the set is connected by an edge to any other vertex.
  - Cliques can overlap $\Rightarrow$ they can share vertices (one or more)
  - A clique in an otherwise sparse network indicates a highly cohesive subgroup.
- **k-plex:** of size n is a maximal subset of n vertices within a network such that each vertex is connected by an edge to atleast $n - k$ of the others.
  - k-plex can also be overlapping.
  - It is a useful concept in social network analysis but there is no solid rule what value k should take.

Group of nodes

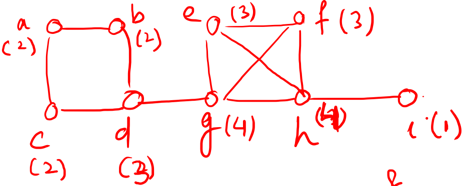
Group of nodes

- **k-core:** is a *maximal* subset of vertices such that each is connected to atleast k-other in the subset.
- **k-core decomposition:** useful in obtaining the core-periphery structure of the network.
- degree of each node can be used to separate a network into distinct portions called shells.
  - periphery: low-degree outer shells
  - core: high degree (connectivity) inner shell

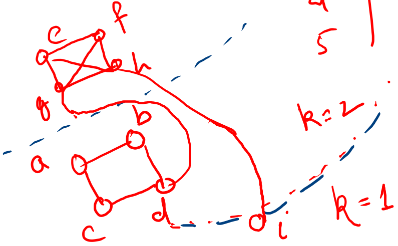
Algorithm

- start with $k = 0$ (singletons)
- Recursively remove all nodes of degree k, until there are no more left.
- The removed nodes make up the k-shell and the remaining nodes make up the $k + 1$-core
- If there are no more nodes left in the core, terminate else, increment k for the next iteration.

Can two k-cores overlap?

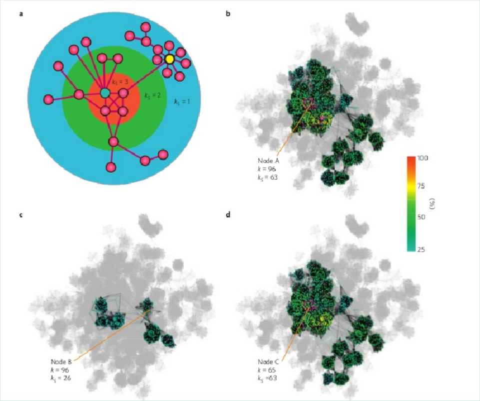


iteration#	k=1	k=2	k=3
1	{i}		
2		{a}	
3		{a, b, c}	
4		{a, b, c, d}	
5			{e, f, g, h}



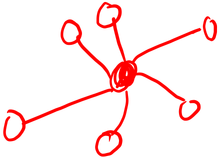
Group of nodes

k-core

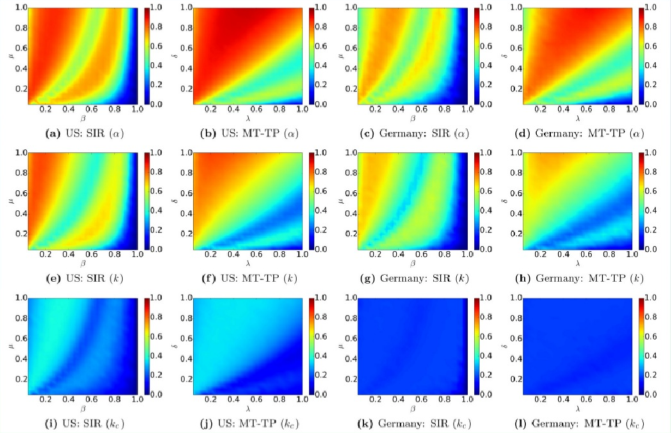


Hubs may not be important in spread.[1]

[1] Kitsak *et. al. Identification of influential spreaders in complex networks, Nature Physics* **6**, *2010.*



Group of nodes



Which is the most influential spreader Influential Spreader [2]

[2] Arruda *et. al.*, *Role of centrality for the identification of influential spreaders in complex networks*, Phys. Rev. E **90**, 032812 (2014)

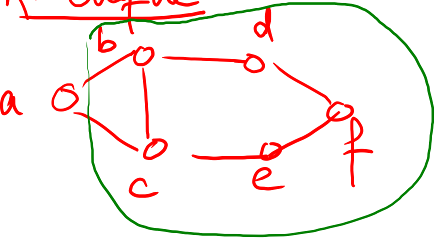
Group of nodes

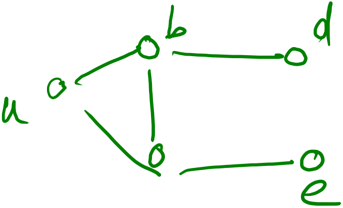
Distance based.

Group of nodes

- **k-clique:** is a *maximal* subset of vertices such that each is no more than a distance k away from any of the others via the edges of the network.
  - For $k = 1$: ordinary cliques.
  - not a very well-behaved one, since a k-clique by this definition need not be connected via paths that run within the subset
- **k-clan:**
- **k-club:**

k-clique





k-clan

k-club.

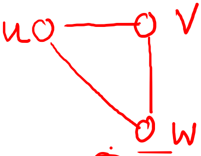
Group of nodes

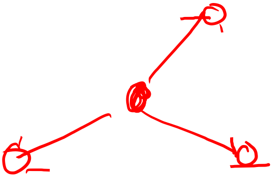
Components and k-component

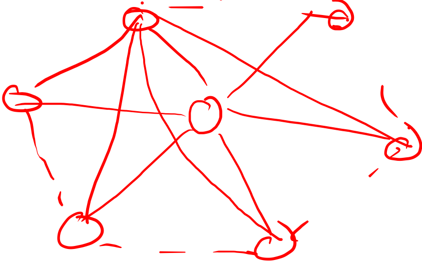
- **Component:** A component in an undirected network is a maximal subset of vertices such that each is reachable by some path from each of the others.
- **k-component (k-connected component):** is a *maximal* subset of vertices such that each is reachable from each of the others by at least k *vertex-independent* paths.
  - **vertex-independent**: two paths are vertex-independent if they share none of the same vertices except the starting and the ending vertices.
  - number of vertex-independent paths between two vertices is equal to the size of the *vertex cut set* between the same two vertices.
  - k-component is connected with the idea of network robustness.

Transitivity

Clustering



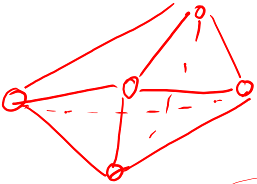




Group of nodes.

Transitivity

Clustering Coefficient



triangle.

Groups of nodes

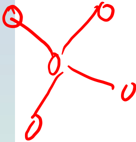
Transitivity

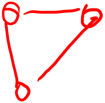
- Transitivity is most important in social networks
- Simplest relation: Connected by an edge
- Perfext transitivity only occurs in networks where each component is perfectly connected subgraph or clique.
- **Clustering coefficient**: Measure of partial transitivity

$$C = \frac{\text{number of closed paths of length 2}}{\text{number of paths of length 2}}$$

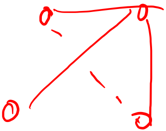
$$\boxed{C = \frac{6 \times \text{number of } \Delta' s}{\text{number of paths of length 2}}}$$

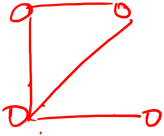
$$\boxed{C = \frac{3 \times \text{number of } \Delta' s}{\text{number of connected triplets}}}$$





$$\frac{|E|}{\binom{n}{2}}$$

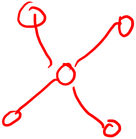


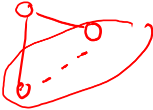


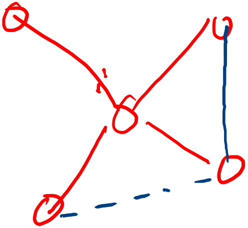
C

local Clustering Coeff.

$$C_i = \frac{\text{no. of edges between the neighbors}}{\binom{k_i}{2}}$$







$C_i = 0$

$$C_i = \frac{1}{\binom{4}{2}} = \frac{1}{6}$$

$$C_i = \frac{2}{\binom{4}{2}} = \frac{1}{3}$$

$$\langle C \rangle = \frac{1}{N} \sum_i C_i$$

Groups of nodes

Transitivity

- **Local Clustering**: Clustering coefficient for a single vertex $\Rightarrow$ Clustering coefficient of a node is the fraciton of the node's neighbor that are connected to each other.

$$\boxed{C(i) = \frac{\tau(i)}{\tau_{max}(i)} = \frac{\tau(i)}{\binom{k_i}{2}}}$$

- Average (global) clustering coefficient:

$$\boxed{C = \frac{\sum_{i:k_i>1} C(i)}{N_{k>1}}}$$

- Clustering coefficient in directed graphs: depends on the kind of triangles that are relevant to a specific case.
- High clustering coefficient $C >> p$, where

$$p = \frac{|E|}{\binom{n}{2}}$$

Groups of nodes

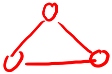
Structural holes

- Local clustering coefficient is low for vertex with high degree.
- Local clustering coefficient is related with the concept of *structural holes*.
- While it is common for the neighbors of a vertex to be connected among themselves, it happens sometimes that these expected connections between neighbors are missing. These missing links are called structural holes.
- Structural holes can be a bad thing with regards to efficient spread of information.
- Can be good for vertex i $\Rightarrow$ high importance.

Clustering Coefficient

↓

Transitive relationship



Network level → number of triangles.

node level

$$C_i = \frac{\text{total number links between neighbors of } i}{\binom{k_i}{2}}$$

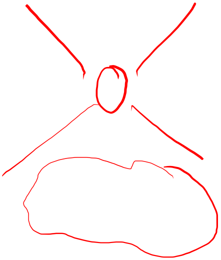
$$C_{WS} = \frac{1}{N}\sum_i C_i.$$

Watts & Strogatz

Social networks — Structural holes.

highly correlated

Betweenness Centrality



Groups of nodes

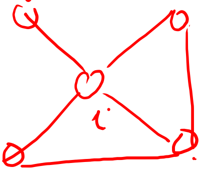
Redundancy

- The redundancy R_i of vertex i is the mean number of connections from a neighbor of i to other neighbors of i.
- Redundancy of node i ranges from 0 to $k_i - 1$.
- $C(i) = \frac{R_i}{k_i - 1}$

Redundancy (R_i) is the mean number of connections between neighbors of i.

$$R_i = \frac{1}{4}\left[0 + 1 + 2 + 1\right]$$

$$= 1$$



total number of connections between neighbors of vertex i

$$\frac{k_i R_i}{2}$$

$$C_i = \frac{k_i R_i/2}{\binom{k_i}{2}} = \frac{R_i}{k_i - 1}$$

Groups of nodes

Reciprocity

- In directed networks there can be loops of size 2.
- Reciprocity measures the frequency of loops of size 2 in a directed network.

$$\boxed{r = \frac{1}{m} \sum_{ij} A_{ij} A_{ji} = \frac{1}{m} \mathrm{Tr} A^2}$$

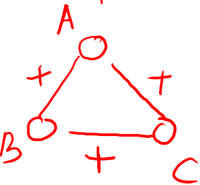
Directed Graphs.

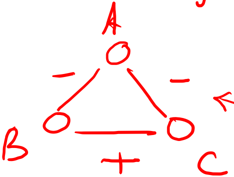
Reciprocity.

$$R_i = \frac{1}{m} \sum_{j} A_{ij} A_{ji} = \frac{1}{m} \mathrm{Tr}\left[A^2\right]$$

Signed Networks.

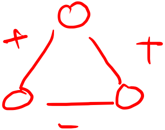
friends and enemy.

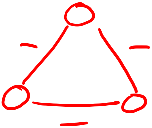




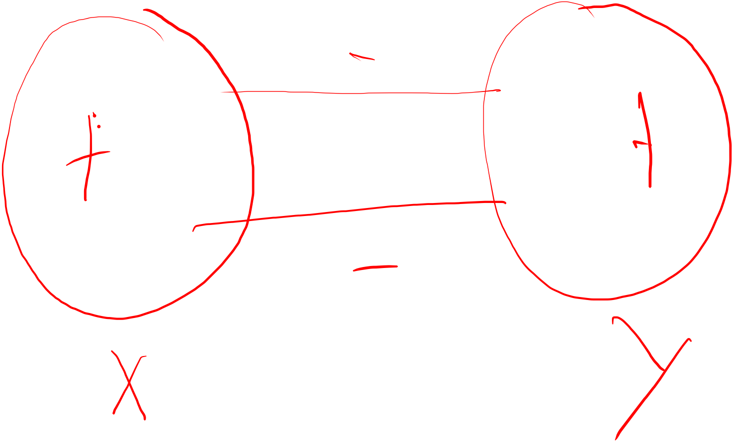
enemy of an enemy is a friend.

Structurally stable.





Structurally unstable


X
Y

Groups of nodes

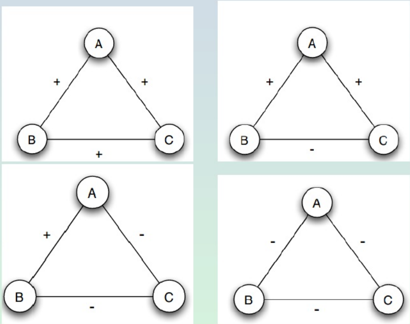
Signed networks

- In social networks, edges can be annotated as positive or negative to signify relationship
  - positive edge $\Rightarrow$ friendship
  - negative edge $\Rightarrow$ antagonism (people interact in a negative way)
  - connection between local and global network properties.

Groups of nodes

Structural Balance

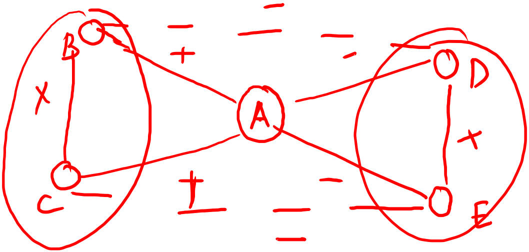
- based on theories in social psychology (Heider) and extended to graphs (Cartwright & Hararay)
- In a group of 3 certain combinations of $+/-$ are socially more plausible.



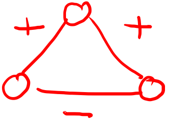
Groups of nodes

Cartwright-Harary Theorem

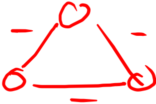
If a labeled complete graph is balanced, then either all pairs of nodes are friends, or else the nodes can be divided into two groups, X and Y, such that every pair of people in X like each other, every pair of peoplein Y like each other, and everyone in X is the enemy of everyone in Y.

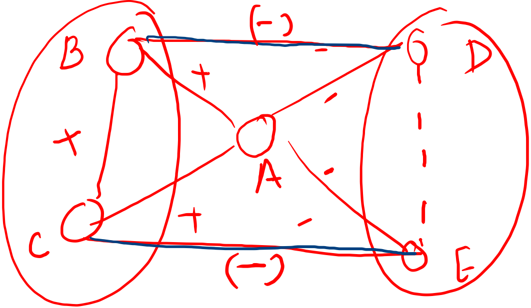


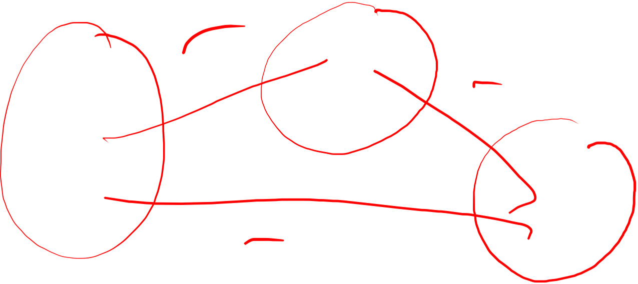
Strong Structural Balance.



Weak Structural Balance




B
C
A
D
E
+
+
+
-
(-)
(-)


-
-

Structural holes.