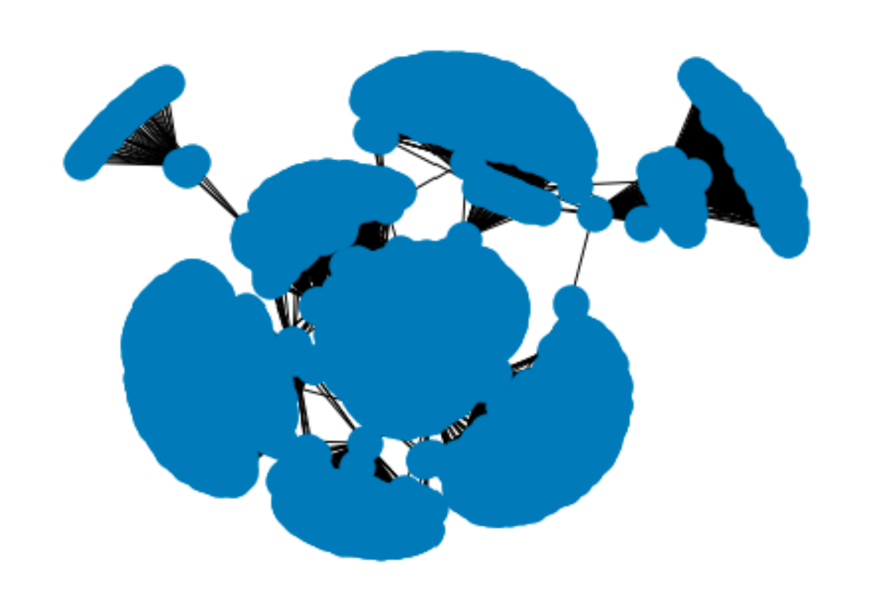
The goal of this project is to evaluate connecting community’s role of a central social network. A central node is generally an important person, a lead, an influencer, or head of a cult. We can study the question of if the relationship is one-sided or if the connections directly communicate with each other. Also, additional roles of a central node can be investigated based on the connecting nodes. For example, if a center node is a criminal, the surrounding nodes are generally criminals as well and they are connected around crime. In this project, we will study the Facebook dataset and how the nodes are connected to each other. The Facebook dataset forms a large network, therefore, it is computationally heavy to study. We will. Minimize the network size to be able to investigate it while maintaining the characteristics of the network. We will do this by visualizing the network and choosing a fraction of the network. Then, we will select the significant influencers of the network by using centrality metrics. We will apply clustering to find smaller networks as well. Further, we will run a graphical model to determine if we can create a community without influencer’s impact.

The Facebook dataset corresponds to a large network with 3663 nodes and 53498 edges. We can see from the following figure that the network forms various communities. These communities could be formed based on different features such as gender, nationality, interests, location, nationality, etc. However, we cannot determine based on the given dataset, which features are included in the study.



We can see from the figure that the network is very large and might be too large for a computer analysis even though we can visualize it. Therefore, we need to sample the dataset for the analysis. If we don’t sample, we may not be able to analyze the dataset using a simple computer.

One method of choosing a smaller subset is connected component. However, the results show that the network is only one connected component. Another method is to select one community to study. We will pick a dense cluster to get the most out of the experiment using a centrality metric: degree and betweenness. The results which are shown in a table demonstrate that 107 is the most connected node. Therefore, we will choose a network incorporating 107. The smaller subset has 333 nodes and 5038 edges.

The following table shows the top 10 highest scoring node based on degree centrality.

Table

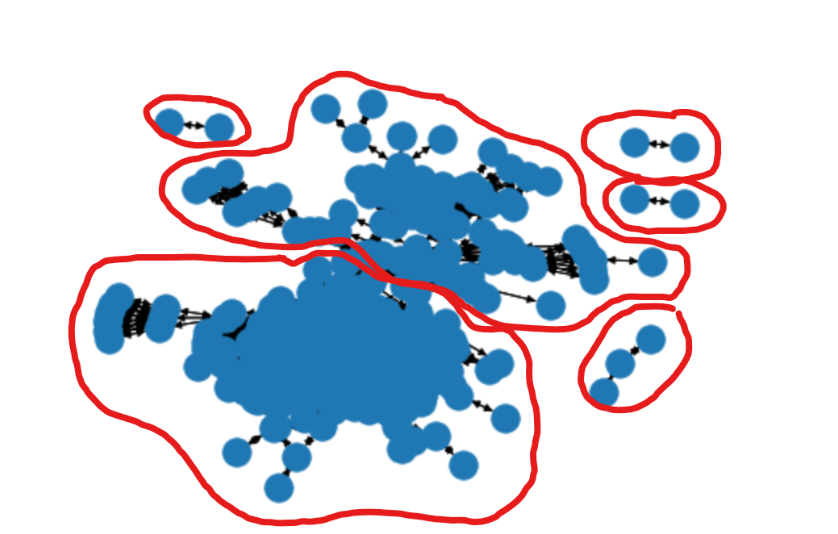
Description automatically generated

The following table shows the top 10 highest scoring node based on degree centrality

Table

Description automatically generated

The following figure shows the potential samples of the dataset. From these potential subnetworks, we will choose one containing 107.



Finding influencers

Influencers are defined to be those who impact others and able to connect people. Hence, I think betweenness centrality is an appropriate metric. Metrics involving directed would be neglected in this case since this is an undirected graph. On the other hand, degree centrality may be inappro-

priate due to its simplicity and closeness being insufficient (influencers are not necessarily ’close’ to their followers). Additionally, influencers may help create a community through them, which is equiva- lent to whether the node is within the paths connecting other nodes. Therefore, betweenness is chosen.

Top 1% of betweenness centrality weight is then chosen. These nodes were chosen because their weights significantly greater than others. While most nodes having a weight smaller or equal than 0.05, these nodes’ weights were 0.266, 0.246, which are five times higher. The nodes include:

• 277 • 175 • 19 • 23

In addition, community detection algorithm has been run with this graph. The community detec- tion is meant to detect subsets of nodes that are more densely connected. Results are visualized in Figure 6b.