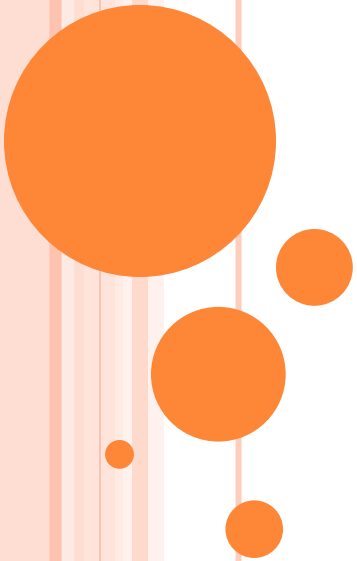


Social network analysis

Module 1

Introduction

- Social Network Analysis (SNA) is a research technique that analysis the social structures that emerges from the combination of relationships among members of a given population
- SNA is the ***study of social relations among a set of actors***.
- The key difference between network analysis and other approaches is the ***focus on relationships between actors*** *rather than the attributes of individual actors*.

- Network analysis takes a global view on social structures based on the belief that ***types and patterns of relationships*** emerge from individual connectivity and that the presence of such types and patterns have substantial effects on the network and its constituents.
- The network structure provides opportunities and imposes constraints on the individual actors by determining the transfer or flow of resources across the network.

- When trying to predict the performance of individuals in a scientific community by some measure, a traditional social science approach would dictate to look at the attributes of the researchers such as the amount of grants they attract, their age, the size of the team they belong to etc.
- A statistical analysis would then proceed by trying to relate these attributes to the outcome variable, i.e. the number of publications

- A network analysis study would focus on the interdependencies within the research community.
- For example, one would look at the patterns of relationships that scientists have and the potential benefits or constraints such relationships may impose on their work.
- For example, one may hypothesize that certain kinds of relationships arranged in a certain pattern may be beneficial to performance compared to the case when that pattern is not present.

- The patterns of relationships may not only be used to explain individual performance but also to hypothesize their impact on the network itself .
- Attributes typically play a secondary role in network studies

- SNA is a different approach to social phenomena and therefore requires a new ***set of concepts and new methods*** for ***data collection and analysis***.
- Network analysis provides a ***vocabulary*** for describing social structures, provides formal ***models*** that capture the common properties of all networks and a ***set of methods*** applicable to the analysis of networks in general.

- The ***concepts and methods*** of network analysis are grounded in a formal description of networks as ***graphs***.
- Methods of analysis primarily originate from ***graph theory*** as these are applied to the graph representation of social network data.
- Network analysis also applies statistical and probabilistic methods and to a lesser extent algebraic techniques.

Why SNA

- Enables us to segment data based on user behaviour.
- Understands natural groups that have formed
  - Topic
  - Personal characteristics
- Understand who are the important people in these groups

Analysing

- Data collection methods
  - Surveys
  - Interviews
  - Observations
- Analysis
  - Computational analysis of matrices

Development of SNA

- The field of Social Network Analysis today is the result of the convergence of several streams of applied research in ***sociology, social psychology*** and ***anthropology***.
- Many social ***psychologists*** of the 1940s found a formal description of social groups when trying to explain processes of group communication.
- In the mid-1950s ***anthropologists*** have found network representations useful in generalizing actual field observations.

- ***Researchers*** from Harvard looked at the ***workgroup behavior*** (e.g. communication, friendships, helping, controversy) at a specific part of the factory, the bank wiring room.
- The investigators noticed that workers themselves used specific terms to describe who is in "***our group***".
- The researchers tried to understand how such terms arise by reproducing in a visual way the group structure of the organization as it emerged from the individual relationships of the factory workers.

- In another study of mixed-race city in the Southern US researchers looked at the network of overlapping "***cliques***" defined by race and age.
- They also went further than the Hawthorne study in generating hypotheses about the possible connections between cliques.
  - Example: they noted that lower-class members of a clique are usually only able to connect to higher-class members of another clique through the higher-class members of their own clique

- Each of the early studies used a different set of concepts and different methods of representation and analysis of social networks.
- From the 1950s network analysis began to converge around the unique world view that distinguishes network analysis from other approaches to sociological research.
- The term “social network” has been introduced by Barnes in 1954.

- This convergence was facilitated by the adoption of a graph representation of social networks by Moreno.
- According to Moreno "a ***sociogram*** was a visual representation of social networks as a ***set of nodes connected by directed links***".
- The nodes represented ***individuals*** and the edges stood for ***personal relations***.

- The field of network analysis has been growing steadily from the beginning, there have been two developments in the last two decades that led to an explosion in network literature.
- 1. Advances in information technology brought a wealth of electronic data and significantly increased analytical power.
- 2. The methods of SNA are increasingly applied to networks other than social networks.
- This advancement is based on the discovery that many networks in nature share a number of commonalities with social networks

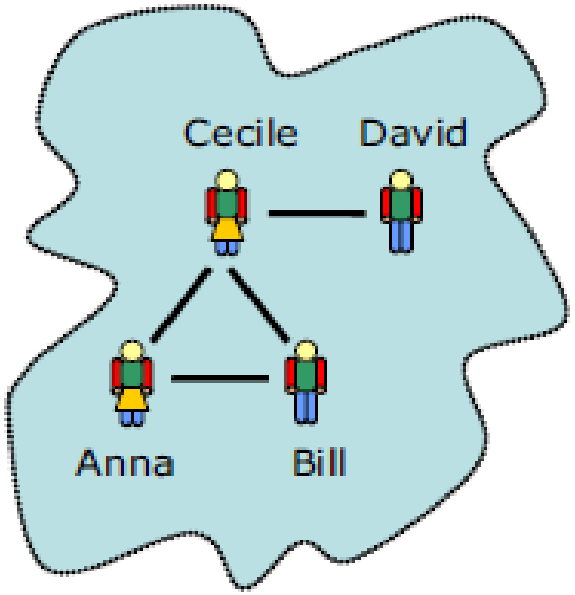
Key concepts and measures in network analysis

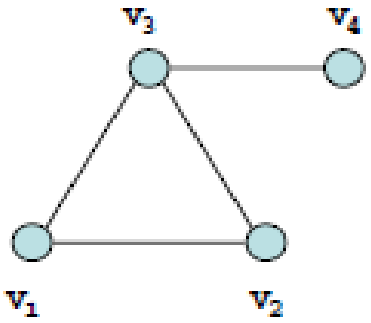
- Social Network Analysis has developed a set of concepts and methods specific to the analysis of social networks.
- Will proceed from the global structure of networks toward the measurement of ego-networks (personal networks), i.e. from the macro level to the micro level of network analysis.

THE GLOBAL STRUCTURE OF NETWORKS

- A social network can be represented as a graph G = (V,E) where V denotes the finite set of vertices and E denotes a finite set of edges such that $E \subseteq V \times V$.
- Each graph can be associated with its characteristic matrix

$$M := (m_{i,j})_{n*n} \text{ where } n = |V|, m_{i,j} = \left\{ \begin{array}{l|l} 1 & (v_i, v_j) \in E \\ 0 & otherwise \end{array} \right\}.$$





	v_1	v_2	v_3	v_4
v_1	0	1	1	0
v_2	1	0	1	0
v_3	1	1	0	1
v_4	0	0	1	0

Graph based representation of real world networks.

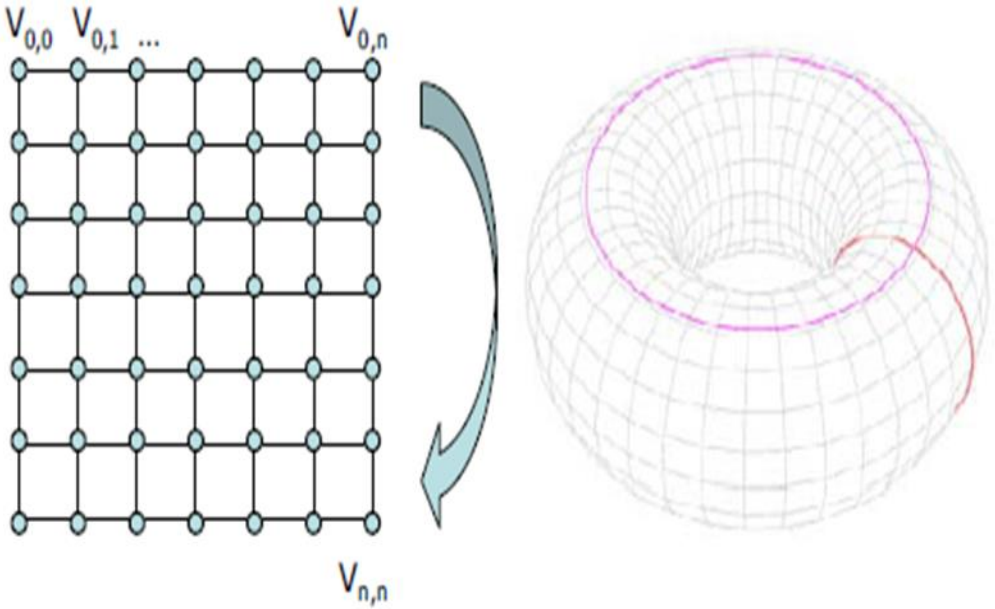
- The structure of social networks came from an experiment by the American psychologist Stanley Milgram .
- Milgram went out to test the common observation that no matter where we live, the world around us seems to be small: we routinely encounter persons not known to us who turn out to be the friends of our friends.
- Milgram thus not only wanted to test whether we are in fact all connected but he was also interested in what is the average distance between any two individuals in the social network of the American society.

- Milgram calculated the average of the length of the chains and concluded that the experiment showed that on average Americans are no more than six steps apart from each other.
- i.e ***Six degree of Separation***
- Milgram estimated the size of the average shortest path of the network, which is also called ***characteristic path length.***

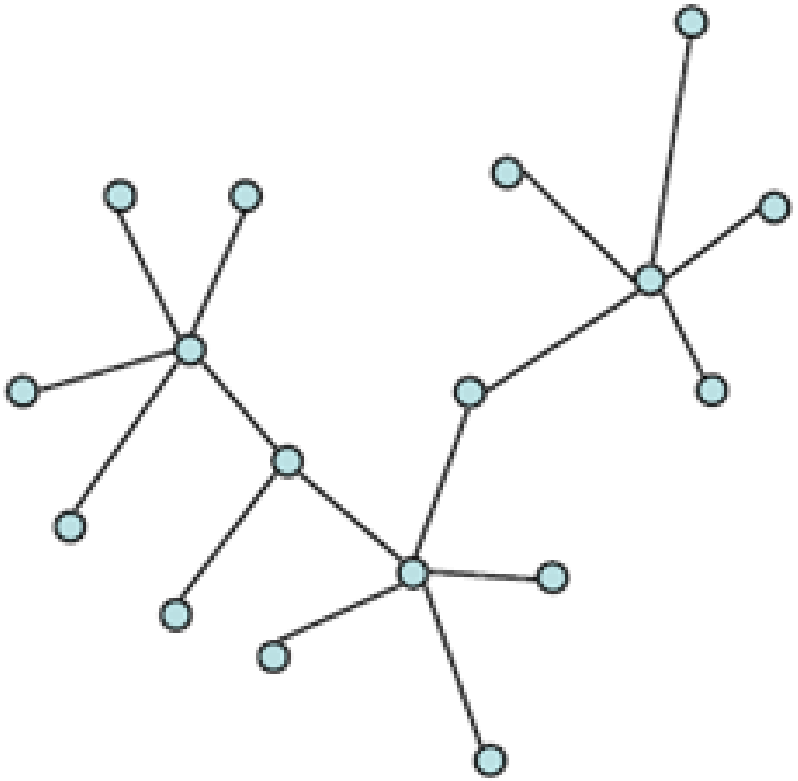
- Every vertex is connected to the next vertex and no vertex is repeated on the path.
- The shortest path between two vertices Vs and Vt is a path that begins at the vertex Vs and ends in the vertex Vt and contains the least possible number of vertices.
- The shortest path between two vertices is called a ***geodesic.***

- The longest geodesic in the graph is called the ***diameter*** of the graph.
- This is the maximum number of steps that is required between any two nodes.
- The ***average shortest path*** is the average of the length of the geodesics between all pairs of vertices in the graph.

- A practical impact of Milgram's finding is that we can exclude certain kind of structures as possible models for social networks.
- ***2D Lattice Model***



- Tree



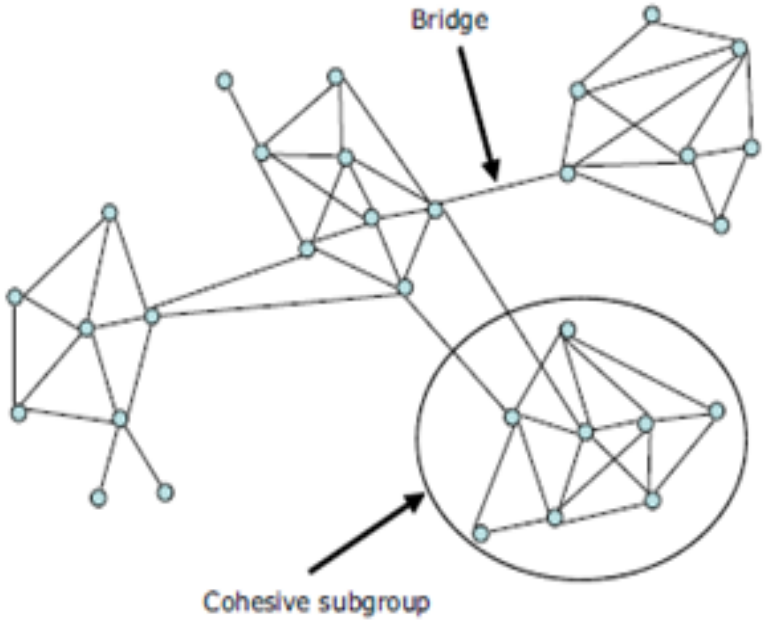
- Clustering is to measure the degrees of nodes to decide which nodes in a graph tends to be clustered together.
- Clustering for a single vertex can be measured by the actual number of the edges between the neighbors of a vertex divided by the possible number of edges between the neighbors.
- When taken the average over all vertices, the measure is known as ***clustering coefficient***.
- The clustering coefficient of a tree is zero, which is easy to see if we consider that there are no triangles of edges (triads) in the graph.

- The random graph model proposed by the Hungarian mathematicians Erdos and Renyi.
- A random graph can be generated by taking a set of vertices with no edges connecting them.
- Subsequently, edges are added by picking pairs of nodes with equal probability.
- This way we create a graph where each pair of vertices will be connected with an equal probability.

- Random graphs are interesting in the sense that they are examples of generative models.
- That is, random graphs are not (only) defined by what they are but also how they arise, i.e. the process of growing such a graph.
- These kinds of processes are also at the center point of interest for the field of complex networks.

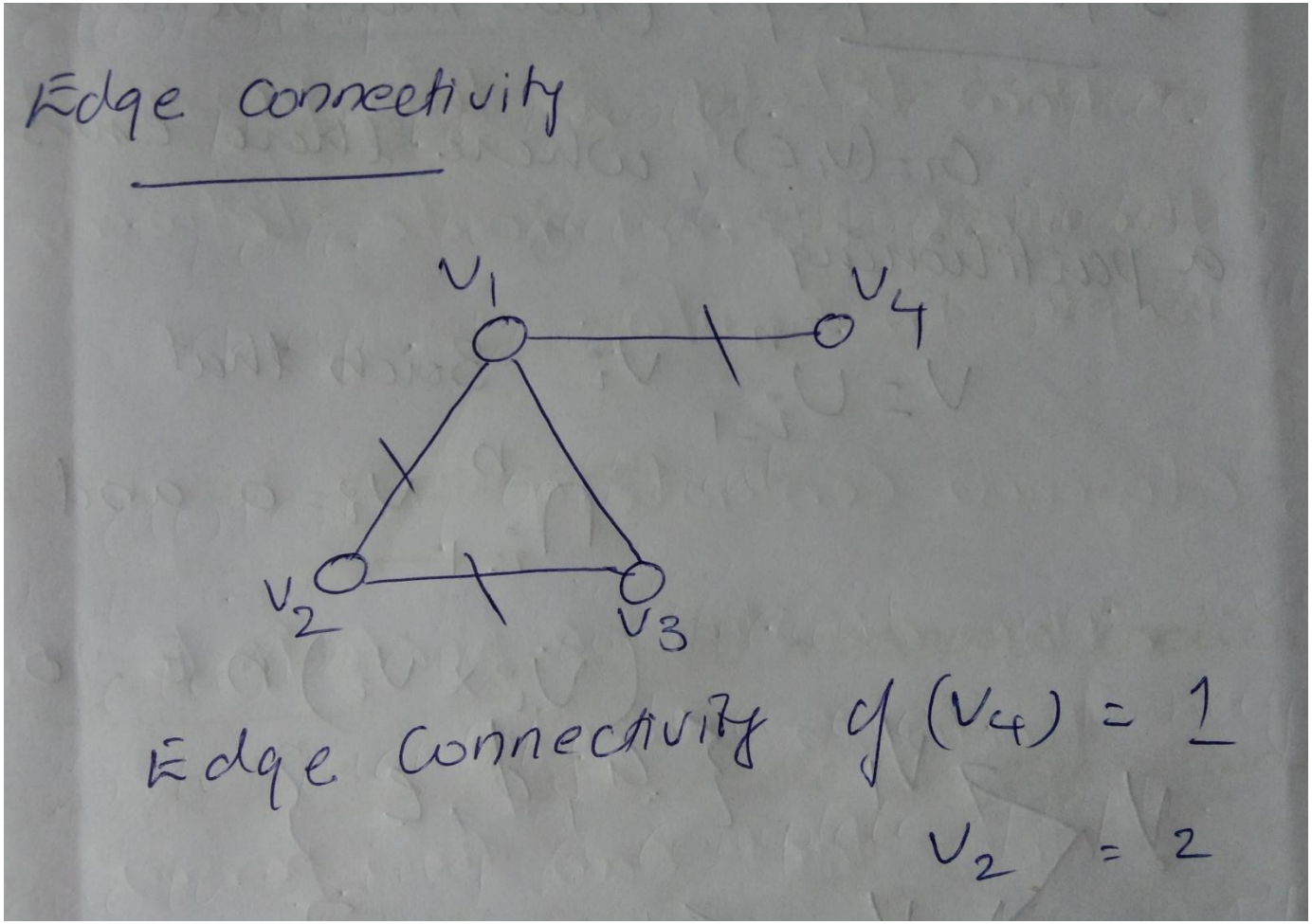
The Macro-Structure of Social Networks

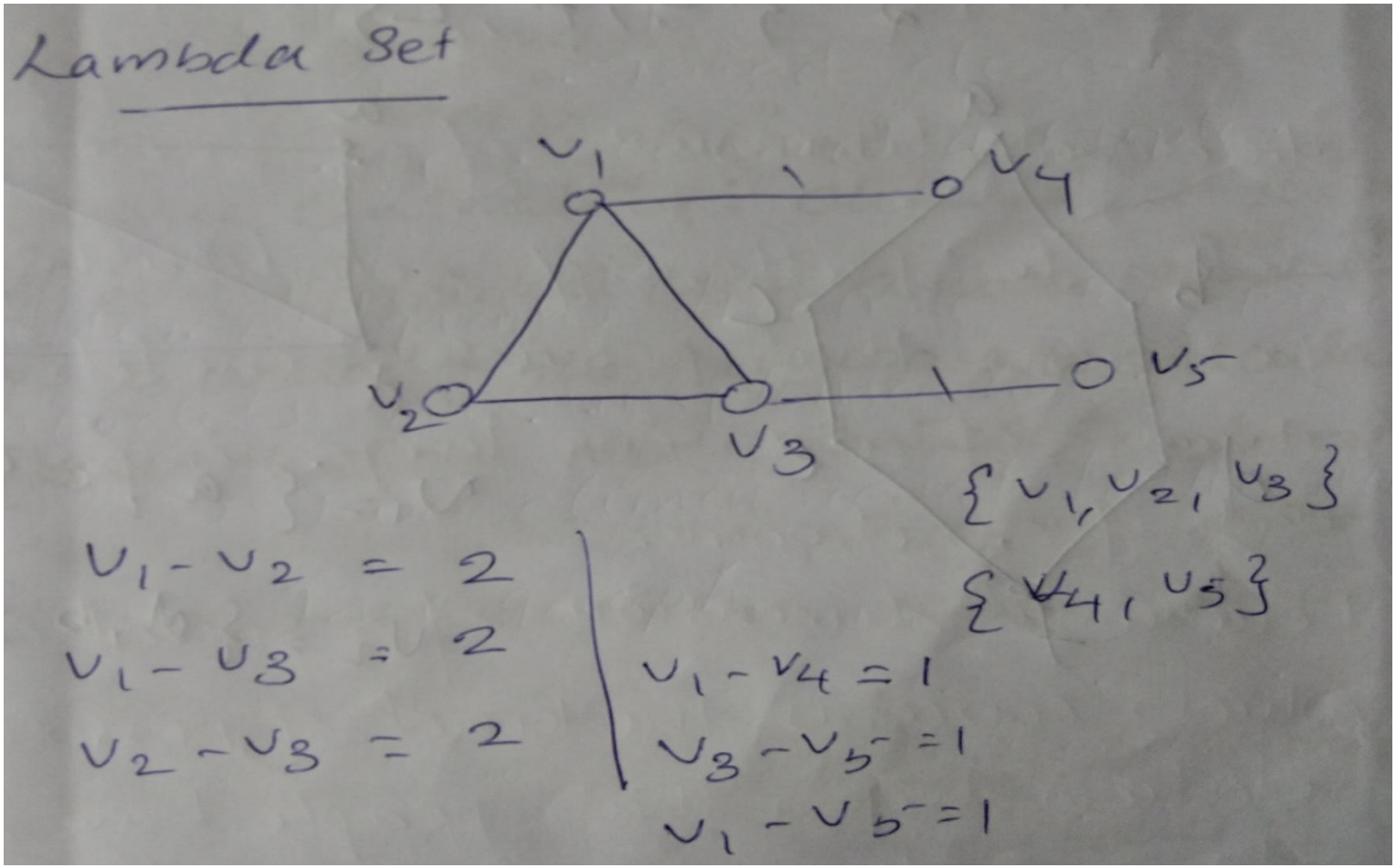
- To find the global characteristics of social networks.
- Most real world networks show a structure where densely connected subgroups are linked together by relatively few bridges



- Various clustering algorithms exist for creating disjunction or overlapping subsets of nodes based on different ***definitions of a what a subgroup is or how to find one***.
- Several definitions are based on the observations that ***subgroups are densely connected and their members are close to each other in the graph***.
- Other definitions constrain subgroups by putting limits on ***the maximum path length between the members***

- ***Cohesiveness*** is to compare the density of ties within the group to the density of ties between members of the subgroup and the outside.
- ***Edge connectivity*** of two vertices Vi and Vj is the minimum number of lines that need to be removed from a graph in order to leave no path between the two vertices.
- A ***lambda-set*** is defined as a set of nodes where any pair of nodes from the set has a larger edge connectivity than any pair of nodes where one node is from within the set and the other node is from outside the set.

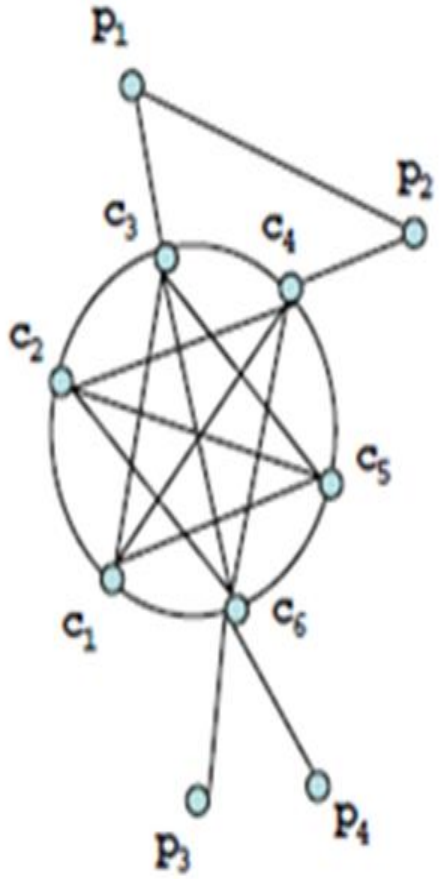

Edge connectivity
v_1
v_4
v_2
v_3
Edge connectivity of $(v_4) = 1$
$v_2 = 2$


Lambda Set
v_1
v_4
v_2
v_3
v_5
$\{v_1, v_2, v_3\}$
$\{v_4, v_5\}$
$v_1 - v_2 = 2$
$v_1 - v_3 = 2$
$v_2 - v_3 = 2$
$v_1 - v_4 = 1$
$v_3 - v_5 = 1$
$v_1 - v_5 = 1$

- Mark Newman introduced the ***edge - betweenness clustering method***.
- Instead of focusing on the density of subgroups, this algorithm targets ***the ties that connect them***.
- The ties that are in between groups can be spotted by calculating their betweenness.
- The ***betweenness*** of an edge is calculated by taking the set of all shortest paths in the graph and looking at what fraction of them contains the given edge.

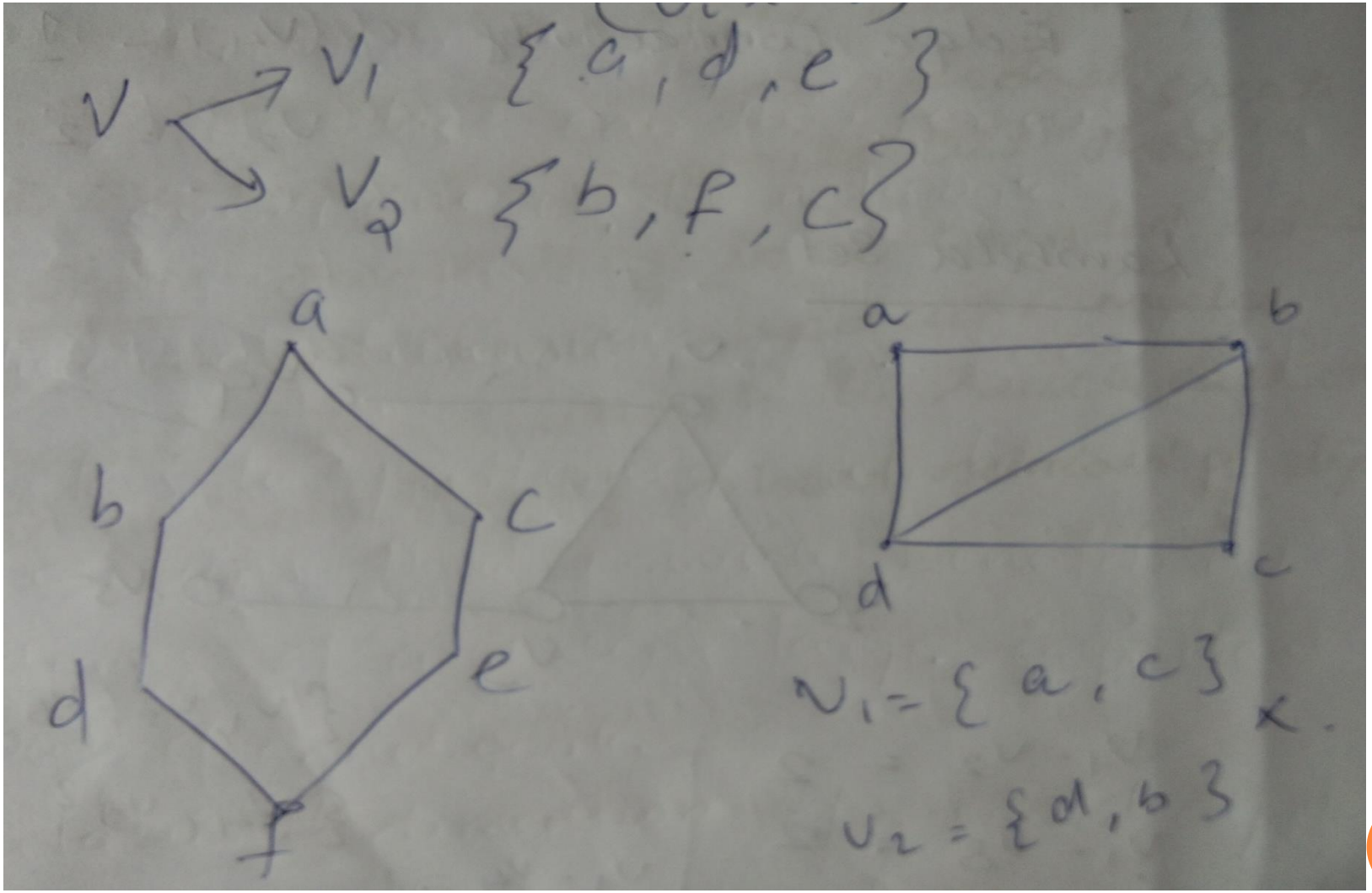
- An edge between clusters has a much higher betweenness than edges inside clusters because all shortest paths between nodes in the different clusters have to go through the given edge.
- By progressively removing the edges with the highest betweenness the graph falls apart in distinct clusters of nodes

- A ***Core-Periphery (C/P) structure*** is one where nodes can be divided in two distinct subgroups:
  - Nodes in the core are densely connected with each other and
  - The nodes on the periphery, which are loosely connected to the core
- The matrix form of a core periphery structure is 
-

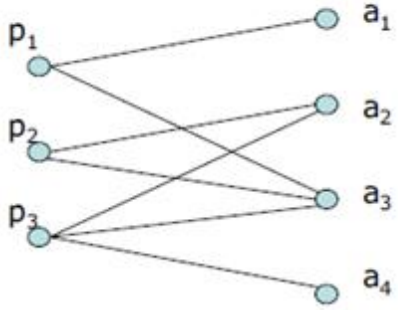


c_1	c_2	c_3	c_4	c_5	c_6	p_1	p_2	p_3	p_4
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	1	0	0	0
1	1	1	1	1	1	0	1	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	0	0	1	1
0	0	1	0	0	0	0	1	0	0
0	0	0	1	0	0	1	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	1	0	0	0	0

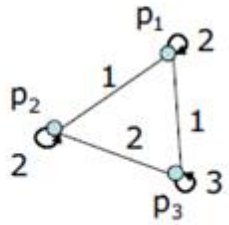
- ***Affiliation networks*** contain information about the relationships between two sets of nodes: ***a set of subjects and a set of affiliations***.
- An affiliation network can be formally represented as a ***bipartite graph***, also known as a two-mode network.
- In general, an n-partite graph or n-mode network is a graph G = (V,E) where there exists a partitioning
- $V=_{i=1}\cup^{n} Vi$ such that $_{i=1}\cap^{n}$ $Vi=0$ and $(Vi \times Vi) \cap E=0$
- The set of vertices is divided into n disjoint sets and there are no edges between vertices belonging to the same set.


V
V1 {a, d, e}
V2 {b, f, c}
a
b
c
d
e
f
a
b
d
c
V1 = {a, c}
V2 = {d, b}

- There are some methods to analyze affiliation networks.
- ***One -mode network.***
  - This transformation considers the overlaps between the affiliations as a measure of tie strength between the actors.



	p_1	p_2	p_3	a_1	a_2	a_3	a_4
p_1	0	0	0	1	0	1	0
p_2	0	0	0	0	1	1	0
p_3	0	0	0	0	1	1	1
a_1	1	0	0	0	0	0	0
a_2	0	1	1	0	0	0	0
a_3	1	1	1	0	0	0	0
a_4	0	0	1	0	0	0	0



	p_1	p_2	p_3
p_1	2	1	1
p_2	1	2	2
p_3	1	2	3

Personal Networks

- ***Social capital*** is a measure of the value of resources, both tangible (e.g., public spaces, private property) and intangible (e.g., actors, human capital, people), and the impact that these relationships have on the resources involved in each relationship, and on larger groups.

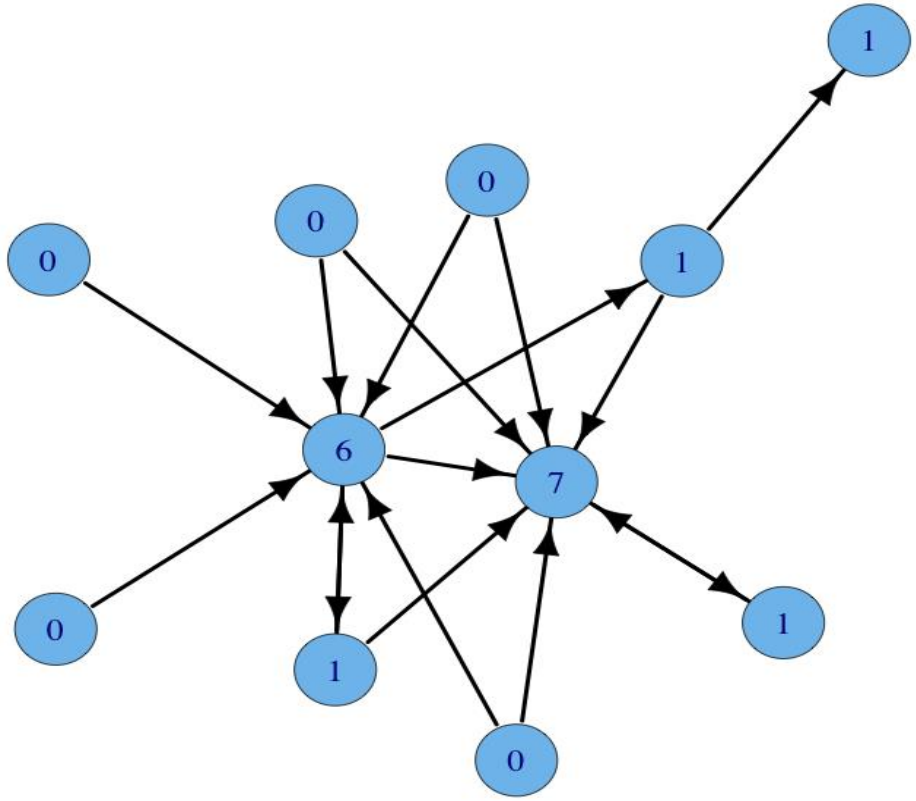
- According to the framework of Nahapiet and Ghoshal ,social capital in relation to its function in the development of knowledge and knowing capability, which they call ***intellectual capital.***
- They suggest that various aspects of social capital enable the creation of intellectual capital by allowing exchanges to take place that lead to the combination of knowledge possessed by individuals.

- Nahapiet and Goshal define social capital as the sum of actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit.
- They identify three main dimensions of social capital:
  - Structural dimension ,
  - Relational dimension and
  - Cognitive dimensions

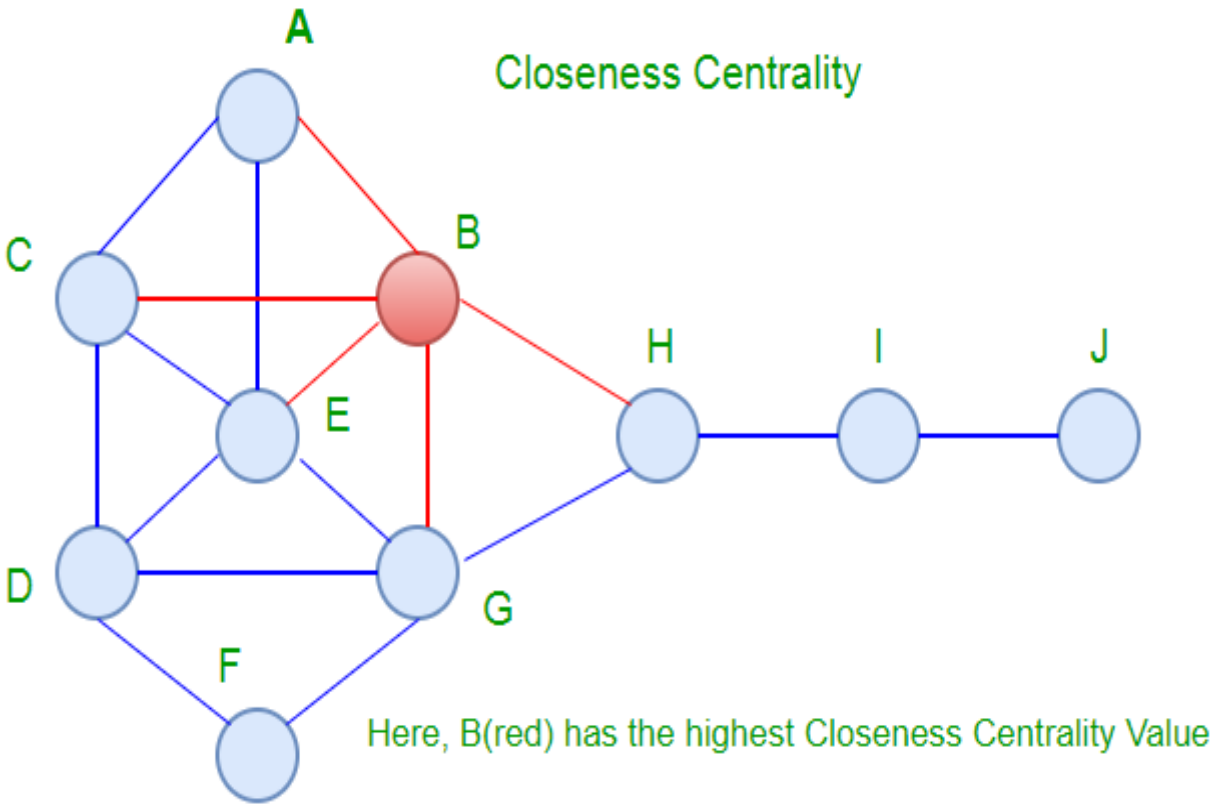
- The ***structural*** dimension of social capital refers to patterns of relationships or positions that provide benefits in terms of accessing large, important parts of the network.
- Common to structural measures of social capital is that they put a single node in the center and provide a measure for the node based on his connectivity to other nodes .

- Centrality
  - It is a measure indicating the importance of a node in a network.
    - Degree centrality
    - Closeness centrality
    - Betweenness centrality

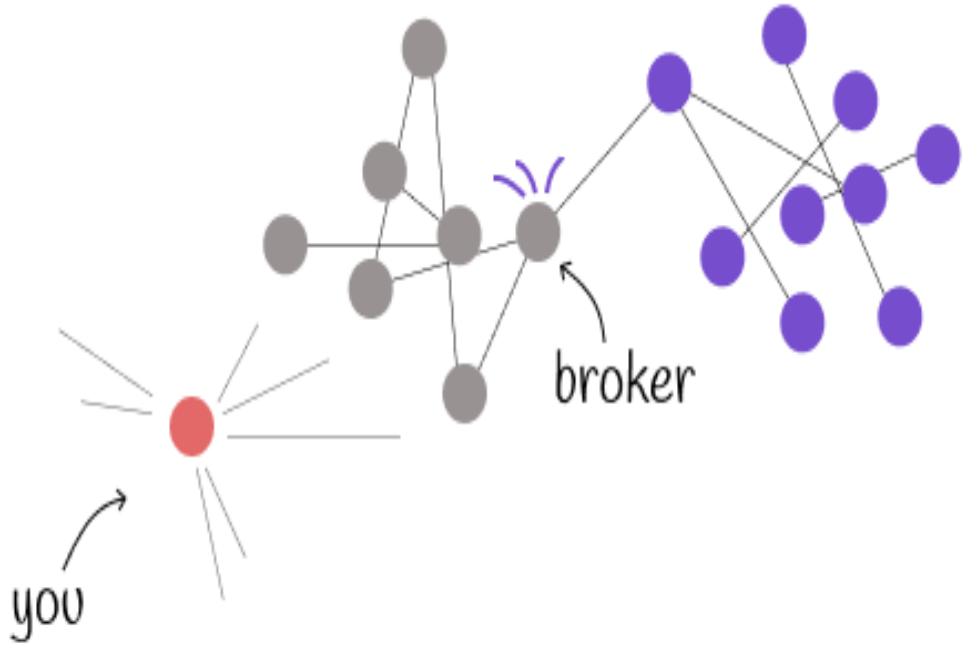
- ***Degree centrality*** is the number of (incoming, outgoing or all) links of a node.
- It is the first way to calculate the nodes that are most potential to determine other nodes which is present in the network.
- This measure is based on the idea that an actor with a large number of links has wider and more efficient access to the network, less reliant on single partners and because of his many ties often participates in deals as a third-party or broker.
  - How many people can reach this person directly?


1
0
0
0
1
6
7
0
1
1
0

- ***Closeness centrality*** is obtained by calculating the average (geodesic) distance of a node to all other nodes in the network.
- Closeness centrality characterizes the reach of the ego to all other nodes of the network.
- In larger networks it makes sense to constrain the size of the neighborhood in which to measure closeness centrality.
  - How distinct a node is to the other nodes in a network?
  - How fast can this person reach everyone in the network?

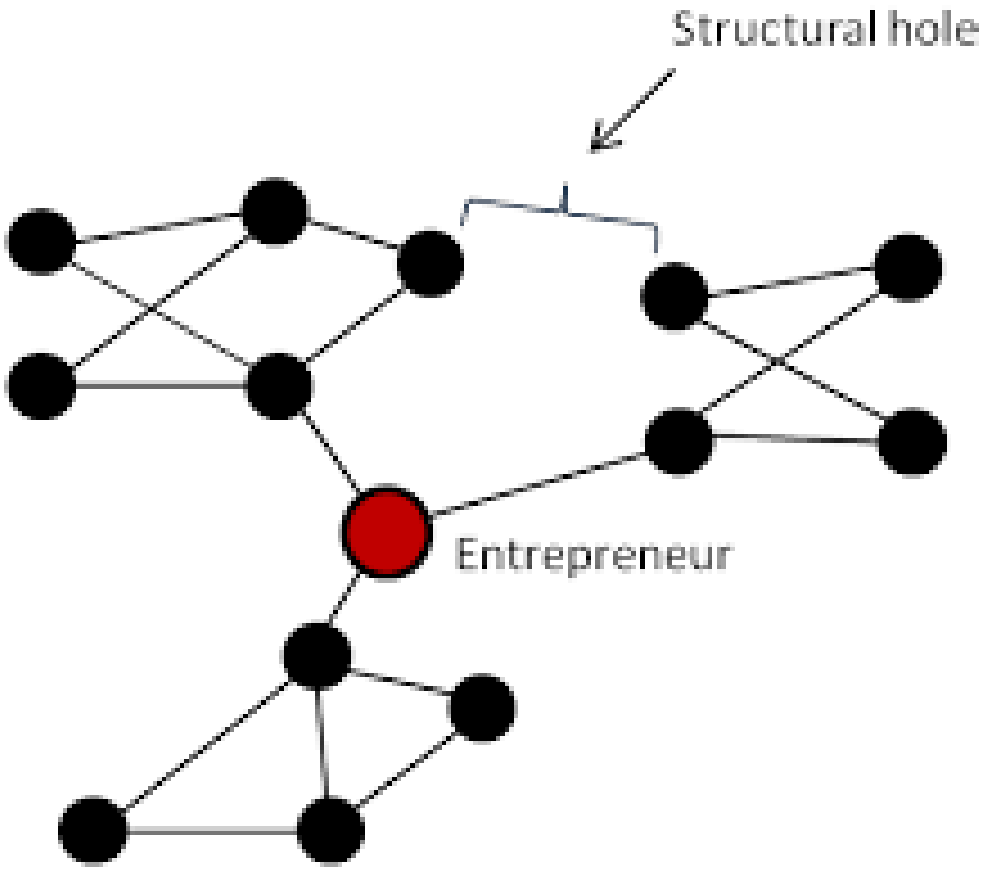

A
Closeness Centrality
B
C
H
I
J
E
D
G
F
Here, B(red) has the highest Closeness Centrality Value

- A ***broker*** gains advantage by standing in between disconnected communities.
- The ties spanning communities tend to be sparse compared to the dense networks of cohesive subgroups and as a result there are typically few bridges across communities.
- Brokers controlling these bridges are said to be in an advantageous position especially because of the value attributed to the information flowing across such ties.


broker
you

- The measure of ***betweenness centrality*** identifies broker positions by looking at the extent to which other parties have to go through a given actor to conduct their dealings.
- ***Betweenness*** is defined as the proportion of paths — among the geodesics between all pairs of nodes—that pass through a given actor.
  - How likely is this person to be the most direct route between two people in the network?

- A ***structural hole*** occurs in the space that exists between closely clustered communities.
- According to Ronald Burt, a broker gains advantage by bridging such holes.
- Burt suggests that structural holes show information benefits in three forms:
  - Access to large, disparate parts of the network (knowing where to turn for information, who can use some information),
  - Timing (receiving information sooner than others)
  - Reputation through referrals


Structural hole
Entrepreneur

- ***Relational*** dimension of social capital concerns the kind of personal relationships that people have developed with each other through a history of interaction .
- The relationships of pairs of actors who occupy similar network positions in similar network configurations may significantly differ based on their past interactions and thus their possibilities for action might also differ.

- The ***cognitive dimension*** of social capital refers to those resources providing shared representations, interpretations and systems of meaning .
- Cognitive ties are based on the existence of shared languages, signs and narratives, which facilitate the exchange of knowledge.