Homework Assignment -4

Nithish Raghunandanan
Matriculation No: 03667351

Answers

1. It is a directed graph containing edges without weights.

2. Number of strongly connected components in the graph: 5736
   Nodes belonging to the biggest strongly connected component: 154

3. Distribution of the betweenness centrality values over the nodes:

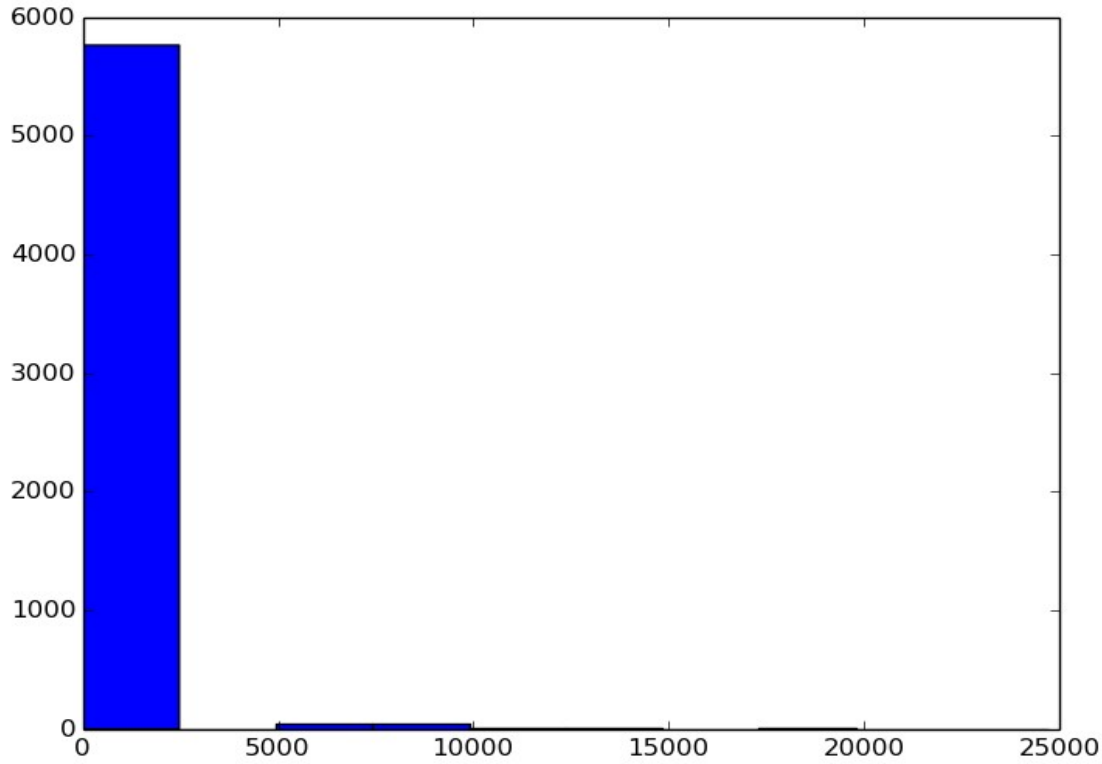

Illustration 1: Visualisation of the Betweenness Centrality Values(Bin Size:10)

For most of the nodes in the network, the betweenness centrality value is in the range of 0 to 10. Then, there is a small number of nodes(around 60) that have their centrality values in the range of 140-170 range. Then, there are few nodes that have their centrality values in the range of 5000-6000. There are few(1-2) nodes that have their centrality value higher than 10000 & 20000.

4. Diameter of the network: 7
   Density of the network: 0.000788246509705

Analysis of the Results

It is not surprising that the graph of social connections is directed and not weighted as the attractions between persons may not be reciprocated by the other person. Also, there is no weight for any person in the social network.

Initially, I was surprised to see that there was only a single strongly connected component in

the clustering based on strongly connected components that had more than one node. But, it can be explained by the fact that the largest strongly connected component is the participants of the exercise. We were asked to rate our social ties with all the other participants of the exercise. So, there is a social connection between all the participants of the exercise in both directions. In the case of all other nodes(from Facebook friend lists), it was just one directional connection between the nodes and the node that rated the tie strength unlike the bidirectional connection(edge) needed for the strong connection. This will also include those participants of the exercise who registered for it but did not rate the other participants. This is why I think that even though there were 177 people whom I rated, there are only 154 people in the largest strongly connected network.

The betweenness centrality values are a measure of how central a node is in a network. It is a measure of the total number of shortest paths from all vertices to the other vertices that pass through a given node. It is also not surprising that it is low(0-10) for a majority of the vertices as they are connected to just a single node(their facebook friend participating in the exercise). The nodes with high betweenness centrality values are most likely to be in the biggest strongly connected network.

Another interesting result obtained from the result is that any two people(nodes) in the network can be connected in 7 or fewer steps. It is almost close to an earlier experiment that said that any two people on the world can be connected in six or fewer steps[1]. It is comparable as this is just a partial network with not complete information about all nodes(e.g., facebook friends of first level network that is not participating in the exercise). And it is also not surprising to see a low density for the network as many of the networks are not completely described(e.g., Facebook friends) reducing the number of possible edges.

References:

[1] https://en.wikipedia.org/wiki/Six_degrees_of_separation

Clustering

I did the clustering based on the community infomap clustering. It is a clustering based on the map equation[1]. The cluster identified communities(clusters) based on the information flow in the network. The clustering does not depend on how the network is formed. It just analyses the sructure of the network and how it affects the information flow in the network. It is a non-deterministic algorithm in the sense that the number of clusters detected could vary in each clustering. So, I got clusters in the range of 800-1000 while clustering using this algorithm.

Working

Initially, all the nodes are placed into different clusters. Then, neighbouring nodes and/or clusters are clustered together in a random sequential way if they reduce the map equation. This process continues until any further clustering does not decrease the map equation further. Also, the clusters once formed are not destroyed. As its a random walk, the clustering could vary in different runs of the clustering algorithm.

References:

[1] The map equation by M. Rosvall, D. Axelsson and C.T. Bergstrom