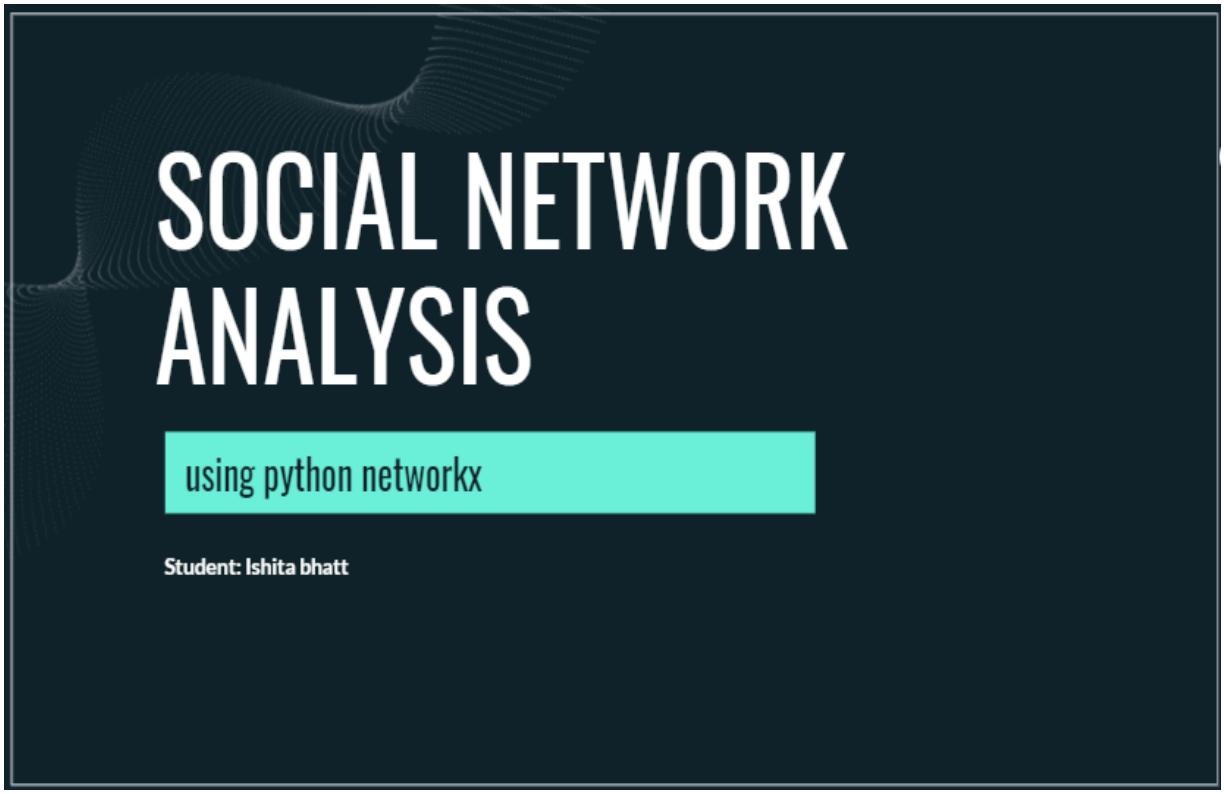


Roll no.-2017647 class roll no.-9

Acknowledgement-

First of all, I would like to thank our institution “Graphic Era deemed to be university” for giving us an opportunity to work on such great projects.

I am also thankful to our mentor Ashwini sir for guiding us through the project.

MOTIVATION-

I chose this project because I feel like this project will help me understand more about languages like python, and its different user-friendly libraries.

The project also helps me understand a little more about how to work with a dataset and use it to analyse different types of networks. As I really like to work with data, this project was an amazing chance to learn more about data and how they work, here, specifically for a social network.

Table of contents-

SOCIAL NETWORK AND IT'S ANALYSIS

ANY NETWORK WITH CONNECTIONS BETWEEN INDIVIDUALS, WHERE THE CONNECTIONS CAPTURE THE RELATIONSHIP BETWEEN THEM IS **A SOCIAL NETWORK.**

SOCIAL NETWORK ANALYSIS, A RESEARCH METHOD DEVELOPED PRIMARILY IN SOCIOLOGY AND COMMUNICATION SCIENCE, FOCUSES ON PATTERNS OF RELATIONS AMONG PEOPLE AND AMONG GROUPS SUCH AS ORGANIZATIONS AND STATES.

SNA HAS ITS ORIGIN IN BOTH SOCIAL SCIENCE AND IN BROADER FIELDS OF **NETWORK ANALYSIS AND GRAPH THEORY**.

NETWORK ANALYSIS IS THE FORMULATION AND SOLUTION OF PROBLEMS THAT HAVE A NETWORK STRUCTURE (USUALLY IN A GRAPH).

GRAPH THEORY PROVIDES A SET OF ABSTRACT CONCEPTS AND METHODS FOR ANALYSIS OF GRAPHS.

SNA IS UNIQUE AND NOT JUST ONLY METHODOLOGY AS IT GIVES US A PERSPECTIVE ON HOW SOCIETY WORKS .IT CENTERS ON RELATIONS BETWEEN INDIVIDUALS, GROUPS OR SOCIAL INSTITUTIONS.

A social network has two basic elements and an optional third:

• a set of nodes(aka vertices)

- objects within the network.

• edges or sets of ties (aka relations, connections, edges, arcs

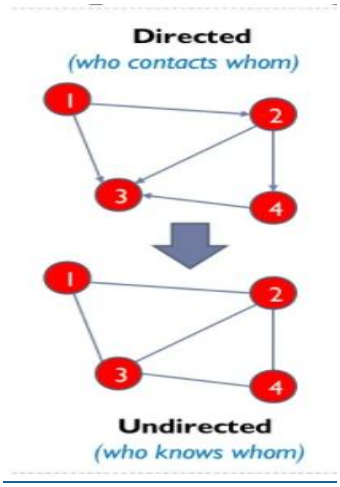
-ties can be directed or undirected.

WHY AND WHEN TO USE SNA

- TO STUDY A SOCIAL NETWORK AND TO IMPROVE EFFECTIVENESS OF THE NETWORK.
- TO VISUALISE DATA SO AS TO UNCOVER PATTERNS IN RELATIONS OR INTERACTIONS.
- TO FOLLOW THE PATHS THAT THE INFORMATION FOLLOWS IN SOCIAL NETWORKS.
- TO IDENTIFY CAUSES OF DYSFUNCTION IN A NETWORK AND PROMOTE COHESIVENESS AND GROWTH.
- FOR A QUANTITATIVE RESEARCH, ALSO VALUABLE FOR THE QUALITATIVE RESEARCH.

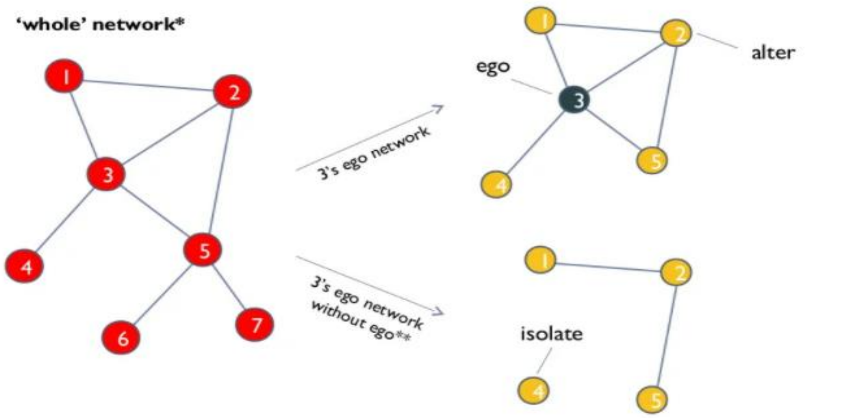
TYPES OF NETWORKS AND ITS TIE STRENGTH

- UNDIRECTED GRAPH NETWORK- HAVE EDGES THAT DO NOT HAVE A DIRECTION.
- DIRECTED GRAPH NETWORK- IS A SET OF VERTICES AND A COLLECTION OF DIRECTED EDGES THAT EACH CONNECTS AN ORDERED PAIR OF VERTICES.

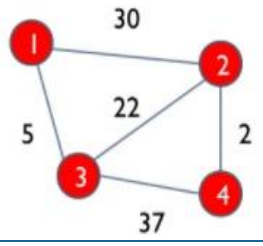


- EGO NETWORK- CONSIST OF A FOCAL NODE ("EGO") AND THE NODES TO WHOM EGO IS DIRECTLY CONNECTED TO (THESE ARE CALLED "ALTERS") PLUS THE TIES, IF ANY, AMONG THE ALTERS.

- WHOLE NETWORK–CONSISTS OF ALL NECESSARY AND IMPORTANT NODES IN A NETWORK. EGO NETWORKS ARE FORMED FROM WHOLE NETWORK. WHOLE NETWORKS ALSO ARE ALMOST IMPOSSIBLE TO FIND OR FORM.



- STRENGTH: EDGE WEIGHTS–ALL THE CONNECTED NODES MAY HAVE SOME WEIGHTS (WHICH MAY BE DEFINED AS SIGNIFICANCE OF THE CONNECTION BETWEEN 2 NODES),



ANALYSING A SOCIAL NETWORK

- Degree centrality is a measure of the number of connections an individual node has. Someone might be said to be more popular or important if they have high degree centrality.
- Betweenness centrality reveals the people that bridge disparate groups of nodes. They are the hubs that enable

communication between people who are not directly connected.

- Eigenvector centrality is a technique for measuring the influence of nodes in a network. It relies on the overlaying of weights onto all nodes in the network, where the weight value is associated with some characteristics of interest, such as expertise in a given field.

COHESION

- Graph density is a measure of the frequency of connections in the actual graph as compared to a graph in which the maximum possible number of connections could occur.
- Path analysis, like degree centrality, is concerned with many aspects of the edges between nodes.
- A Cluster is a tightly connected group within itself, with fewer or perhaps no ties to other parts of the network. Analysis of the structural cohesion of a cluster is concerned with the nodes within the cluster that act as bridges. These are the nodes with high betweenness centrality within the cluster, which if removed, would cause the cluster to fall apart.
- Reciprocity implies a directionality to a relationship. A reciprocal relationship can be represented by one or two edges, where one edge points at both nodes, or two edges represent the bidirectional nature of the relationship. Thus a social network that reflects reciprocity between nodes is by definition a directed graph.

CASE STUDY

Problem Statement - Analysing a given social network for finding out the:

1. Most influential
2. Most important

3. Best connector

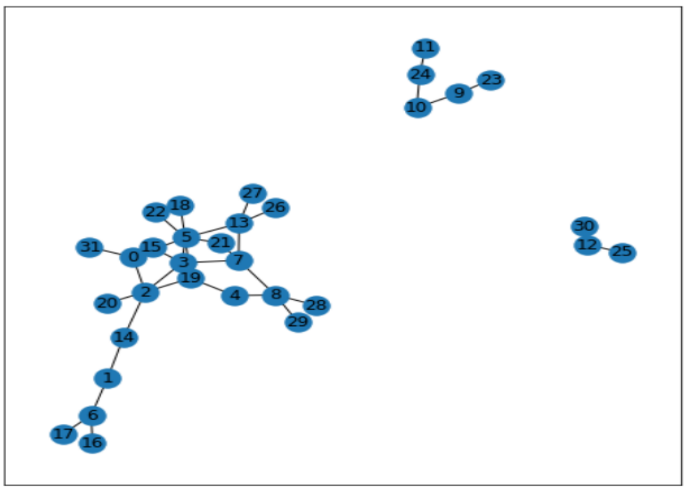
I will be using the platform Kaggle and the language python to analyse. The libraries I have mainly used are matplotlib.pyplot and networkx .

The dataset used is a random dataset (or edge list) which is the relation between which 2 people.

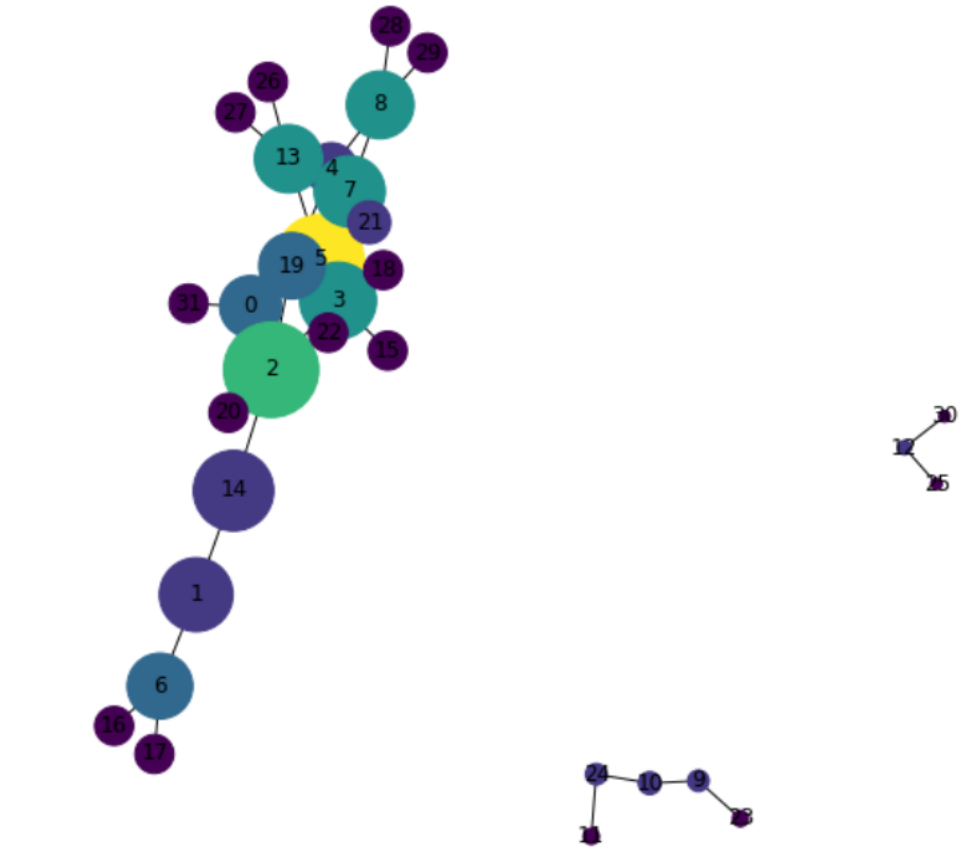
0	2
0	5
1	6
1	14
2	3
2	14
3	15
3	5
4	8
5	13
7	13
3	7
6	16
6	17
5	18
2	19
5	19
4	19

2	20
5	21
7	21
8	7
5	22
9	23
9	10
10	24
11	24
12	25
13	26
13	27
8	28
8	29
12	30
0	31

Graph-



We can also visualize the network such that the node colour varies with **Degree** and node size with **Betweenness Centrality**.



The table below lists the five node labels with the highest centrality measures.

DEGREE CENTRALITY	EIGENVECTOR CENTRALITY	BETWEENESS CENTRALITY
5	5	2
2	3	5
3	2	14
8	7	3
13	13	1

The table shows the nodes in ascending order. The degree centrality indicates the most influential, eigenvector centrality indicates the most important link and betweenness centrality indicates the best connector.

Here, we can see that some nodes are common between Degree Centrality, which is a measure of degree, and Betweenness Centrality

which controls the information flow. It is natural that nodes that are more connected also lie on shortest paths between other nodes. The node 5 is an important node as it is crucial according to all three centrality measures that we had considered.

The density for undirected graphs is

$$d = 2m/n(n-1),$$

and for directed graphs is

$$d = m/n(n-1),$$

where n is the number of nodes and m is the number of edges in G.

The density of the undirected graph using the formula:

0.0570409982174688

AVERAGE CLUSTERING of the graph is 0.0.

Conclusion-

From the above dataset graphs and table, we can conclude some things:

- The edge list or dataset is undirected, so the graph formed is an example of undirected graph.
- The graph shows there is at least 3 clusters of network.
- In the second graph it shows us that 5 is the most important node as its colour is yellow and the colour of the nodes change according to their degree and betweenness centrality.
- We can also figure out that 5 is the most important node as, if we compare all 3 central measures from the table.
- Average clustering is 0.0 and density of the graph is approximately 0.057.

REFERENCE-

Social Network Analysis & Mapping in Python with NetworkX - DataCamp

Social Network Analysis with NetworkX- Working with a DataSet - YouTube

Social Network Analysis - an overview | ScienceDirect Topics

https://www.slideshare.net/gcheliotis/social-network-analysis-3273045