Unit I

1

INTRODUCTION TO SOCIAL NETWORK ANALYSIS (SNA)

Unit Structure

1.0 OBJECTIVE

- To gain knowledge about Social Network Analysis
- To learn various relations
- To use graph theory for Social Network Analysis
- To find shortest distance using Dijkstra's algorithm

1.1 INTRODUCTION

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them.

The SNA structure is made up of node entities, such as humans, and ties, such as relationships. The advent of modern thought and computing facilitated a gradual evolution of the social networking concept in the form of highly complex, graph-based networks with many types of nodes and ties. These networks are the key to procedures and initiatives involving problem solving, administration and operations.

1.2 INTRODUCTION TO NETWORKS AND RELATIONS

SNA can be used by a group as a process of "learning and understanding the (formal and informal) networks that operate in a given field" (Hovland, 2007). This extensive form of mind-mapping allows the group not only to identify networks, but also to highlight the patterns of information exchange within the network. Although networks are often created to pass information from one individual to another, and over time the content is also shared with a wider network, it also gradually grows to take in other outside contact and networks. An SNA focuses on the structure of the relationships that weave between people and organisations within a network.

Within every network there are starting points and branching points. Each individual who forms a network, whether it us within a team or an organisation, will exchange information with others who share the same or common beliefs and ideas. They will be likely to share information with friends, partners, and relatives if they find the information interesting. Maximising the appeal of this information can increase traffic to the sites or pages of the distributing organisation.

What is a 'social network'?

A social network is made up of what are called 'nodes' (points) and 'links', all of which are then identifiable categories of analysis. These include people, groups, and organisations – which are usually the main priority and concern for any type of social examination. Links in this type of analysis which focus on the 'collective' include social contacts and

exchangeable information. It has been argued for some time that organisations are embedded in networks of larger social processes, which they influence, and which also influence them (Hovland, p.10).

How do we use SNA?

"A range of methods can be used, including ethnography, participant observation, key informant interviews, semi-structured interviews, 'snowball' sampling, focus groups, and content analysis of the media" (Schelhas and Cerveny, 2002, Social Network Analysis for Collaboration in Natural Resource Management).

"The aim is to construct a 'map' of the linkages that exist between people in this field" (Hovland, 2005).

"Social network analysis is the mapping and measuring of relationships and flows between people, groups, organisations, computers or other information/knowledge processing entities" (Valdis Krebs, 2002). Social Network Analysis (SNA) is a method for visualizing our people and connection power, leading us to identify how we can best interact to share knowledge.

1.2.1 Analyzing relationships to understand people and groups

The science of Social Network Analysis (SNA) boils down to one central concept—our relationships, taken together, define who we are and how we act. Our personality, education, background, race, ethnicity—all interact with our pattern of relationships and leave indelible marks on it. Thus, by observing and studying these patterns we can answer many questions about our sociality.

What is a relationship? In an interpersonal context, it can be friendship, influence, affection, trust—or conversely, dislike, conflict, or many other things.

1.2.2 Binary and valued relationships

Relationships can be binary or valued: "Max follows Alex on Twitter" is a binary relationship while "Max retweeted 4 tweets from Alex" is valued. In the Twitter world, such relationships are easily quantified, but in the "softer" social world it's very hard to determine and quantify the quality of an interpersonal relationship.

A useful stand-in for strength of an interpersonal relationship is frequency of communication. Besides being objectively measurable, frequency of communication has been found by scientists to reflect accurately on the emotional content, and amount of influence in a relationship. This would, of course, not be true in many contexts (and you, my dear reader, are probably busy coming up with counterexamples right now)—but in many cases, for the lack of better data, frequency of communication works.

1.2.3 Symmetric and asymmetric relationships

It is easy to see that some relationships are asymmetric by nature. Teacher/student or boss/employee roles presume a directionality of a relationship, and do not allow for a symmetric tie back. Following on Twitter and LiveJournal is directional by definition—but a follow-back tie can exist, thus symmetrizing the relationship

Other relationships are symmetric. Facebook friends and LinkedIn connections require mutual confirmation—the software forces a symmetry even when the real human relationship is asymmetric.

In the real world, friendships and romantic relationships are asymmetric, as much as we would like them not to be that way. Hence, we struggle with unrequited love, one-sided friendships and other delusions of popularity. Given good data, we can study these phenomena using SNA—but such data would be very difficult to obtain and subject to self-reporting and other biases.

1.2.4 Multimode relationships

Finally, we should mention that relationships can exist between actors of different types—Corporations employ People, Investors buy stock in Corporations, People possess Information and Resources, and so on. All of these ties are described as *bimodal* or *2-mode*

1.3 USING GRAPH THEORY FOR SOCIAL NETWORKS ANALYSIS

What is graph theory?

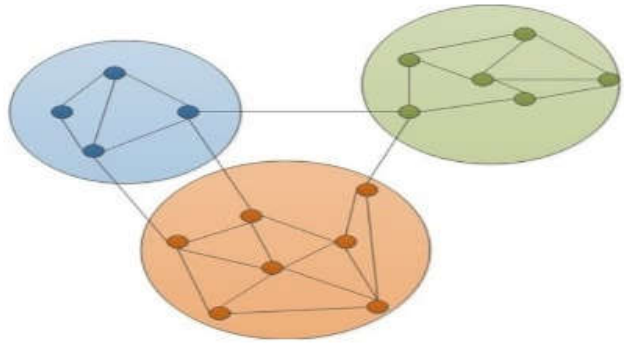
The first thing you should know is that a graph is a mathematical structure that allows you to represent everyday problems in a graphic way. In addition, network theory allows you to represent only one type of relationship (simple representation), but it also allows you to represent more than one type (in that case, it would be called multiple).

Even eminent figures such as the founder of Facebook, Mark Zuckenberg, have spoken of “social graphs” to represent the connections and relationships that users of the social network have.

Graph theory is a branch of mathematics, the same branch that is also used in computer science. It is based on both discrete and applied mathematics. In this way, it manages to encompass different concepts.

Applying graph theory to social networks

Let’s think about the commercial strategy that any telecommunication company seeking to know the composition of the links would carry out. You would be interested in knowing which people we usually talk to and thus adapt your commercial strategy to offer personalized offers and/or rates.



In addition to this, applying the networks to social networks can work to adapt the products to the real needs, making them appear at the right time.

When we talk about networks applied to social networks the most common is that they are used to "detect communities". Thanks to the algorithms we can see characteristics, attributes and relationships that match within a group. When we analyze the subnetworks, we can see the vertices that are most related to each other, and also how they relate to the rest of the vertices.

If we look at the graph above, we can see that three different communities have been detected, in which we can assume that all the members of the same community have characteristics or attributes that coincide.

1.3.1 Adjacency matrices

Every network can be expressed mathematically in the form of an adjacency matrix (Figure 4). In these matrices the rows and columns are assigned to the nodes in the network and the presence of an edge is symbolised by a numerical value. By using the matrix representation of the network, we can calculate network properties such as degree, and other centralities by applying basic concepts from linear algebra (see later in the course).

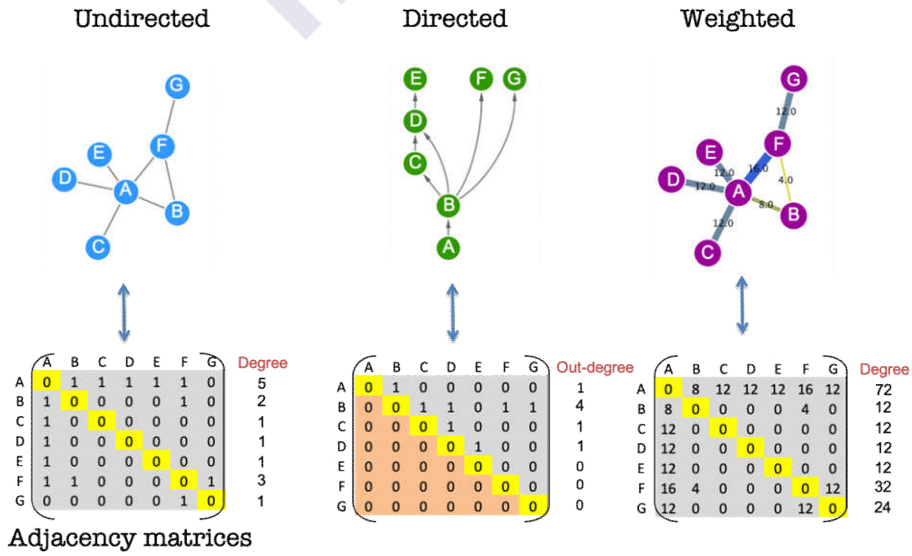


Figure 4 Graphs by edge type and their adjacency matrices.

A network with undirected, unweighted edges will be represented by a symmetric matrix containing only the values *1* and *0* to represent the presence and absence of connections, respectively.

Directed and weighted networks can make use of different numerical values in the matrix to express these more complex relationships. The sign of the values, for example, is sometimes used to indicate stimulation or inhibition.

1.3.2 Edge-lists

An **edgelist** is usually formatted as a table where the first two columns contain the IDs of a pair of nodes in the network that have a tie between them. Optional additional columns may contain properties of the relationship between the nodes (e.g., the value of a tie). Any pair of nodes that does not have a tie between them is usually not included in an edgelist. This property is what makes edgelists a more efficient network data storage format than sociomatrices (see below). Unobserved edges can be encoded in edgelist format by including "NA" in the value column. Here's an example of a simple edgelist table with a value column:

Ego	Alter	Value
Harry	Hermione	5
Harry	Ron	6
Hermione	Ron	5

The columns in an edgelist table are usually ordered "Ego" (often the person who completed the interview or who was the subject of a focal follow) followed by "Alter" (the person that the focal individual named or interacted with). In the case of undirected network data, the ordering of these columns does not matter, but in directed data, it does. igraph and statnet software will encode directed edgelist data with the arrow pointing from the first to the second column, so if the ties you recorded have reversed directionality from ego to alter (e.g., the alter gave something to ego) you should flip the order of the columns before converting the data to a network in order to get the edges properly directed (if you want your network to show the direction that support flows in the network). Most directed ties are straightforward to interpret but sometimes it gets complicated, depending on your research design.

The final column in our imaginary edgelist contains a value for the edge. It might be the number of years the pair have known each other, or some measure of the strength or quality of their relationship.

1.3.3 Adjacency lists

An adjacency list is-- a hybrid between an adjacency matrix and an edge list. An adjacency list is an array of linked lists that serves the purpose of representing a graph. What makes it unique is that its shape also makes it

easy to see which vertices are adjacent to any other vertices. Each vertex in a graph can easily reference its neighbors through a linked list.

Due to this, an adjacency list is the most common representation of a graph. Another reason is that graph traversal problems often require us to be able to easily figure out which nodes are the neighbors of another node. In most graph traversal interview problems, we don't really need to build the entire graph. Rather, it's important to know where we can travel (or in other words, who the neighbors of a node are).

1.3.4 Graph traversals and distances

Graph traversal (also known as graph search) refers to the process of visiting (checking and/or updating) each vertex in a graph. Such traversals are classified by the order in which the vertices are visited. Tree traversal is a special case of graph traversal. For any two locations in a spatial network, their network distance is the length of the shortest path between these two locations along the network. The shortest path is computed based on the travel weight, such as travel distance or travel time, of network edges

1.3.5 Depth-first traversal

Depth-first search (DFS) is an algorithm for traversing or searching tree or graph data structures. The algorithm starts at the root node (selecting some arbitrary node as the root node in the case of a graph) and explores as far as possible along each branch before backtracking.

1.3.6 Breadth-first traversal paths and walks

Breadth-first search (BFS) is an algorithm for searching a tree data structure for a node that satisfies a given property. It starts at the tree root and explores all nodes at the present depth prior to moving on to the nodes at the next depth level.

1.3.7 Dijkstra's algorithm

Dijkstra's algorithm is very similar to Prim's algorithm for minimum spanning tree. Like Prim's MST, we generate a *SPT (shortest path tree)* with a given source as a root. We maintain two sets, one set contains vertices included in the shortest-path tree, other set includes vertices not yet included in the shortest-path tree. At every step of the algorithm, we find a vertex that is in the other set (set of not yet included) and has a minimum distance from thesource. Below are the detailed steps used in Dijkstra's algorithm to find the shortest path from a single source vertex to all other vertices in the given graph.

Algorithm

1) Create a set *sptSet* (shortest path tree set) that keeps track of vertices included in the shortest-path tree, i.e., whose minimum distance from the source is calculated and finalized. Initially, this set is empty.

2) Assign a distance value to all vertices in the input graph. Initialize all distance values as INFINITE. Assign distance value as 0 for the source vertex so that it is picked first.

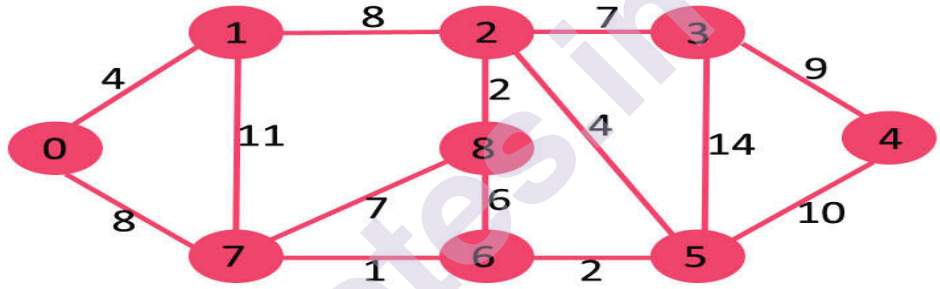
3) While *sptSet* doesn't include all vertices

a) Pick a vertex u which is not there in *sptSet* and has a minimum distance value.

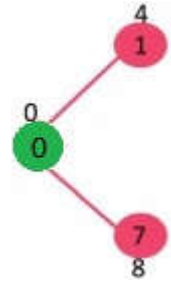
b) Include u to *sptSet*.

c) Update distance value of all adjacent vertices of u. To update the distance values, iterate through all adjacent vertices. For every adjacent vertex v, if the sum of distance value of u (from source) and weight of edge u-v, is less than the distance value of v, then update the distance value of v.

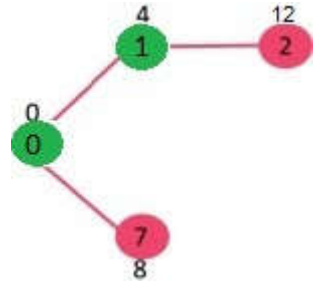
Let us understand with the following example:



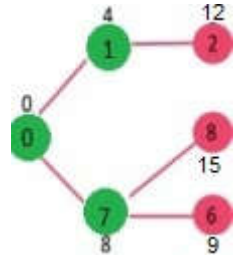
The set *sptSet* is initially empty and distances assigned to vertices are {0, INF, INF, INF, INF, INF, INF, INF} where INF indicates infinite. Now pick the vertex with a minimum distance value. The vertex 0 is picked, include it in *sptSet*. So *sptSet* becomes {0}. After including 0 to *sptSet*, update distance values of its adjacent vertices. Adjacent vertices of 0 are 1 and 7. The distance values of 1 and 7 are updated as 4 and 8. The following subgraph shows vertices and their distance values, only the vertices with finite distance values are shown. The vertices included in SPT are shown in green colour.



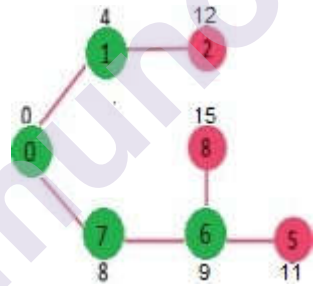
Pick the vertex with minimum distance value and not already included in SPT (not in sptSET). The vertex 1 is picked and added to sptSet. So sptSet now becomes {0, 1}. Update the distance values of adjacent vertices of 1. The distance value of vertex 2 becomes 12.



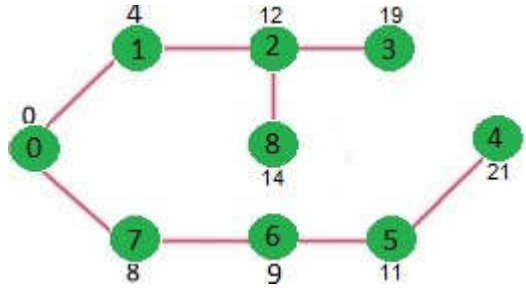
Pick the vertex with minimum distance value and not already included in SPT (not in sptSET). Vertex 7 is picked. So sptSet now becomes {0, 1, 7}. Update the distance values of adjacent vertices of 7. The distance value of vertex 6 and 8 becomes finite (15 and 9 respectively).



Pick the vertex with minimum distance value and not already included in SPT (not in sptSET). Vertex 6 is picked. So sptSet now becomes {0, 1, 7, 6}. Update the distance values of adjacent vertices of 6. The distance value of vertex 5 and 8 are updated.

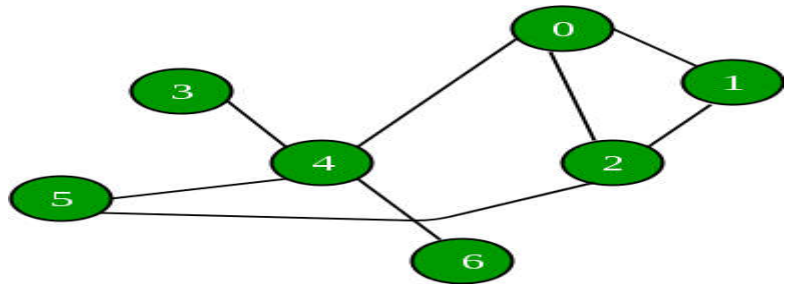


We repeat the above steps until *sptSet* includes all vertices of the given graph. Finally, we get the following Shortest Path Tree (SPT).



1.3.8 Graph Distance and Graph Diameter

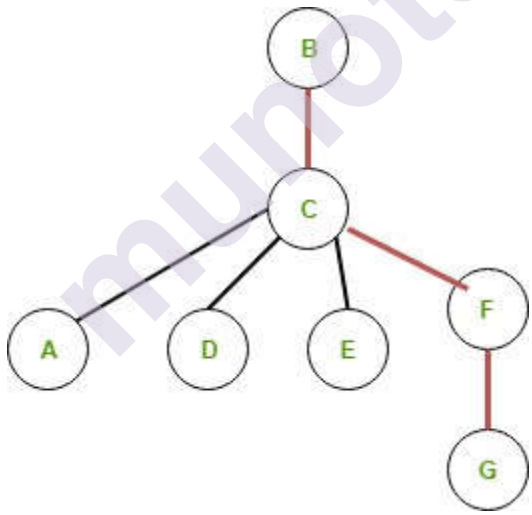
The distance between two vertices in a graph is the number of edges in a shortest or minimal path. It gives the available minimum distance between two edges. There can exist more than one shortest path between two vertices.



Shortest Distance between 1 - 5 is **2** 1 → 2 → 5

Diameter of graph –

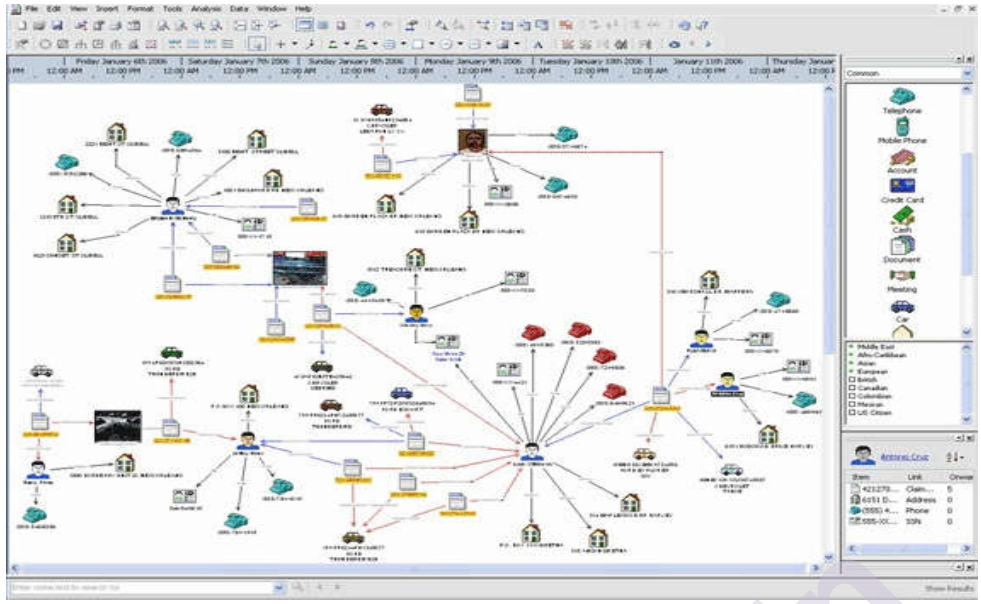
The diameter of graph is the maximum distance between the pair of vertices. It can also be defined as the maximal distance between the pair of vertices. Way to solve it is to find all the paths and then find the maximum of all.



Diameter: **3** BC → CF → FG

1.3.9 Social networks vs. link analysis

Another cousin to SNA is Link Analysis (LI). Some of you may have used LI in business intelligence or law enforcement work or seen it on TV. "Without a Trace" uses link analysis in every episode; "Numbers" and "Law and Order" resort to it on occasion.Link analysis is in many ways similar to SNA—both talk about relationships in terms of nodes and edges (Figure 1-1) and both try to derive the idea of who is more important in a network by analyzing the whole network, not individual events.

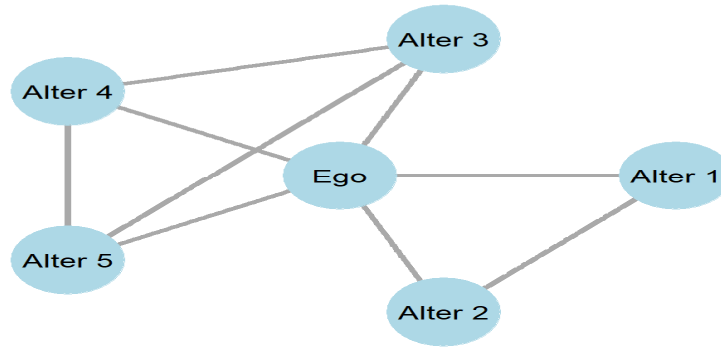


However, LI allows for a mixing of different node and edge types in the same network—i.e, "**A** *gave* **$300** to **B** to *procure* **drugs** for **C**". In this example, **bold** words are nodes, or actors, and *italic* are actions, or edges. The problem is understanding on a quantitative level whether the act of giving money is different than the act of procuring drugs—and thus LI relies on human-level understanding of language and is qualitative in its pure form.

Most link analysis tools, including Analyst's Notebook and Palantir, include qualitative data gathering and tools for qualitative decision-making, and these are excellent and utilized widely in a number of communities. However, the application of quantitative metrics centrality measures is dangerous because mixing nodes and edges of different meanings (e.g., money and telephone calls) produces a result that is mathematically invalid. Unfortunately, this does not stop the software from computing these metrics.

1.4 EGO-CENTRIC AND SOCIO-CENTRIC DENSITY

Ego-centric networks (or shortened to "ego" networks) are a particular type of network which specifically maps the connections of and from the perspective of a single person (an "ego"). For example, if you were to ask someone to name their friends (or any other type of "alter," which is defined as someone who is not the ego), they will tell you who their friends are. However, they will not tell you who the friends of other people are. While in Figure 1.3 and 1.4 we had graphs of complete networks, in this case, we are only getting a small part of the overall social network.



Sociocentric network analysis involves the quantification of relationships between people within a defined group – a classroom of children, a board of directors, the residents of a village or town, the trading partners in a bloc of nations.

1.5 LET US SUM UP

- Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory
- Graph theory is a branch of mathematics, the same branch that is also used in computer science. It is based on both discrete and applied mathematics.
- The distance between two vertices in a graph is the number of edges in a shortest or minimal path.
- Ego-centric networks are a particular type of network which specifically maps the connections of and from the perspective of a single person.
- Sociocentric network analysis involves the quantification of relationships between people within a defined group.
- In Dijkstra's algorithm to find the shortest path from a single source vertex to all other vertices in the given graph.

1.6 LIST OF REFERENCES

- Social Network Analysis: Methods and Applications by Katherine Faust and Stanley Wasserman
- Social network analysis by John Scott
- The SAGE Handbook of Social Network Analysis

1.7 BIBLIOGRAPHY

- https://www.techopedia.com/definition/3205/social-network-analysis-sna
- https://www.oreilly.com/library/view/social-network-analysis/9781449311377/ch01.html

- https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/introduction-to-graph-theory/graph-theory-adjacency-matrices/
- https://eehh-stanford.github.io/SNA-workshop/data-import.html#edgelists
- https://algodaily.com/lessons/implementing-graphs-edge-list-adjacency-list-adjacency-matrix
- https://www.geeksforgeeks.org/dijkstras-shortest-path-algorithm-greedy-algo-7/
- https://www.geeksforgeeks.org/graph-measurements-length-distance-diameter-eccentricity-radius-center/
- https://bookdown.org/omarlizardo/_main/2-10-ego-centric-networks.html
- Social network analysis by John Scott
- The SAGE Handbook of Social Network Analysis

1.8 UNIT END EXERCISE

1. Explain Social Network Analysis.

2. Briefly Explain

    a) Adjacency matrices

    b) Edge-lists

    c) Adjacency lists

    d) Graph traversals and distances

    e) Depth-first traversal

    f) Breadth-first traversal paths and walks

3. Explain Dijkstra's algorithm

4. Short notes on: Graph distance and graph diameter

5. Explain Social networks vs. link analysis

6. Explain Ego-centric and socio-centric density.

2

NETWORKS, CENTRALITY AND CENTRALIZATION IN SNA

Unit Structure

2.0 OBJECTIVES

After going through this unit, you will be able to:

- Understand Network structure and its components
- Explain Connectivity components of networks
- Define Network centrality components like density, reachability connectivity and reciprocity
- Describe the Structure holes in the networking structure
- Analyse of the network structure using the SNA structures cliques, components
- Practice Google page ranking algorithm

2.1. UNDERSTANDING NETWORKS

Network consists of nodes and edges. Nodes are also called actors or vertices. Edges are also called links or ties. In social Network Analysis nodes represent people. The actors establish the relationship through edges with each other. The figure 2.1.1. represents the structure of network.

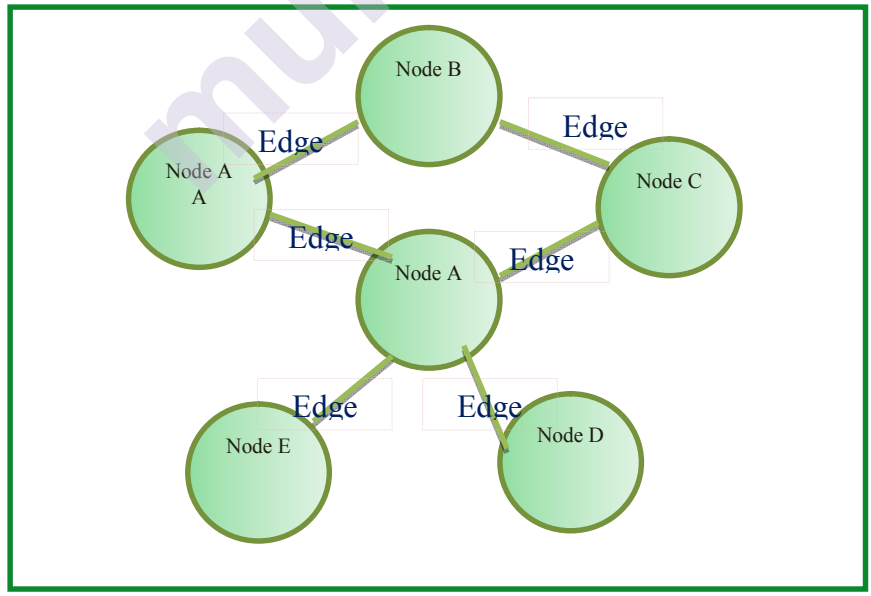


Figure 2.1.1. Structure of Network

2.1.1 Density

The proportions of all the possible ties or connections denote the density of a netwo lculated as

$$\textbf{Density} = \frac{\textit{Sum of Ties}}{\textit{Number of possible ties}}$$

The ratio between the total number of ties with the total number of possible ties gives the density of the network. The density drives the social capita and social constraints among a network.

Knoke and Kuklinski (1982) selected a subset of 10 organizations with two relationships from 95 organizations with 13 different types of relationships. In this the Money exchange is recorded in KNOKM, information exchange in KNOKI which is used for analysis here. In the Matrix 1 represents no ties and 2 represents there exists a relationship.

When calculating density self-ties are ignored and cohesion is taking onto consideration. The density of Matrix #1 is different from density of Matrix #2 which shows the density depends on the relationship exists within the actors based on the criteria selected.

Matrix#1		KNOKI							
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
COUN	0	1	0	0	1	0	1	0	1	0
COMM	1	0	1	1	1	0	1	1	1	0
EDUC	0	1	0	1	1	1	1	0	0	1
INDU	1	1	0	0	1	0	1	0	0	0
MAYR	1	1	1	1	0	0	1	1	1	1
WRO	0	0	1	0	0	0	1	0	1	0
NEWS	0	1	0	1	1	0	0	0	0	0
UWAY	1	1	0	1	1	0	1	0	1	0
WELF	0	1	0	0	1	0	1	0	0	0
WEST	1	1	1	0	1	0	1	0	0	0
		Density			0.49					
		Standard Deviation			0.4999					

Table 2.1.1. .A part of KNOKI Data Set

Table 2.1.2. A part of KNOKM Data Set

Matrix#2		KNOKM							
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
COUN	1	0	1	0	1	0	0	1	1	1
COMM	0	0	1	0	0	0	0	0	0	0
EDUC	0	0	0	0	0	0	0	1	0	0
INDU	0	1	1	0	0	0	1	1	1	0
MAYR	0	1	1	0	0	0	0	1	1	0
WRO	0	0	0	0	0	0	0	0	0	0
NEWS	0	1	0	0	0	0	0	1	0	0
UWAY	0	0	0	0	0	0	0	0	1	1
WELF	0	0	1	0	0	0	0	1	0	0
WEST	0	0	0	0	0	0	0	0	0	0
Density			0.23							
Standard Deviation			0.42							

2.1.2. Reachability

If any actor is reached by another actor then there exists a reachability of nodes. In this case we can trace a target actor/node from a given source node. This does not consider the number of actors lies in the middle of the path way. In directed, asymmetric data the source can reach the target but the target actor cannot reach the source actor. For instance there is a directed tie from actor A to B, now B can be reachable from A, but A cannot be reachable from B.

Table 2.1.3

Matrix #3

	A	B	C	D	E
A	0	1	1	0	1
B	1	1	0	0	1
C	0	0	1	1	1
D	1	1	1	0	0
E	1	0	1	1	0

Table 2.1.4

Matrix #3

	A	B	C	D	E
A	0	1	0	0	1
B	1	0	0	0	1
C	0	0	0	1	1
D	1	0	0	0	0
E	1	0	0	1	0

In the given Table 2.1.3. B and C are reachable in all cases if they are undirected. In the Table 2.1.4 it is not possible B and C are not reachable in some cases.

2.1.3. Connectivity

Adjacency refers the direct connection between two actors, it may be directed or un-directed. If there are more path ways connect two actors then there is a high connectivity to reach the actors from different paths.

Table 2.1.5 Point connectivity of KNOKE information exchange

Matrix#5										
	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1	5	5	3	4	5	1	6	4	4	3
2	5	8	3	5	8	1	6	5	3	4
3	3	3	4	4	3	1	4	3	3	3
4	5	5	3	5	5	1	5	4	3	4
5	5	8	3	5	8	1	6	5	3	5
6	1	1	1	1	1	1	2	1	2	1
7	5	6	3	5	6	1	6	4	2	3
8	5	5	3	5	5	1	5	5	4	4
9	3	3	3	3	3	1	3	3	3	3
10	4	5	3	4	5	1	4	4	3	5

From the Table 2.1.5 the Actor 6 is having weak connectivity. If any Actor refuses to send message to one then it is difficult for Actor 6 to get most of the information.

2.1.4. Reciprocity

SymmetricDyadic data: denotes that a pair of actors that may be connected or not. With directed data there are four possible dyadic relationships. A and B are two actors. In directed data there are four possible dyadic relationship/ties. They are A sends to B, B sends to A, Both sends each other, Both do not have any connection. A network which has more number of null or reciprocated ties over asymmetric connections are more equal or stable network.

Degree of Reciprocity:

In a population data there are two ways of indexing the degree of reciprocity. They are i) Dyad Method ii) Arc Method

i) Dyad Method:

This calculates the proportion of pairs of actors that are reciprocated. In the *Figure 2.1.2.* the A actor and B actor are reciprocated. But the Actor B and Actor C are not reciprocated. Possible reciprocations are AB, BC, AC (3 Numbers). But available is only AB (1 Number). Therefore the degree is 1/3 ie 0.3333.

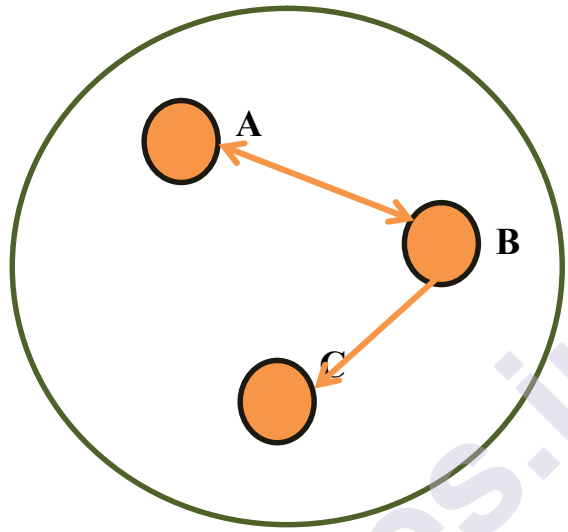


Figure 2.1.2.

ii) Arc Method:

This method focuses on relations. This calculates the percentage of all possible ties in the parts of available reciprocated structures. The degree is defined as number of ties that are involved in reciprocal relations relative to the total number of actual ties. In this example it is 2/3 ie 0.667. The number of ties involved are two. They are AB and BA. The number of actual ties are three. They are AB, BA and BC.

2.1.5. Group-external and Group-internal ties in networks

The E-I (external - internal) index takes the number of ties of group members to outsiders, subtracts the number of ties to other group members, and divides by the total number of ties. The resulting index ranges from -1 to +1. Since this measure is concerned with any connectionbetween members, the directions of ties are ignored.

The E-I index can be applied at three levels: the entire population, each group, and each individual. That is, the network as a whole (all the groups) can be characterized in terms of the bounded-ness and closure of its subpopulations. We can also examine variation across the groups in their degree of closure; and, each individual can be seen as more or less embedded in their group. The range of possible values of the E-I index is restricted by the number of groups, relative group sizes, and totalnumber of ties in a graph.

Often this range restriction is quite severe, so it is important to re-scale the coefficient torange between the maximum possible degree of "external-ness" (+1) and the maximum possible degree of *internalness.*

2.1.6. Ego networks

Subnetworks that are centered on a certain node is called Ego Networks. In Facebook and LinkedIn these are describes as " Your Network". But we can access only our network. To compare Ego Networks of various people a large dataset is required. Ego networks are derived using Breath-First Search and the depth is limited.

***Network Distance*:** The distance between links/ties. For example a link means Dev is the friend Sharma.Sastri is the friend of friend of Dev which is distance of 2. Rao is the friend of friend of friend of Dev which is having the network distance of 3.

Neighbourhood: It is the collection of ego and all nodes to whom ego has a connection at some path length. In social networkanalysis, the "neighbourhood" is almost always one-step; that is, it includes only ego and actors that are directly adjacent.

N-step neighbourhood: It expands the definition of the size of ego's neighbourhood by including all nodes to whom ego has aconnection at a path length of N, and all the connections among all of these actors. Neighbourhoods of greater path length than 1 (i.e. egos adjacent nodes) are rarely used in social network analysis.

An "out" neighbourhood include all the actors to whom ties are directed from ego. An "in"neighbourhood include all the actors who sent ties directly to ego.

2.1.7. Extracting and visualizing ego networks

Extraction of ego networks is simple as NetworkX network analyser provides a ready-made function to do the job.

```
>>>net.ego_graph(cc,'justinbieber')

<networkx.classes.multigraph.MultiGraph object at 0x1ad54090>
```

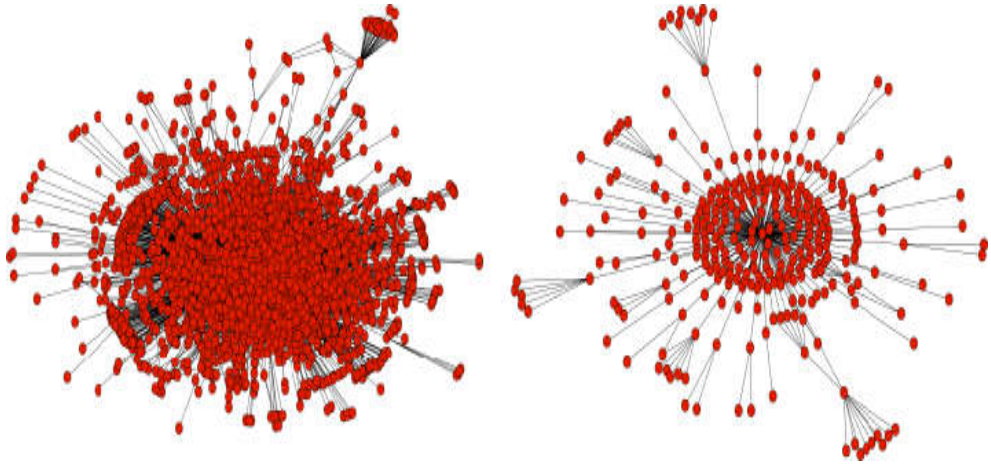


Figure 2.1.3. Justin Bieber is in the Egypt Retweet dataset. His ego network

.The ego_graph function returns a NetworkX graph object, and all the usual metrics such as degree, betweenness can be computed on it. Knowing the size of an ego network is important to understand the reach of the information that a person can transmit.

Clustering coefficient: It is a metric that measures the proportion of friends they are also friends with each other. This metric can be applied to entire networks In ego networks, the interpretation is dense ego networks with a lot of mutual trust have a high clustering coefficient. Star networks with a singlebroadcast node and passive listeners have a low clustering coefficient.

Ego networks in the Egypt data:

we need to convert the ego network from a Multi-graph to a simple Graph

>>>bieb = net.Graph(net.ego_graph(cc,'justinbieber', radius=2))

>>>len(bieb)

22

>>>net.average_clustering(bieb)

0.0

>>>ghonim= net.Graph(net.ego_graph(cc,'Ghonim', radius=2))

>>>len(ghonim)

3450

>>>net.average_clustering(ghonim)

0.22613518489812276

Not only does WaelGhonim have a vastly larger retweet network (despite having 100 times fewer followers then Bieber), his ego network is a network of trust where people retweet messages from him and from each other—a network where a revolutionary message can easily spread and be sustained.

2.1.8. Structural holes

The term "structural holes" is coined by Ronald Buttto refer the positional advantage/disadvantage of individuals that result from how they are embedded in neighbourhoods. Imagine a network of three actors (A, B, and C), in which each is connected to each of the others

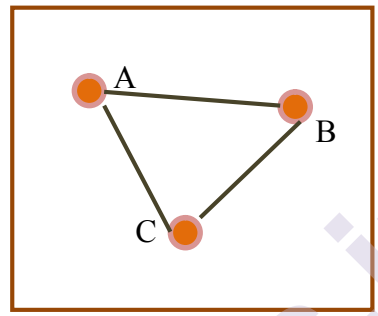


Figure 2.1.4. Three actor network with no structural holes

In the figure 2.1.4., Suppose that actor A in wanted to influence or exchange with another actor. Assume that both B and C may have some interest in interacting or exchanging, as well. Actor A will not be in a strong bargaining position in this network, because both of A's potential exchange partners (B and C) have alternatives to treating with A; they could isolate A, and exchange with one another.

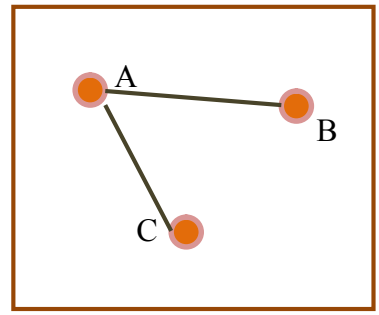


Figure 2.1.5. Three actor network with structural holes

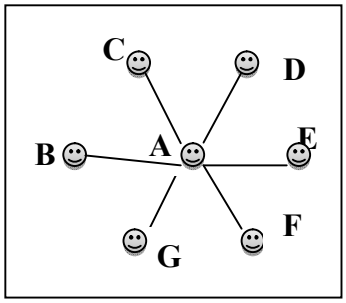
Now imagine that we open a "structural hole" between actors B and C, as in Figure 2.1.5. That is, a relation or tie is "absent" such thatB and C cannot exchange (perhaps they are not aware of one another, or there are very high transaction costs involved in forming atie).

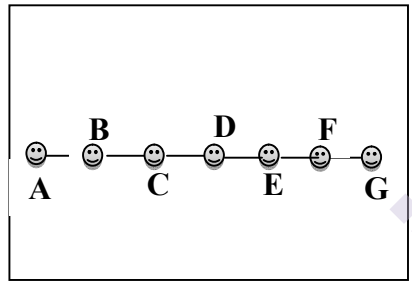
In this situation, actor A has an advantaged position as a direct result of the "structural hole" between actors B and C. Actor A hastwo alternative exchange partners; actors B and C have only one choice, if they choose to

(or must) enter into an exchange.Real networks, of course, usually have more actors.

2.2. CENTRALITY

Actors those who are connected with more number of actors in the network has the maximum centrality. This type of actors are having advantages as acting like bridges, third parties, getting more information, transferring more information and having more alternative paths. The effective measure of an actor's centrality and power potential is their degree.





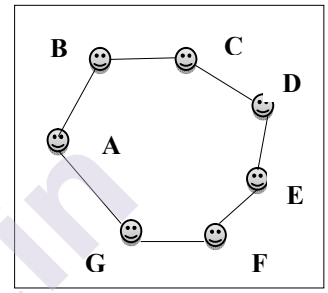


Figure 2.2.1.STAR network Figure 2.2.2. Line Network Figure 2.2.3. Ring Network

2.2.1 Degree of centrality

In the star network given in Figure 2.2.1, actor A has more opportunities and alternatives than other actors. If actor D chooses to not provide A with a resource, the A has a number of other choices to get it; however, if D does not prefer to exchange with A, then D cannot do any exchange.When the number of ties of an actor is more then, the power of actors also will be more. they have. In the star network, Actor A has degree six, all other actors have degree one. This logic underlies measures of centrality and power based on *actor degree*, which we will discuss below. Actors who have more ties have greater opportunities because they have choices. This autonomy makes them lessdependent on any specific other actor, and hence more powerful.

In the line network given in Figure 2.2.2, matters are a bit more complicated. The actors at the end of the line (A and G) are actually at a structural disadvantage, but all others are apparently equal.

Now, consider the ring network given in Figure 2.2.3. in terms of degree. Each actor has exactly the same number of alternative trading partners (or degree), so all positions are equally advantaged or disadvantaged.

Generally, though, actors that are more central to the structure, in the sense of having higher degree or more connections, tend to have favoured positions, and hence more power.

2.2.2. Closeness and betweenness centrality

Actor A is closer to other actors than any other actor. Hence it is powerful compared to all other actors. Power can be applied by direct bargaining and exchange. Actors who are able to reach other actors at shorter path lengths, or who are more reachable by other actors at shorter path lengths have favoured positions. This structural advantage can be translated into power. In the star network, actor A is at a geodesic distance of one from all other actors; each other actor is at a geodesic distance of two from all other actors (but A). This logic of structural advantage underlies approaches that emphasize the distribution of closeness and distance as a source of power.

In actor closeness in ring network each actor lies at different path lengths from the other actors, but all actors have identical distributions of closeness, and again would appear to be equal in terms of their structural positions.

In the line network, the middle actor (D) is closer to all other actors than are the set C, E, the set B, F, and the set A, G. Again, the actors at the ends of the line, or at the periphery, are at a disadvantage.

Betweenness:

The reason that actor A is advantaged in the star network is because actor A lies *between* each other pairs of actors, and no other actors lie between A and other actors. If A wants to contact F, A may simply do so. If F wants to contact B, they must do so by way of A. This gives actor A the capacity to broker contacts among other actors ie to extract "service charges" and to isolate actors or prevent contacts.

The third aspect of a structurally advantaged position then is in being between other actors. In the ring network, each actor lies between each other pair of actors. Actually, there are two pathways connecting each pair of actors, and each third actor lies on one, but not on the other of them. Again, all actors are equally advantaged or disadvantaged. In the line network, our end points (A,G) do not lie between any pairs, and have no brokering power. Actors closer to the middle of the chain lie on more pathways among pairs, and are again in an advantaged position.

2.2.3. Local and global centrality

A central point(Actor) was one which was 'at the centre' of a number of connections, an actor with more direct contacts with other actors. The point centrality is measured as by the degrees of the various points in the graph. The degree centrality is the local centrality measure which counts the number of ties connected by each node and points at individuals who can quickly connect with the network. Since It is a local measure it does not consider rest of the network and the importance of its value depends on the size of the network.

To calculate popular centrality measures – centrality degree, the igraph package of CRAN in R is used.

```
# n = number of nodes/actors, m = the number of ties/edges
> erdos.gr <- sample_gnm(n=10, m=25)
>
> plot(erdos.gr)
> erdos.gr <- sample_gnm(n=15, m=30)
> plot(erdos.gr)
> degree.cent <- centr_degree(erdos.gr, mode = "all")
>
>
> degree.cent$res
 [1] 5 4 3 8 2 4 6 2 5 2 4 3 6 4 2
```

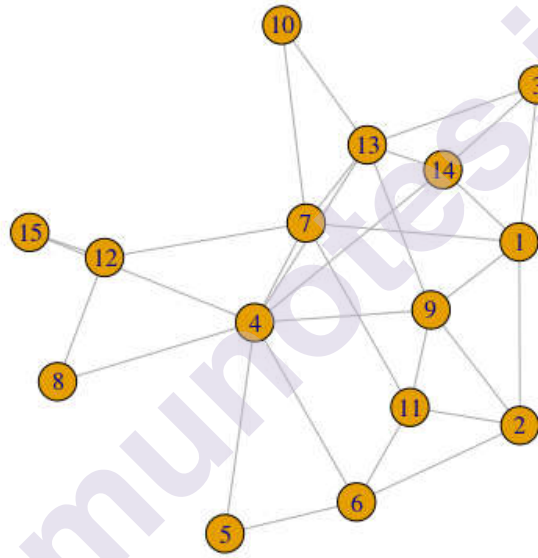


Figure 2.2.4. Undirected graph

From the random generated un-directed graph given in Figure 2.2.4 It is an undirected network, a graph with bidirectional edges in contrast with a directed graph in which the direction of an edge from one vertex to another is considered, with 15 nodes and 30 edges.

The degree centrality of this graph would be calculated using centr.degree() function. The 5 th node has the highest centrality.

Global centrality measures, considers the whole of the network. The most widely used global centrality measures is closeness centrality. This measure scores each node based on their closeness to all other nodes within the network.

It calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths and is useful for finding the individuals who are best placed to influence the entire network most quickly. closeness() function in igraph can be used to findout the global centrality.

2.2.4. Centralization and graph centers

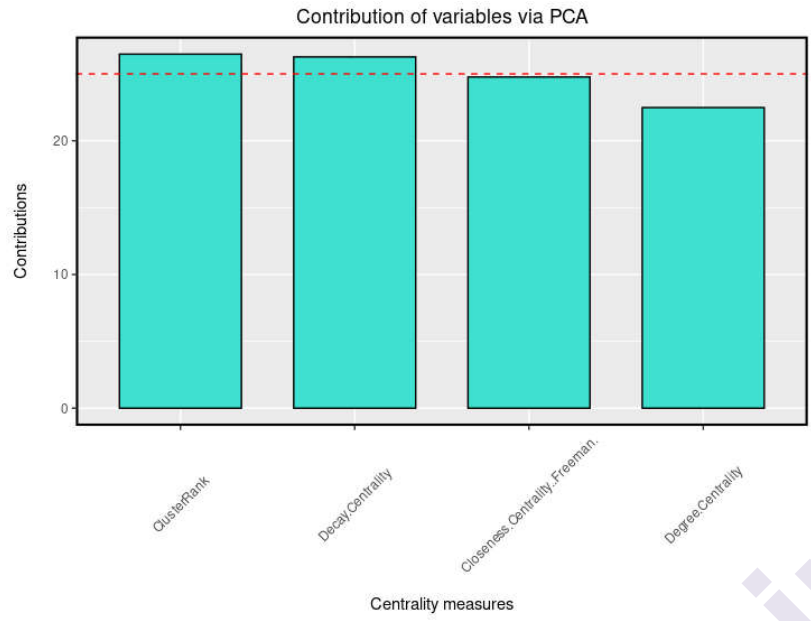
The whole graph centralized structure centralization point can be examined. The concepts of density and centralization refer to differing aspects of the overall 'compactness' of a graph. Density describes the general level of cohesion in a graph; centralization describes the extent **to** which this cohesion is organized around particular focal actors. Centralization and density, therefore, are important complementary measures. The result of centrality calculation will vary according to the centrality type.

```
> library(CINNA)
> data("zachary")
> plot(zachary)
> pr_cent<-proper_centralities(zachary)
 [1] "subgraph centrality scores"
 [2] "Topological Coefficient"
 [3] "Average Distance"
 [4] "Barycenter Centrality"
 [5] "BottleNeck Centrality"
 [6] "Centroid value"
 [7] "Closeness Centrality (Freeman)"
 [8] "ClusterRank"
 [9] "Decay Centrality"
[10] "Degree Centrality"
"…
 …
[42] "Load Centrality"
[43] "Flow Betweenness Centrality"
[44] "Information Centrality"
[45] "Dangalchev Closeness Centrality"
[46] "Group Centrality"
[47] "Harmonic Centrality"
[48] "Local Bridging Centrality"
[49] "Wiener Index Centrality"
```

Figure 2.2.5 Centralities determined through R code

There are various methods suggested by the researchers to find out the graph. In general graph center is calculated using I the closeness centrality. To make more focused and accurate the degree and betweenness centrality is calculated and the graph center actor is determined. There are 49 centralities driven through the R code, it is given in figure 2.2.5. Using sample centrality the result of 4 centralies is given in the figure 2.2.6.This example uses Zachary dataset.

```
>calculate_centralities(zachary, include = pr_cent[7:10])%>%
pca_centralities(scale.unit = TRUE)
```



2.2.5. Notion of importance within network

A degree-based measure of point centrality, therefore, corresponds to the intuitive notion of how well connected a point is within its local environment. Because this is calculated simply in terms of the number of actors to which a particular point is adjacent, ignoring any indirect connections it may have, the degree can be regarded as a measure of local centrality. The simplest notion of closeness is, perhaps, that calculated from the 'sum distance', the sum of the geodesic distances to all other actors in the graph (Sabidussi, 1966). If the matrix of distances between actors in an undirected graph is calculated, the sum distance of a point is its column or row sum in this matrix (the two values are the same). A point with a low sum distance is 'close' to a large number of other actors, and so closeness can be seen as the reciprocal of the sum distance. In a directed graph, of course, paths must be measured through lines which run in the same direction, and, for this reason, calculations based on row and column sums will differ.

2.2.6. Google PageRank algorithm

PageRank centrality is determined through incoming links. PageRank was originally developed for indexing webpages, but can be applied to social networks as well, as long as they are directed graphs

PageRank algorithm

PageRank is an iterative process, otherwise known as an *anytime algorithm. It is* scaled between 0 and 1 and represents the likelihood that a person following links (i.e., traversing the network, “surfing” the web, etc) will arrive at a particular page or encounter a particular person. A 0.5 probability is commonly interpreted as a “50% chance” of an event.

Hence, a PageRank of 0.5 means there is a 50% chance that a person clicking on a random link will be directed to the document with the 0.5

PageRank.: **Simplified version of PageRank Algorithm**

1. Assume a small network of four nodes: (A)lice, (B)ob, (C)arol, and (D)avid.

2. Initially, assign equal probability to A, B, C, and D:

$PR (A) =PR (B) =PC(C) =PR (D) =0.25$.

3. If B, C, and D only link to A, A's PageRank would be computed as

$PR(A)=PR(B)+PR(C)+PR(D)=0.25+0.25+0.25=0.75$.

4. If a page has multiple outgoing links (*outdegree> 1*) then its PageRank contribution is equally divided by all of the link targets.

5. Suppose that page B has a link to page C as well as to page A, while page D has links to all three pages. The value of the link-votes is divided among all the outbound links on a page. Thus, page B gives a vote worth 0.125 to page A and a vote worth 0.125 to page C. Only one third of D's PageRank is counted for A's PageRank (approximately 0.083).

6. In the general case, PageRank can be computed as

$PR(N)=Sum_{i \in nodes} (PR(i)/out_degree(i))$.

7. Repeat calculation of PageRank for all nodes until the values stabilizes

2.3. ANALYSING NETWORK STRUCTURE

A network structure represents a group of nodes or actors or objects and relationships or ties between them. Network is also termed as graph in mathematics. The behaviour of nodes and ties are analysed using various metrics. For example if we study about Twitter users then target users are known as nodes and the followers or followings are considered as relationships or ties. The analysis of this relationships and behavious play vital in many applications. These analysis are carried through different network structures like cliques, clans and factions.

2.3.1. Bottom-up approaches using cliques

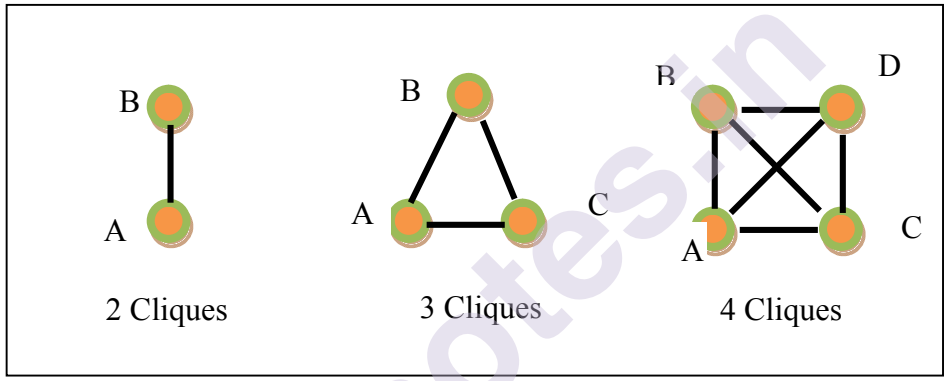
Networks are composed of groups or sub-graphs. Two actors and a tie which connects both actors form a "group." A clique extends the dyad by adding members who are tied to all of the members in the group to it. The size of cliques and clique-like subgroups build a map total network. In this Bottom-up approach, first we focus on individual actor and then analyse

how they are embedded in the web of overlapping groups in the larger structure.

Example: The information on the strength, cost, or probability of relations is available then bottom-up thinking to find maximal groups can be applied.

Cliques

A clique is a sub-set of a network in where the actors of the network are more closely connected to one another compared with the connection with other actors of the network. Friendship network based on age, gender and ideology among human groups are some example for cliques. Dyad is the smallest group. A clique is the maximum number of actors who have all possible ties present among themselves. Different type of cliques are given in figure 2.3.1. The maximum possible actors are included in a "Maximal complete sub-graph".



2.3.2. N-cliques

N-Cliques is defined as a clique with length of the path to connect two sub-cliques N. Sometimes in sub-networks there may be some members are not so tightly or closely connected may exist. This approach is used to find long and complex groupings. Two methods are recommended to relax the constraints of clique definition.to make it general.

Method 1: Define an actor as a member of a clique if they are connected to every other member of the group at a distance at least two. This takes the concept friend of friend. In the given example A, B, C can be connected with E through D. It takes two ties CD and DE to connect with E. Hence the sub-structure is called N- Clique. Here N represents the length of the path to make connection which is two. Figure 2.3.2 shows the structure of N-Clique.

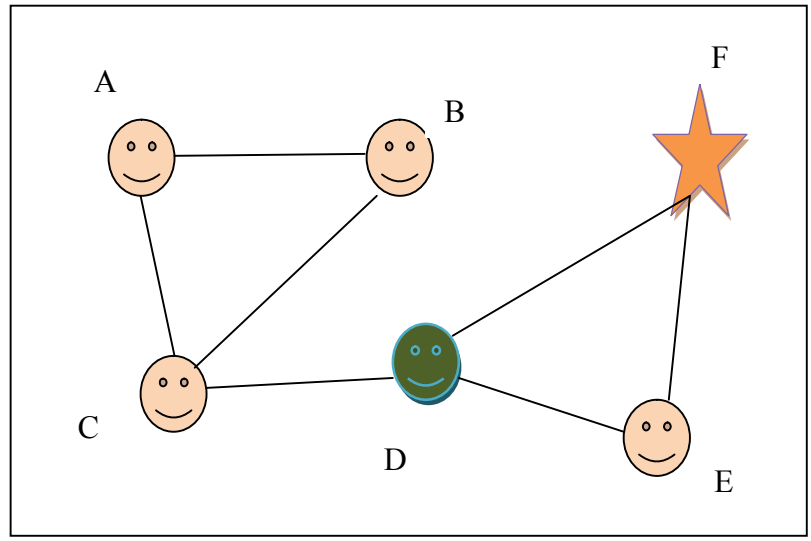
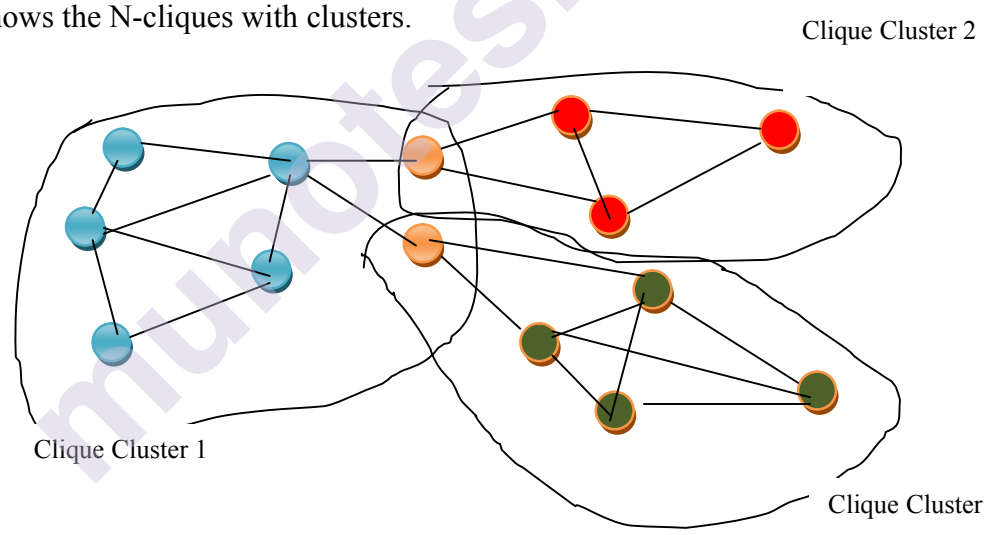


Figure 2.3.2. N-Clique

Method 2: Inner circle of actors for larger groupings also considered as N Cliques. This follows clustering and co-membership concept. Figure 2.3.3 shows the N-cliques with clusters.



Figures 2.3.3. Clustering in Cliques

2.3.3. N-clans

The main issue of N-Cliques is there is a possibility of actors who are not the members of the particular N Cilque may be within the group. In the sociological applications it would create a problematic scenario

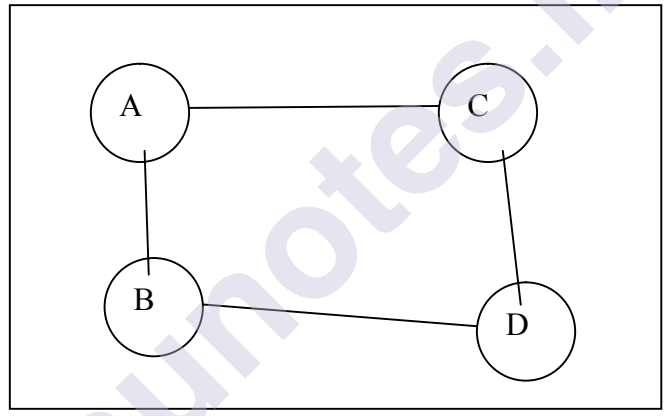
This issue of N-clique is resolve by introducing the additional constraint like it should satisfy that the total span or path distance between any two members of N-clique.

To overcome this problem, some analysts have suggested restricting N-cliques by insisting that the total span or path distance between any two members of an N-clique also satisfy a condition. The additional restriction has the effect of forcing all ties among members of an N-clique to occur by way of other members of the n-clique

The n-clan approach is defined as that all the ties among actors occur through other members of the group.

2.3.4. K-plexes

To relax the Clique constraints it is allowed to consider the actors who have ties with members of cliques except some k members. This concept relaxes the strong assumptions of the "Maximal Complete Sub-Graph". For example, if A has ties with B and C, but not D; while both B and C have ties with D, all four actors could fall in clique under the K-Plex approach. The figure 2.3.4 shows the structure of K-plex. This approach says that a node is a member of a clique of size n if it has direct ties to n-k members of that clique.



Figures 2.3.4. K-Plex

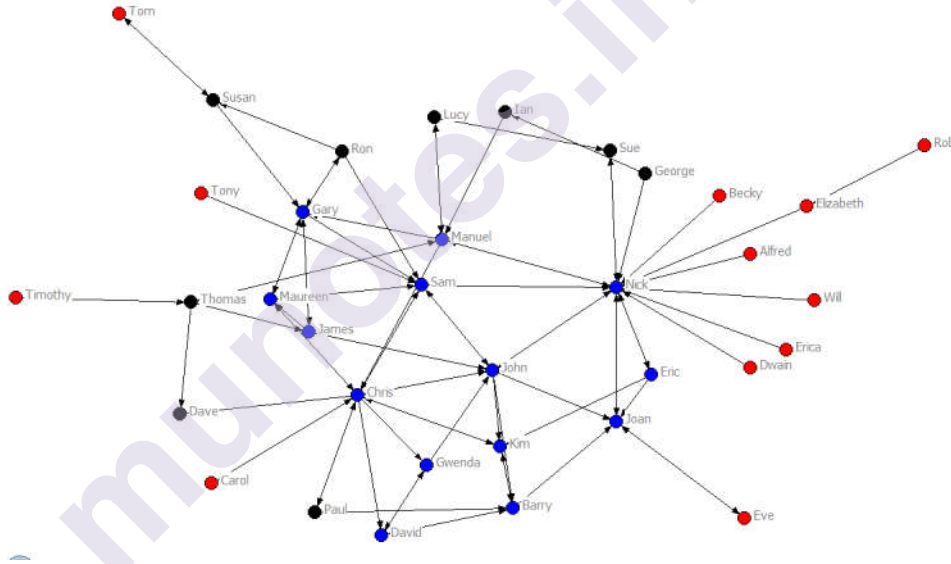
The k-plex approach would seem to have quite a bit in common with the n-clique approach, but k-plex analysis often gives quite a different picture of the sub-structures of a graph. Rather than the large and "stringy" groupings sometimes produced by n-clique analysis, k-plex analysis tends to find relatively large numbers of smaller groupings. This tends to focus attention on overlaps and co-presence (centralization) more than solidarity and reach.

The K-plex method of defining cliques tends to find "overlapping social circles" when compared to the maximal or N-clique method. The k-plex approach to defining sub-structures makes a good deal of sense for many problems. It requires that members of a group have ties to (most) other group members – ties by way of intermediaries (like the n-clique approach) do not quality a node for membership. The picture of group structure that emerges from k-plex approaches can be rather different from that of n-clique analysis. Again, it is not that one is "right" and the other "wrong." Depending on the goals of the analysis, both can yield valuable insights into the sub-structure of groups.

2.3.5. K-cores

K-cores contains a maximal group of actors who are all tied with a number of say, k number of different individuals of the group. The k-core approach is allowing actors to join the group if they are connected to k members, regardless of how many other members they may not be connected to the actor. By varying the value of k (that is, how many members of the group do you have to be connected to), different pictures can emerge. K-cores can be (and usually are) more inclusive than k-plexes. And, as k becomes smaller, group sizes will increase.

K-core analysis helps to identify the parts of a network that are more connected than others. The number of immediate ties of a node is noted as k. A k-core of 1 refers to all nodes that have a degree of 1 or more, ie. all nodes are connected in the network. A k-core of 2 refers to the subset of all nodes that have a degree of 2 or more, etc. Figure 2.3.5. shows three k-cores with three different color nodes..



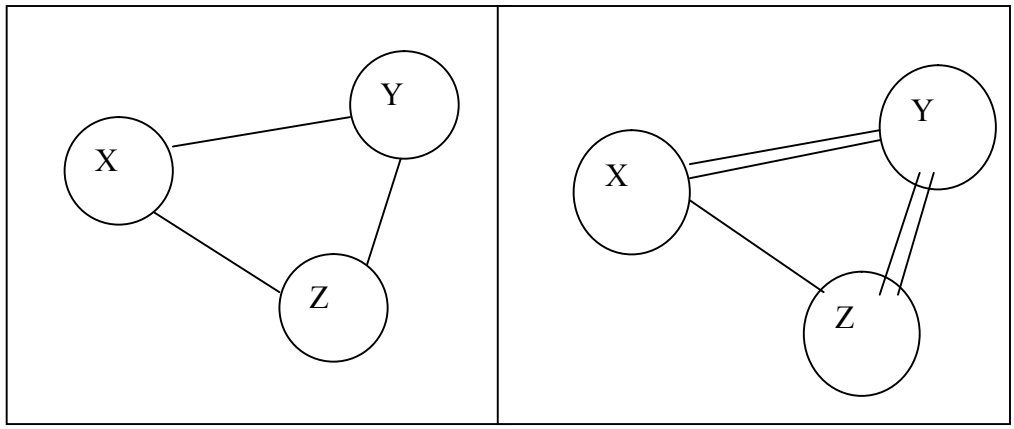
Figures 2.3.5 Three K.Cores

The k-core is used in the applications like, if an actor has ties to a sufficient number of members of a group, they may feel tied to that group, though they may not be knowing many members of the group..

2.3.6. F-groups

F-Groups

F-groups identifies maximal groups made up of "strongly transitive" and "weakly transitive" triads. A strong tie triad is formed when, if there is a tie XY and a tie YZ, there is also a tie XZ that is equal in value to the XY and YZ ties. A weakly transitive triad is formed if the ties XY and YZ are both stronger than the tie XZ, but the tie XZ is greater than some cut-off value.



Figures 2.3.6. F-Groups

2.3.7 Top-down approaches

The overlaps and composition of components make the overall structure of networks. In the bottom-up approach the actors build networks using dynamic processes. Instead if the entire network is considered for the analysis then it leads to top-down approach. Here rather than dyad, the whole structure identifies the substructures as parts that are locally denser than the field as a whole. In a sense, this more macro lens is looking for "holes" or "vulnerabilities" or "weak spots" in the overall structure or solidarity of the network. These holes or weak spots define lines of division or cleavage in the larger group, and point to how it might be de-composed into smaller units. This top-down perspective leads us to think of dynamics that operate at the level of group-selection, and to focus on the constraints under which actors construct networks.

Some methods are used to define the divisions and "weak spots" in a network. The most common approaches are Components, Blocks and cutpoints, Lambda sets and bridges and Factions

2.3.7.1. Components

Components of a graph are sub-graphs that are connected within, but disconnected between sub-graphs. If a graph contains one or more "isolates," these actors are components. More interesting components are those which divide the network into separate parts, and where each part has several actors who are connected to one another.

For directed graphs we can define two different kinds of components. A weak component is a set of nodes that are connected, regardless of the direction of ties. A strong component requires that there be a directed path from A to B in order for the two to be in the same component.

2.3.7.2. Blocks and cut-points

Removing the node from a graph is used to identify the weak point in the graph. When a node is removed from a graph if the graph is divided into separate parts that are dis-connected from each other, then that particular

node or actor is called cut-point. Theses cutpoints act as agents or brokers among the disconnected graphs. The divisions into which cut-points divide a graph are called *blocks* or bicomponent. The maximal non-separable blocks can be found through locating the cutpoints.

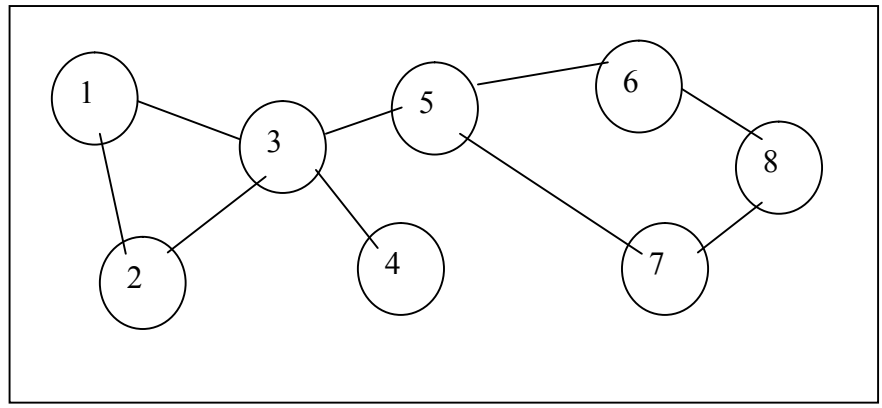


Figure 2.3.7. Blocks and cut-points

In this Figure 2.3.7. If actor 3 is removed then the graph will be separated into 3 blocks with (1, 2), (4), and (5, 6, 7, 8). Here the cutpoint is node 3.

2.3.7.4. Lambda sets and bridges

Removing the connection from a graph is used to separate the graph into blocks. The Lambda set approach ranks each of the relationships in the network in terms of importance by evaluating how much of the flows among actors in the net go through each link. It then identifies sets of relationships which, if disconnected, would most greatly disrupt the flow among all of the actors. The math and computation is rather extreme, though the idea is fairly simple. The lambda set idea has moved us quite far away from the strict components idea. Rather than emphasizing the "decomposition" or separation of the structure into un-connected components, the lambda set idea is a more "continuous" one. It highlights points at which the fabric of connection is most vulnerable to disruption.

2.3.7.5. Factions

Imagine a society in which each person was closely tied to all others in their own subpopulation (that is, all sub-populations are cliques), and there are no connections at all among sub-populations (that is, each sub-population is a component). Most real populations do not look like this, but the "ideal type" of complete connection within and complete disconnection between sub-groups is a useful reference point for assessing the degree of *factionalization* in a population. If we took all the members of each "faction" in this ideal-typical society, and put their rows and columns together in an adjacency matrix, we would see a distinctive pattern of "1-blocks" and "0-blocks." All connections among actors within a faction would be present, all connections between actors in different factions would be absent.

The "Final number of errors" can be used as a measure of the "goodness of fit" of the "blocking" of the matrix. This count (27 in this case) is the sum of the number of zeros within factions plus the number of ones in the non-diagonal blocks (ties between members of different factions, which are supposed to be absent in the ideal type). Since there are 49 total ties in our data, being wrong on the locations of 27 is not a terribly good fit. It is, however, the best we can do with four "factions." The four factions are identified, and we note that two of them are individuals (10, 9), and one is a dyad (3, 6).

The "blocked" or "grouped" adjacency matrix shows a picture of the solution. We can see that there is quite a lot of density "off the main diagonal" where there shouldn't be any. The final panel of the results reports the "block densities" as the number of ties that are present in blocks as proportions of all possible ties. This approach corresponds nicely to the intuitive notion that the groups of a graph can be defined by a combination of local high density, and the presence of "structural holes" between some sets of actors and others. The picture then not only identifies actual or potential factions, but also tells us about the relations among the factions, potential allies and enemies, in some cases.

2.4 SUMMARY

This chapter discusses about understanding networks, Centrality and analysing the network structure.

The understanding network structure explains how the density, reachability, connectivity and reciprocity of a network are calculated. More over the ego networks concept and benefits and demerits of structured holes are given in this chapter. The role of centralization is important in social network analysis. The different centralizations like closeness, betweenness, degree of centralization are used to find the importance of the nodes and the strength of the relationships. Alson with structures the sub-structure in terms of groupings or cliques are used in network analysis. The number, size, and connections among the sub-groupings in a network can tell us a lot about the likely behavior of the network as a whole. All the aspects of sub-group structure can be very relevant to predicting the behavior of the network as a whole. Certain individuals may act as "bridges" among groups, others may be isolates; some actors may be cosmopolitans, and others locals in terms of their group affiliations. Such variation in the ways that individuals are connected to groups or cliques can be quite consequential for their behavior as individuals.

2.5 REFERENCE FOR FURTHER READING

1. Introduction to Social Network Methods: Robert A. Hanneman, Mark Riddle, University of California, 2005, Published in digital form and available at http://faculty.ucr.edu/~hanneman/nettext/index.html.

2. Social Network Analysis for Startups- Finding connections on the social web: Maksim Tsvetovat, Alexander Kouznetsov, O'Reilly Media, 2011.

3. Social Network Analysis- 3rd edition, John Scott, SAGE Publications, 2012.

4. Mark S. Handcock, David Hunter, Carter T. Butts, Steven M. Goodreau and Martina Morris. 2003 statnet: An R package for the Statistical Modeling of Social Networks http://www.csde.washington.edu/statnet

5. Vladimir Batagelj and Andrej Mrvar (2006), Pajek datasets http://vlado.fmf.uni-lj.si/pub/networks/data/.

6. Krackhardt and Stern (1988) developed a very simple and useful measure of the group embedding based on comparing the numbers of ties within groups and between groups

7. Getting Started in Social Network Analysis with NETDRAW, Bruce Cronin University of Greenwich Business School, Occasional Paper 01/15, January 2015 brought to you by CORE View metadata, citation and similar papers at core.ac.uk

8. Structural Holes, The Social Structure of Competition, Ronald S. Burt

9. www.analytictech.com

10. https://www.datacamp.com/

2.6. MODEL QUESTIONS

1. How do you calculate Density and reachability?

2. Compare Connectivity with reciprocity.

3. Explain Ego Networks with a real time application.

4. How do you calculate Centralization and graph centers? Explain.

5. Write the Google PageRAnk Algorithm.

6. How do N-cliques and N-clans "relax" the definition of a clique?

7. Explain about K-plexes and K-cores.

8. How does the idea of a "block" relax the strict definition of a component?

9. Explain the cutpoints with its advantages and disadvantages.

10. Discuss the bottom-up network structures in detail.

3

MEASURES OF SIMILARITY AND STRUCTURAL EQUIVALENCE IN SNA

Unit Structure

3.1 OBJECTIVES

At the end of this unit, student will able to

- Describe network roles and positions, defining equivalence or similarity
- Illustrate the concept of Brute Force and Tabu Search
- Explain the concept of similarity and dissimilarity measures
- Compare and contrast between different types of distances such as Euclidean, Manhattan and squared distances.

3.2 INTRODUCTION

1. The ways that structural analysts look at network data. They look at patterns in the overall structure (e.g. connectedness, density, etc.) and the embeddedness of each actor (e.g. geodesic distances, centrality). A second major way of going about examining network data by looking for "sub-structures," or groupings of actors that are closer to one another than they are to other groupings. For example, we looked at the meaning of "cliques" "blocks" and "bridges" as ways of thinking about and describing how the actors in a network may be divided into sub-groups on the basis of their patterns of relations with one another.

2. The central node of a "star" network is "closer" to all other members than any other member A clique as a "maximal complete sub graph" sounds tough, but, again, is easy to grasp. It is simply the biggest collection of folks who all have connections with everyone else in the group. Again, the idea is not difficult to grasp, because it is really quite concrete: we can see and feel cliques.

3. The patterns of relations among social actors: the analysis of "equivalence classes." Being able to define, theorize about, and analyze data in terms of equivalence is important because we want to be able to make generalizations about social behavior and social structure. That is, we want to be able to state principles that hold for all groups, all organizations, all societies, etc. To do this, we must think about actors not as individual unique persons (which they are), but as examples of categories -- sets of actors who are, in some defined way, "equivalent." As an empirical task, we need to be able to group together actors who are the most similar, and to describe what makes them similar; and, to describe what makes them different, as a category, from members of other categories.

4. Sociological thinking uses abstract categories routinely. "Working class, middle class, upper class" are one such set of categories that describe social positions. "Men and Women" are really labels for categories of persons who are more similar within category than between category -- at least for the purposes of understanding and predicting some aspects of their social behavior. When categories like these are used as parts of sociological theories, they are being used to describe the "social roles" or "social positions" typical of members of the category.

5. Many of the category systems used by sociologists are based on "attributes" of individual actors that are in common across actors. If state that "European-American males, ages 45-64 are likely to have relatively

high incomes" I am talking about a group of people who are demographically similar -- they share certain attributes (maleness, European ancestry, biological age, and income). Structural analysis is not particularly concerned with systems of categories (i.e. variables), that are based on descriptions of similarity of individual attributes (some radical structural analysts would even argue that such categories are not really "sociological" at all). Structural analysts seek to define categories and variables in terms of similarities of the patterns of relations among actors, rather than attributes of actors. That is, the definition of a category, or a "social role" or "social position" depends upon its relationship to another category. Social roles and positions, structural analysts argue, are inherently "relational."

6. What is a "worker?" We could mean a person who does labor (an attribute, actually one shared by all humans). A more sociologically interesting definition was given by Marx as a person who sells control of their labor power to a capitalist. Note that the meaning of "worker" depends upon a capitalist -- and vice versa. It is the relation (in this case, as Marx would say, a relation of exploitation) between occupants of the two role that defines the meaning of the roles.

7. The point is: to the structural analyst, the building blocks of social structure are "social roles" or "social positions." These social roles or positions are defined by regularities in the patterns of relations among actors, not attributes of the actors themselves. We identify and study social roles and positions by studying relations among actors, not by studying attributes of individual actors. Even things that appear to be "attributes of individuals" such as race, religion, and age can be thought of as short-hand labels for patterns of relations. For example, "white" as a social category is really a short-hand way of referring to persons who typically have a common form of relationships with members of another category -- "non-whites." Things that might at first appear to be attributes of individuals are really just ways of saying that an individual falls in a category that has certain patterns of characteristic relationships with members of other categories.

3.3 APPROACHES TO NETWORK POSITIONS AND SOCIAL ROLES

1. Because "positions" or "roles" or "social categories" are defined by "relations" among actors, we can identify and empirically define social positions using network data. In an intuitive way, we would say that two actors have the same "position" or "role" to the extent that their pattern of relationships with other actors is the same. But, there are a couple things about this intuitive definition that are troublesome.

2. First, what relations to we take into account, among whom, in seeking to identify which actors are similar and which are not? The relations that I have with the university (as "Professor") are similar in some ways to the relations that my students have with the university: we are both governed by many of the same rules, practices, and procedures. The relations I have

with the university are very different from those of my students in some ways (e.g. the university pays me, students pay the university). Which relations should count and which ones not, in trying to describe the roles of "professor" and "student?" Indeed, why am I examining relations among my students, me, and the university, instead of including, say, members of the state legislature? There is no simple answer about what the "right relations" are to examine; and, there is no simple answer about who the relevant set of "actors" are. It all depends upon the purposes of our investigation, the theoretical perspective we are using, and the populations to which we would like to be able to generalize our findings. Social network data analytic methods are of little use in answering these conceptual questions.

3. The second problem with our intuitive definition of a "role" or "position" is this: assuming that I have a set of actors and a set of relations that make sense for studying a particular question, what do I mean that actors who share the same position are similar in their pattern of relationships or ties? The idea of "similarity" has to be rather precisely defined. Again, there is no single and clear "right" answer for all purposes of investigation. But, there are rigorous ways of thinking about what it means to be "similar" and there are rigorous ways of actually examining data to define social roles and social positions empirically. These are the issues where there are some ways in which widely used methods can provide guidance.

3.3.1 Defining equivalence or similarity

1. What do we mean when we say that two actors have "similar" patterns of relations, and hence are both members of the same role or social position? Network analysis most broadly defines two nodes (or other more elaborate structures) as similar if they fall in the same "equivalence class." Frankly, that's no immediate help. But it does say that there is something that would cause us to say two actors (or other structures) are members of a "class" that is different from other "classes."

2. Now it becomes a question of what features of an actor's position place them into a "class" with other actors? In what way are they "equivalent?"

3. There are many ways in which actors could be defined as "equivalent" based on their relations with others. For example, we could create two "equivalence classes" of actors with out-degree of zero, and actors with out-degree of more than zero. Indeed, a very large number of the algorithms examined group sets of actors into categories based on some commonality in their positions in graphs.

4. Three particular definitions of "equivalence" have been particularly useful in applying graph theory to the understanding of "social roles" and "structural positions." We will look at these in the next three chapters on "structural equivalence," "automorphic equivalence," and "regular equivalence." Of these, "automorphic" has rarely been used in substantive work.

5. The basic ideas of these three kinds of equivalence are easily explained with three types of equivalence class as structural, automorphic and regular equivalence.

6. Figure given below depicts the Wasserman-Faust Network

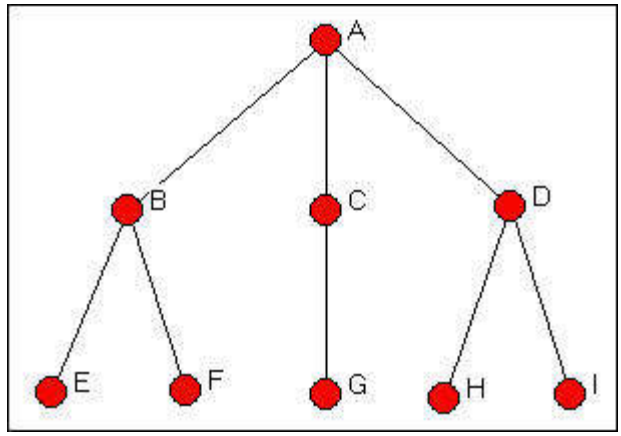


Fig 1 Wasserman-Faust Network

3.3.2 Structural Equivalence

1. In this type of equivalence there if two nodes are said to be exactly equivalent if they have the same relationships to all other nodes. It means nodes follow same pattern as root node.

2. It is very specific as two actors must be exactly substitutable in order to be structurally equivalent.

3. In fig 1 there are seven structural equivalence classes as follows

3.1 There is no actor who has exactly the same set of ties as actor A (ties to B, C, and D), so actor A is in a class by itself.

3.2 The same is true for actors B, C, and D. Each of these actors has a unique set ties to others, so they form three classes, each with one member.

3.3 E and F, however, fall in the same structural equivalence class. Each has a single tie; and that tie is to actor B. Since E and F have exactly the same pattern of ties with all other actors, they are structurally equivalent.

3.4 Actor G, again, is in a class by itself. its profile of ties with the other nodes in the diagram is unique.

3.5 Finally, actors H and I fall in the same structural equivalence class. That is, they have exactly the same pattern of ties to all other actors.

4. Actors that are structurally equivalent are in identical "positions" in the structure of the diagram. Whatever opportunities and constraints operate on one member of a class are also present for the others. The nodes in a structural equivalence class are, in a sense, in the same position with regard to all other actors.

5. Because exact structural equivalence is likely to be rare (particularly in large networks), we often are interested in examining the degree of structural equivalence, rather than the simple presence or absence of exact equivalence.

6. Structural equivalence is the "strongest" form of that network analysts usually consider. If we soften the requirements just a bit, we can often find some interesting other patterns of equivalence.

3.3.3 Automorphic Equivalence

1. The idea of structural equivalence is powerful because it identifies actors that have the same position, or who are completely substitutable. But, even intuitively, you can probably imagine other "less strict" definitions of what it means for two actors to be similar or equivalent.

2. Suppose that the graph in figure 1 described a franchise group of hamburger restaurants. Actor A is the central headquarters, actors B, C, and D are the managers of three different stores. Actors E and F are workers at one store; G is the lone worker at a second store; H and I are workers at the third store.

3. Even though actor B and actor D are not structurally equivalent (they do have the same boss, but not the same workers), they do seem to be "equivalent" in a different sense. Both manager B and D report to a boss (in this case, the same boss), and each has exactly two workers. These are different people, but the two managers seem somehow equivalent. If we swapped them, and also swapped the four workers, all of the distances among all the actors in the graph would be exactly identical. In fact, actors B and D form an "automorphic" equivalence class.

4. In above figure there are actually five automorphic equivalence classes: {A}, {B, D}, {C}, {E, F, H, I}, and {G}. These classes are groupings who's members would remain at the same distance from all other actors if they were swapped, and, members of other classes were also swapped.

5. The idea of automorphic equivalence is that sets of actors can be equivalent by being embedded in local structures that have the same patterns of ties -- "parallel" structures. Large scale populations of social actors (perhaps like hamburger restaurant chains) can display a great deal of this sort of "structural replication." The faces are different, but the structures are identical.

3.3.4 Regular Equivalence

1. Two nodes are said to be regularly equivalent if they have the same profile of ties with members of other sets of actors that are also regularly equivalent. This is a complicated way of saying something that we recognize intuitively.

2. Two mothers, for example, are "equivalent" because each has a certain pattern of ties with a husband, children, and in-laws (for one example -- but one that is very culturally relative). The two mothers do not have ties

to the same husband (usually) or the same children or in-laws. That is, they are not "structurally equivalent." Because different mothers may have different numbers of husbands, children, and in-laws, they will not be automorphically equivalent. But they are similar because they have the same relationships with some member or members of another set of actors (who are themselves regarded as equivalent because of the similarity of their ties to a member of the set "mother").

3. In above figure there are three equivalence classes as first is actor A; the second is composed of the three actors B, C, and D; the third is composed of the remaining five actors E, F, G, H, and I.

4. The easiest class to see is the five actors across the bottom of the diagram (E, F, G, H, and I). These actors are regularly equivalent to one another because a) they have no tie with any actor in the first class (that is, with actor A, each has a tie with an actor in the second class (either B or C or D). Each of the five actors, then, has an identical pattern of ties with actors in the other classes.

5. Actors B, C, and D form a class because they each have a tie with a member of the first class (that is, with actor A) and b) they each have a tie with a member of the third class. B and D actually have ties with two members of the third class, whereas actor C has a tie to only one member of the third class; this doesn't matter, as there is a tie to some member of the third class.

6. Actor A is in a class by itself, defined by a) a tie to at least one member of class two and b) no tie to any member of class three.

7. As with structural and automorphic equivalence, exact regular equivalence may be rare in a large population with many equivalence classes. Approximate regular equivalence can be very meaningful though, because it gets at the notion of which actors fall in which social roles, and how social roles (not role occupants) relate to one another.

3.3.5 Finding Equivalence sets

1. The formal definition says that two actors are regularly equivalent if they have similar patterns of ties to equivalent others. Consider two men. Each has children (though they have different numbers of children, and, obviously have different children). Each has a wife (though again, usually different persons fill this role with respect to each man). Each wife, in turn also has children and a husband (that is, they have ties with one or more members of each of those sets). Each child has ties to one or more members of the set of "husbands" and "wives."

2. What is important in identifying actors is that each “husband” have atleast one tie to a person in the “wife” category and at least one person in the “child” category. That is husband are equivalent to each other because each has similar ties to some member of the sets of wives and children.

3. But there would seem to be a problem with this fairly simple definition. If the definition of each position depends on its relations with other positions , where do we start?

4. There are a number of algorithms that are helpful in identifying regular equivalence sets. UCINET provides some methods that are particularly helpful for locating approximately regularly equivalent actors in valued, multi-relational and directed graphs.

5. Consider the Wasserman-Faust example network. However that this is a picture of order-giving in a simple hierarchy. That is all ties are directed from the top of the diagram and moves towards downwards as shown in below figure, where we will find a regular equivalence characterization of this graph.

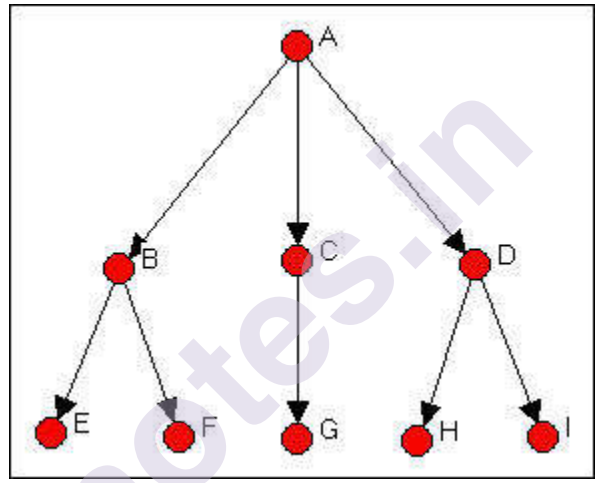


Fig 2 Directed Tie Version of the Wasserman- Faust Network

6. For a first step, characterize each node as either a "source" (an actor that sends ties, but does not receive them), a "repeater" (an actor that both repeats and sends), or a "sink" (an actor that receives ties, but does not send). The source is A; repeaters are B, C, and D; and sinks are E, F, G, H, and I. There is a fourth logical possibility. An "isolate" is a node that neither sends nor receives ties. Isolates form a regular equivalence set in any network, and should be excluded from the regular equivalence analysis of the connected sub-graph.

7. Consider the three "repeaters" B, C, and D. In the neighborhood (that is, adjacent to) actor B are both "sources" and "sinks." The same is true for "repeaters" C and D, even though the three actors may have different numbers of sources and sinks, and these may be different (or the same) specific sources and sinks. We cannot define the "role" of the set {B, C, D} any further, because we have exhausted their neighborhoods.

8. Now consider our "sinks" (i.e. actors E, F, G, H, and I). Each is connected to a source (although the sources may be different). We have already determined, in the current case, that all of these sources (actors B, C, and D) are regularly equivalent. So, E through I are equivalently connected to equivalent others. We are done with our partitioning.

9. The result of {A} {B, C, D} {E, F, G, H, I} satisfies the condition that each actor in each partition have the same pattern of connections to actors in other partitions. The permuted adjacency matrix is shown in table 1

Table 1 Permuted Wasserman-Faust network to show regular equivalence classes

	A	B	C	D	E	F	G	H	I
A	-	1	1	1	0	0	0	0	0
B	0	-	0	0	1	1	0	0	0
C	0	0	-	0	0	0	1	0	0
D	0	0	0	-	0	0	0	1	1
E	0	0	0	0	-	0	0	0	0
F	0	0	0	0	0	-	0	0	0
G	0	0	0	0	0	0	-	0	0
H	0	0	0	0	0	0	0	-	0
I	0	0	0	0	0	0	0	0	-

Here 1 Means there is tie between two nodes and zero means no connection between two nodes and – indicates same node cannot have tie between them

10 Table 2 presents Block image of regular equivalence classes in directed Wasserman-Faust network

	A	B,C,D	E,F,G,H,I
A	-	1	0
B,C,D	0	-	1
E,F,G,H,I	0	0	-

Here {A} sends to one or more of {BCD} but to none of {EFGHI}. {BCD} does not send to {A}, but each of {BCD} sends to atleast one of {EFGHI}. None of {EFGHI} send to any of {A} or of {BCD}.

11. For directed binary graphs, the neighborhood search method applied in wasserman-Faust network works quite well. For binary graphs that are not directed, usually the geodesic distance among actors is computed and used instead of raw adjacency. For graphs with valued relations (Strength, cost, probability) a method for identifying approximate regular equivalence was developed by white and Reitz.

3.3.6 Brute Force and Tabu Search

1.With binary data, numerical algorithms are used to search for classes of actors that satisfy the mathematical definitions of automorphic equivalence. When the new graph and the old graph have the same distances among nodes, the graphs are isomorphic and the swapping that done identifies the isomorphic sub-graphs.

2. One approach to binary data, "all permutations" (Netowrk->Roles & Positions>Automorphic>All Permutations) literally compares every possible swapping of nodes to find isomorphic graphs with even a small graph i.e method is nothing but a brute force method. An alternative approach with the same intent ("optimization by tabu search")(Network>Roles & Positions>Exact>optimization) can much more quickly sort nodes in to a user-defined number of partitions in such a way as to maximize automorphic equivalence.

3. When we have measures of the strength, cost or probability of relations among nodes(i.e valued data), exact automorphic equivalence is far less likely. It is possible however, to identify classes of approximately equivalent actors on the basis of their profile of distance to all other actors. The "equivalence of distances" method (Network>Roles & Positions> Automorphic>Maxsim) Produces measures of the degree of automorphic equivalence for each pair of nodes, which can be examined by clustering and scaling methods to identify approximate classes.

4. Brute Force- All Permutations

4.1 The automorphisms in a graph can be identified by the brute force method of examining every possible permutation of the graph. With a small graph, and a fast computer, this is a useful thing to do. Basically, every possible permutation of the graph is examined to see if it has the same tie structure as the original graph. For graphs of more than a few actors, the number of permutations that need to be compared becomes extremely large.

4.2 Let's use Networks>Roles& Positions>Automorphic>All Permutations to search the Wasserman-Faust Network shown in figure below

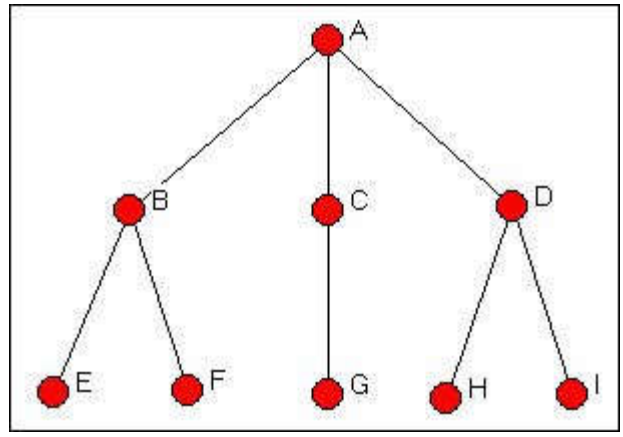


Figure 3 Wasserman-Faust Network for Automorphic Permutations

Here there are five types of orbits as 1) Orbit 1 consist of only one Node that is A 2) Orbit 2 consist of two nodes as B and D as they has the same pattern of distance i.e they comprising of two leaves node as E,F, H,I. 3) Orbit 3 comprising of Node c 4) Orbit 4 Comprising of 4 nodes as E,F,H,I 5) Orbit 5 comprising of Node G . Note that automorphism classes identify groups of actors who have the same pattern of distance from other actors, rather than sub-structures as in case of Node B and D.

5. Tabu Search- Optimization

5.1 For larger graphs, direct search for all equivalences is impractical both because it is computationally intensive and because exactly equivalent actors are likely to be rare.

5.2 Network>Roles & Positions >Exact>optimization provides a numerical tool for finding the best approximations to a user-selected number of automorphism classes. In using this method, it is important to explore a range of possible number of partitions to determine how many partitions are useful in order to re-run the algorithm a number of times to insure that a global, rather than local minimum has been found.

5.3 The method begins by randomly allocating nodes to partitions. A measure of badness of fit is calculated as the sum of squares for each row and each column within each block, along with calculating the variance of these sums of squares. Then Sum of variance is calculated across the block to construct a measure of badness of fit. Search continues to find allocation of actors to partitions that minimizes this badness of fit statistic.

5.4 Here we are using the Knoke bureaucracies information exchange network data for calculations of automorphisms. In the Knoke information data there are no exact automorphisms.

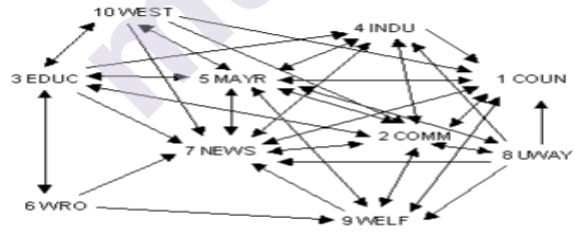


Fig 4 Knoke Information Data Network

Here for this fit of automorphic equivalence models is given in terms of mean value as

Partition	Fit
2	4.366
3	4.054
4	3.912

5	3.504
6	3.328

So in automorphic equivalence between nodes is actually operates on the profile of distance between nodes or actors, Here in above table Partition 2 and 3 are equidistant as they are having mean value as 4.366 and 4.054.

3.3.7 Equivalence of Distances

1 Maxsim

1. When we have information on the strength, cost, or probability of relations (i.e. valued data), exact automorphic equivalence could be expected to be extremely rare. But, since automorphic equivalence emphasizes the similarity in the profile of distances of actors from others, the idea of approximate equivalence can be applied to valued data.

2. Network>Roles & Positions>Automorphic>MaxSim generates a matrix of "similarity" between shape of the distributions of ties of actors that can be grouped by clustering and scaling into approximate classes. The approach can also be applied to binary data, if we first convert the adjacency matrix into a matrix of geodesic near-ness (which can be treated as a valued measure of the strength of ties).

3. The algorithm begins with a (reciprocal of) distance or strength of tie matrix. The distances of each actor re-organized into a sorted list from low to high, and the Euclidean distance is used to calculate the dissimilarity between the distance profiles of each pair of actors.

4. The algorithm scores actors who have similar distance profiles as more automorphically equivalent. Again, the focus is on whether actor u has a similar set of distances, regardless of which distances, to actor v. Again, dimensional scaling or clustering of the distances can be used to identify sets of approximately automorphically equivalent actors.

5. Example -Line Network

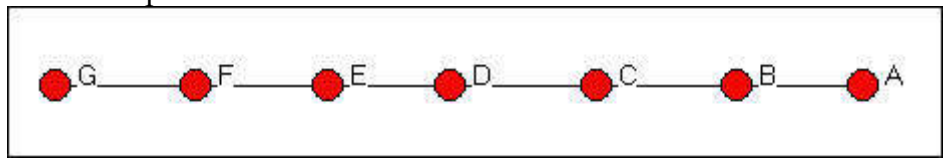


Fig 5 Line Network Graph

Here automorphic equivalence of geodesic distances in the line network is given as shown in below figure

NOTE: Binary adjacency matrix converted to reciprocals of geodesic distances.

Distances Among Actors

		1 A	2 B	3 C	4 D	5 E	6 F	7 G
1	A	0.00	4.56	5.13	5.27	5.13	4.56	0.00
2	B	4.56	0.00	1.64	1.94	1.64	0.00	4.56
3	C	5.13	1.64	0.00	0.71	0.00	1.64	5.13
4	D	5.27	1.94	0.71	0.00	0.71	1.94	5.27
5	E	5.13	1.64	0.00	0.71	0.00	1.64	5.13
6	F	4.56	0.00	1.64	1.94	1.64	0.00	4.56
7	G	0.00	4.56	5.13	5.27	5.13	4.56	0.00

HIERARCHICAL CLUSTERING OF (NON-)EQUIVALENCE MATRIX

	D	E	C	B	F	A	G
Level	4	5	3	2	6	1	7
0.000	.	XXX	XXX	XXX			
0.707	XXXXX	XXX	XXX				
1.714	XXXXXXXXX	XXX					
4.784	XXXXXXXXXXXXX						

Fig 6 Binary Adjacency Matrix converted to reciprocals of geodesic distances

Here in above first the matrix is converted in to a geodesic distance matrix where distance between actors is given. So according to matrix first the actor 3 and actor 5 has same geodesic distance of 5.13, second the actor 2 and 6 has of 4.56, third 1 and 7 has same geodesic distance and cluster out actor is 4 whose distance is not matching with any of these actors. Then at last step Euclidean distance between these lists is calculated as a measure of the non-automorphic-equivalence and hierarchical clustering is applied.

6. Example 2- Maxsim method has applied on donors data of California political campaigns, where strength of ties is measured among the actors with the number of positions in campaigns they have in common when either contributed.

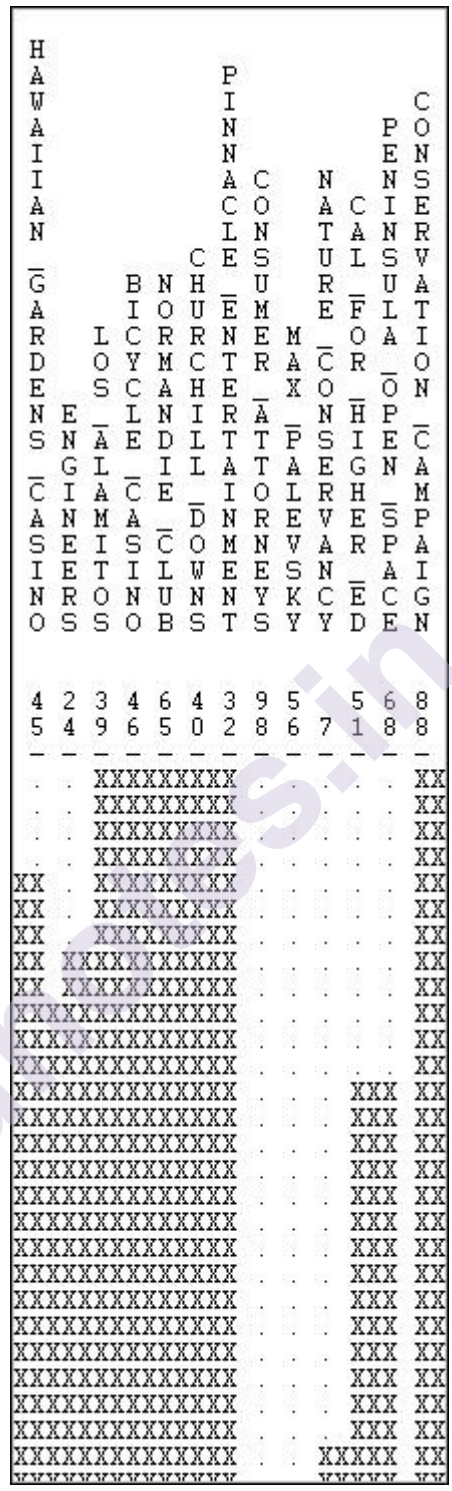


Fig 6 Truncated California automorphic equivalence

Here the above the figure represents only small part of large data set which shows that a number of non-Indian casinos and race-tracks cluster together and separately from some other donors who are primarily concerned with education and ecological issues.

7. The identification of approximate equivalence classes in valued data can be helpful in locating groups of actors who have a similar location in

the structure of the graph as a whole. By emphasizing distance profiles, however, it is possible to finds classes of actors that include nodes that are quite distant from one another, but at a similar distance to all the other actors. That is, actors that have similar positions in the network as a whole.

3.4.1 Valued Relations

1 Pearson Correlations, Covariance and Cross-Products

1. Valued Relations- a)A common approach for indexing the similarity of two valued variables is the degree of linear association between the two. Exactly the same approach can be applied to the vectors that describe the relationship strengths of two actors to all other actors.b)As with any measures of linear association, linearity is a key assumption. It is often wise, even when data are at the interval level (e.g. volume of trade from one nation to all others) to consider measures with weaker assumptions (like measures of association designed for ordinal variables).

2. Pearson, Correlation

2.1 The correlation measure of similarity is particularly useful when the data on ties are "valued," that is, tell us about the strength and direction of association, rather than simple presence or absence.

2.2 Pearson correlations range from -1.00 (meaning that the two actors have exactly the opposite ties to each other actor), through zero (meaning that knowing one actor's tie to a third party doesn't help us at all in guessing what the other actor's tie to the third party might be), to +1.00 (meaning that the two actors always have exactly the same tie to other actors - perfect structural equivalence).

2.3 Pearson correlations are often used to summarize pair-wise structural equivalence because the statistic (called "little r") is widely used in social statistics. If the data on ties are truly nominal, or if density is very high or very low, correlations can sometimes be a little troublesome, and matches (see below) should also be examined.

2.4 Figure shown below the the correlations of the ten Knoke organization's profiles of in and out information ties. We are applying correlation, even though the Knoke data are binary. The UCINET algorithm Tools >Similarities will calculate correlations for rows or columns.

	1	2	3	4	5	6	7	8	9	10
1	1.000	0.447	-0.000	0.775	0.293	0.258	0.467	0.775	1.000	0.500
2	0.447	1.000	-0.447	0.447	0.655	0.293	0.333	0.745	0.333	0.378
3	-0.000	-0.447	1.000	0.258	-0.293	-0.149	0.600	-0.333	0.447	0.258
4	0.775	0.447	0.258	1.000	0.293	-0.258	0.745	0.775	0.775	0.775
5	0.293	0.655	-0.293	0.293	1.000	0.000	0.218	0.488	0.218	0.378
6	0.258	0.293	-0.149	-0.258	0.000	1.000	-0.447	-0.149	0.149	0.067
7	0.467	0.333	0.600	0.745	0.218	-0.447	1.000	0.600	0.745	0.258
8	0.775	0.745	-0.333	0.775	0.488	-0.149	0.600	1.000	0.600	0.149
9	1.000	0.333	0.447	0.775	0.218	0.149	0.745	0.600	1.000	0.600
10	0.500	0.378	0.258	0.775	0.378	0.067	0.258	0.149	0.600	1.000

Fig 7 Pearson Correlations of rows for Knoke Information Network

2.5 We can see, for example, that node 1 and node 9 have identical patterns of ties; there is a moderately strong tendency for actor 6 to have ties to actors that actor 7 does not, and vice versa.

2.6 The Pearson correlation measure does not pay attention to the overall prevalence of ties (the mean of the row or column), and it does not pay attention to differences between actors in the variances of their ties. Often this is desirable - to focus only on the pattern, rather than the mean and variance as aspects of similarity between actors.

2.7 Covariance Matrix- we might want our measure of similarity to reflect not only the pattern of ties, but also differences among actors in their overall tie density. Tools>similarities will also calculate the covariance matrix.

2.8 Cross product- If we want to include differences in variances across actors as aspects of (dis)similarity, as well as means, the c*ross-product* ratio calculated in Tools>Similarities might be used.

2.Euclidean, Manhattan and Squared Distances

1. An alternative approach to linear correlation(and its relatives) is to measure the "distance" or "dissimilarity" between the tie profiles of each pair of actors. Several "distance" measures are fairly commonly used in network analysis particularly the Euclidean distance or squared Euclidean distance.

2. These measures are not sensitive to the linearity of association and can be used with either valued or binary data.

3. Figure below shows the Euclidean distances among the Knoke organizations calculated using Tools>Dissimilarities and Distances>Std Vector Dissimilarities/distances

	1	2	3	4	5	6	7	8	9	10
1	0	2	2	1	2	2	1	1	0	1
2	2	0	2	2	1	2	2	1	2	2
3	2	2	0	2	2	2	1	2	2	2
4	1	2	2	0	2	2	1	1	1	1
5	2	1	2	2	0	2	2	1	2	2
6	2	2	2	2	2	0	2	2	2	2
7	1	2	1	1	2	2	0	1	1	2
8	1	1	2	1	1	2	1	0	1	2
9	0	2	2	1	2	2	1	1	0	1
10	1	2	2	1	2	2	2	2	1	0

Fig 8 Euclidean distance in sending for Knoke information network

4. The Euclidean distance between two vectors is equal to the square root of the sum of the squared differences between them. That is the distance between Actor A and Actor C is subtracted from the distance of actor B to Actor C, then their difference is squared. This is then repeated across all

the other actors (D,E,F etc) and summed. The square root of the sum is then taken.

5. A closely related measure is the “Manhattan” or block distance between the two vectors. This distance is simply the sum of the absolute difference between the actors ties to each alter, summed across the alters.

3.4.2 Understanding clustering-agglomerative and divisive clusters

1. Agglomerative hierarchical clustering of nodes on the basis of the similarity of their profiles of ties to other cases provides a "joining tree" or "dendogram" that visualizes the degree of similarity among cases - and can be used to find approximate equivalence classes.

2.Tools>Cluster>Hierarchical proceeds proceeds by initially placing each case in its own cluster. The two most similar cases (those with the highest measured similarity index) are then combined into a class. The similarity of this new class to all others is then computed on the basis of one of three methods.

2.On the basis of the newly computed similarity matrix, the joining/recalculation process is repeated until all cases are "agglomerated" into a single cluster. The "hierarchical" part of the method's name refers to the fact that once a case has been joined into a cluster, it is never re-classified. This results in clusters of increasing size that always enclose smaller clusters.

3. The "Average" method computes the similarity of the average scores in the newly formed cluster to all other clusters; the "Single-Link" method (a.k.a. "nearest neighbor") computes the similarities on the basis of the similarity of the member of the new cluster that is most similar to each other case not in the cluster.

4. The "Complete-Link" method (a.k.a. "farthest neighbor") computes similarities between the member of the new cluster that is least similar to each other case not in the cluster. The default method is to use the cluster average; single-link methods will tend to give long-stringy joining diagrams; complete-link methods will tend to give highly separated joining diagrams.

5.The hamming distance in information sending in the Knoke network was computed and the results were stored as a file. This file was then input to Tools>cluster>Hierarchical. The “Average” method was to be used, and that the data were “dissimilarities”. The results are shown in figure below

```
************************************************************
Matrix KNOKI in dataset Dissim-KNOKBUR
************************************************************

HIERARCHICAL CLUSTERING

                                 1
 Level    6 3 2 5 8 4 7 1 9 0
------    - - - - - - - - - -
0.0000     .   .   .   .   .   .   XXX .
1.0000     .   .   XXX .   XXX XXX .
1.2222     .   .   XXX .   XXXXXXX .
1.6667     .   .   XXXXX XXXXXXX .
2.1111     .   .   XXXXX XXXXXXXXX
2.8815     .   .   XXXXXXXXXXXXXXX
3.3091     .   XXXXXXXXXXXXXXXXX
4.2158     XXXXXXXXXXXXXXXXXXX


Measures of cluster adequacy

                     1       2       3       4       5       6       7       8
                ------  ------  ------  ------  ------  ------  ------  ------
  1      Eta   -0.308 -0.422 -0.578 -0.634 -0.687 -0.547 -0.487
  2        Q   -0.118 -0.143 -0.159 -0.173 -0.171 -0.074 -0.044
  3 Q-prime   -0.132 -0.167 -0.190 -0.216 -0.228 -0.111 -0.087
  4      E-I    1.000  0.970  0.894  0.848  0.727  0.015 -0.409


Size of each cluster, expressed as a proportion of the total population clustered

               1     2     3     4     5     6     7     8
           ----- ----- ----- ----- ----- ----- ----- -----
  1  CL1  0.200 0.200 0.400 0.400 0.500 0.800 0.900 1.000
  2  CL2  0.100 0.200 0.200 0.300 0.300 0.100 0.100
  3  CL3  0.100 0.100 0.100 0.100 0.100 0.100
  4  CL4  0.100 0.200 0.100 0.100 0.100
  5  CL5  0.100 0.100 0.100 0.100
  6  CL6  0.100 0.100 0.100
  7  CL7  0.100 0.100
  8  CL8  0.100
  9  CL9  0.100
 10 CL10
```

Fig 9 Clustering of Hamming distances of information sending in the Knoke Network

6. The first graphic shows that nodes 1 and 9 were the most similar and joined first. The graphic by the way can be rendered as a more polished dendrogram using Tools>dendrogram>draw on data saved from the cluster tool. At the next step, there are three clusters(cases 2 and5, 4 and 7 and 1 and 9).

7. The joining continues until (at the 8th step) all cases are agglomerated in to a single cluster. This gives a clear picture of the similarity of cases and the groupings or classes of cases. But there are really eight pictures here (one for each step of the joining). Which is the “right” Solution?.

8. Again, there is no single answer. Theory and a substantive knowledge of the processes giving rise to the data are the best guide. The second panel "Measures of cluster adequacy" can be of some assistance. There are a number of indexes here, and most will (usually) give the similar answers.

9. As we move from the right (higher steps or amounts of agglomeration) to the left (more clusters, less agglomeration) fit improves. The E-I index is often most helpful, as it measures the ratio of the numbers of ties within the clusters to ties between clusters. Generally, the goal is to achieve classes that are highly similar within, and quite distinct without. Here, one might be most tempted by the solution of the 5th step of the process (clusters of 2+5, 4+7+1+9, and the others being single-item clusters).

10. To be meaningful, clusters should also contain a reasonable percentage of the cases. The last panel shows information on the relative sizes of the clusters at each stage.

3.4.3 Binary Relations

1 Matches: Exact, Jaccard, Hamming

1. If the information that we have about the ties among our actors is binary, correlation and distance measures can be used, but may not be optimal. For data that are binary, it is more common to look at the vectors of two actor's ties, and see how closely the entries in one "match" the entries in the other.

2. Matches: Exact

2.1 Figure below shows the result for the columns relation of the Knoke bureaucracies

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1 COUN	1.000	0.625	0.625	0.625	0.625	0.250	0.625	0.750	0.625	0.500
2 COMM	0.625	1.000	0.250	0.625	1.000	0.125	0.875	0.250	0.375	0.375
3 EDUC	0.625	0.250	1.000	0.500	0.250	0.625	0.500	0.750	0.625	0.750
4 INDU	0.625	0.625	0.500	1.000	0.625	0.500	0.500	0.750	0.500	0.625
5 MAYR	0.625	1.000	0.250	0.625	1.000	0.125	0.875	0.250	0.375	0.250
6 WRO	0.250	0.125	0.625	0.500	0.125	1.000	0.125	0.625	0.375	0.875
7 NEWS	0.625	0.875	0.500	0.500	0.875	0.125	1.000	0.250	0.625	0.250
8 UWAY	0.750	0.250	0.750	0.750	0.250	0.625	0.250	1.000	0.750	0.750
9 WELF	0.625	0.375	0.625	0.500	0.375	0.375	0.625	0.750	1.000	0.375
10 WEST	0.500	0.375	0.750	0.625	0.250	0.875	0.250	0.750	0.375	1.000

Fig 10 Proportion of Matches for Knoke Information receiving

2.2 These results show similarity in a way that is quite easy to interpret. The number 0.625 in the cell 2,1 means that in comparing actor no 1 and actor no 2, they have the same tie(present or absent) to other actors 62.5% of the time. The measure is particularly useful with multi-category nominal measures of ties, it also provides a nice scaling for binary data.

2.3 In some networks connections are very sparse. Indeed, if one were looking at ties of personal acquaintance in very large organizations, the data might have very low density. Where density is very low, the "matches" "correlation" and "distance" measures can all show relatively little variation among the actors, and may cause difficulty in discerning structural equivalence sets (of course, in very large, low density networks, there may really be very low levels of structural equivalence).

3. Jaccard

3.1 One approach to solve problem of matches and coefficient is to use jaccard method which states that to calculate the number of times that both actors report a tie (or the same type of tie) to the same third actors as a percentage of the total number of ties reported. That is, we ignore cases

where neither X or Y are tied to Z, and ask, of the total ties that are present, what percentage are in common.

Percent of Positive Matches (Jaccard coefficients)

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	1.00									
2	0.54	1.00								
3	0.46	0.31	1.00							
4	0.60	0.54	0.42	1.00						
5	0.50	0.93	0.38	0.50	1.00					
6	0.18	0.27	0.11	0.18	0.25	1.00				
7	0.58	0.64	0.54	0.55	0.60	0.08	1.00			
8	0.67	0.46	0.50	0.67	0.43	0.20	0.38	1.00		
9	0.67	0.36	0.50	0.55	0.33	0.11	0.64	0.56	1.00	
10	0.40	0.43	0.44	0.60	0.36	0.38	0.31	0.50	0.36	1.00

Fig 11 Jaccard Coefficient for information receiving profiles in Knoke network

3.2 Again the same basic picture emerges. The uniqueness of actor no 6, though is emphasized. Actor six is more unique by this measure because of the relatively small number of total ties that it has -- this results in a lower level of similarity when "joint absence" of ties are ignored. Where data are sparse, and where there are very substantial differences in the degrees of points, the positive match coefficient is a good choice for binary or nominal data.

4.Hamming Distance

4.1 The hamming distance is the number of entries in the vector for one actor that would need to be changed in order to make it identical to the vector of the other actor. These differences could be either adding or dropping a tie, so the Hamming distance treats joint absence as similarity.

		1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	COUN	0	3	3	3	3	6	3	2	3	4
2	COMM	3	0	6	3	0	7	1	6	5	5
3	EDUC	3	6	0	4	6	3	4	2	3	2
4	INDU	3	3	4	0	3	4	4	2	4	3
5	MAYR	3	0	6	3	0	7	1	6	5	6
6	WRO	6	7	3	4	7	0	7	3	5	1
7	NEWS	3	1	4	4	1	7	0	6	3	6
8	UWAY	2	6	2	2	6	3	6	0	2	2
9	WELF	3	5	3	4	5	5	3	2	0	5
10	WEST	4	5	2	3	6	1	6	2	5	0

Fig 12 Hamming distance of information receiving in Knoke Network

Summary

In this section we studied about various methods described above that are used in social network analysis to find out the strength between two ties or nodes in form of geodesic distance, regular equivalence, structural

equivalence, automorphic equivalence, Valued relations and Binary relations and how the distance between them is measured using pearson correlation, covariance, agglomerative clustering, exact, jaccard and hamming distances.

References

[1] “Introduction to Social Network Methods” by Robert A. Hanneman University of California

Questions

Q1.How are network roles and social roles different from network "sub-structures" as ways of describing social networks?

Q2. Explain the differences among structural, automorphic, and regular equivalence.

Q3. Actors who are structurally equivalent have the same patterns of ties to the same other actors. How do correlation, distance, and match measures index this kind of equivalence or similarity?

Q4. If the adjacency matrix for a network can be blocked into perfect sets of structurally equivalent actors, all blocks will be filled with zeros or with ones. Why is this?

Q5. If two actors have identical geodesic distances to all other actors, they are (probably) automorphically equivalent. Why does having identical distances to all other actors make actors "substitutable" but not necessarily structurally equivalent?

Q6. Regularly equivalent actors have the same pattern of ties to the same kinds of other actors -- but not necessarily the same distances to all other actors, or ties to the same other actors. Why is this kind of equivalence particularly important in sociological analysis?

4

TWO-MODE NETWORKS FOR SNA

Unit Structure

4.0. OBJECTIVES

After going through this unit, you will be able to:

- Explicate Two-mode Networks and its applications
- Comprehend Bi-partite data structure
- Compare the applications of SVD , factor and correspondence analysis
- Analyse the methods of qualitative analysis
- Describe the importance of affiliation and attribute networks

4.1. UNDERSTANDING TWO-MODE NETWORKS

Nowadays, more data in the network are in 2-mode structure. This means that it represents two different types of actors and ties to define the connections between the one group of actors with other group of actors. This two-mode data network analyses the Macro-Micro relationships between the actors.

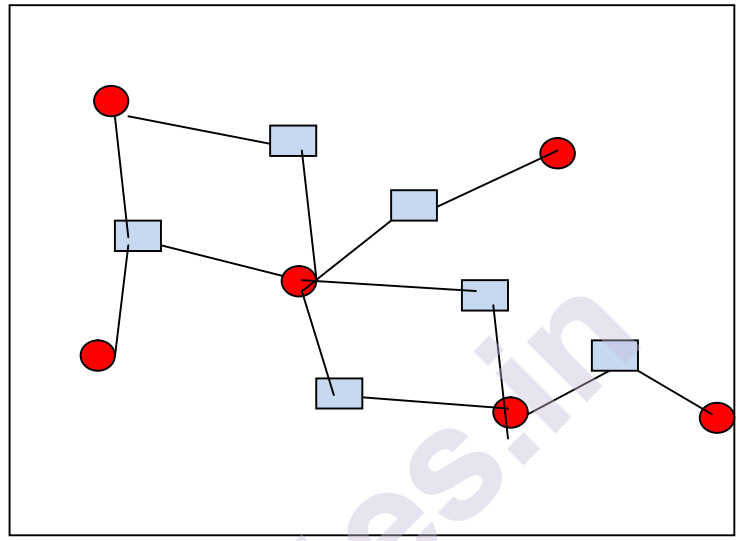


Figure 4.1.1.1.

In figure 4.1.1.1. two types of actors, one set of actors represented by circle with red colour and another set of actors by rectangle with blue colour are connected through the ties. In this the red circles belong to one group of actors and the blue rectangles belong to another group of actors.

Among the two types of actors one is macro actor, who plays major role in the society and having relationships with themselves. The other is micro actors, who plays the roles with the macro actors and in certain occasions they are as well interconnected with themselves. These macro and micro actors establish the ties between them. This structure is termed as two-mode network.

The table 4.1.1.1. represents the matrix form of ***Davis data*** (Davis et al., Homans 1950). This data is collected by the author over nine-month period by closely watching and observing the social activities of 18 women in Southern women's club.

During that period, various subsets of these women had met in a series of 14 informal social events. This data shows the list of events E1..E14 attended by the women given in the data table. The women attended the various activities like going to a store, attending a meeting of a club, a church supper, a party, a meeting of an association etc.

Table 4.1.1.1. Davis Southern Women's' Matrix data

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	1	1
DOROTHY	0	0	0	0	0	0	0	1	1	1	0	1	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

From the data choices of parties attended by the women as macro structures may affect the choices of the individual women. These types of data are two-mode data. The women with one set and the activities with another set, how the women are tied up with activities can be derived from the data. This type of activity is also known as macro-micro activity.

In this chapter, the concepts are explained through data which describe the contributions of a small number of large donors to campaigns supporting and opposing ballot initiatives in California during the period 2000 to 2004. There are 44 members are taken for data analysis. There are 44 initiatives for the donors. Hence the data set hastowo modes such as i) Donors ii) Initiatives

Two different forms of the data are used.

i) Valued Data

The relations between donors and initiatives using a simple ordinal scale are described by this type of data. The following table shows the details about the actor code.

Action	**Actor code**
Contribution towards opposing a particular initiative	-1
No Contribution	0
Contribution towards in support of the initiative	+1

ii) Binary Data

The binary data specifies the contribution in the campaign on each initiative with binary values as given in the table.

Donor Contribution	**Actor code**
Contributed	+1
Not Contributed	0

4.1.1. BI-PARTITE DATA STRUCTURES

The rectangular data matrix is used for storing 2 mode data. The actors represented in rows and events represented in columns. Figure 4.1.1.2. shows a portion of the valued data set used for the analysis.

Table 4.1.1.2. Rectangular data array of California political donations data

		1 OR	2 OR	3 OR	4 OR	5 OR	6 OR	7 OR	8 OR	9 OR	10 OR
1	TeachersAssn	0	0	0	0	0	0	-1	1	-1	-1
2	Cahuallas	0	-1	0	0	0	0	0	0	0	0
3	Morongos	0	0	0	0	0	0	0	0	0	0
4	PacTel	0	0	0	0	0	0	0	0	0	0
5	Hastings	0	0	0	0	0	0	0	1	0	0
6	Walton	0	0	0	0	0	0	0	1	0	0
7	Dems	0	0	0	0	0	-1	-1	1	0	0
8	Engineers	0	0	0	0	0	0	0	0	0	0
9	Bing	0	0	0	0	0	0	0	0	-1	-1
10	Restaurants	0	-1	0	0	0	0	-1	0	0	0
11	ServiceWorkers	0	-1	1	1	1	0	-1	1	-1	-1
12	Perenchio	0	-1	0	0	0	0	0	0	-1	-1
13	Hospitals	0	0	0	0	0	0	0	0	0	0
14	SchoolEmp	0	0	0	0	0	-1	-1	1	0	1
15	Reiner	0	0	0	1	0	-1	0	0	-1	-1
16	Builders	0	0	0	0	0	0	-1	1	0	0
17	StateEmp	0	0	0	0	0	-1	-1	0	0	0
18	CFT	0	0	0	0	0	0	-1	0	-1	-1
19	Fisher	0	-1	0	0	0	0	-1	1	0	0
20	Republicans	1	0	0	0	0	1	-1	0	0	0
21	AFSCME	0	0	1	0	0	0	-1	0	0	0
22	Intel	0	0	1	0	0	0	0	1	0	0
23	Chevron	0	-1	1	0	1	0	-1	0	0	0

From the given table 4.1.1.2.:

Donors contributed donations in opposition to:

The ballot initiatives 7, 9, and 10 ((Having more minus -1 values)

Donors contributed a donation supporting of:

The ballot initiative 8.(Having more +1 values)

Common approach to two-mode data : It is to converted into **two one-mode data** sets, and examine relations within each mode separately. For example, create a data set of actor-by-actor ties, measuring the strength of the tie between each pair of actors by the number of times that they contributed on the same side of initiatives, summed across the 40-some initiatives. A one-mode data set of initiative-by-initiative ties can be created and the coding the strength of the relation as the number of donors that each pair of initiatives had in common.

Using the suitable tool one-mode data sets are created from a two-mode rectangular data array. A retrieval technique is used to convert the two mode dataset into affiliations of one mode valued data set.

The row mode (actors) is selected. The cross-product method for binary data (Table 4.1.1.3) takes each entry of the row for actor A, and multiplies it times the same entry for actor B, and then sums the result. With binary data, each product is 1 only if both actors were "present" at the event, and the sum across events yields the number of events in common - a valued measure of strength.

Table 4.1.1.3. Actor-by-actor tie strengths

		1 Te	2 Ca	3 Mo	4 Pa	5 Ha	6 Wa	7 De	8 En	9 Bi	10 Re
1	TeachersAssn	16	2	5	3	6	5	10	2	6	2
2	Cahuallas	2	7	4	2	2	1	3	2	2	3
3	Morongos	5	4	10	3	2	1	3	2	3	1
4	PacTel	3	2	3	6	2	1	3	2	2	2
5	Hastings	6	2	2	2	9	7	4	3	2	2
6	Walton	5	1	1	1	7	7	3	1	2	1
7	Dems	10	3	3	3	4	3	18	3	4	1
8	Engineers	2	2	2	2	3	1	3	6	1	2
9	Bing	6	2	3	2	2	2	4	1	8	-1
10	Restaurants	2	3	1	2	2	1	1	2	-1	8
11	ServiceWorkers	11	3	3	0	3	2	10	2	4	3
12	Perenchio	4	3	4	3	3	2	1	2	4	2
13	Hospitals	5	2	4	3	2	1	4	2	2	2
14	SchoolEmp	11	3	3	3	6	5	14	2	3	2
15	Reiner	8	2	5	3	3	2	5	2	6	1
16	Builders	7	3	5	4	9	7	5	3	3	4
17	StateEmp	7	3	4	3	2	1	10	4	3	2
18	CFT	13	1	4	3	5	4	9	2	5	2
19	Fisher	3	2	1	1	5	5	2	0	2	3
20	Republicans	1	1	2	1	0	0	0	0	1	2
21	AFSCME	10	3	4	3	4	3	11	3	4	2
22	Intel	3	1	1	2	6	6	2	-1	2	2
23	Chevron	4	4	5	2	4	3	3	2	4	4

However in this case, the cross-product method to valued data is used to convert two-mode into one-mode network.

Actors involved status	**Initiative**	**Result**
Neither Actors donated	0,0 (0 * 0)	0
One donated , One not donated	a) 1,0 (+1 * 0) b) (-1,0) (-1 * 0)	0 & No tie
Both donated (in the same direction)	a) (1,1) or (-1,-1) b) (+1 * +1) or (-1 * -1)	+1 & Positive tie
Both donated (in opposite direction)	a) (1, -1) (+1 * -1) b) (-1,1) (-1,+1)	-1 & Negative tie

The minimums method examines the entries for the two actors at each event, and selects the minimum value. For binary data, the result is the same as the cross-product method

For valued data, the minimums method the tie between the two actors is equal to the weaker of the ties of the two actors to the event. This approach is commonly used when the original data are measured as valued.

Illustration: The teachers association participated in 16 campaigns. The association took the same position on issues as the Democratic party (actor 7) ten more times than taking opposite (or no) position. The restaurant association (node 10) took an opposite position to Mr. Bing (node 9) more frequently than supporting (or no) position.

Resulting one-mode matrices of actors-by-actors and events-by-events are: valued matrices. This indicating the strength of the tie based on co-occurrence.

Two-mode data are sometimes stored in the "bipartite" matrix. A bipartite matrix is formed by adding the rows as additional columns, and columns as additional rows. For example, a bipartite matrix of the donors data would have 68 rows (the 23 actors followed by the 45 initiatives) by 68 columns (the 23 actors followed by the 45 initiatives). The two actor-by-event blocks of the matrix are identical to the original matrix; the two new blocks (actors by actors and events by events) are usually coded as zeros. The tool converts two-mode rectangular matrices to two-mode bipartite matrices. In the tool the data to be entered are

i) Two-mode dataset

ii) Value to fill with in mode ties

iii) Make the result symmetric or not

iv) Output dataset

The *value to fill within-mode ties* usually zero in the developed tool, so that actors are connected only by co-presence at events, and events are connected only by having actors in common. Algorithm for one-mode data is applied to get the result.

4.1.2. VISUALIZING TWO-MODE DATA

Graphs can be used to visualize 2-mode data. Both actors and events are treated as nodes, and lines are used to show the connections of actors to events (there will be no lines from actors to actors directly, or from events to events).. Figure 4.1.2.1 shows one rendering of the California donors data in it's valued form.

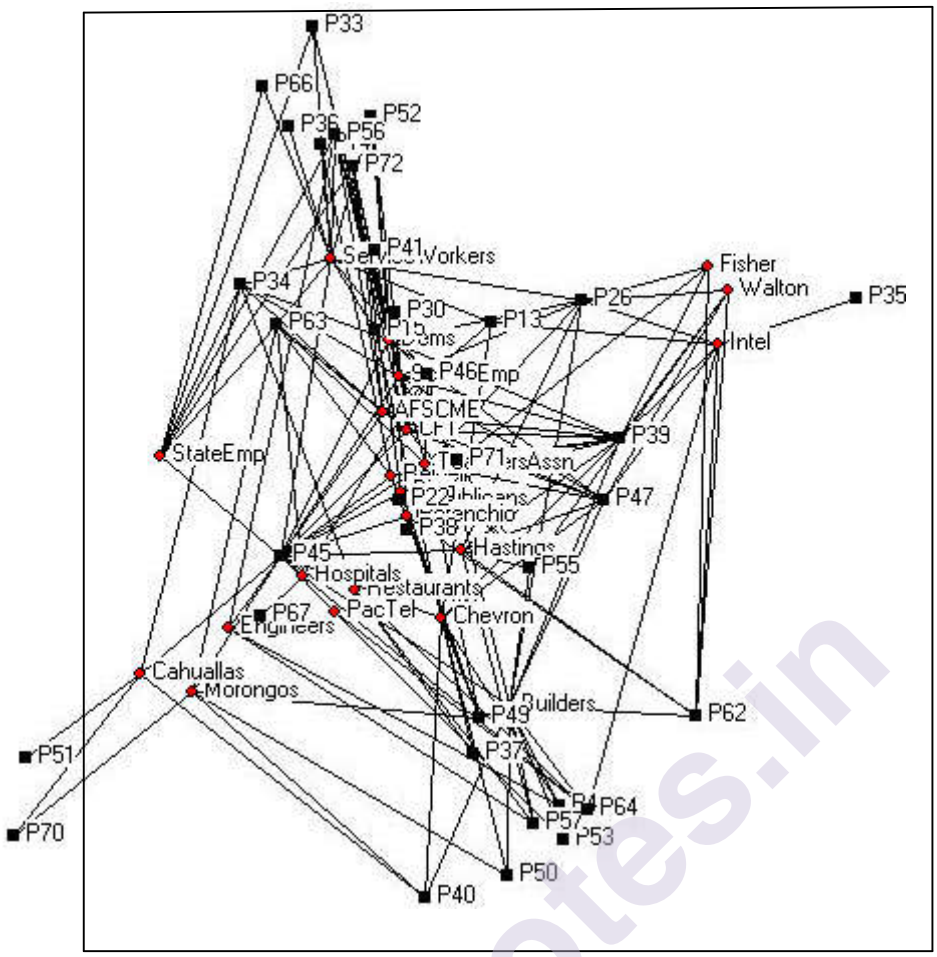


Figure .4.1.2.1. Two-mode valued network of California donors and initiatives

Findings from map:

i) Actors that are close together, are connected because they have similar profiles of events. For example, the Cahualla and Morongo Indians in the lower left corner.

ii) The two tribes were jointly involved in initiatives about gambling (P70) and environment (P40).

Numeric Methods Captures

i) Clustering of actors based on events

ii) Co-presence of Actors bring the events together

Final result is Bundles of (Clusters of) actors or events

4.1.3. QUANTITATIVE ANALYSIS

This is an approach that emphasizes statistical and mathematical analysis to hep to find out the real dimension of the problem. This method mainly focuses on numbers or data. In social network analysis this quantitative

approach is mainly used to find out the various types of relationships between the actors. In addition this approach helps to determine the solutions for the various issues and used to make valuable decisions. In this chapter two types of Quantitative analysis i) SVD ii) Factor analysis are illustrated.

4.1.3.1. Two-mode Singular value decomposition (SVD) analysis

Factor and Component Analysis: The approach of locating, or scoring, individual cases in terms of their scores on factors of the common variance among multiple indicators.

Scale or Index: Done in terms of participation of the actors in the events. It is applied either to actors or to events. The events can be scaled in terms of the patterns of co-participation of actors, but weight the actors according to their frequency of co-occurrence.

Joint variance dimension can be determined and the actors and events are mapped into the same space. This gives information about the

a) Actors those are similar in terms of their participation in events.

b) Events that are similar in terms of what actors participate in them.

c) Actors and Events that are located near.

Clusters of actors and events that are similarly located may form meaningful types or domains of social action.

Interpretation of the fundamental factors or dimensions would result in why the actors and events are having the ties.

Two-mode SVD analysis

Singular Value Decomposition (SVD) : It is a method of identifying the factors underlying two-mode valued data. The method of extracting factors (singular values) differs somewhat from conventional factor and components analysis, so for factoring results both SVD and 2-mode should be examined.

Example for SVD: Input a matrix of 23 major donor by 44 California ballot initiatives. Each actor is scored as -1 if they contributed in opposition to the initiative, +1 if they contributed in favour of the initiative, or 0 if they did not contribute. The resulting matrix is valued data that can be examined with SVD and factor analysis; however, the low number of contributors to many initiatives, and the very restricted variance of the scale are not ideal.

Table 4.1.2.2 Two-mode scaling of California donors and initiatives by Single Value decomposition: Singular values

SINGULAR

FACTOR	VALUE	PERCENT	CUM %	RATIO	PRE	CUM PRE
1:	9.204	15.2	15.2	1.564	0.334	0.334
2:	5.886	9.7	24.9	1.295	0.146	0.479
3:	4.544	7.5	32.4	1.112	0.087	0.566
4:	4.085	6.7	39.1	1.064	0.070	0.637
5:	3.838	6.3	45.5	1.141	0.062	0.699
6:	3.364	5.6	51.0	1.107	0.048	0.746
7:	3.040	5.0	56.0	1.153	0.039	0.785
8:	2.637	4.4	60.4	1.069	0.029	0.814
9:	2.467	4.1	64.5	1.021	0.026	0.840
10:	2.416	4.0	68.4	1.030	0.025	0.865
11:	2.346	3.9	72.3	1.071	0.023	0.888
12:	2.190	3.6	75.9	1.072	0.020	0.908
13:	2.042	3.4	79.3	1.084	0.018	0.926
14:	1.885	3.1	82.4	1.087	0.015	0.941
15:	1.734	2.9	85.3	1.042	0.013	0.953
16:	1.663	2.7	88.0	1.159	0.012	0.965
17:	1.435	2.4	90.4	1.028	0.009	0.974
18:	1.396	2.3	92.7	1.135	0.008	0.982
19:	1.230	2.0	94.7	1.129	0.006	0.988
20:	1.090	1.8	96.5	1.082	0.005	0.993
21:	1.007	1.7	98.2	1.635	0.004	0.997
22:	0.616	1.0	99.2	1.258	0.002	0.999
23:	0.489	0.8	100.0		0.001	1.000
	60.604	100.0				

Above table shows the “singular values" extracted from the rectangular donor-by-initiative matrix using the standard tool. The "singular values" are similar to "eigenvalues" in the more common factor and components scaling techniques.

Result: The joint "space" of the variance among donors and initiatives is not well captured by a simple characterization.

Issue: If we could easily make sense of the patterns with ideas like "left/ right" and "financial/moral" as underlying dimensions, there would be only a few singular values that explained substantial portions of the joint variance. This result tells us that the ways that actors and events "go together" is not clean, simple, and easy in this case.

Solution: To solve the issue how the events and donors are "scaled" or located on the underlying dimensions, the ballot initiatives in Table 4.1.2.2.. shows the location, or scale scores of each of the ballot proposition on the first six underlying dimensions of this highly multi-dimensional space.

Table 4.1.2.2. SVD of California donors and initiatives: Scaling of initiatives

Row Scores

		1	2	3	4	5	6
1	P1	−0.002	−0.012	−0.004	−0.043	−0.059	−0.155
2	P12	−0.090	0.101	−0.214	−0.019	0.413	−0.204
3	P13	0.097	−0.050	0.023	0.049	−0.261	−0.107
4	P14	0.058	0.048	0.130	0.088	−0.091	−0.020
5	P15	0.054	−0.008	0.087	0.030	−0.251	−0.048
6	P22	−0.125	−0.108	0.097	0.069	−0.117	−0.211
7	P25	−0.302	0.039	0.098	0.075	0.344	0.158
8	P26	0.227	−0.173	−0.216	0.273	−0.073	0.067
9	P28	−0.158	−0.048	−0.341	−0.209	−0.090	0.089
10	P30	−0.121	−0.028	−0.433	−0.219	−0.081	0.105
11	P31	0.002	0.012	0.004	0.043	0.059	0.155
			• • •				
39	P67	0.006	0.010	0.010	0.011	−0.011	−0.025
40	P68	−0.014	0.012	−0.059	0.091	−0.021	0.027
41	P70	0.023	−0.020	0.079	−0.152	−0.028	0.031
42	P71	0.032	−0.051	0.065	0.043	0.006	−0.032
43	P72	0.144	0.130	−0.162	−0.116	0.129	−0.095
44	P1A	−0.012	0.024	−0.055	0.134	0.038	0.182

First dimension: Locate initiatives supporting public expenditure for education and social welfare toward one pole, and initiatives supporting limitation of legislative power toward the other though interpretations like this are entirely subjective.

Second & Higher Dimensions: This specifies that initiatives can also be seen as differing from one another in other ways. But, the results locate or scale the donors along the same underlying dimensions. These loadings are shown in table 4.1.2.3..

Table 4.1.2.3. SVD of California donors and initiatives: Scaling of donors

Column Scores

		1	2	3	4	5	6
1	Morongos	0.133	−0.069	0.270	−0.372	0.082	−0.091
2	Republicans	−0.020	−0.072	−0.019	−0.177	−0.227	−0.523
3	Restaurants	0.053	−0.105	0.019	−0.097	−0.342	0.225
4	ServiceWorkers	0.333	0.296	0.256	0.383	−0.569	0.056
5	Fisher	0.098	−0.287	−0.036	0.200	−0.200	0.018
6	Perenchio	0.104	−0.112	0.505	0.100	0.109	0.406
7	Hastings	0.169	−0.267	−0.141	0.183	0.102	0.218
8	Walton	0.120	−0.232	−0.170	0.271	0.104	0.072
9	Chevron	0.161	−0.340	0.138	−0.261	−0.394	−0.217
10	Reiner	0.201	−0.012	0.334	−0.025	0.222	−0.124
11	Cahuallas	0.077	−0.047	0.090	−0.249	−0.188	0.196
12	TeachersAssn	0.363	0.038	0.077	0.154	0.196	−0.236
13	CFT	0.321	0.093	0.153	0.226	0.209	−0.125
14	Engineers	0.061	0.022	0.020	−0.202	0.025	0.320
15	Intel	0.092	−0.353	−0.144	0.231	−0.010	−0.073
16	Builders	0.229	−0.533	−0.095	−0.148	0.035	0.087
17	PacTel	0.076	−0.060	0.041	−0.220	0.199	0.106
18	Hospitals	0.130	−0.001	0.088	−0.176	0.157	0.023
19	Bing	0.134	−0.020	0.225	0.017	0.180	−0.277
20	StateEmp	0.248	0.254	−0.064	−0.276	−0.063	0.227
21	AFSCME	0.310	0.104	−0.146	−0.154	−0.030	−0.127
22	SchoolEmp	0.344	0.119	−0.417	−0.039	0.036	0.053
23	Dems	0.341	0.202	−0.313	−0.119	0.028	0.031

Result Analysis :In the positive end of dimension one, the Democratic party, public employees and teachers unions are found; at the opposite pole, Republicans and some business and professional groups are found.

Map: The locations of the actors and events in a scatterplot are visualized and defined by scale scores on the various dimensions. The map in Figure 4.2.1.1. shows the results for the first two dimensions of this space.

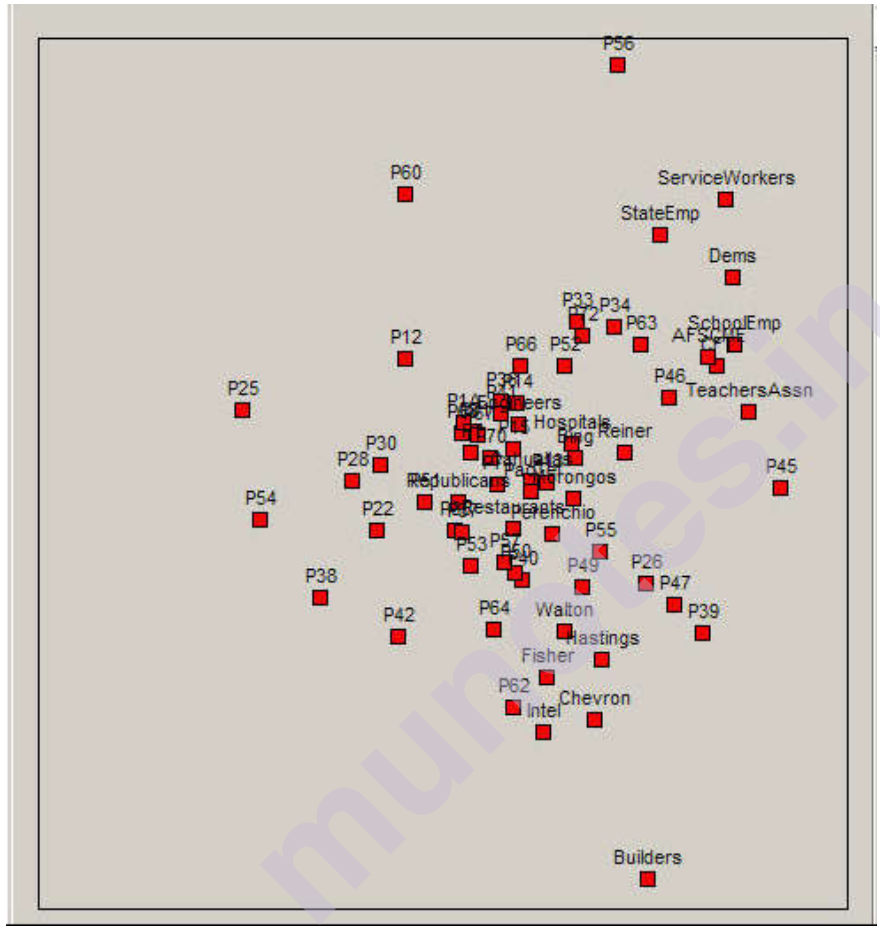


Figure 4.2.1.1 SVD of California donors and initiatives: Two-dimensional map

Result Discussion: First dimension :(left-right in the figure) seems to have its poles "anchored" by differences among the initiatives;

Second dimension (top-bottom) seems to be defined more by differences among groups (with the exception of proposition 56).

The result produces some interesting clusters that show groups of actors along with the issues that are central to their patterns of participation. The Democrats and unions cluster (upper right) along with a number of particular propositions in which they were highly active (e.g. 46, 63). Corporate, building, and venture capitalist cluster (more loosely) in the lower right, along with core issues that formed their primary agenda in the initiative process (e.g. prop. 62).

4.1.3.2. Two-mode factor analysis

Factor analysis provides an alternative method to SVD to the same goals such as identifying underlying dimensions of the joint space of actor-by-event variance, and locating or scaling actors and events in that space. The method used by factor analysis to identify the dimensions differs from SVD. Table 4.1.3.1. shows the eigenvalues (by principle components) calculated using the tool.

Table. 4.1.3.1. Eigen values of two-mode factoring of California donors and initiatives

EIGENVALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO
1:	8.321	18.9	18.9	1.640
2:	5.073	11.5	30.4	1.116
3:	4.545	10.3	40.8	1.160
4:	3.919	8.9	49.7	1.117
5:	3.509	8.0	57.7	1.152
6:	3.046	6.9	64.6	1.410
7:	2.160	4.9	69.5	1.106
			• • •	
29:	0.000	0.0	100.0	1.248
30:	0.000	0.0	100.0	1.509
31:	0.000	0.0	100.0	1.239
32:	0.000	0.0	100.0	
	44.000	100.0		

Solution: It is different from SVD, and considerable dimensional complexities are given in the joint variance of actors and events.

Simple characterizations of the underlying dimensions (e.g. "left/right") do not provide very accurate predictions about the locations of individual actors or events. The factor analysis method does produce lower complexity than SVD.

The scaling of actors on the first three factors given in the following table 4.1.3.2.. The first factor, by this method, produces a similar pattern to SVD. At one pole are Democrats and unions, at the other lie many capitalist groups. There are, however, some notable differences (e.g. AFSCME).

Table. 4.1.3.2.. Loadings of donors

Unrotated Factor Loading

		1	2	3
1	Morongos	-0.26	-0.26	-0.39
2	Republicans	0.05	0.54	-0.37
3	Restaurants	0.22	-0.24	0.39
4	ServiceWorkers	0.47	-0.49	0.29
5	Fisher	0.31	-0.52	0.39
6	Perenchio	-0.54	-0.39	0.10
7	Hastings	-0.41	-0.10	-0.02
8	Walton	0.21	0.49	0.52
9	Chevron	-0.44	0.37	-0.33
10	Reiner	-0.28	0.47	-0.33
11	Cahuallas	0.26	0.26	0.39
12	TeachersAssn	0.77	0.13	-0.17
13	CFT	0.58	0.05	-0.25
14	Engineers	-0.21	0.09	0.42
15	Intel	0.59	-0.52	0.35
16	Builders	-0.38	-0.40	0.27
17	PacTel	-0.75	-0.28	0.21
18	Hospitals	0.10	0.57	0.33
19	Bing	-0.38	-0.17	-0.08
20	StateEmp	0.34	0.38	-0.24
21	AFSCME	-0.63	0.01	-0.05
22	SchoolEmp	0.44	0.02	-0.20
23	Dems	0.62	0.35	0.06

Table. 4.1.3.3. Loadings of events

	1	2	3
1	0.001	0.020	-0.104
2	0.088	0.101	0.008
3	-0.043	-0.084	0.039
4	-0.027	-0.126	-0.004
5	-0.057	-0.127	0.006
6	-0.091	-0.021	-0.045
7	-0.043	0.019	-0.025
8	-0.003	0.002	0.151
9	0.026	0.111	-0.018
10	0.048	0.112	-0.003
11	-0.001	-0.020	0.104
12	0.100	-0.025	0.005
	• • •		
41	0.005	0.016	-0.106
42	-0.038	-0.029	0.020
43	0.117	0.047	0.001
44	-0.007	-0.023	0.144

Table. 4.1.3.3. shows the loadings of the events. The patterns here also have some similarity to the SVD results, but do differ considerably in the specifics.

4.1.3.3. Two-mode correspondence analysis

For binary data, the use of factor analysis and SVD is not recommended. Factoring methods operate on the variance/covariance or correlation matrices among actors and events. When the connections of actors to events is measured at the binary level (which is very often the case in network analysis) correlations may seriously understate covariance and make patterns difficult to separate.

As an alternative for binary actor-by-event scaling, the method of correspondence analysis can be used.

Correspondence analysis

i) It operates on multi-variate binary cross-tabulations

ii) It's distributional assumptions are better suited to binary data.

Example:

The political donor and initiatives data are dichotomized by assigning a value of

i) 1 if an actor gave a donation either in favour or against an initiative,

ii) Assigning a zero if they did not participate in the campaign on a particular initiative.

The partisanship has been given more attention rather than simple participation. Two data sets - one based on opposition or not, one based on support or not are created and two separate correspondence analyses are carriedout.

Table. 4.1.3.3. shows the location of events (initiatives) along three dimensions of the joint actor-event space identified by the correspondence analysis method.

Table. 4.1.3.3. Event coordinates for co-participation of donors in California initiative campaigns

Row Scores

		1	2	3
1	P1	3.096	3.662	-1.898
2	P12	0.177	-0.319	0.554
3	P13	-0.263	-0.079	0.170
4	P14	-0.104	-0.304	-0.502
5	P15	0.070	-0.436	0.183
		• • •		
40	P68	3.404	-1.607	1.357
41	P70	2.682	-1.763	1.446
42	P71	-0.266	0.420	0.389
43	P72	-0.323	0.009	-0.301
44	P1A	3.250	1.027	-0.270

Result: Since these data do not reflect partisanship, only participation, it reflects. However, that this method can be used to locate the initiatives along multiple underlying dimensions that capture variance in both actors and events.

Table 4.1.3.4. shows the scaling of the actors.

Table 4.1.3.4. Actor coordinates for co-participation of donors in California initiative campaigns

Column Scores

		1	2	3
1	Morongos	2.189	-1.017	0.794
2	Republicans	1.991	2.316	-1.110
3	Restaurants	-0.230	-0.241	-0.026
4	ServiceWorkers	-0.131	-0.283	-0.384
5	Fisher	-0.339	0.633	0.659
6	Perenchio	-0.100	0.162	0.201
7	Hastings	-0.590	0.531	0.833
8	Walton	-0.669	0.800	0.966
9	Chevron	0.220	-0.269	0.599
10	Reiner	-0.004	-0.102	-0.203
11	Cahuallas	1.260	-1.213	0.899
12	TeachersAssn	-0.244	-0.008	0.047
13	CFT	-0.103	-0.032	-0.067
14	Engineers	-0.400	0.068	0.201
15	Intel	-0.571	0.443	0.403
16	Builders	-0.112	-0.062	0.445
17	PacTel	-0.248	-1.472	-1.722
18	Hospitals	-0.146	-1.176	-1.415
19	Bing	-0.280	-0.124	0.251
20	StateEmp	-0.093	0.009	-0.637
21	AFSCME	-0.196	-0.090	-0.220
22	SchoolEmp	-0.277	0.150	-0.165
23	Dems	-0.169	0.329	-0.259

The first dimension has some similarity to the Democrat/union versus capitalist poles. But this difference means that the two groupings tend to participate in different groups of initiatives.. Visualization is the best approach to finding meaningful patterns.

Figure 4.1.3.1.shows the plot of the actors and events in the first two dimensions of the joint correspondence analysis space.

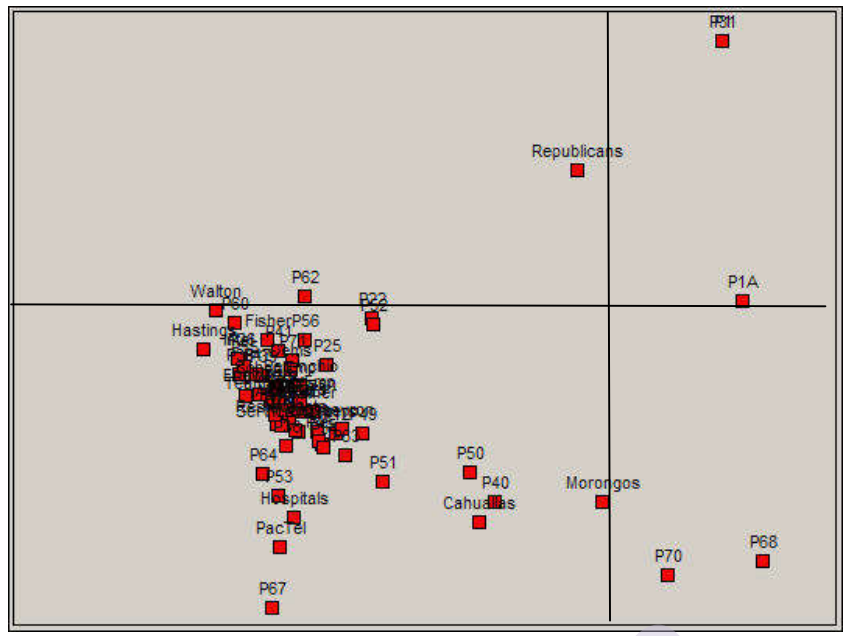


Figure 4.1.3.1. Correspondence analysis two-dimensional map

Result: In the lower right there are some propositions regarding Indian casino gambling represented by 68 and 70. The other two propositions regarding ecological/conservation issues are represented by 40 and 50. Two of the major Native American Nations (the Cahualla and Morongo band of Mission Indians) are mapped together. The result is showing that there is a cluster of issues that "co-occur" with a cluster of donors- actors defining events, and events defining actors.

4.1.4. QUALITATIVE ANALYSIS

Actors and events are co-presence with each other. In the case of either an actor was, or wasn't present, and the incidence matrix is binary, there will be some issue when data are parsed in correspondence analysis.

Block Modeling: This is an alternative method for correspondence analysis. It works directly on the binary incidence matrix by trying to permute rows and columns to fit, as closely as possible, idealized images. This method does not involve any of the distributional assumptions that are made in scaling analysis.

4.1.4.1. Two-mode core-periphery analysis

The core-periphery structure is a typical pattern that divides both the rows and the columns into two classes. One of the blocks on the main diagonal is a high-density block which is known as core block; the other block on the main diagonal is a low-density block. The core-periphery model is indifferent to the density of ties in the off-diagonal blocks.

Core: When the core-periphery model is applied to actor-by-actor data the model identifies a set of actors with high density of ties among themselves known as **core**. This model shares many events in common. The "core" consists of a partition of actors that are closely connected to each of the

events in an event partition; and simultaneously a partition of events that are closely connected to the actors in the core partition.

Periphery: In another set of actors who have very low density of ties among themselves known as **periphery** by having few events in common.

Comparison: Between Core and Periphery

S.No.	Core	Periphery
1	Actors are able to coordinate their actions	Actors cannot coordinate their actions.
2	Actor are at a structural advantage in exchange relations with actors in the periphery.	No structural advantage with core
3	It is a cluster of frequently co-occurring actors and events.	It consists of a partition of actors who are not co-incident to the same events; and a partition of events that are disjoint because they have no actors in common.

Numerical methods using tools are used to search for the partition of actors and of events that comes as close as possible to the idealized image. Table 4.1.4.1. shows a portion of the results of applying this method to participation (not partisanship) in the California donors and initiatives data.

Table 4.1.4.1. Results of participation in the California donors and initiatives data

```
Starting fitness: 0.475
Final fitness: 0.508

Blocked Adjacency Matrix

                    2   3 3 3 1     3 1 2 3 3 2 4 2 1   1 1 2         2 2 1
                    3 2 6 1 8 7   8 0 0 2 2 5 4 3 1 8   5 2 0 4 6 5 3 5 6 1 1
                    P 1 P P P P P P P P P P P P P P P   P P P P P P P P P P P
                   ---------------------------------------------------------------
 17   TeachersAssn | 1   1 1   1 1 1 1 1 1 1   1   1 1 |                   1
 15           Dems | 1       1 1 1 1   1 1 1 1 1   1 |   1 1   1
 14       Builders | 1     1   1 1     1 1 1 1   1 1 |                   1 1
 22 ServiceWorkers | 1 1 1   1 1 1 1   1 1 1       1 | 1 1   1   1 1
 16      SchoolEmp | 1     1 1 1 1 1 1 1   1 1 1 1 |   1     1
 19          AFSCME | 1   1     1 1   1   1 1   1 1 |               1
  9            CFT | 1   1 1     1   1 1 1 1   1   1 1 |
                   ---------------------------------------------------------------
  2       Morongos |     1           1   1                             1 1
  7       StateEmp |     1   1 1 1   1   1 1     1       |   1     1
 10      Perenchio |   1       1     1 1 1   1       1   |                   1
  1         PacTel |                 1   1               |
 12         Walton |       1       1         1 1     1   |
 13        Hastings |       1       1     1   1 1     1   |
  3      Hospitals | 1   1           1   1               |
  4      Engineers |                     1 1       1     |
  5      Cahuallas |   1                 1               |
  6    Restaurants |   1           1               1     |
 18         Reiner |     1           1 1 1           1   |       1 1     1
  8           Bing |                   1 1         1   1   |                   1
 20          Intel |               1         1 1 1       1   |           1
 21        Chevron |   1           1       1   1     1 1 |             1 1 1 1
 11         Fisher |   1           1 1       1 1       1   |
 23    Republicans |               1         1 1           |         1         1 1
                   ---------------------------------------------------------------


Density matrix

        1     2
    ----- -----
  1 0.658 0.179
  2 0.260 0.135
```

Genetic algorithm is the numerical search method used by core and periphery. The measure of goodness of fit is stated in terms of a "fitness" score where 0 means bad fit, 1 means excellent fit. The goodness of the result by examining the density matrix is at the end of the output. If the block model was completely successful, the 1,1, block should have a density of one, and the 2, 2 block should have a density of zero.

Result Discussion: The blocked matrix shows a "core" composed of the Democratic Party, a number of major unions, and the building industry association who are all very likely to participate in a considerable number of initiatives (proposition 23 through proposition 18). The remainder of the actors are grouped into the periphery as both participating less frequently, and having few issues in common. A considerable number of issues are also grouped as "peripheral" in the sense that they attract few donors, and these donors have little in common. The upper right) that core actors in upper right do participate to some degree (.179) in peripheral issues. In the lower left the peripheral actors participate somewhat more heavily (.260) in core issues.

4.1.4.2. Two-mode Factions Analysis

Factions: Groupings that have high density within the group, and low density of ties between groups. This method is an alternative block model.

The subgroups factions choice in the tool fits this block model to one mode data.for any number of specified factions. The two-mode choice fits the same type of model to two-mode data for only two factions.

Factions model applied to one-mode actor data : Identifies two clusters of actors who are closely tied to one another by attending all of the same events, but very loosely connected to members of other factions and the events that tie them together

. ***Factions model applied to one-mode event data*** Identifies events that are closely tied by having exactly the same participants.

The two-mode option in the tool applies the same approach to the rectangular actor-by-event matrix. This locates joint groupings of actors and events that are as mutually exclusive as possible. Figure 4.1.4.2. shows the results of the two mode factions block model to the participation of top donors in political initiatives.

Table 4.1.4.2. Two mode factions model of California $1M donors and ballot initiatives (truncated)

Starting fitness: 0.104
Final fitness: 0.303
Correlation to ideal: 0.303

Blocked Adjacency Matrix

		34P	32P	25P	31P	10P	06P	07P	08P	36P	37P	30P	05P	51P	12P	28P	43P	32P	15P	54P	40P	06P	21P	23P	06P	13P	29P
1	PacTel											1			1												1
2	Morongos		1							1		1			1										1		
3	Hospitals									1		1			1											1	
4	Engineers								1					1	1			1									
16	SchoolEmp	1		1	1	1	1			1		1		1	1	1	1	1								1	
17	TeachersAssn	1	1	1	1		1	1	1	1		1		1	1	1		1								1	
7	StateEmp					1	1		1	1		1			1		1	1									
8	Bing		1		1						1				1	1	1										
9	CFT		1		1	1		1				1		1	1	1		1								1	
10	Perenchio	1		1		1				1		1	1		1	1							1				
11	Fisher	1					1	1				1			1		1						1				
12	Walton	1	1		1			1				1			1												
13	Hastings	1	1		1			1			1		1		1	1											
14	Builders	1	1	1	1		1	1			1		1	1	1	1		1							1		1	1
15	Dems		1			1	1	1		1		1	1		1	1	1	1				1	1			1		
21	Chevron		1	1			1							1	1	1		1		1			1	1	1	1		
19	AFSCME		1				1		1	1		1			1	1	1	1							1		1	
18	Reiner			1		1	1			1		1			1	1				1							
23	Republicans					1	1						1					1								1		
20	Intel	1	1					1				1			1		1	1							1			
6	Restaurants							1			1						1					1	1					
22	ServiceWorkers				1		1	1	1	1				1	1			1	1	1	1			1	1		1	1
5	Cahuallas														1								1				

Density matrix

	1	2
1	0.401	0.119
2	0.212	0.299

Two measures of goodness-of-fit are available.

i) ***Fitness score:*** It is the correlation between the observed scores such as 0 or 1.The scores that should be present in each block.

ii)***Densities in the blocks*** : It gives goodness of fit. For a factions analysis, an ideal pattern is dense 1- blocks along the diagonal and zero-blocks off the diagonal.

Result Discussion: The fit of the two factions model is not as impressive as the fit of the core-periphery model. This suggests that an image of California politics as one of two separate and largely disjoint issue-actor spaces is not as useful as an image of a high intensity core of actors and issues coupled with an otherwise disjoint set of issues and participants.

The blocking itself also is not very appealing, placing most of the actors in one faction (with modest density of .401). The second faction is small, and has a density (.299) that is not very different from the off-diagonal blocks.

4.1.6. AFFILIATION NETWORKS

Persons A and B are both members of a club. They can form an open triad, or a structural hole, but it is infered that if A and B are members of the same club (Figure 4.1.6.1), they may know each other; and the triad is closed. This is a weak inference. To make a more concrete case, it should be considered that if they were members of the club at the same time, or if the club has multiple chapters in different cities, etc.

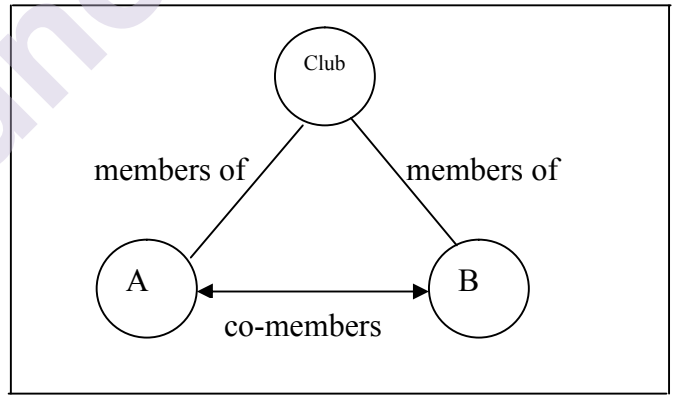


Figure 4.1.6.1. Triadic closure and co-membership

Consider the same people are members of more then one club as shown in the top of **Figure 4.1.6.2.** where nodes E, F, and H are co-members in 2 clubs. This presents a stronger association between the people, having a common group identity. The co-memberships can be accumulated until the connections are real, and weigh the inferred links accordingly.

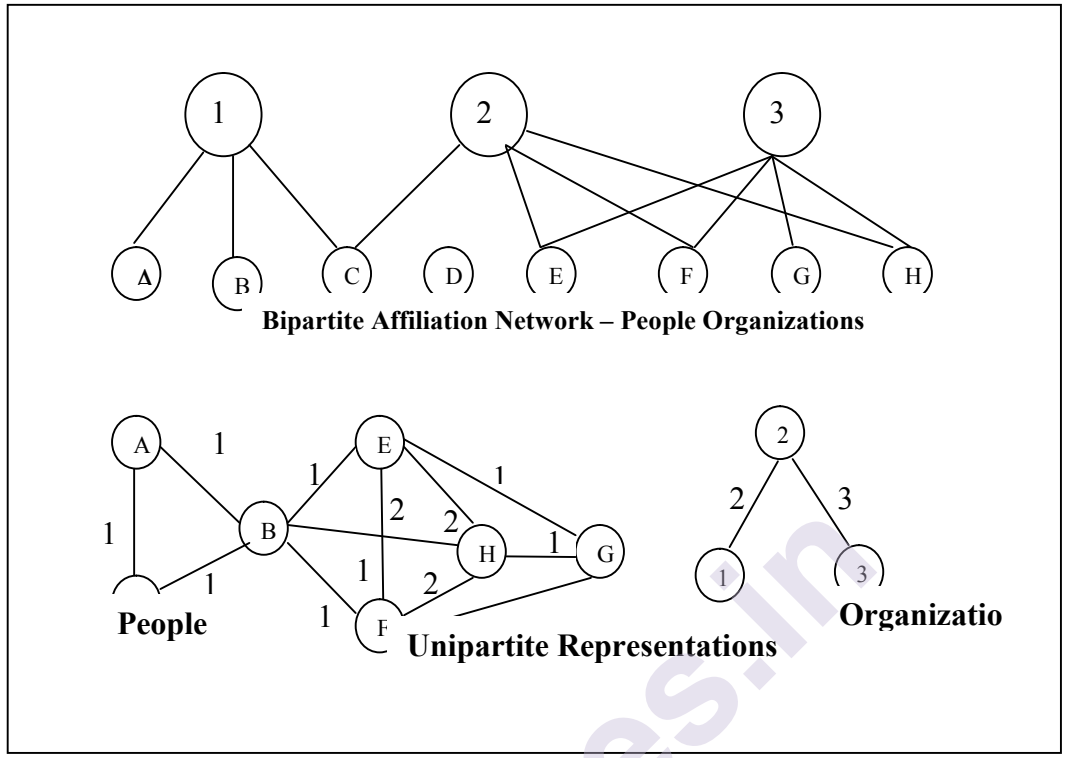


Figure . 4.1.6.2 Creating an affiliation network from a 2-mode network

Figure . 4.1.6.2. shows two resulting projected networks i) A network of people where links were determined through co-membership in groups ii) A network of groups where links were determined by comembership of people. To create these networks, count the comemberships for every one of the people or for every one of the clubs.

These networks can be used for all social network analysis, but are particularly for analysis with the island method and clustering techniques. The reason is these networks are essentially networks of similarities or correlations.

4.1.7 ATTRIBUTE NETWORKS

This an application of 2-mode network analysis is based on the idea of *homophily meaning that similar,* the idea that people who share interests or attributes are more likely to talk to each other and form ties than people who are very different.

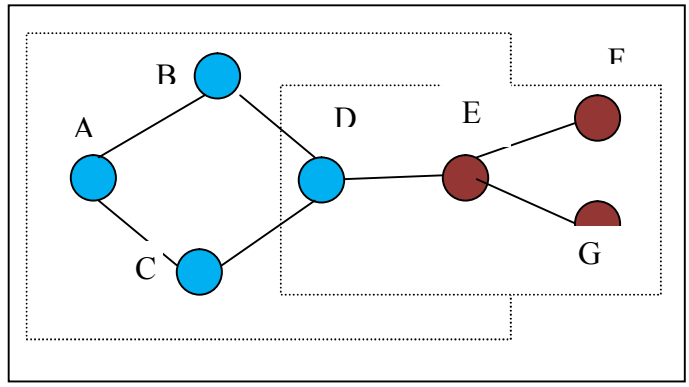


Figure 4.1.7.1. Common Network (Election Hashtags and People)

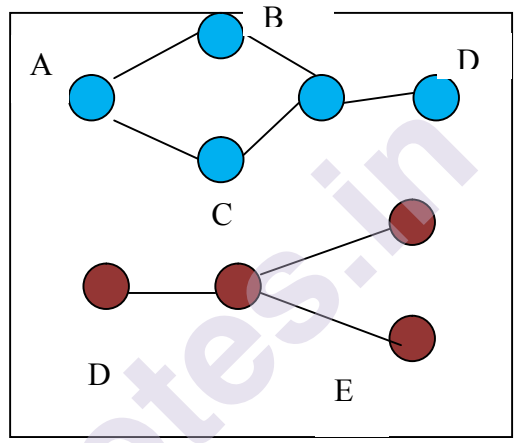


Figure 4.1.7.2. Attribute network

(People Network, Hash tag Network)

When people become more tightly connected, they become more similar in their views, but up to a certain limit. However, if one wants to build a "suggest a friend" mechanism for their online social network, treating an attribute or interest matrix as a 2-mode network can be a useful mechanism. Every one of the pieces of information (tags, keywords, etc.) could be treated as a node in a 2-mode network, compute a person-to-person affiliation network from it, and apply the island method or clustering to find potential groupings of people. Then, to suggest friends, pick the top links in the affiliation network.

An inverse affiliation network, attributes through people, could provide very interesting insights as well. For example in mapping political discourse on Twitter, the tweets from several thousand people containing the hashtag#election can be extracted, build a 2-mode network from people to hashtags, and compute a #hashtag through-people affiliation network. From the figure 4.1.7.1 the election related tweets are given. The tweets are given as blue circle for# hashtags and the people nodes are specified in brown circles. If they are taken in attribute network the resultant Figure (4.7.1.2) contains the people with election #hashtags and others separately. In such networks, clusters will act as proxies for entire

areas of discourse and will separate people with election hash tag and not. The exploration of clusters may yield an idea of divisions inside supporting campaign in the tweets.

4.3. SUMMARY

Two-mode data gives interesting possibilities for gaining insights into macro-micro or agent-structure relations. With two-mode data, the macro-structures (events) pattern the interactions among agents (or not) can be determined. The actors define and create macro structures by their patterns of affiliation with them also explained through the illustration.. In addition, to describe patterns of relations between actors and structures simultaneously.

In this chapter some of the typical ways in which two-mode data arise in social network analysis, and the data structures that are used to record and manipulate two-mode data are examined. The utility of two-mode graphs (bi-partite graphs) in visualizing the "social space" defined by both actors and events also exhibited in this chapter.

The methods for trying to identify patterns in two-mode data that might better help us describe and understand why actors and events "fit together" in the ways they do. One class of methods derives from factor analysis and related approaches. They can also be useful to identify groups of actors and the events that "gotogether" when viewed through the lens of latent abstract dimensions.

Another class of methods is based on block modeling. The goal of these methods is to assess how well the observed patterns of actor-event affiliations fit some prior notions of the nature of the "joint space" .To the extent that the actor-event affiliations can be usefully thought of in these ways, block models also then allow us to classify types or groups of actors along with the events that are characteristic of them.

Another important topic discussed in this chapter is affiliation and attribute networks. The affiliation networks are used to examine the comembership relations whereas the attribute networks are used to cluster the networks based on the attributes.

For illustration the Davis data and California Teachers Association data is used in this chapter.

4.4. REFERENCE FOR FURTHER READING

1. Introduction to Social Network Methods: Robert A. Hanneman, Mark Riddle, University of California, 2005, Published in digital form and available at http://faculty.ucr.edu/~hanneman/nettext/index.html.

2. Social Network Analysis for Startups- Finding connections on the social web: Maksim Tsvetovat, Alexander Kouznetsov, O'Reilly Media, 2011.

3. Social Network Analysis- 3rd edition, John Scott, SAGE Publications, 2012.

4. Mark S. Handcock, David Hunter, Carter T. Butts, Steven M. Goodreau and Martina Morris. 2003 statnet: An R package for the Statistical Modeling of Social Networks http://www.csde.washington.edu/statnet

5. Vladimir Batagelj and Andrej Mrvar (2006), Pajek datasets http://vlado.fmf.uni¬ lj.si/pub/networks/data/.

6. Krackhardt and Stern (1988) developed a very simple and useful measure of the group embedding based on comparing the numbers of ties within groups and between groups

7. Getting Started in Social Network Analysis with NETDRAW, Bruce Cronin University of Greenwich Business School, Occasional Paper 01/15, January 2015 brought to you by CORE View metadata, citation and similar papers at core.ac.uk

8. Structural Holes, The Social Structure of Competition, Ronald S. Burt

9. www.analytictech.com

10. https://www.datacamp.com/

11. https://networkdata.ics.uci.edu/netdata/html/davis.html

12. Finding Social Groups: A Meta-Analysis of the Southern Women Data1 Linton C. Freeman University of California, Irvine

4.5. MODEL QUESTIONS

1. What is two-mode network? Explain with example.

2. How the bipartite network is managed? Give example.

3. What are the two alternative methods used in bipartite network?

4. State the purpose of visualising data.

5. Describe the quantitative analysis in social network using suitable example.

6. Compare SVD with two-mode factor analysis

7. Illustrate how the results of correspondence analysis can be interpreted

8. Write short note on quality analysis.

9. Explain the affiliation networks briefly.

10. Summarize the purpose of attribute networks.