Module 2

CHAPTER 2

Social Network Structure, Measures and Visualization

University Prescribed Syllabus

Social Network Structure, Measures & Visualization

Basics of Social Network Structure - Nodes, Edges & Tie Describing the Networks Measures - Degree Distribution, Density, Connectivity, Centralization, Tie Strength & Trust Network Visualization - Graph Layout, Visualizing Network features, Scale Issues.

Social Media Network Analytics - Common Network Terms, Common Social Media Network Types, Types of Networks, Common Network Terminologies, Network Analytics Tools.

2.1 BASICS OF SOCIAL NETWORK STRUCTURE

- Social Media is technology which is based on computers to share ideas, information, thoughts and any other form of expression. Social media contains lots of information in terms of connections, texts, audio, video etc. The problem is how we can visualize this information?

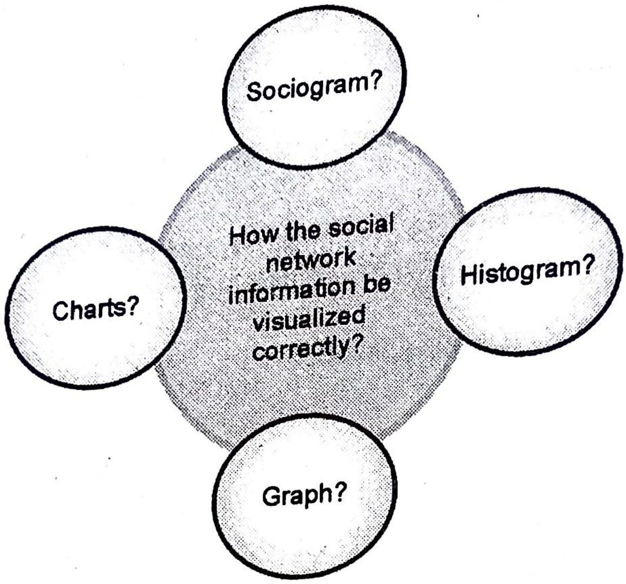

Fig. 2.1.1 : Representation of Social Media information

- Fig. 2.1.1 represents different ways by which we can represent the information. The majority of these metaphors are equally counted as sociograms. Sociogram is a graph database which is used to map groups's social connections and portray individual connections with group networks.
- This type of representation is easy for understanding and provides detailed information about the actual relations modeled in the data. In sociogram actors (people in facebook network) are represented as nodes and links are used to represent connections amongst actors.

A network or graph is a set of nodes and edges.

- Knowing the nodes and edges is all that is needed to analyze a social network. However, edges can have a number of additional features, which can be used in analysis. Edges can be labeled.
- The label describes something about the relationship between the people. It could name the relationship (e.g., sister, mother, cousin), or some information about the relationship. Fig 2.1.3 indicates connection between different students in a network who are connected through the social media groups.

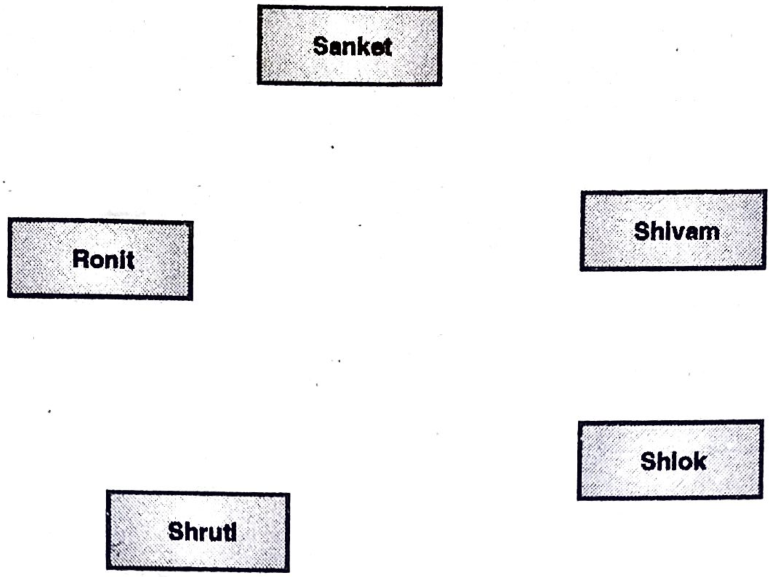

Fig. 2.1.2 : A five student network, each student denoted by a node

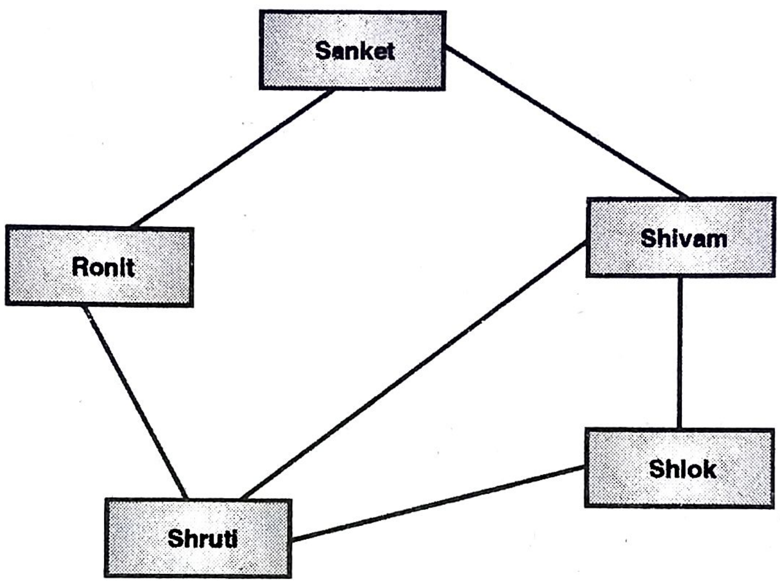

Fig. 2.1.3 : The edges connect students who were in same software group

- Edges can be weighted or valued. The weight is a number that indicates numerical information about a relationship. Often, this is the strength of a relationship, but it can come from a variety of sources and indicate many things.
- Often, this is the strength of a relationship, but it can come from a variety of sources and indicate many things. For all the edges in this graph, the weight is 1, because they have only been in one group with each other, except the edge between Shivam and Shruti. Since both are in 3 common social media groups and they are

  1. Java Programming 2. Software Testing
  3. Software Development

so the weight between them would be 3. Weights can be shown as numeric lab on the edges, or the edges can be drawn thicker to show the greater weig Fig. 2.1.5 shows both of these options.

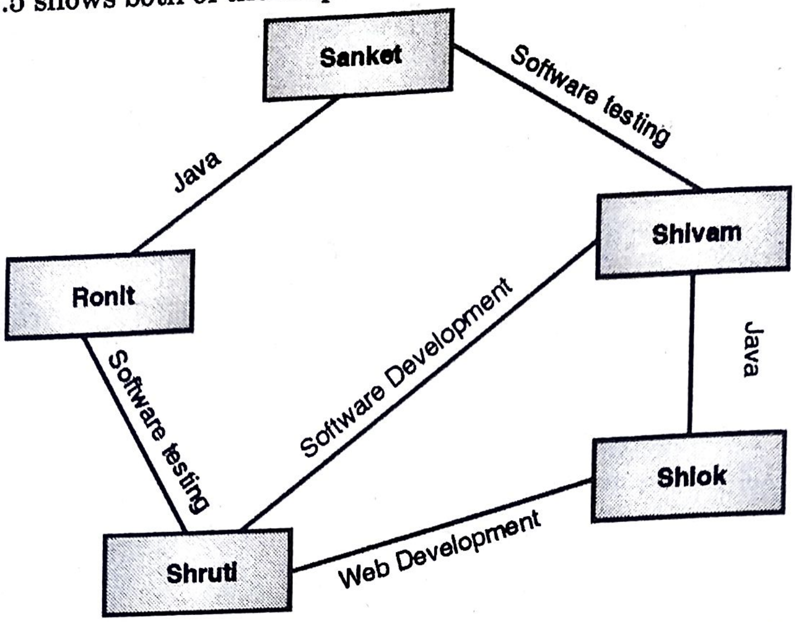

Fig. 2.1.4 : A labeled graph where the edges indicate at least one group that the student have been in together

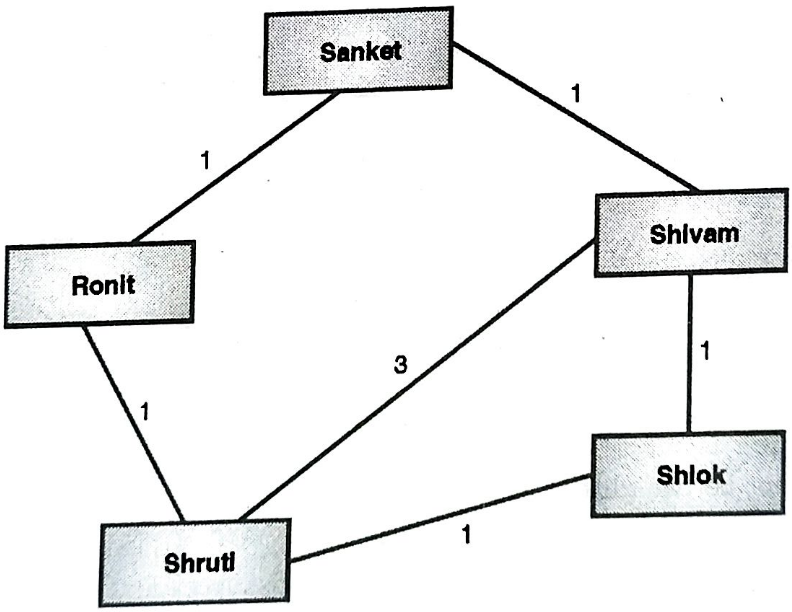

Fig. 2.1.5 : A weighted graph where weights are indicated both as numbers and by the thickness of the edge. In this graph, weight indicates how many groups the actors have been in together

- Edges can also be either *directed* or *undirected*. An undirected edge indicates mutual relationship, whereas a *directed* edge indicates a relationship that on node has with the other that is not necessarily reciprocated. The type of edge use defines the network as either a *directed network* or an *undirected network*.

- The example of a social network based on social groups is an undirected network. It is undirected because if two persons are in a group together, there is no notion of a one-way relationship. Undirected edges are drawn as simple lines between nodes, as is seen in Figs. 2.1.3 - 2.1.5.
- If we were to build a network of email communication, then we could have a directed network. Person A may send an email to Person B without receiving a reply. In that case, we would draw an edge indicating the one-way relationship from A to B, but not the reverse.
- In a directed network, edges can be reciprocated. Person A may email Person C, and C may reply. In this case, we want the edge to indicate a relationship in both directions. When showing a directed net- work, edges have arrowheads to show the direction of the relationship.
- If there is a relationship that is reciprocated, it is either drawn with a line that has arrows on both ends, or by two directed edges as shown in Fig. 2.1.6. An undirected edge is never used in a directed network.
- In a directed graph, the number of possible edges is double that of an (other- wise identical) undirected graph. Between any two nodes, there can be only one edge in an undirected graph, but two possible edges in a directed one.

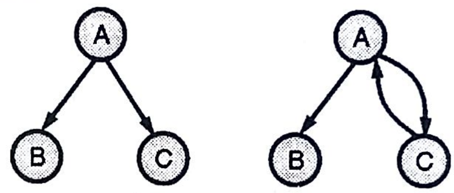

Fig. 2.1.6

- Refer Fig 2.1.6. Two ways of drawing a directed network. The edge from A to B is directed only one way. The edge from A to C goes in both directions and can be drawn either as one edge with two arrow heads (left) or as two edges pointing in opposite directions (right)

☞ Adjacency lists

- An *adjacency list*, also called an *edge list*, is one of the most basic and frequently used representations of a network. Each edge in the network is indicated by listing the pair of nodes that are connected. For example, the adjacency list for the above network in Fig. 2.1.3 is as follows :

Sanket - Ronit
Sanket - Shivam
Ronit - Shruti
Shivam - Shlok
Shruti - Shivam
Shruti - Shlok

Fig. 2.1.7 : Adjacency list for network in Fig. 2.1.3

- Each line contains one pair of nodes. In this example, the names of the nodes are separated by hyphen characters, but you could also use tabs or other characters as a separator.
- The order of these lines does not matter since there is no concept of order in networks.
- Since this network is undirected, the order of the node names in each pair is irrelevant, too. The current list has "Sanket - Ronit" as the first entry, but it would have the same meaning if we reversed the order to "Ronit - Sanket."
- However, if the network is directed, this would not be true. In a directed network, the order of the node names is important. If a pair is listed as "Node A, Node B" in a directed network, it means there is a relationship from Node A to Node B. The reverse relationship is not implied, but it can be indicated by including another line listing "Node B, Node A." If both pairs are listed, it means there is a relationship in both directions.
- Adjacency lists can also include additional information about the edges. This is included on the same line as the two node names, and usually follows them. An edge weight is a common value to see included in an adjacency list. Again using the students network example and the edge weights from Fig. 2.1.5, the list would be written as follows :

Sanket - Ronit 1
Sanket - Shivam 1
Ronit - Shruti 1
Shivam - Shlok 1
Shruti - Shivam 3
Shruti - Shlok 1

Fig. 2.1.8 : Adjacency list with weight information

- Edge labels can also be included in an adjacency list in the same way.

Adjacency matrix

- An alternative to the adjacency list is an *adjacency matrix*. In an adjacency matrix, a grid is set up that lists all the nodes on both the X-axis (horizontal) and the Y-axis (vertical).
- Nodes are considered as rows as well as columns. Then, values are filled into the matrix to indicate if there is or is not an edge between every pair of nodes. Typically, a 0 indicates no edge and a 1 indicates an edge.

Table 2.1.1

The Adjacency Matrix for student's Network					
	Sanket	**Ronit**	**Shruti**	**Shivam**	**Shlok**
Sanket	0	1	0	1	0
Ronit	1	0	1	0	0
Shruti	0	1	0	1	1
Shivam	1	0	1	0	1
Shlok	0	0	1	1	0

- Notice a couple of things about this matrix.
  - First, the diagonal is all zeros because there are no edges between a node and itself in our example. Some net- works do allow for self-loops. For example, in an email network, if a person emails himself, there could be a link from one node to itself, and thus there would be a 1 on the diagonal.
  - Second, the matrix is symmetric.
- The numbers in the first row are the same as the numbers in the first column. The numbers in the second row are the same as the numbers in the second column.
- This is because the graph is undirected. Just as in the adjacency list, where the order of pairs in an undirected graph didn't matter.

Table 2.1.2

Notice that the diagonal, indicating students link to himself/ herself is always zero					
	Sanket	**Ronit**	**Shruti**	**Shivam**	**Shlok**
Sanket	0	1	0	1	0
Ronit	1	0	1	0	0
Shruti	0	1	0	1	1
Shivam	1	0	1	0	1
Shlok	0	0	1	1	0

If we have a directed network, the matrix will not necessarily be symmetric.

- In the examples we have seen so far, we have been recording a 1 in the matrix to indicate an edge is present, and a 0 when there is no edge. This scheme can be altered to show the weight of an edge as well. To do this, we replace the 1 with the edge weight. Using the values from Fig. 2.1.5, we would have a weight of 3 between Shruti and Shivam. The matrix would look like this:

Table 2.1.3

The Adjacency Matrix for student's Network with edges weight					
	Sanket	**Ronit**	**Shruti**	**Shivam**	**Shlok**
Sanket	0	1	0	1	0
Ronit	1	0	1	0	0
Shruti	0	1	0	3	1
Shivam	1	0	3	0	1
Shlok	0	0	1	1	0

❑ What is a social media node?

- In network science, nodes are treated as actors (the dots or circles on the graph) and edges as relationships (the lines on the graph).
- Several actors can be represented by nodes. For example, web pages are treated as nodes in internet networks and in case of any social network peoples are treated as nodes.

❑ What is the edge in social media networks?

In mathematical terms, edge refers to the line that connects more than one node.

❑ What is the tie in social media networks?

- The edge that exists between two nodes is called a tie. For example, in the case of facebook network friendship between two peoples is represented by edge. Two types of tie are undirected and directed.

(1) **Undirected tie :** Undirected tie represents the relationship between two nodes which is the same in both directions. For example, if two people are friends on facebook, then it can be represented as an undirected tie.

(2) **Directed tie :** Directed tie represents edges between two nodes in which the relationship is uneven. For example, if Kavya is following Sachin on instagram, then there will be a directed tie from node Kavya to node Sachin.

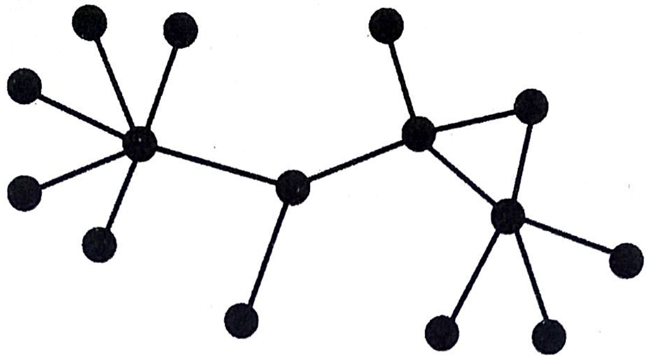

Fig. 2.1.9 : Sample Network Structure

- A Sociogram or simply graph consists of set of (V, E) where,

  V is set of nodes or actors and

  E is set of Edges/ties/links

Table 2.1.4 : Graph terminologies

Origin	Actors	Relationships
Mathematics	Vertex	Edge
Sociogram	Node	Tie/Link

2.2 DESCRIBING THE NETWORKS MEASURES

A number of measures can be used to describe the structure of a network as a whole.

2.2.1 Degree Distribution

- For degree distribution, let's first understand what is the degree of a node.
- Degree of a node is how many nodes are connected to a particular node in a network. The degree distribution P(k) of a network is then defined to be the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and nk of them have degree k, we have

$$P(k) = \frac{nk}{n}$$

Pdeg (k) = fraction of nodes in the graph with degree k.

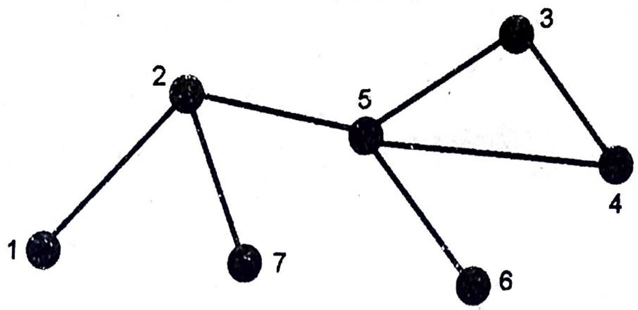

Fig. 2.2.1 : A Simple Network to understand degree distribution

- For this undirected network, degree of each node is

  Degree of node (1) = k1 = 1

  Degree of node (2) = k2 = 3

  Degree of node (3) = k3 = 2

  Degree of node (4) = k4 = 2

  Degree of node (5) = k5 = 4

  Degree of node (6) = k6 = 1

  Degree of node (7) = k7 = 1

- Find out the no. of nodes with degree 1, 2 , 3 and so on

  No. of Nodes with degree 1 are: 3 (node 1, node 6 and node 7)

  No. of Nodes with degree 2 are: 2 (node 3 and node 4)

  No. of Nodes with degree 3 are: 1 (node 2)

  No. of Nodes with degree 4 are: 1 (node 5)

- And degree distribution is calculated as

  P (1) = nk/n = 3/7

  P (2) = nk/n = 2/7

  P (3) = nk/n = 1/7

  P (4) = nk/n = 1/7

  And all other P(k) as zero

2.2.2 Density

- Density refers to the connections between the nodes. Density can be defined as total no. of edges divided by total possible no. of edges in the graph.

  Density = Total no. of edges/ total possible no. of edges

No. of possible edges are calculated as

Total possible no. of edges = N (N-1)/2

Where,

N = total no. of nodes in graph

Density will be 100%, if all possible connections are there in a graph.

Example 2.2.1 : If there are 9 nodes in a graph and no edges are 13. Find the density of Network?

☑ **Solution :**

Density is defined as

Density = Total no. of edges/ total possible no. of edges

Given,

Nodes in a graph: N = 9

No. of edges are: 13

To calculate possible connections from each node are

$$\text{Total possible no. of edges} = \frac{N(N-1)}{2}$$

$$= \frac{9 * 8}{2} = 36$$

Therefore, density can be calculated as

Density = Total no. of edges/ total possible no. of edges

Density = 13/ 36 = 0.36

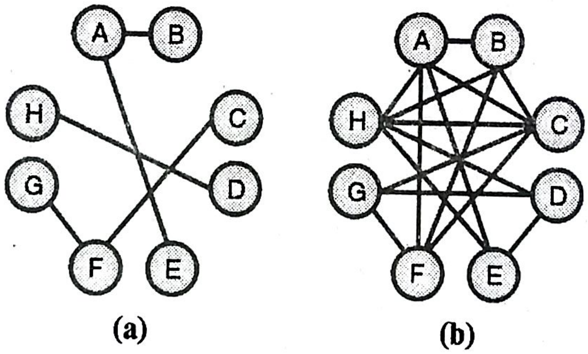

Fig. 2.2.2 : Network (a) on the left has fewer edges than network (b) on the right. Since they both have the same number of nodes and thus the same number of possible edges, network (a) is denser

2.2.3 Connectivity

- Density measures the percentage of possible edges in a graph. *Connectivity*, also known as *cohesion*, measures how those edges are distributed. Connectivity is a count of the minimum number of nodes that would have to be removed before the graph becomes disconnected; that is, there is no longer a path from each node to every other node.

- In Fig. 2.2.3, the connectivity is 1 because removing node B, C, or D would disconnect the graph. Since removing any one of those nodes disconnects the graph, the connectivity is 1. In Fig. 2.2.4, the connectivity is 2. Removing any one node would not break the graph into two parts, but there are several options for removing two nodes that would. For example, removing nodes E and F would separate G from the rest of the graph. If we removed B and D instead, node A would become separated.

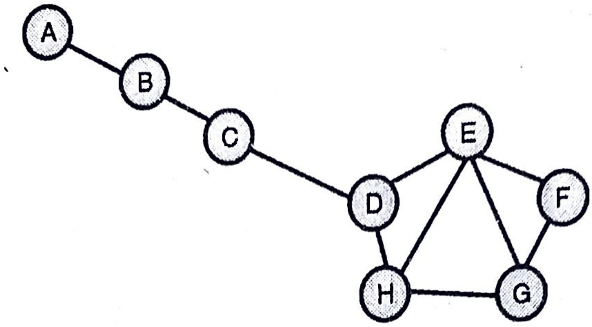

Fig. 2.2.3 : A Network with connectivity as 1

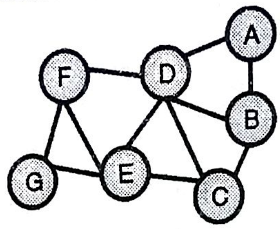

Fig. 2.2.4 : A Network with connectivity as 2

2.2.4 Centralization

- Graph representation is used while mining social media. This representation describes the friendship or the interactions over social media.
- For the graph representation, there is a question-
  Which nodes are central (famous individuals) in the network?
- The question can be answered with the help of centrality network measures. With the help of centrality we can find central nodes in a network.
- Importance of a node in a network is defined by centrality.

Several types of centrality are :

Degree Centrality

- The person who has many connections is considered as important. Same concept is transformed into measure as degree centrality. The degree centrality measures the no. of connections a node has with a network. Degree centrality is calculated at node level.
- For undirected graph, the degree centrality Cd for node vi is

$$Cd\ (vi) = di$$

Where di denotes the degree of node vi.

- Degree is nothing but the number of adjacent edges of a node in a graph or network.
- For directed graphs, in-degree or out-degree is used. We can also use the combination as the degree centrality value:

$$Cd\ (vi) = din \qquad \text{(prestige)}$$
$$Cd\ (vi) = dout \qquad \text{(gregariousness),}$$
$$Cd\ (vi) = din + dout$$

- Popularity of node is calculated by considering in-degrees, also called as PROMINENCE is and its value shows prominence or prestige.
- For measuring whether a person is sociable or not, out-degrees are used and it's called the gregariousness of a node.
- If we are using both in-degrees and out-degrees, edge directions are ignored and its similar to an undirected graph.
- Comparison of centrality values across the network is not allowed by this measure, so degree centrality is normalized.

Closeness centrality

- Closeness centrality indicates how close a node is to all other nodes in the network. It is calculated as the average of the shortest path length from the node to every other node in the network. Consider Fig. 2.2.3.
- Let's start by computing the average shortest path length of node D.

Table 2.2.1 : Shortest Path From node D

Node	Shortest Path From D	Path
A	3	D-C-B-A
B	2	D-C-B
C	1	D-C
E	1	D-E
F	2	D-E-F
G	2	D-H-G
H	1	D-H
Average	1.71	

By taking the average of all shortest paths from D, closeness centrality for node D is 1.71.

Betweenness centrality

- Betweenness centrality measures how important a node is to the shortest paths through the network. To compute betweenness for a node N, we select a pair of nodes and find all the shortest paths between those nodes. Then we compute the fraction of those shortest paths that include node N.
- If there were five shortest paths between a pair of nodes, and three of them went through node N, then the fraction would be 3 / 5 = 0.6. We repeat this process for every pair of nodes in the network. We then add up the fractions we computed, and this is the betweenness centrality for node N.

Eigenvector Centrality

- Degree centrality takes into consideration no of nodes connected to the node i.e. neighbors of the node. But in real life having more friends is not important, rather having more important friends indicates a strong signal.
- Degree centrality is generalized to eigenvector centrality by adding importance to its neighbors. In case of directed graphs, incoming edges are counted. Eigenvector centrality is defined for both directed and undirected graphs.

2.3 TIE STRENGTH & TRUST

Tie Strength

- To analyze tie strength in social network analysis, the network must include relationship information. In small networks, especially if data is hand-collected, it may be feasible to ask each person to rate the strength of their tie to each person.
- By necessity, larger networks require a mechanism for measuring tie strength. There is no single factor that defines a strong or weak tie, but a number of predictors can be combined to estimate the strength of a relationship.
- The strength of the tie is measured with several characteristics. The characteristics are

  (1) The length of time,

  (2) The intensity of emotions,

  (3) The attachment (reciprocal services), and

  (4) The mutual assistance (intimacy) which characterize the tie
- Time can include the amount of time people spend with each other, the duration of their relationship (i.e., how long they have known each other), and how frequently they see one another.

- Emotional intensity is indicated by the closeness of a relationship; close friends or family members are likely to be strong ties, while more casual friends and acquaintances would be weaker ties.
- Intimacy, or mutual confiding, relates to people sharing secrets or intimate personal details with one another. The more of this information they exchange, the closer their relationship is likely to be.
- Reciprocal Services are favors that people do for one another. They may be personal (e.g., pet sitting or picking up someone's dry cleaning), financial (e.g., loaning money), professional (e.g., putting people in contact with one another), or otherwise.

☞ Intimacy

- Number of days since their last communication
- Number of friends in common
- Number of "intimate" words in their communications, as determined by software that automatically analyzes text.

☞ Intensity

- Number of words exchanged on one another's walls
- Depth of email threads in their inboxes (i.e., how many messages were sent back and forth in a conversation)

☞ Reciprocal services

- Number of links shared on one another's wall
- Applications the users had in common (presumably because they could be working together within the application context)

☞ Social distance

- Age difference
- Difference in the number of educational degrees
- Difference in the number of occupations

Together, these variables were used to try to predict the tie strength of two people's relationship.

There are two types of ties

(1) Weak social tie, the link between individuals who are not interacting frequently. The people who do not have regular interaction.

(2) Strong social ties, the link between the individuals who interact frequently such as link between close friends.

Trust

- The person being trusted is expected to do the "right" thing. This usually means she will act with the other person's best interests in mind and/or take actions that benefit the other person.
- The person making the decision about whether or not to trust someone is con-sidering more than just her expectations about the other person's actions. She must also decide if she is willing to take some risk by putting her trust in the other person.
- That may be a small risk or a large one. Receiving and acting on a poor movie recommendation may only waste a few hours of time, but making a large loan that is not repaid can have major effects.
- So can asking for a recommendation letter from someone who will not write a good one.
- All of these ideas can be condensed down into several important factors. First, the person doing the trusting must make herself vulnerable and take some risk by trusting the other person.
- Second, she takes that risk because she believes the other person will act well or behave in a way that will benefit her.
- Vulnerability, risk, and positive expectations of the other person are the core components of the trust relationship.
- Thus, as a definition of trust, we can say the following: A person trusts another if she is willing to take a risk based on her expectation that the trusted person's actions will lead to a positive outcome.

2.4 NETWORK VISUALIZATION

Graph Layout

- What can make a good graph or layout is not always clear. Graph layout depends on analyst want to see the information and following points
  1. Which network is going to be viewed?
  2. What are the different features of the network?
- There are some guidelines proposed by researchers and they are
  a. Each node in a graph must be visible.
  b. Count the degree of each node
  c. Each link must be easily followed from source to destination
  d. Clusters and if any outliers are there they must be easily identified.

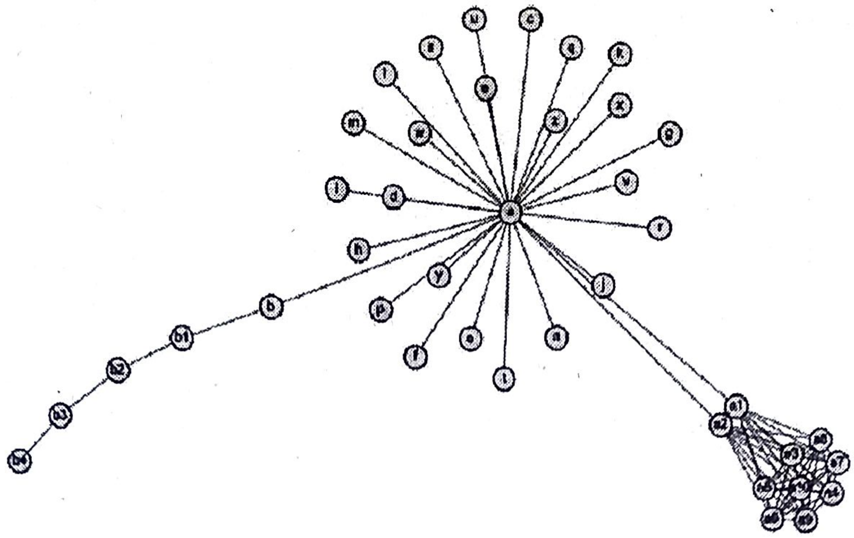

Fig. 2.4.1 : A sample of network visualization

Random layout

- When the visualization tool takes the data, it is placed randomly. This technique is called random layout and does not provide much insights as data is placed randomly.
- Patterns, insights are difficult with the help of these layouts. We cannot answer which node has the highest degree. Fig. 2.4.2 shows the same network from Fig. 2.4.1 presented in a random layout.

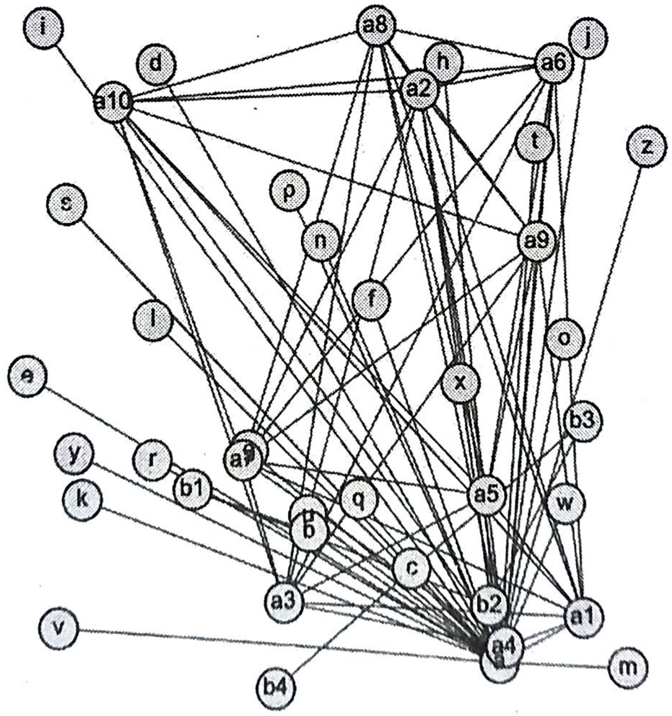

Fig. 2.4.2 : Random Layout

☞ Circular layout

- Circular layout connects the nodes in a circular pattern. When the no. of nodes are greater in network it is placed closely to each other.
- Circular position nodes are then connected by edges. Another way is to place nodes in grid manner.

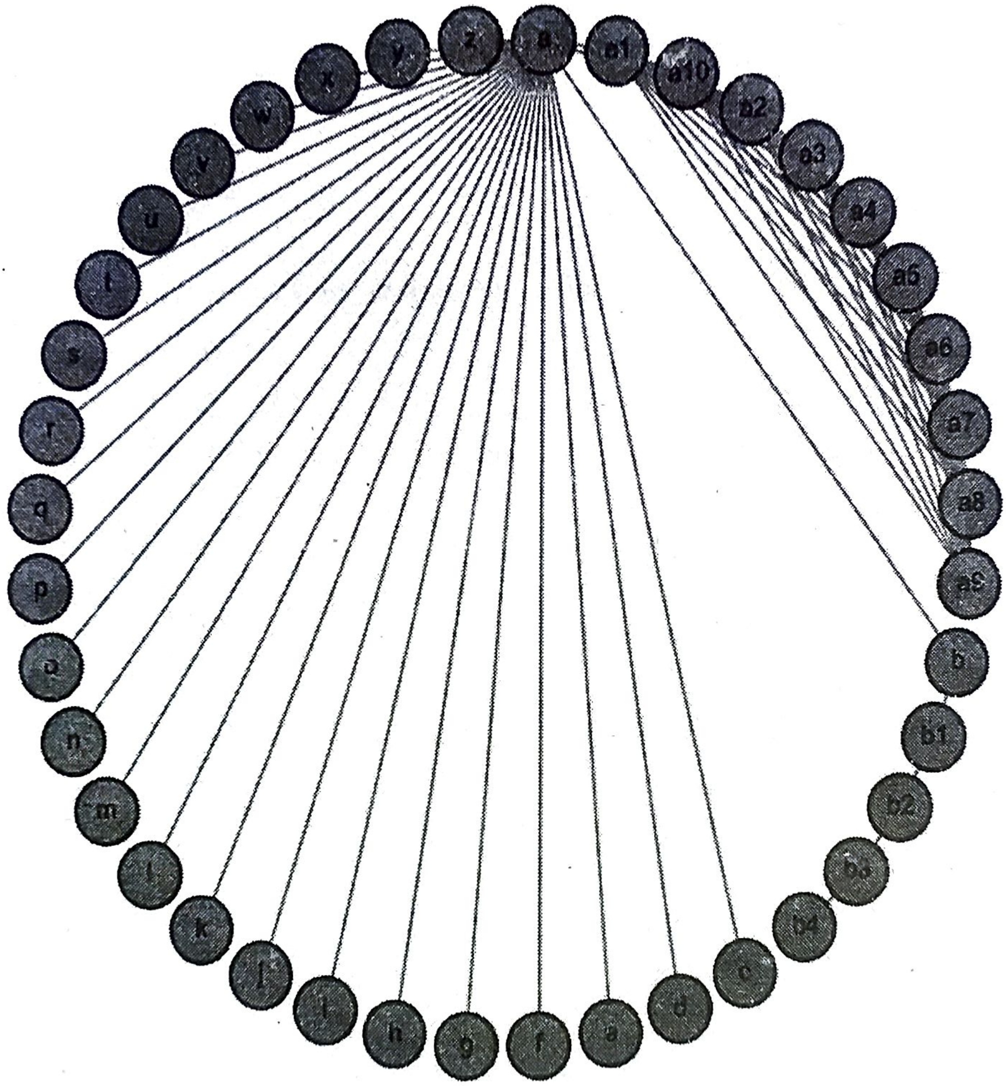

Fig. 2.4.3 : Circular layout of graph of fig. 2.3.1

☞ Grid Layout

- Figure shows an example of grid layout. Note that the degree of node a is clearly high, the cluster of nodes a1 through a10 are obvious, and chain nodes through b4 are clear across the top.

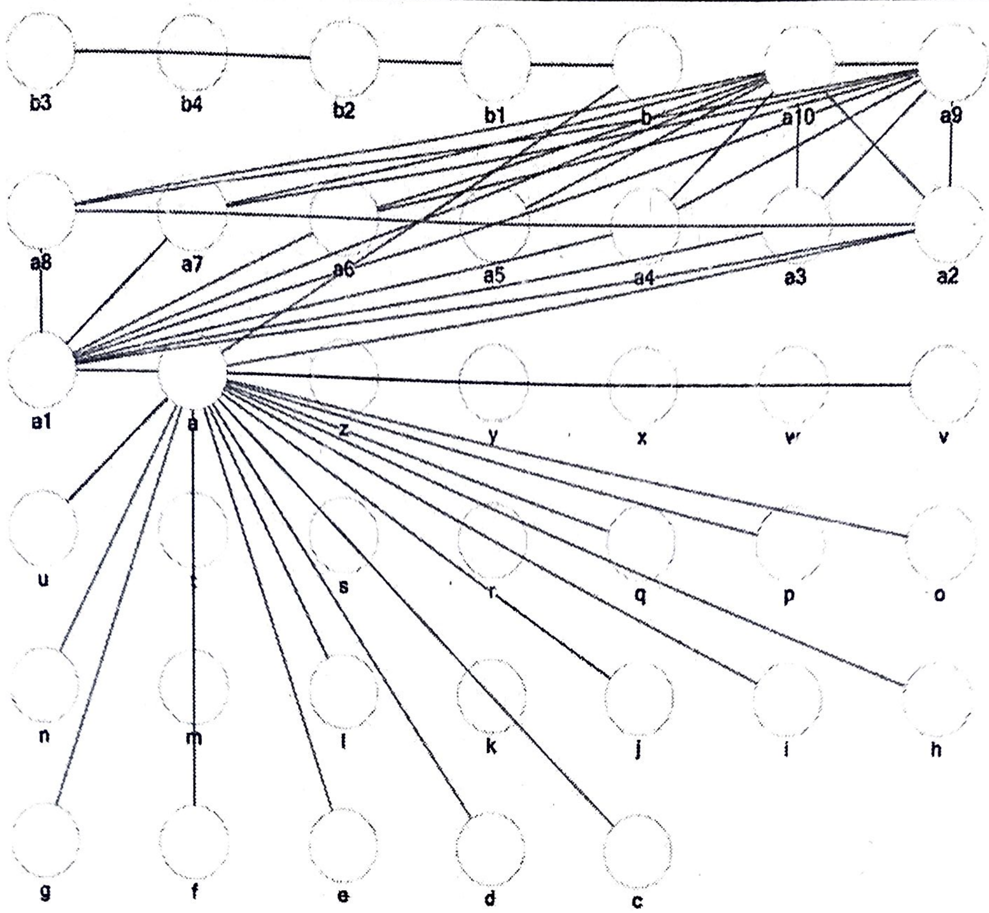

Fig. 2.4.4 : Grid Layout of the nodes in sample graph

2.5 VISUALIZING NETWORK FEATURES

- Graph layout accomplishes the placement of nodes and edges in a graph. There are other features like edge weight, node properties, labels and clusters can also be visualized.
- Like graph layout, there are several options for this. We are going to focus on features in this section.

Labels

- Labels are very hard to set as we have to give labels to nodes as well as edges. When the graphs are small we can provide labels to both nodes and edges but it becomes too small to visualize.
- Consider a network with 92 nodes (Fig. 2.5.1), which are relatively small as compared to social media networks. If nodes represent peoples and edges represent connections then giving labels to nodes and edges is difficult as well as not clearly visible.
- Nodes as well as edge labels are hard to give because of slanting edges or alignment of edges. There are still some options by which labels can be seen. To show the labels for nodes and edges of interest or keeping a box around the text.

Fig. 2.5.1 : A network of YouTube videos with the node labels shown.

☞ Size, Shape and Color

- There are so many metrics available like centrality, degree etc. These metrics can be added by size, color or both. Highest degree node can be shown in the graph.

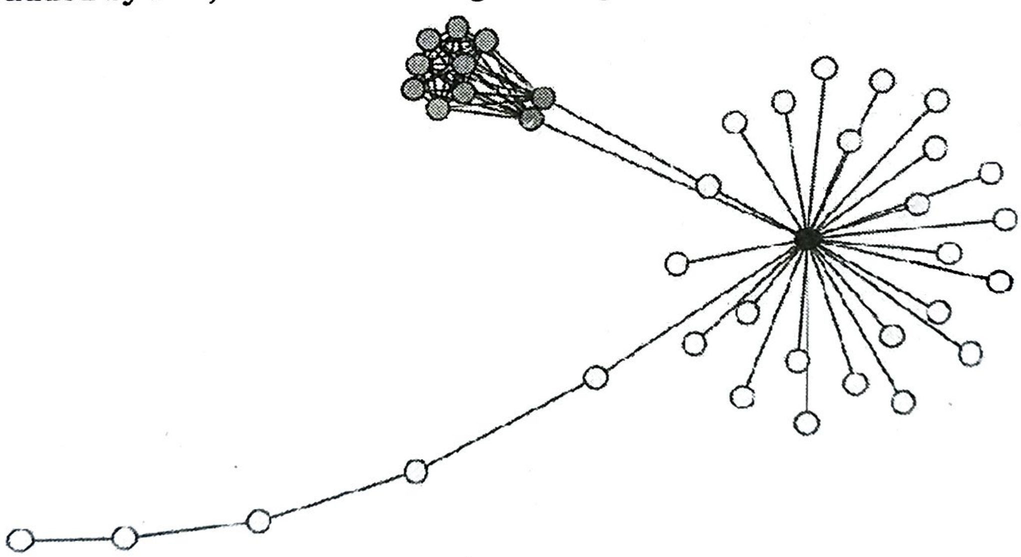

Fig. 2.5.2 : Color-coding nodes according to their degree, with higher degree shown by darker nodes

- Fig. 2.5.2 shows color encoding of node degree. Darker colors indicate nodes with higher degrees, and not surprisingly, node a(central nodes) is the darkest. For clarity, the node labels have been left off this graph.
- Node color is used to indicate friends, groups, family members, co-workers, classmates and so on. Clustering coefficient is added with the help of node size.

When clusters are tight where all nodes are connected to each other, the size of the node is large.

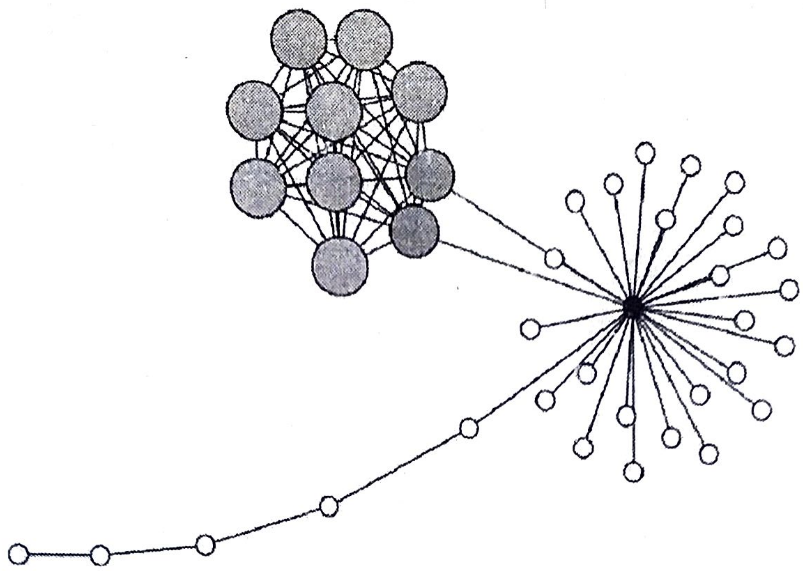

Fig. 2.5.3 : A graph indicating clustering coefficient with node size and degree with node color

Fig. 2.5.3 shows a graph that uses color for degree and size for clustering coefficient.

- Edges can also be treated with color and thickness to indicate the attributes. Edge weights are indicated which denote the frequency of communication, relationship strength. If frequency of communication is greater than edge can be represented as thick edge.

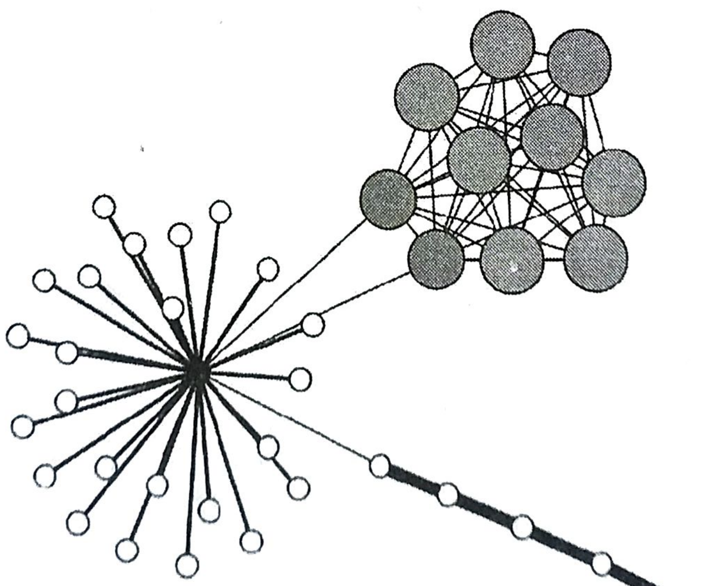

Fig. 2.5.4 : The sample network with edge width indicates the weight on each edge. Note that the central node has medium-strength relationships with most neighbors, but weak ones to the cluster in the upper right and the chain in the lower right. The chain of nodes in the lower right has high weights on the edges connecting them

- Fig. 2.5.4 shows the same example network with weights added to the edges. These are visualized by adjusting the width of the edge. Wider edges indicate stronger relationships.

☞ Larger graph properties

- Larger graph properties can also be encoded in visualizations. For example, clusters are sometimes apparent on their own, but visual properties to indicate them will often clarify a visualization further.
- Fig. 2.5.4 shows a new graph that has two main clusters. This graph is a network of YouTube videos, where nodes represent videos and edges connected videos that share a common tag.
- All of these videos were tagged with the word "cubs". Even without the color coding, the two groups would be relatively easy to see. But using a community detection algorithm that groups nodes into clusters, and then color coding by those clusters, makes it even more apparent. This is shown in Fig. 2.5.5.

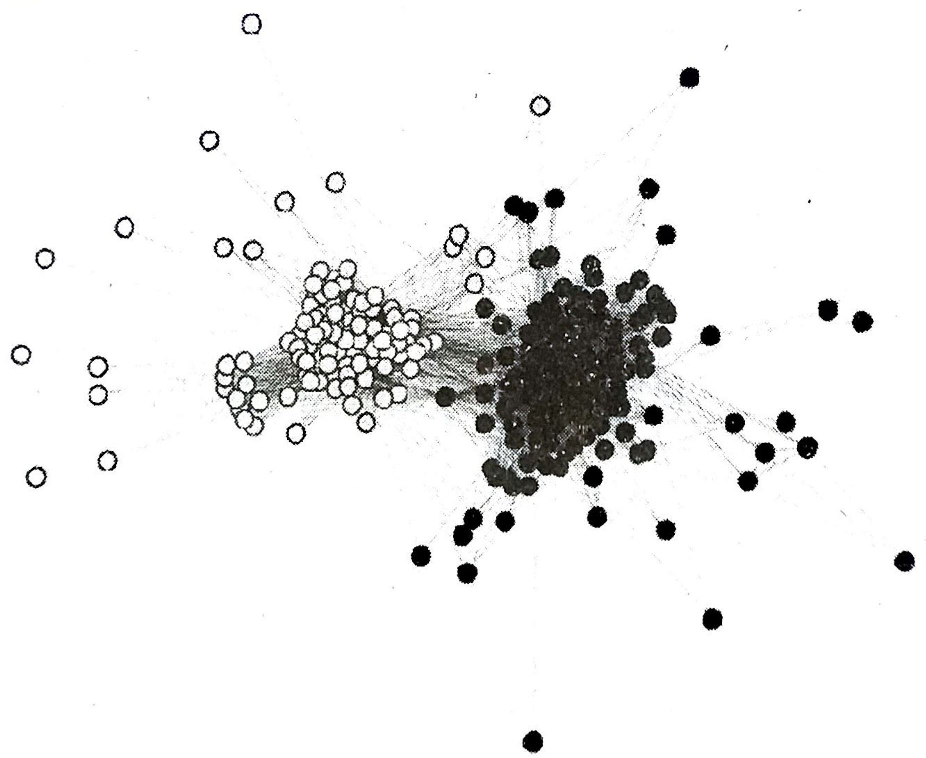

Fig. 2.5.5 : A network of YouTube videos where color indicates the community or cluster to which each node belongs

Scale Issues

- Networks which are discussed so far are with some hundreds no of nodes and thousand no. of edges. Visualization can be better with small networks. But when no. of nodes are added in a graph then visualization of nodes and edges diminishes.

- Generally the networks under 1000 nodes are safest networks

Density

- When there are fewer nodes, then also density is a problem for visualization. Fig. 2.5.4 has the edges filtered so that only those with a weight of 40% or more are visible. However, as this network shows, there are no interesting pat- terns visible with the threshold of 40%; the network is simply too dense.

Filtering for visual patterns

- It is often difficult to see any patterns in very dense networks. One way to compensate for this is to filter the networks when possible. For example, if we take the same network from Fig. 2.5.6 and filter the edges so that they only connect senators who have voted the same way on at least two-thirds of the bills, the pattern changes dramatically. This is shown in Fig. 2.5.6.
- In this figure, two clear clusters emerge, representing the two major political parties. Furthermore, five senators are pulled out from the major party clusters along the center, indicating that they frequently vote with members of both parties.

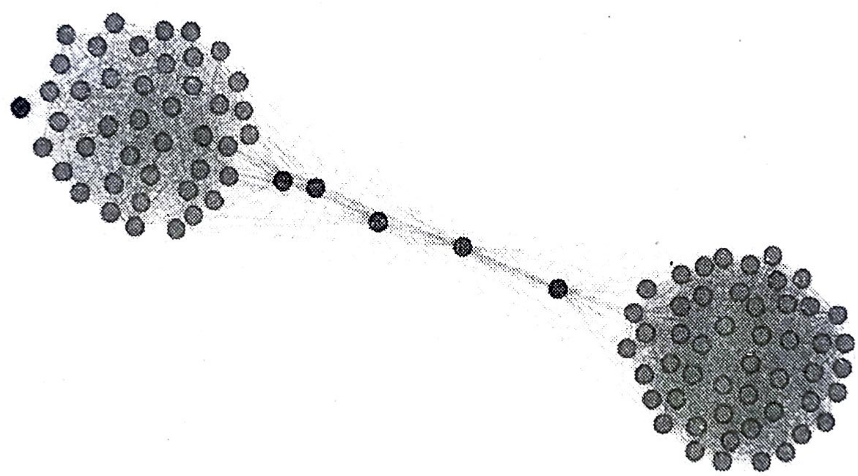

Fig. 2.5.6 : The same network of 2.5.4 is filtered to include only edges between senators who have voted same way on at least two-third of bills

2.6 SOCIAL MEDIA NETWORK ANALYTICS

- Networks are the building blocks of social media and can carry useful business insights. Social media network analytics thus deals with constructing, analyzing, and understanding social media networks.
- Social network analytics can be used for a variety of purposes. It can be employed to identify influential nodes (e.g., people and organizations) or their position in the network, or to understand the overall structure of a network.

- Overall, the purpose of network analysis is to
  - Understand overall network structure; for example, number of nodes, number of links, density, clustering coefficient, and diameter.
  - Find influential nodes and their rankings; for example, degree, betweenness, and closeness centralities.
  - Find important links and their rankings; for example, weight, betweenness, and centrality.
  - Find cohesive subgroups; for example, pinpointing communities within a network.
  - Investigate multiplexity; for example, analyzing comparisons between different link types, such as friends vs. Enemies

☞ Common Network Terms

Network

- At a very basic level, a network is a group of nodes that are connected with Nodes (also known as vertices) can represent anything, including individuals, organizations, countries, computers, websites, or any other entities. Links (also known as ties, edges, or arcs) represent the relationship among the nodes in a network. Networks can also exist among animals.

A social network

- A social network is a group of nodes and links formed by social entities where nodes can represent social entities such as people and organizations. Links represent their relationships, such as friendship and trade relations. Social networks can exist both in the real and online worlds.
- A network among classmates is an example of a real world social network. And a Twitter following network is an example of an online social media network. In a Twitter follow-following network, nodes are the Twitter users, and links among the nodes represents the follow-following relationship (i.e., who is following whom) among the users. The subject of this book is online social networks.

A social network site

- A social network site is a special-purpose software (or social media tool)designed to facilitate the creation and maintenance of social relations.
- Facebook, Google+, and LinkedIn are examples of social network sites.

Social Networking

- The act of forming, expanding, and maintaining social relations is called social networking.
- Using social network sites, users can, for example, form, expand, and maintain online social ties with family, friends, colleagues, and sometimes strangers.

Social network analysis

- Social network analysis is the science of studying and understanding social networks and social networking.
- It is a well established field with roots in a variety of disciplines including Graph Theory, Sociology, Information Science, and Communication Science.

2.6.1 Common Social Media Network Types

There are several types of social media networks; we will discuss a few of them. Social media network types are as follows

Friendship Networks

- The most common type of social media networks are friendship networks, such as Facebook, Google+. Friendship networks let people maintain social ties and share content with people they are closely associated with, such as family and friends.
- Nodes in these networks are people, and links are social relationships (e.g., friendship, family, and activities).

Follow-Following Networks

- In the follow-following network, users follow (or keep track of) other users of interest. Twitter is a good example of follow-following network where users follow influential people, brands, and organizations.
- Nodes in these networks are, for example, people, brands, and organizations, and links represent following relations (e.g., who is following whom). Below are two common Twitter terminologies.
- Following - Following are the people who you follow on Twitter.
- Following someone on Twitter means :
  - You are subscribing to their tweets as a follower.
  - Their updates will appear in your Home tab.
  - That person is able to send you direct messages.
- Followers - Followers are people who follow you on Twitter. If someone follows you, it means that :
  - They will show up in your followers list.
  - They will see your tweets in their home timeline whenever they log in to Twitter.
  - You can send them direct messages.

Fan Network

- A fan network is formed by social media fans or supporters of someone or something, such as a product, service, person, brand, business, or other entity.

- The network formed by the social media users subscribed to your Facebook fan page is an example of a fan network.

☞ Group Network

- Group networks are formed by people who share common interests and agendas. Most social media platforms allow the creation of groups where members can post, comment, and manage in-group activities.
- Examples of social media groups are Twitter professional groups, Yahoo Groups and Facebook groups.

☞ Professional Network

- LinkedIn is a good example of professional networks where people manage their professional identity by creating a profile that lists their achievements, education, work history, and interests.
- Nodes in these networks are, for example, people, brands, and organizations, and links are professional relations

☞ Content Network

- Content networks are formed by the content posted by social media users. A network among YouTube videos is an example of a content network.
- In such a network, nodes are social media content (such as videos, tags, and photos) and links can represent, for example, similarity (content belonging to the same categories that can be linked together).

☞ Dating Network

Dating networks (such as match.com and Tender) are focused on matching and arranging a dating partner based on personal information (such as age, gender, and location) provided by a user

☞ Co Occurrence Network

- Co occurrence networks are formed when two more entities (e.g., keywords, people, ideas, and brands) co occur over social media outlets.
- For example, one can construct a co-occurring network of brand names (or people) to investigate how often certain brands (or people) co-exist over social media outlets. In such networks, nodes will be the brand names and the links will represent the co-occurrence relationships among the brands.

☞ Geo Coexistence Network

- Geo coexistence networks are formed when two more entities (e.g., people, devices, and addresses) coexist in a geographic location.
- In such a network nodes represents entities (e.g., people), and links among them represent coexistence.

☞ Hyperlink Network

- In simple words, hyperlink is a mechanism to move among electronic documents (such as websites). Hyperlinks can be referred to as being either in-links, thus bringing traffic/users to your website) or out-links (i.e., links originating in your website and going out, thus sending traffic to other websites). Hyperlink also forms networks.
- Typically, in these network nodes are websites and links represent referral relationships (in the form of in-links or out-links).

☞ Types of Networks

- From a technical point of view, the above-mentioned networks can be classified in a variety of ways, including
  - based on existence,
  - based on direction of links,
  - based on mode, and
  - based on weights.

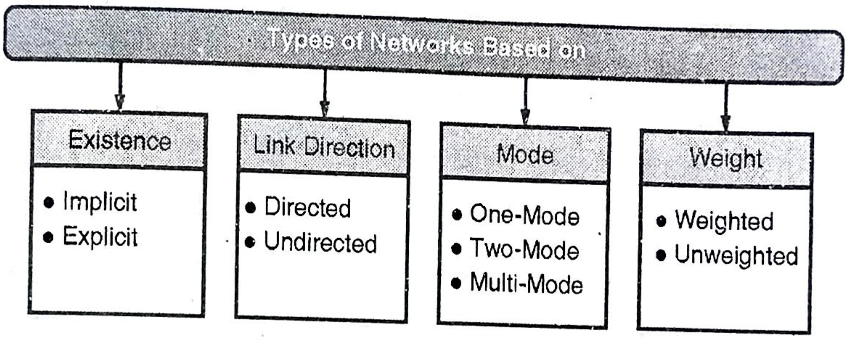

Fig. 2.6.1 : Types of Networks

Based on existence

Based on the way the networks exist online or are constructed, they can be classified as

(1) Implicit networks or (2) Explicit networks

(1) Implicit Networks

- Implicit networks do not exist by default (or are hidden) and need to be intentionally constructed with the help of dedicated tools and techniques.
- Examples of such networks include keyword co occurrence networks, co citation networks, hyperlink networks, etc.
- Constructing and studying implicit networks can provide valuable information and insights

(2) Explicit Networks

- Explicit social media networks exist by default; in other words, they are explicitly designed for social media users to be part of.
- Most social media networks are explicit in nature.
- Examples of explicit social media networks include Facebook friendship network, Twitter follow-following networks, LinkedIn professional networks, YouTube subscribers' network, and bloggers' networks.

Based on direction

Based on the directions of links among the nodes, the networks can be classified as

(1) Directed networks, and
(2) Undirected networks

(1) Directed networks

- A network with directed links among nodes is called a directed network.
- Usually, a link with an arrow is drawn to show the direction of the relationship among the nodes. For example, the Twitter following-following network is a directed network where the direction of the arrow shows who is following whom.

(2) Undirected Network

- In undirected networks, the links among the nodes do not have any direction.
- A Facebook friendship network is an example of an undirected network.

Based On Mode

Based on the composition of nodes, networks can be classified as :

(1) One-mode network,
(2) Two-mode networks, and
(3) Multimode networks

(1) One-Mode Networks

- A one-mode network is formed among a single set of nodes of the same nature.
- A Facebook friendship network is an example of a one-mode network where nodes (people) form network ties (friendships).

(2) Two-Mode Networks

- Two-mode networks (also known as bipartite networks) are networks with two sets of nodes of different classes. In these networks, network ties exist only between nodes belonging to different sets.

- For example, consider the two-mode network, where one set of nodes (circles) could be social media users and another set of nodes (squares) could be participation in a series of events. Users are linked to the events they attended.

(3) Multimode Network

- A multimode network is also possible where multiple heterogeneous nodes are connected together.
- It can be considered as an amalgam of one and two-mode networks.

Based on Weights

Networks can also be classified based on the weight assigned to the links among the nodes. Mainly there are two types of weighted networks :

(1) Weighted networks, and
(2) Unweighted networks

(1) Weighted Networks

- In weighted networks, the links among nodes bear certain weights o indicate the strength of association among the nodes.
- The link (relationship) between, for example, two Facebook friends (nodes) will be thicker if they communicate more frequently.
- Weighted networks can provide rich information, but are difficult to construct.

(2) Unweighted Networks

- In unweighted networks, links among nodes do not bear weights.
- The links only indicate the existence of a relationship and cannot provide clues about the strength of the relationship. Unweighted networks are easy to construct, but may conceal useful information.

Common Node Terminologies

Network terminologies or properties can be divided into two categories :

(1) node-level properties, and
(2) network-level properties.

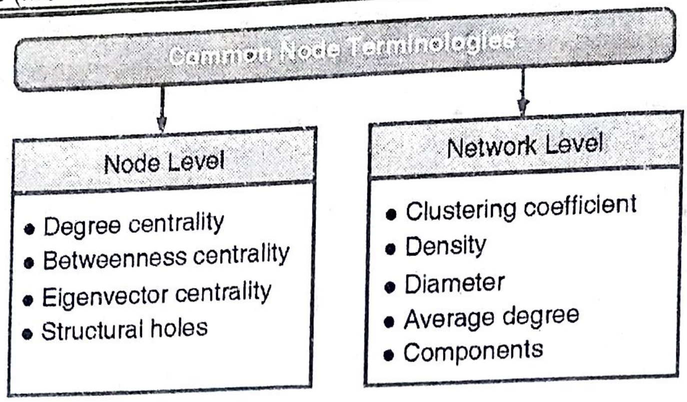

Fig. 2.6.2 : Common Node Terminologies

☞ Node-Level Properties

Node-level properties focus on one node and its position in the network. Some important node properties include degree centrality, betweenness centrality, eigenvector centrality, and structural holes.

Degree Centrality

Degree centrality of a node in a network measures the number of links a node has to other nodes

Betweenness Centrality

Betweenness centrality is related to the centrality (or position) of a node in a network

Eigenvector Centrality

Eigenvector centrality measures the importance of a node based on its connections with other important nodes in a network

Structural Holes

- In a social media network, some users, because of their network position, may have an advantage or disadvantage in terms of opportunities to form and propagate information.
- New ideas and information mostly come from structure holes (or weak ties) that exist in a network. A user with more weak ties can receive novel ideas and information from remote network clusters.

☞ Network Level Properties

Network properties provide insight into the overall structure and health of a network. Important network-level properties include clustering coefficient, density, diameter, average degree, and components.

Clustering Coefficient

The clustering coefficient of a network is the degree to which nodes in a network tend to cluster or group together

Density

The density of a network deals with a number of links in a network. Density can be calculated as the number of links present in a network divided by the number of all possible links between pairs of nodes in a network (for an undirected network, the number of all possible links can be calculated as n (n – 1)/2); where n is the number of nodes in a network).

Components

Components of a network are the isolated sub-networks that connect within, but are disconnected between, sub-networks

Diameter

The diameter of a network is the largest of all the calculated shortest path between any pair of nodes in a network, and it can provide an idea of how long it would take for some information/ideas/message to pass through the network.

Average Degree

The average degree centrality measures the average number of links among nodes in a network.

☞ Network Analytics Tools

- **NodeXL :** NodeXL (an add-in for Microsoft Excel) is the free tool for social network analysis and visualization. It can help you construct and analyze Facebook networks (based on co likes and co comments), Twitter networks (followers, followings, and tweets), and YouTube networks (user network and comments), among others.
- **UCINET :** UCINET is a social network analysis software application for windows operating system. It also includes Netdraw tool for network visualization. It can be downloaded and used for free for90 days :

  https: / / sites.google.com / site / ucinetsoftware / home.
- **Pajek :** Pajek is a software application for analyzing and visualizing large networks (http://mrvar.fdv.uni-lj.si/pajek/). Pajek runs on Microsoft Windows Operating systems and is free for noncommercial use.
- **Netminer :** Netminer (http://www.netminer.com/) is also a software application for large social network analysis and visualization. The application can used be for free for 28 days.

- **Flocker :** Flocker (http://flocker.outliers.es/) is a Twitter real-time retweets and mentions networks analytics tool.
- **Reach :** Reach is an online platform to map hashtag networks and identify the most influential accounts in the Twitter conversation :

  http://www.reachsocial.com/.
- **Mentionmapp :** This online tool is used to investigate Twitter mentions networks (http://mention mapp.com/).

Chapter Ends...

❑❑❑