

Assignment 6

Dibyendu Roy Chaudhuri
MT19034

Question1

Description

Link of the network: <https://snap.stanford.edu/data/email-Eu-core.html>

Description about the Network

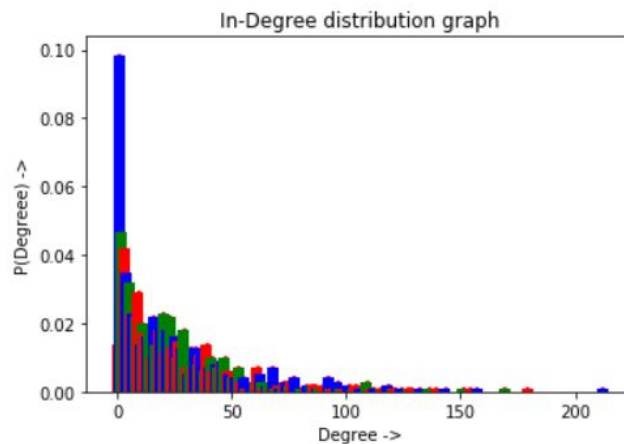
```
Number of nodes- 1005
Number of edges- 25571
Average inDegree- 25.443781094527363
Average outDegree- 25.443781094527363
Max InDegree node id- 160      with inDegree 212.0
Max OutDegree node id- 160     with outDegree 334.0
Average Network Density- 0.025342411448732432
```

Network density= #edges/(#potential edges)

Degree Distribution

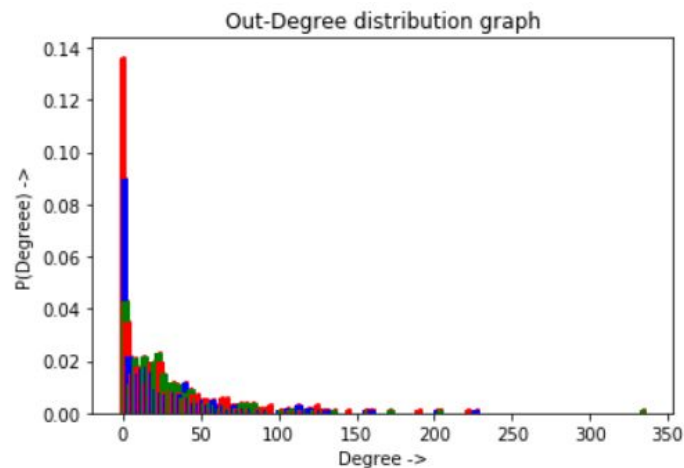
Degree of a directed graph consists of In-degree and Out-degree. We can also say the degree of a node in a directed graph is simply a summation of it's in-degree and out-degree.

In-degree of a node is simply a summation of incoming edges of that node.



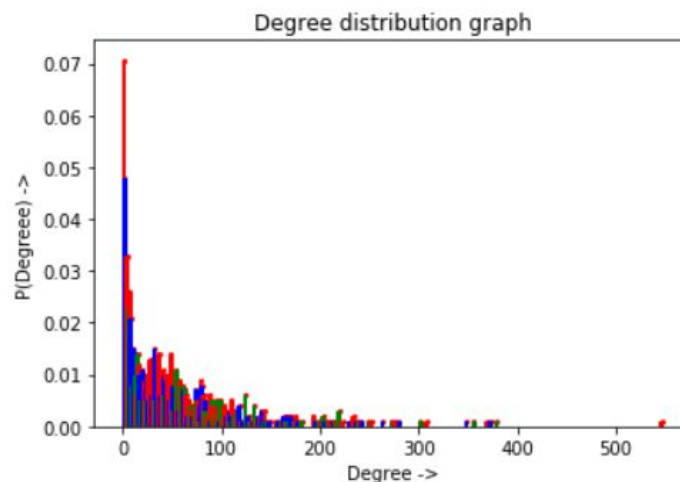
We can see most of the nodes have less in-degree than have a higher degree and the probability of having a particular degree is decreasing with the increasing number of incoming edges.

Out-degree of a node is simply a summation of outgoing edges from that node.



We can see most of the nodes have less out-degree than have a higher degree and the probability of having a particular degree is decreasing with the increasing number of outgoing edges.

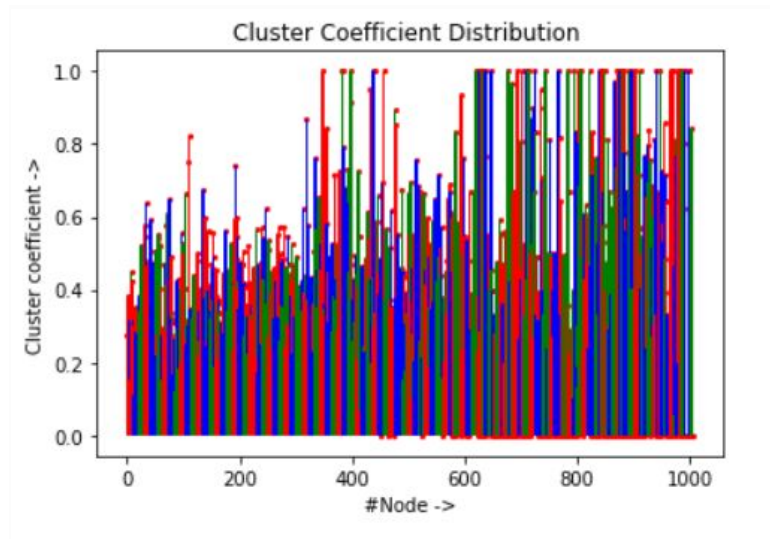
Degree of a node is the summation of its in-degree and out-degree.



We can see most of the nodes have less degree than have a higher degree and the probability of having a particular degree is decreasing with the increasing number of incoming and outgoing edges.

Clustering coefficient

The clustering of a node is the fraction of possible triangles that contain that node.



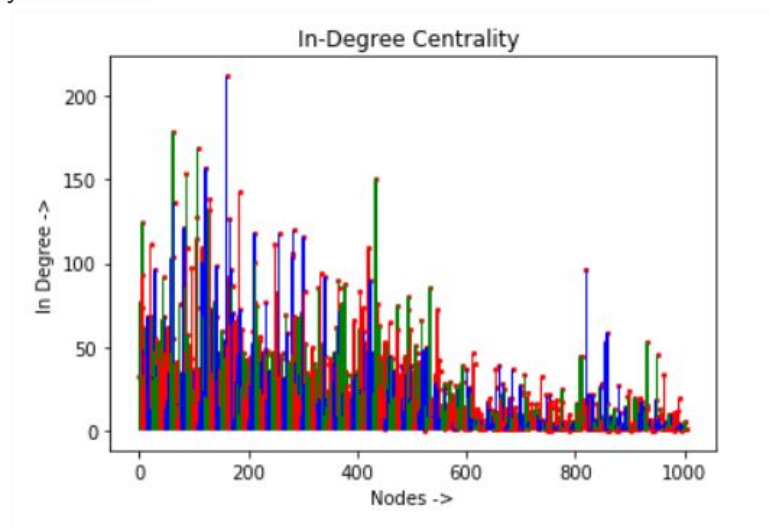
Average cluster coefficient- 0.39935496642215434

Cluster coefficient = $\frac{2 * p}{n(n-1)}$ [p- number of link b/w neighbour, n- #neighbour]

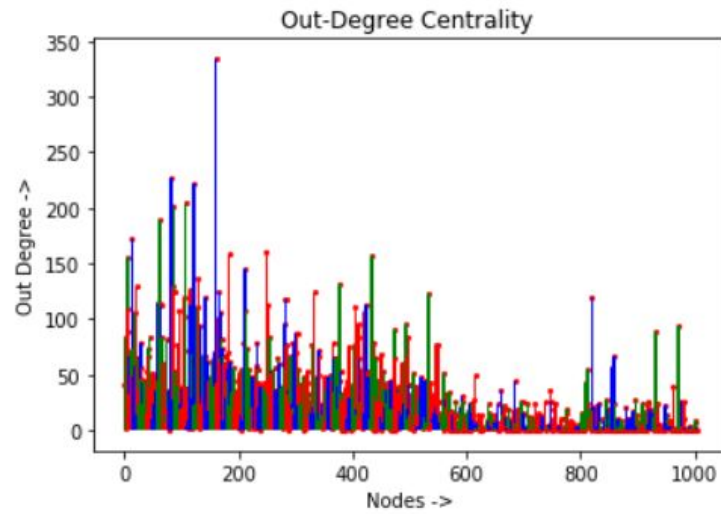
Degree Centrality

We have already discussed two types of degree is present in a directed graph- Indegree and out-degree.

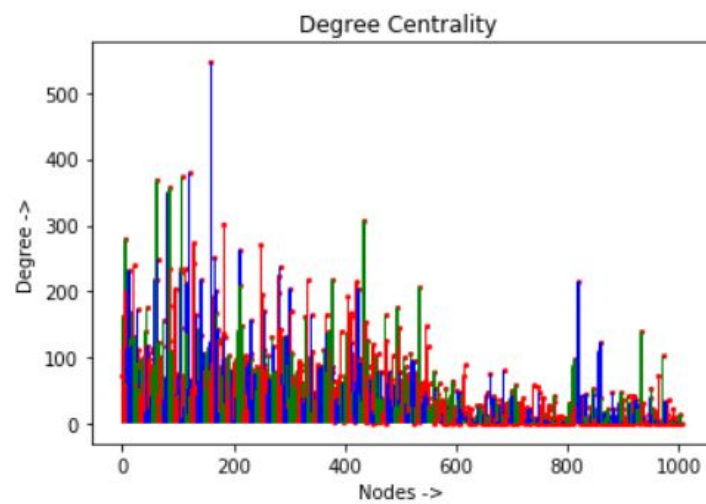
Indegree Centrality of nodes-



Outdegree Centrality of nodes-

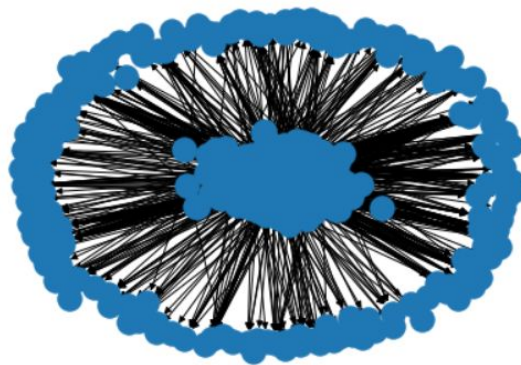


Degree Centrality of nodes-



Question2

Graph Diagram (using networkx library)



Top 10 node using simple PageRank

[1.0, 130.0, 160.0, 532.0, 62.0, 86.0, 107.0, 365.0, 121.0, 227.0]

Top 10 node using Hub score

[160.0, 82.0, 121.0, 107.0, 62.0, 249.0, 434.0, 183.0, 86.0, 114.0]

Top 10 node using simple Auth score

[160.0, 107.0, 62.0, 434.0, 121.0, 183.0, 128.0, 249.0, 256.0, 129.0]

Comparison

Pagerank algorithm considers how many important nodes in a network quote a particular node in the network. On another hand, auth and hub scores are calculated using hits algorithm which considers both incoming and outgoing links from that node. Higher hub score means node links to pages with high authority score. Similarly higher auth score means it is pointed by nodes with high hub score.

In details comparison-

1. Observing the above result, we can say node 1 should be pointed by higher relevant nodes. So I have checked up to the top 20 hub score nodes with the nodes are pointing to node 1 and found that node 82, 121, 142, 21 which have a higher hub score, are pointing to node 1. Node 1 has self-loop which is also helping node 1 to have higher page rank score. A similar trend might be followed by other top nodes in page rank list.
2. Now if consider hub list, then we can see topmost node 160 is actually pointing almost every node present in top 10 auth score list (node 160, 107, 183, 128, 249, 256, 129). According to HITs algorithm, hub scores= summation of auth scores of all nodes, it points. So we can say others node with higher hub score is also following similar kind of trend.
3. Similarly, in the case of authority score, I have found node 160 is actually pointed by node 160 itself, 82, 121, 249, 183, 114 which are having higher hub score and belongs from top 10 nodes in hub score based rank list. According to HITs algorithm, auth scores= summation of hub scores of all nodes which point that node. This kind of pattern might be followed by other top-ranked auth scores nodes.
4. I have also observed some clique patterns which are also helping a few nodes (160, 82, 107, 62, etc) to boost up their ranking.