

SNA Assignment 1 Report

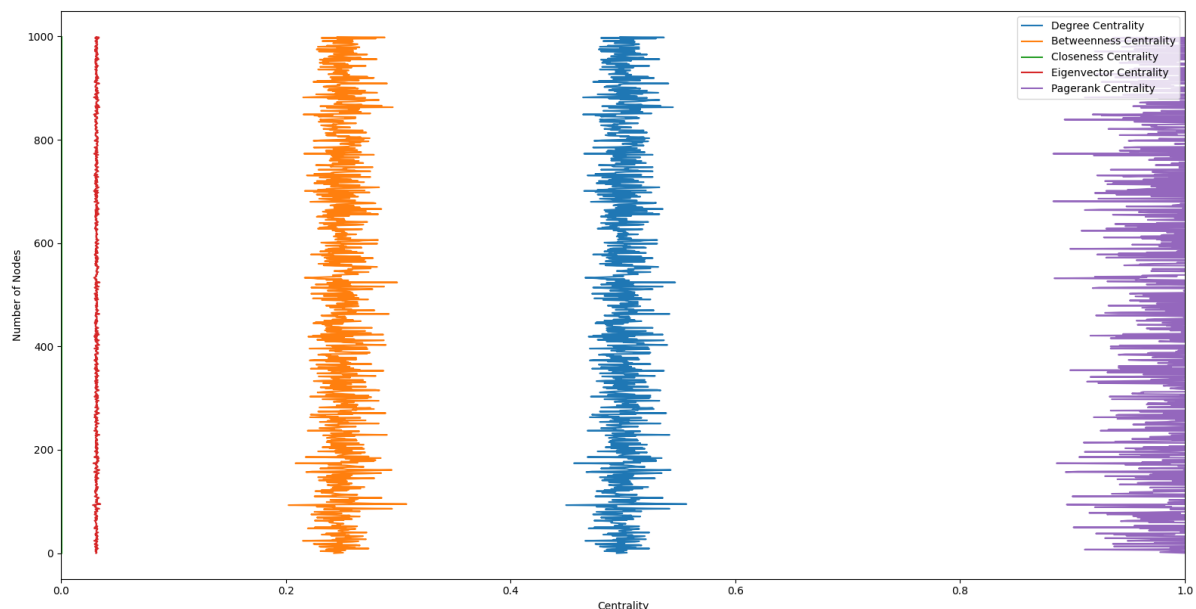
Subhashis Suara

UCSE19012

Methodology

A random undirected graph (of 1000 nodes) was generated using the `gnp_random_graph` function from `networkx` python library. Given centrality measures and clustering coefficient were measured from the generated graph using functions written in the code (link given below). The results of the measures were normalized and plotted as a multi-line graph (Number of nodes vs Centrality values) using `matplotlib.pyplot`. The results of clustering coefficient were normalized and plotted as a bar graph (Frequency vs Clustering) and the random graph was plotted, displaying the local clustering with colouring as per the clustering value (shown in upcoming section).

Comparison between different centrality measures

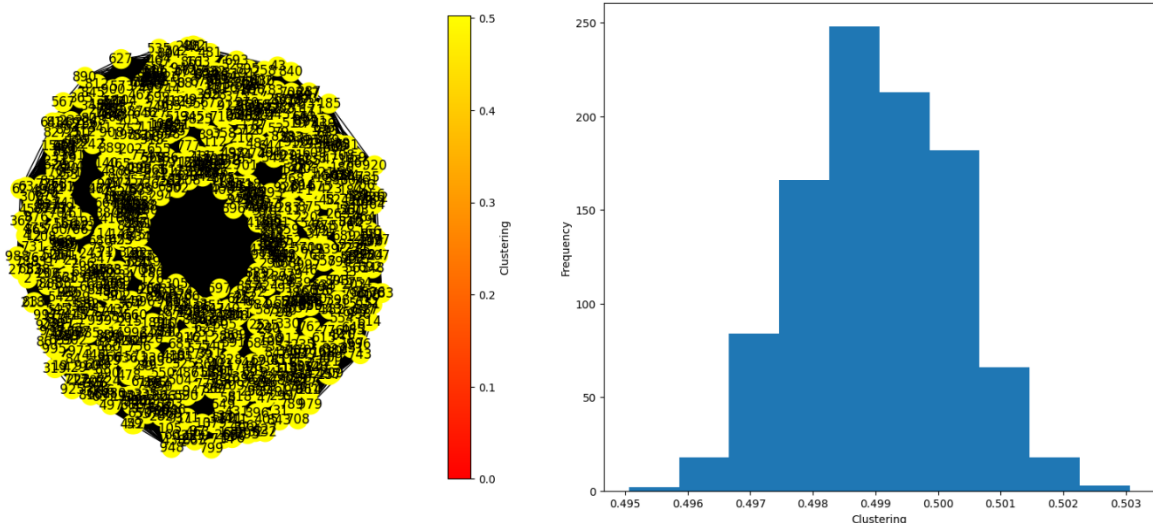


The graph given above represents the various centrality measures for all the nodes of the generated random graph. From the graph we can observe the range in which each centrality measures lies for nodes of the generated random graph. Another interesting observation can be seen below.

```
(env) subha SNA Programming Assignment 1 3.8.6 master 2m 5.92s python .\solution.py
Top 5 nodes with highest degree centrality: [778, 940, 852, 12, 268]
Top 5 nodes with highest betweenness centrality: [778, 940, 852, 12, 480]
Top 5 nodes with highest closeness centrality: [778, 940, 852, 12, 268]
Top 5 nodes with highest eigenvector centrality: [778, 940, 852, 12, 268]
Top 5 nodes with highest pagerank centrality: [268, 687, 940, 12, 778]
```

From the above output, we can see that node 778 has the most centrality, considering all the centrality measures. We can also see the similarity in top 5 nodes with highest centrality in each measure. The similarity signifies certain properties. For example, the similarity can be important for node's ability to diffuse/flow information efficiently. Hence, this was the comparison and analysis between different centrality measures.

Clustering coefficient



We know that the clustering coefficient of a node measures how many vertices tend to cluster together. The histogram given above represents the frequency

distribution of local clustering coefficient values. The global clustering coefficient value is 0.4990342801446112. The clustering coefficient differs from measures of centrality. It is like the aggregate density metric but focused on egocentric networks.

Resources used in this assignment are available here:

https://github.com/subhashissuara/SNA_Programming_Assignment_1