NAME: HARSH TYAGI
REG.NO: 22MCB0022

SOCIAL NETWORK ANALYTICS

ABSTRACT

This project applies social network analysis to study different centrality measures using a graph. We analyzed the network using various measures, including Degree Centrality, Betweenness centrality, Closeness centrality, Eigenvector Centrality, to identify important nodes, subgroups, or patterns of interaction. This project contributes to the growing body of research on social networks and provides new insights.

SOCIAL NETWORK

A social network is a structure made up of individuals or organizations that are connected to each other by social relationships, such as friendship, kinship, common interests, or professional relationships. In a social network, each individual or organization is called a node, and the connections between them are called edges or links. Social networks can be small, like a group of friends, or they can be very large, like Facebook or LinkedIn.

SOCIAL NETWORK ANALYSIS

Social network analysis is the process of investigating social structures through the use of networks and graph theory. This article introduces data scientists to the theory of social networks, with a short introduction to graph theory and information spread. It dives into Python code with NetworkX constructing and implying social networks from real datasets.

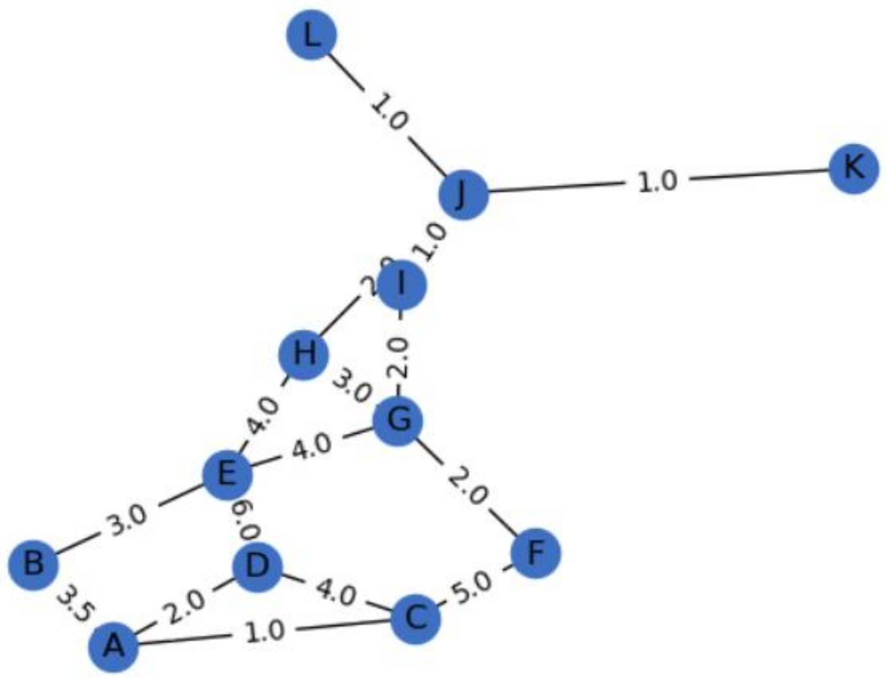
NetworkX:

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. It is used to study large complex networks represented in form of graphs with nodes and edges. Using networkx we can load and store complex networks. We can generate many types of random and classic networks, analyze network structure, build network models, design new network algorithms and draw networks.

TYPES OF GRAPH

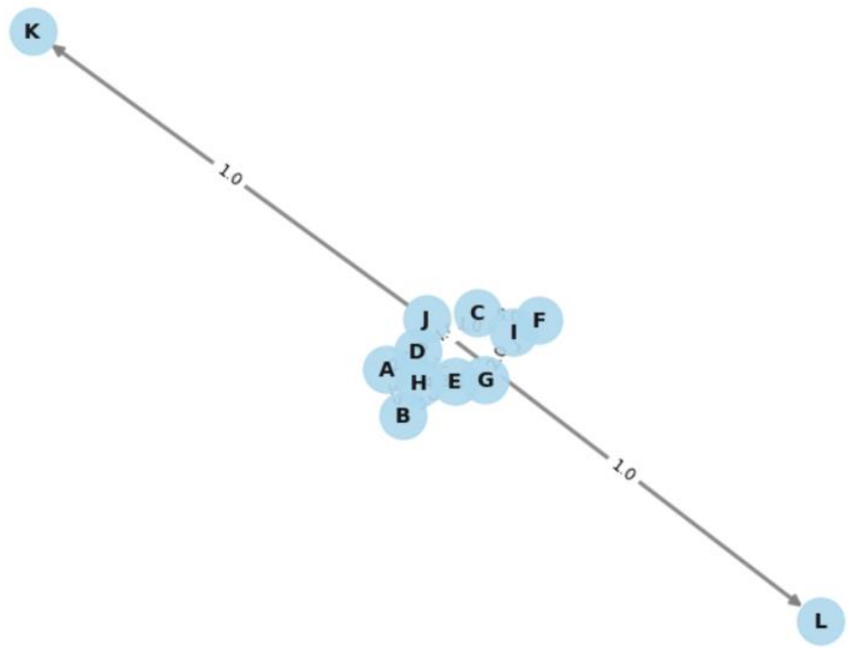
Undirected Graph:

An undirected graph is a graph, i.e., a set of objects (called vertices or nodes) that are connected, where all the edges are bidirectional. An undirected graph is sometimes called an undirected network.



Directed Graph:

A directed graph is a graph, i.e., a set of objects (called vertices or nodes) that are connected, where all the edges are directed from one vertex to another. A directed graph is sometimes called a digraph or a directed network.



Adjacency Matrix:

The adjacency matrix also called the connection matrix, contains rows and columns that represent a simple labeled graph, with 0 or 1 in the position of (Vi, Vj) according to the condition of whether Vi and Vj are adjacent or not.

Centrality Measures:

Centrality measures are scalar values given to each node in the graph to quantify its importance based on an assumption.

1. **Degree Centrality**

Degree Centrality is a measure of node importance in a network based on the number of links or edges that a node has with other nodes in the network. The degree centrality of a node can be calculated using the following formula:
Degree Centrality of Node i = number of edges connected to Node i / (N-1)

where N is the total number of nodes in the network. This formula calculates the degree centrality of a node by dividing the number of edges that connect to the node by the maximum number of possible edges that the node could have in the network, which is N-1 (since a node cannot be connected to itself). The resulting value ranges from 0 to 1, with higher values indicating a higher degree of centrality.

2. **Betweenness centrality:**

Betweenness Centrality is a measure of node importance in a network based on the extent to which a node lies on the shortest paths between other pairs of nodes in the network. The betweenness centrality of node i is calculated as follows:

For each pair of nodes (j,k) in the network, find the shortest path(s) between them.

For each node, i in the network, calculate the fraction of these shortest paths that pass-through i. This fraction is called the betweenness centrality of node i.

The betweenness centrality of node i is given by the following formula:

Betweenness Centrality of Node i = Σ (s, t) $\in$ Φ(i) σ(s, t | i) / σ(s, t)

where Φ(i) is the set of all pairs of nodes in the network that include node i, σ(s, t) is the total number of shortest paths between nodes s and t, and σ(s, t | i) is the number of those paths that pass through node i.

In other words, the betweenness centrality of a node i is the sum of the fraction of all shortest paths in the network that pass-through i. This measure gives an indication of the extent to which a node acts as a "bridge" between other nodes in the network and is often used to identify nodes that play a critical role in the flow of information or resources within the network.

3. Closeness centrality:

Closeness centrality is a measure of centrality in a network that calculates the average shortest distance from each vertex to every other vertex. It is calculated as the reciprocal of the sum of the length of the shortest paths between a node and all other nodes in the graph.

The formula for closeness centrality is 1/ (average distance to all other vertices).

In a connected graph, closeness centrality (or closeness) of a node is calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph.

Closeness centrality measures how central a node is in a network. The more central a node is, the closer it is to all other nodes. Closeness centrality can be used to identify important nodes in a network that are critical for information flow or communication. It can also be used to compare different nodes' importance within a network.

There are three different ways to measure network centrality: degree, closeness, and betweenness. Degree centrality measures how many connections each node has, while betweenness centrality measures how often each node appears on the shortest paths between pairs of other nodes.

4**. Eigenvector Centrality:**

Eigenvector centrality is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Eigenvector centrality is calculated using eigenvectors, which are a set of vectors associated with a linear system of equations. The eigenvector centrality score for each node is proportional to the sum of its neighbors' scores, with weights given by the adjacency matrix.

Mathematically, eigenvector centrality can be represented as:

$Ax = \lambda x$ where A is the adjacency matrix of the graph, x is an eigenvector corresponding to eigenvalue λ, and Ax represents matrix multiplication between A and x.