**MINI-PROJECT**

**NAME: Ayush Om Srivastava**

**UNIVERSITY ROLL NUMBER:2017316**

**CLASS: CE-SPL**

**TOPIC: SOCIAL NETWORK ANALYSIS**

**PROBLEM STATEMENT:**

**Question:** Given an instance of set of nodes in a social network graph, find the centrality of dataset along with polarity.

**Introduction**

A network consists of nodes and ties. Nodes are also called actors or vertices. Based on different representation of nodes and ties in network, there are 4 types of networks:

1. Social networks
2. Biological networks
3. Information networks
4. Technological networks

Social network analysis also known as Organizational network analysis, is a set of tools and processes used for a better understanding the relationship in a network. In social network, nodes of the network are represented by people or organization and links are the relationships between these people and organizations. The tools give you a major advantage when studying a population. It can make analysis much more realistic because you see both the individual and the ways that they are connected at the same time.

Network analysis is usually conducted at 3 levels

1. Dyad level
2. Node level
3. Group/network level

One of the most significant differences between quantitative methods and social network analysis is that quantitative methods assume independence of observation. But the fundamental assumption of social network analysis is interdependence.

Chart

Description automatically generated

**MOTIVATION:**

It was one of the most interesting topics I came across. In the era where almost everyone is on social medias, learning a way to provide insight into social influences within teams, identify cultural issues, etc. can help one understand how team building can change the dynamics of an organization’s social network. It helps to trace back rumors and fake news across the social networks and teaches us how information flows which is critical for improving communication and mobilizing knowledge.

**Tools used:**

Since I am a beginner, I have used simple tools to explain my project as simple and clear as possible.

#Facebook Dataset is acquired from snap Stanford and twitter dataset from Kaggle.

# To represent complex data structure of a graph with various features attached to each node, python-igraph has been used.

#Dictionary data structure is deployed to store the corresponding features of each node.

# VS code is used to execute code with the help of python language and its related libraries.

A picture containing accessory, colorful, decorated

Description automatically generated

**Methodology:**

**STEP 1:**

First step in approach is to define your **focus.** in completing a network analysis, it is beneficial to set your focus. this will involve considering various elements of the analysis:Table

Description automatically generated with medium confidence

**STEP 2:**

One need to decide **what data he/she will use**. SNA can be performed to any data that highlights relationships between things. If the element is gang for example, the approach works best with data that can capture non-criminal as well as criminal links, since a lot of useful information is contained in social links.

**Step 3:**

**Collection of data** is the next step. Network data collection includes

#Observations

#Social network survey and interview

#Archival data

#Application programming interface (API)

API is the best way to collect data since observation, interviews and surveys could not be repeated on a regular basis due to time constraints and financial costs. People are also not a trusted source of information because of faulty memory and social desirability bias.

**Step 4: Analyze**

After successfully collecting the data, SNA entails the network you create to investigate questions that one need to answer. Therefore, there is no predefined way of undertaking the analysis.

For example, by plotting the network’s centrality scores, you can examine the role of the nodes in the network relative to the others in that network.

**Step 5: Validation**

The analysis shown by intelligence may be incomplete or misleading in places, thus skewing the picture. For example, drug dealing may be more visible than gang wars but both activities may well be occurring. To resolve such problems, it is important to validate findings against operational experience. One need to ask themselves

• Do the findings match what is known?

• Is there anything that seems unusual?

• Can any unusual results be explained by issues with the data?

**Application of SNA:**

1. In marketing, SNA can provide marketers with valuable insights for developing communication and branding strategies by building social capital in social networking sites as twitter.
2. SNA has been used as a strategic approach to team building and to understand how team building can change the dynamics of an organization’s social network.
3. SNA is used to shine a light on the social influences within team structures to identify where the cultural problems began and how they spread.
4. SNA is not so much about focusing on individuals. It is about social systemic aspect. How a team collaborates or whether an organization transmits its internal knowledge efficiently. Connectivity is the most important thing.
5. In the fields of change management, culture development, performance management and so many more, insight offered by SNA is a powerful tool. In today’s business environment, a professional working in corporate communications, for example, has a variety of different communication mechanisms. One of those is engaging with specific people, with influencers. SNA helps in identifying the influencers.

**Conclusion:**

Using SNA, we calculated centrality and sentiments of the twitter data set with and accuracy of 0.9.