WSC assignment 3 - Part 1

Network Centrality

Centrality measures identify the most important nodes in a network. The most important nodes can be one of the following:

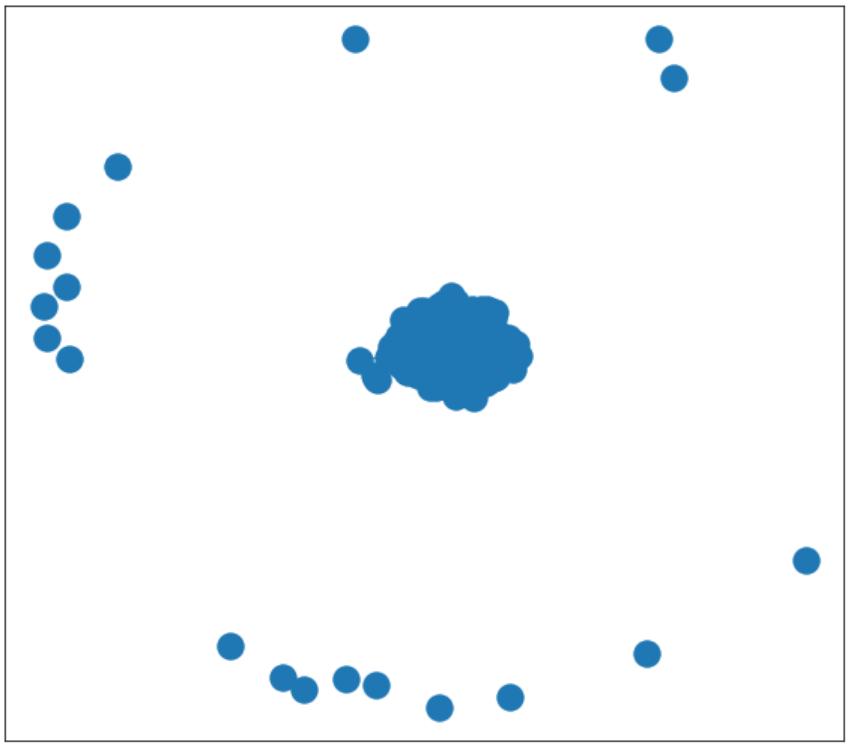
- Influential nodes in a social network.
- Nodes that disseminate information to many nodes or prevent epidemics.
- Hubs in a transportation network.
- Important pages on the Web.
- Nodes that prevent the network from breaking up.

Centrality Measure overview:

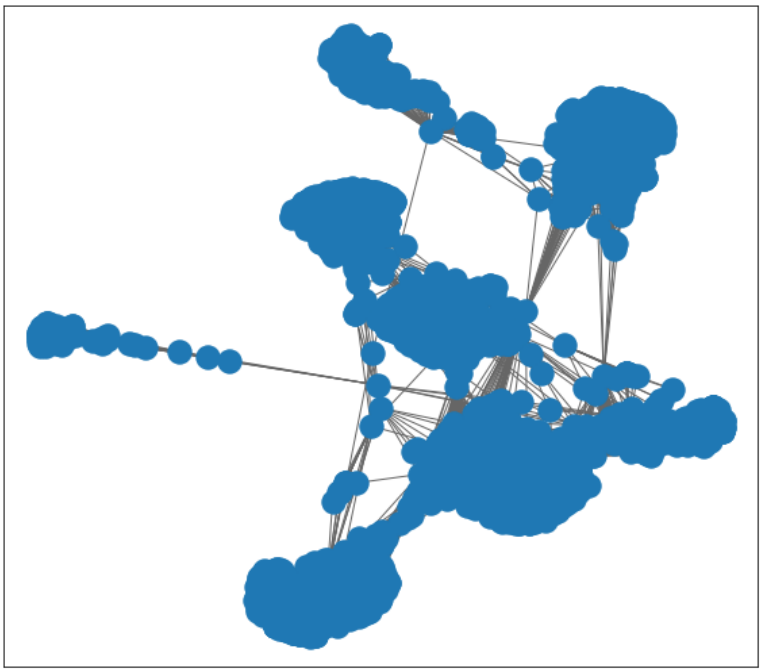
The centrality measures allow us to find the most important nodes in a network. The definition of important node might change depending on our interest. Sometimes we might be interested in finding who are the most influential nodes in a social network, which are good at disseminating information to many other nodes. On other hand, we might be interested in searching the nodes that are good at preventing bad behaviours or epidemics from spreading on a social network. These measures allow us to find nodes that prevent the network from breaking up and if we remove these nodes, they would cause the network to fracture and break up into different components.

Following are the visualisations of dataset used for the analysis

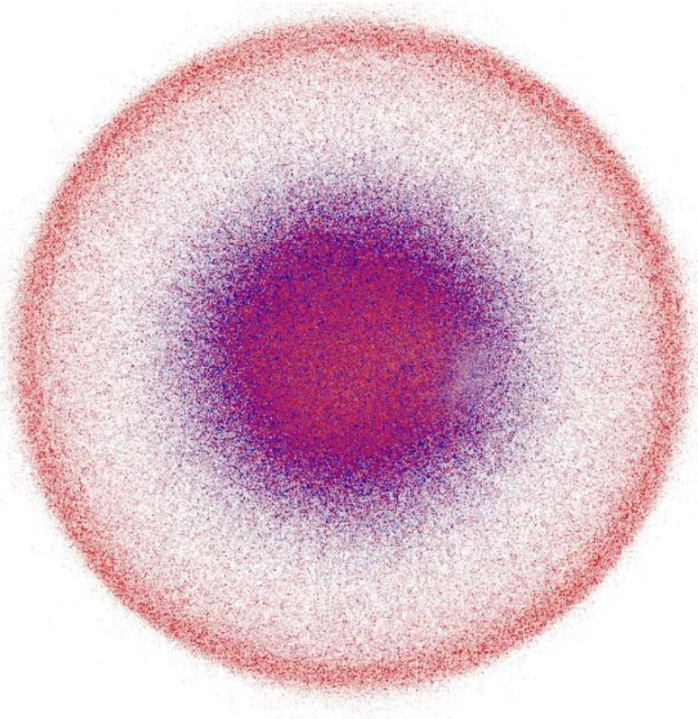
1. Small dataset



2. Medium Dataset: Facebook combined



3. Large Dataset: DBLP community (Visualised using Gephi tool)



Types of Centrality Measures:

Degree Centrality:

Here the assumption is that important nodes have many connections. This is most basic type of centrality measure where we count the neighbours that are directly connected to the node

$C_{deg}(v) = \frac{d_v}{|N|-1}$, where N is the set of nodes in the network and d_v is the degree of node v.

Closeness Centrality:

Here, the assumption is that important nodes are close to other nodes in the Network. In other words, closeness centrality says that nodes who are central are a short distance away from all the other nodes in the network

$C_{close}(v) = \frac{|N|-1}{\sum_{u \in N \setminus \{v\}} d(v,u)}$, where

N = set of nodes in the network,

$d(v,u)$ = length of shortest path from v to u.

Betweenness Centrality:

Here the assumption is that important nodes are the ones who connect other nodes in the network. The metric used to calculate betweenness centrality is distance. Distance between two nodes is the shortest path between them.

$C_{btw}(v) = \sum_{s,t \in N} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}$

$\sigma_{s,t}$ = the number of shortest paths between nodes s and t.

$\sigma_{s,t}(v)$ = the number shortest paths between nodes s and t *that pass through node* v.

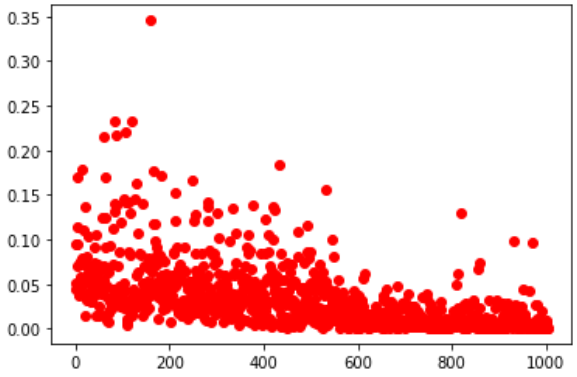
Eigenvector centrality:

Relative scores are assigned 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. A high eigenvector score means that a node is connected to many nodes who themselves have high scores.

Analysis:

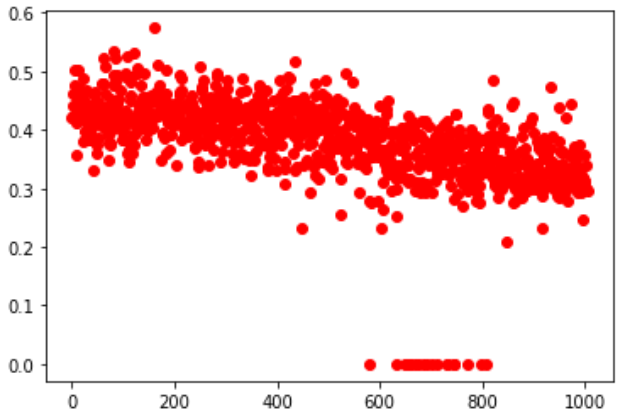
1. Small Graph Analysis:

    a. Degree centrality:



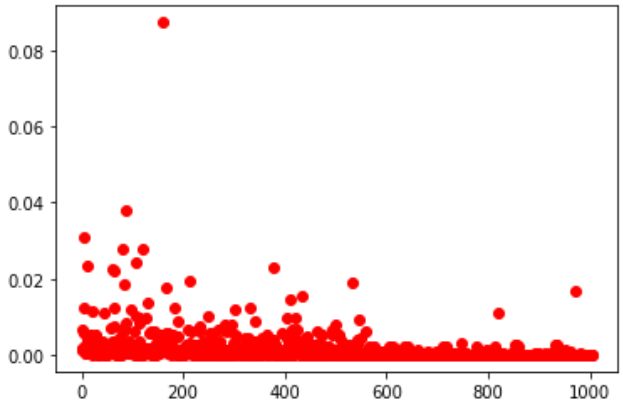
According to degree centrality, following are the ranks of node as per their Importance:
Rank 1: Node no. 160
Rank 2: Node no. 121
Rank 3: Node no. 82
Rank 4: Node no. 107
Rank 5: Node no. 86

    b. Closeness Centrality:



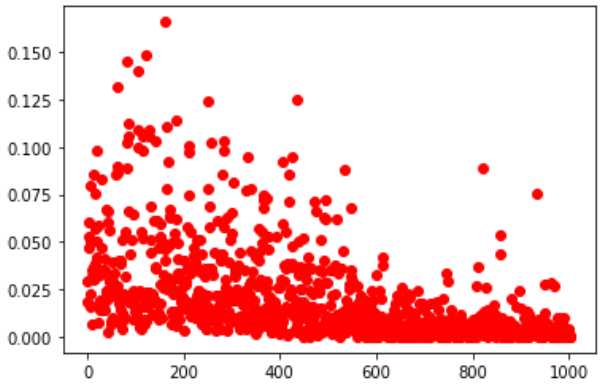
According to closeness centrality, following are the ranks of node as per their importance.
Rank 1: Node no. 160
Rank 2: Node no. 82
Rank 3: Node no. 121
Rank 4: Node no. 107
Rank 5: Node no. 62

c. Betweenness centrality:



According to Betweeness centrality, following are the ranks of node as per their importance.
Rank 1 :160
Rank 2: 86
Rank 3: 5
Rank 4: 82
Rank 5: 121

d. Eigenvector centrality:



According to Eigen vector centrality, following are the ranks of node as per the importance:
Rank 1: 160
Rank 2: 121
Rank 3: 82
Rank 4: 107
Rank 5: 62

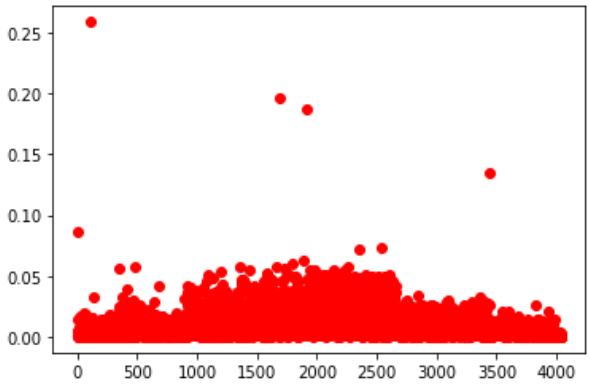
e. Summary

Ranks	Centralities			
	Degree	Closeness	Betweenness	Eigenvector
1	160	160	160	160
2	121	82	86	121
3	82	121	5	82
4	107	107	82	107
5	86	62	121	62

Now that we have seen a number of different ways of finding central nodes in a network, let's compare the different centrality measures, and see how nodes are ranked differently. Node no. 160 is the most important node in all the centrality measures. Node no 121 is present in all the top 5 rankings of centrality measures, but its rank differs. In closeness centrality, node ranks from 2 to 5 are changed. The node no 86 is no more an important node and it is replaced by node 62. This means, node 62 is closer to other nodes compared to node 86. In Betweenness centrality, node 5 is ranked 3. Node 5 is particularly interesting node as it didn't appear in Degree or closeness rankings. This means, Node 5 is neither among the closest nor has high degree, but most of the shortest path in the network pas through node 5. In Eigenvector centrality, the results are similar to degree centrality. The only difference is that the rank 5: 86 in closeness centrality was replaced by rank 5: 62 in Eigenvector centrality

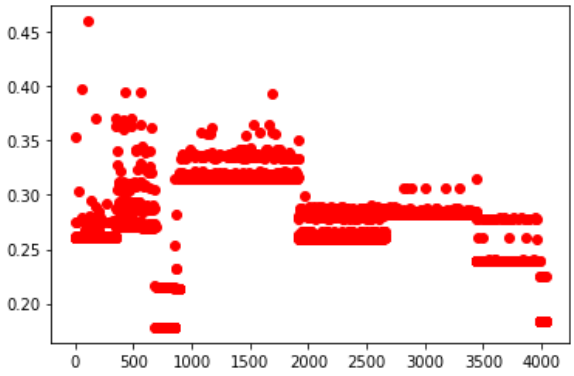
2. Medium sized network:

   a. Degree Centrality:



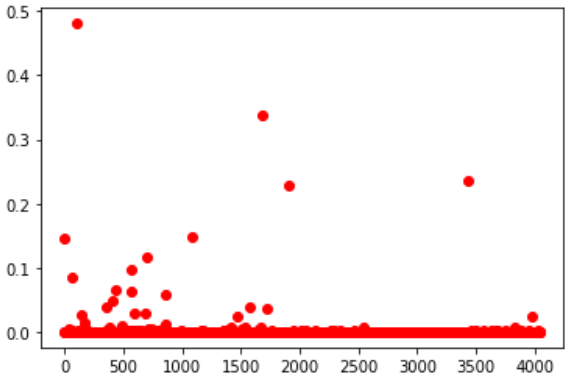
According to degree centrality, following are the ranks of node as per their importance:
Rank 1: Node no. 107
Rank 2: Node no. 1684
Rank 3: Node no. 1912
Rank 4: Node no. 3437
Rank 5: Node no. 0

   b. Closeness Centrality:



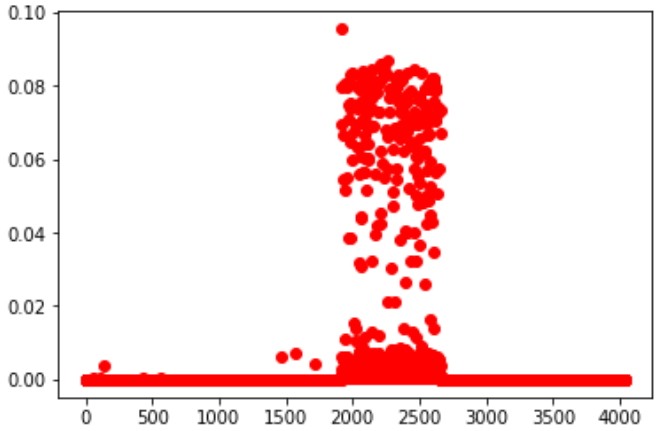
According to closeness centrality, following are the ranks of node as per their importance:
Rank 1: 107
Rank 2: 58
Rank 3: 428
Rank 4: 563
Rank 5: 1684

c. Betweenness Centrality:



According to closeness centrality, following are the ranks of node as per their importance.
Rank 1: 107
Rank 2: 1684
Rank 3: 3437
Rank 4: 1912
Rank 5: 1085

d. Eigenvector centrality:



According to closeness centrality, following are the ranks of node as per their importance:
Rank 1: 1912
Rank 2: 2266
Rank 3: 2206
Rank 4: 2233
Rank 5: 2464

e. Summary

Ranks	Centralities			
	Degree	Closeness	Betweenness	Eigenvector
1	107	107	107	1912
2	1684	58	1684	2266
3	1912	428	3437	2206
4	3437	563	1912	2233
5	0	1684	1085	2464

As per our observation, there are several different ways of finding central nodes in a network, let's compare the different centrality measures, and see how nodes are ranked differently. Node no. 107 is the most important node in 3 out of 4 centrality measures, interestingly, node does not appear in eigenvector centrality rankings. Node no 1912 and 1684 are present in 3 out of 4 of the top 5 rankings of centrality measures, but their rank differs. In closeness centrality, node ranks from 2 to 5 are changed. The node no 1912,3437,0 are no more important nodes and they are replaced by other nodes. In Betweenness centrality, node 5 is ranked 3. Node 58,428 and 563 are no more important and they are replaced by other nodes. Node no 185 is new entrant in betweenness centrality which didn't appear in degree and closeness centrality. In Eigenvector centrality, the results are particularly interesting as it introduces 4 new nodes which didn't appear in other centrality measures.