Assignment 3 Report
Group 80
Devansh Joshi -2019302
Mohit Bhar- 2019256
Hardik Dudeja- 2019422

Q1 Dataset: The dataset that we have used for the purpose of this assignment is the p2p-Gnutella04.txt dataset. This is a sequence of snapshots of the file sharing network that is part of the Gnutella peer-to-peer file sharing network. There are a total of 9 snapshots of Gnutella network that were collected in August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts. Following are reported values for the dataset:

1. Number of Nodes

```
In [4]: graph.getNodes(Graph)
Out[4]: 10879
```

2. Number of Edges

```
In [5]: graph.getEdges(Graph)
Out[5]: 39993
```

3. Average In-degree

```
In [6]: graph.calculate_average_indegree(Graph)
Out[6]: 3.676165088703006
```

4. Average Out-degree

```
In [7]: graph.calculate_average_outdegree(Graph)
Out[7]: 3.676165088703006
```

5. Node with Max In-degree

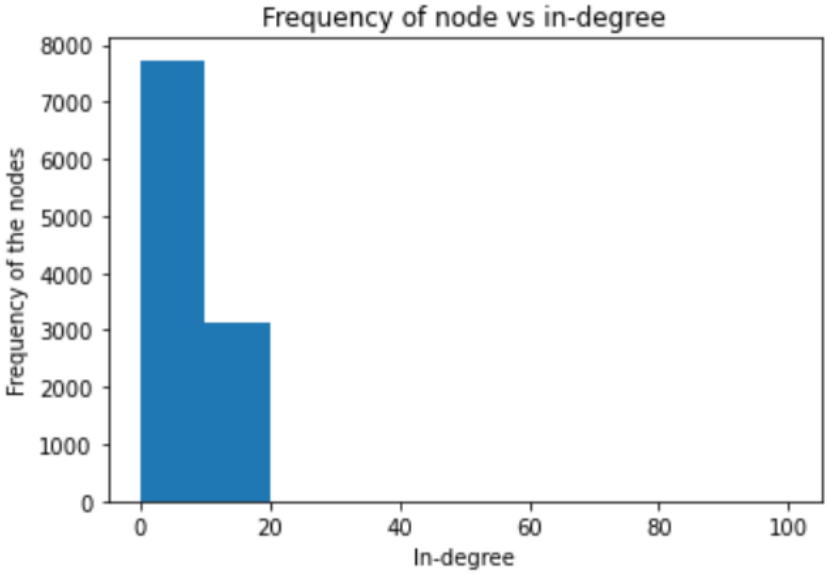
```
In [8]: graph.get_node_max_indegree(Graph)
Out[8]: 3109
```

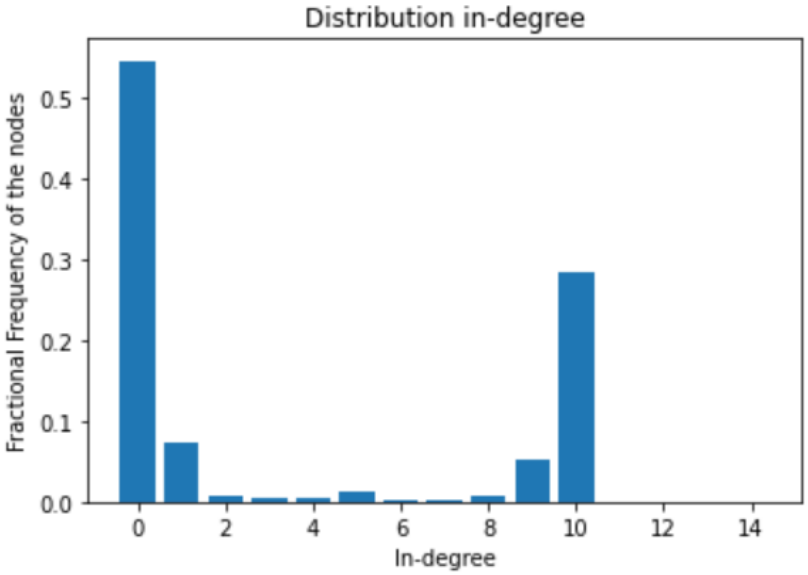
6. Node with Max Out-degree

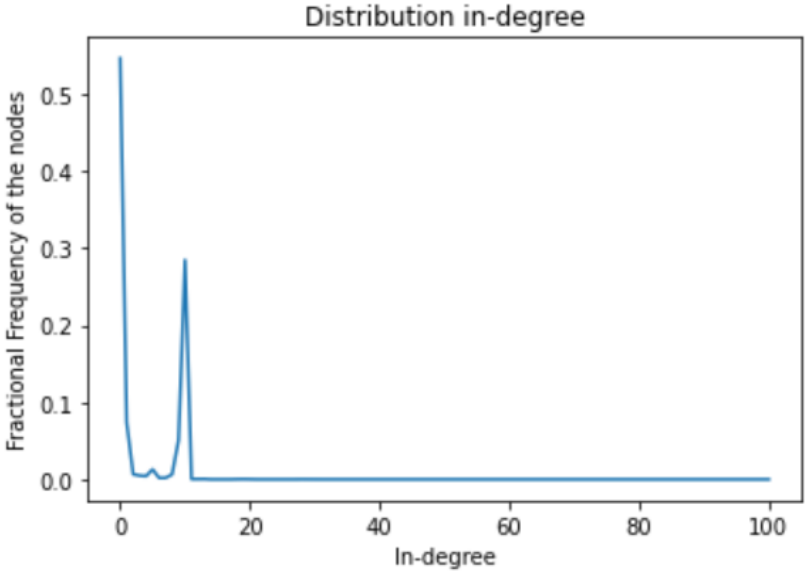
```
In [9]: graph.get_node_max_outdegree(Graph)
Out[9]: 1054
```

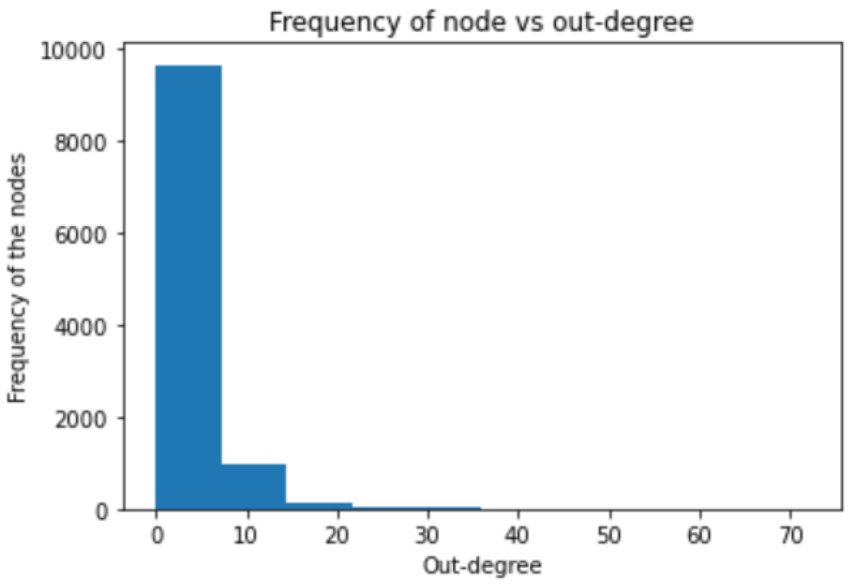
7. The density of Network

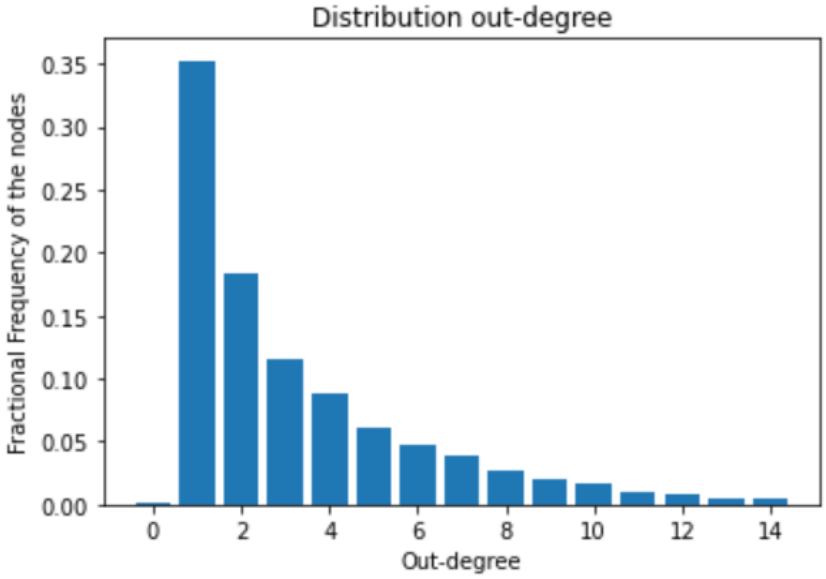
```
In [10]: graph.getDensity(Graph)
Out[10]: 0.0003379449428849978
```

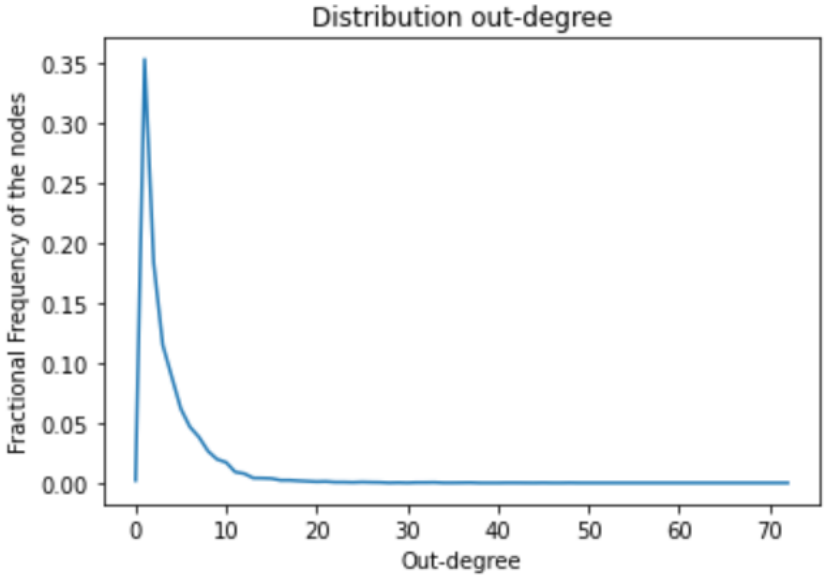

Frequency of node vs in-degree
Frequency of the nodes
In-degree
0
1000
2000
3000
4000
5000
6000
7000
8000
0
20
40
60
80
100

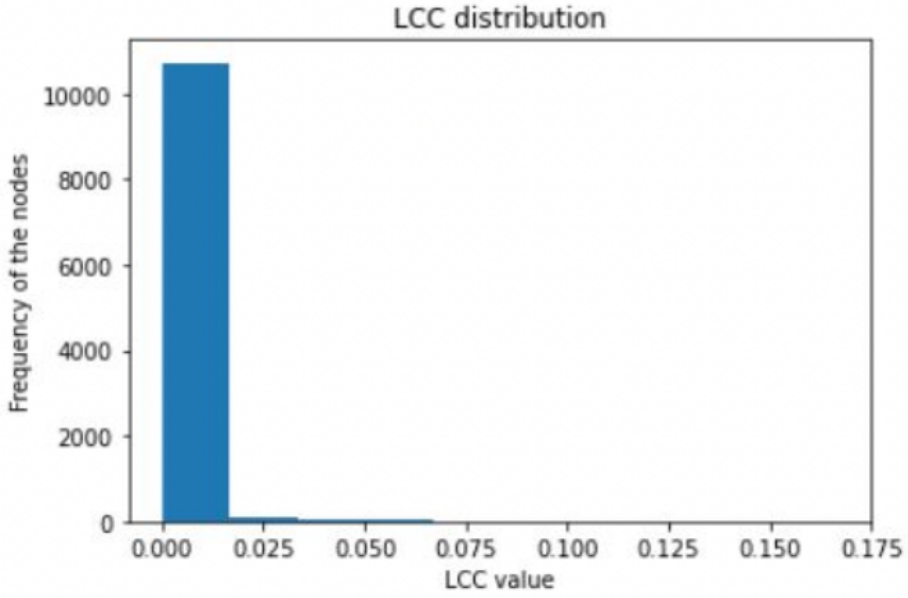

Distribution in-degree
Fractional Frequency of the nodes
In-degree
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
10
12
14

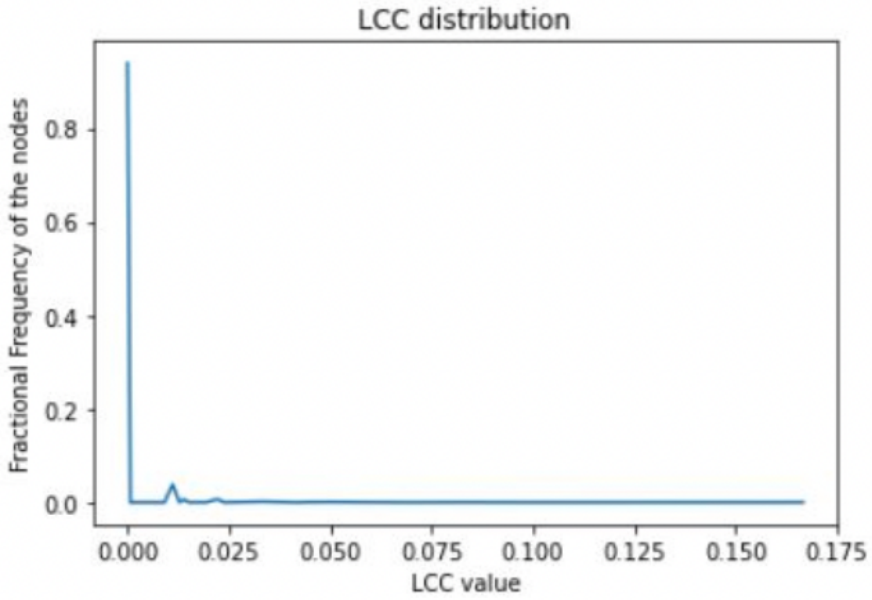

Distribution in-degree
Fractional Frequency of the nodes
In-degree
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100


Frequency of node vs out-degree
Frequency of the nodes
Out-degree
0
2000
4000
6000
8000
10000
0
10
20
30
40
50
60
70


Distribution out-degree
Fractional Frequency of the nodes
Out-degree
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
2
4
6
8
10
12
14


Distribution out-degree
Fractional Frequency of the nodes
Out-degree
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50
60
70


LCC distribution
Frequency of the nodes
10000
8000
6000
4000
2000
0
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
LCC value


LCC distribution
Fractional Frequency of the nodes
0.8
0.6
0.4
0.2
0.0
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
LCC value

The formula for calculating the local clustering coefficient for directed graph that we have used is:

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

Here e_{jk} represents edges and v_j represents vertices. k_i is the number of neighbors of a vertex. Neighbourhood N_i is the set of immediately connected neighbors of a vertex.

Q2)

1. To calculate page rank score for each node we have made use of networkx library's nx.pagerank() function.

```
Node: 0 PageRank: 0.0001212693780956343
Node: 1 PageRank: 8.474289155891317e-05
Node: 2 PageRank: 0.000135506439805258
Node: 3 PageRank: 0.0001219618933066092 7
Node: 4 PageRank: 0.00040545528647798135
Node: 5 PageRank: 0.00010332713155916155
Node: 6 PageRank: 0.00015777029580475385
Node: 7 PageRank: 0.00011447643862052635
Node: 8 PageRank: 6.534193828083182e-05
Node: 9 PageRank: 0.00019076342433952504
Node: 10 PageRank: 9.360042658524542e-05
Node: 11 PageRank: 0.000158859619125404
Node: 12 PageRank: 6.991481096217797e-05
Node: 13 PageRank: 0.00012180888819699102
Node: 14 PageRank: 7.709255301035561e-05
Node: 15 PageRank: 8.925464003635525e-05
Node: 16 PageRank: 0.0001406513327483791
Node: 17 PageRank: 9.716816820329565e-05
Node: 18 PageRank: 0.00019860626247021734
Node: 19 PageRank: 8.349401944357422e-05
Node: 20 PageRank: 0.00010332446055873575
Node: 21 PageRank: 6.532282317913018e-05
Node: 22 PageRank: 9.663850391796315e-05
Node: 23 PageRank: 8.137877401335503e-05
Node: 24 PageRank: 6.532282317913018e-05
...
Node: 10877 PageRank: 0.00010627974763847013
Node: 10873 PageRank: 6.525254041396984e-05
Node: 10874 PageRank: 5.499573860784478e-05
Node: 10875 PageRank: 5.966904008181786e-05
Node: 10876 PageRank: 5.966904008181786e-05
```

2. To calculate authority and hub score for each node we have made use of networkx library's nx.hits() function.

```
Node: 0 Authority: 0.0003318884021838684 Hub: 1.979812884036766e-05
Node: 1 Authority: 1.20775491128382e-05 Hub: 2.0466092998118155e-05
Node: 2 Authority: 3.5345120891267626e-05 Hub: 0.0
Node: 3 Authority: 4.489500024168235e-05 Hub: 2.1049074679667525e-05
Node: 4 Authority: 0.00012098292969830178 Hub: 0.0
Node: 5 Authority: 3.653956891306698e-05 Hub: 0.0
Node: 6 Authority: 5.2536544630668404e-05 Hub: 0.0
Node: 7 Authority: 4.284645902348583e-05 Hub: 0.0
Node: 8 Authority: 1.975262930461583e-06 Hub: 2.4188235115863908e-05
Node: 9 Authority: 0.00010601276604399316 Hub: 0.0
Node: 10 Authority: 1.6060100372377594e-05 Hub: 2.427111028416567e-05
Node: 11 Authority: 0.00012430023819660696 Hub: 0.0
Node: 12 Authority: 3.519898071444847e-06 Hub: 7.682257200319916e-06
Node: 13 Authority: 3.80917560781302e-05 Hub: 0.0
Node: 14 Authority: 2.5866075215289027e-05 Hub: 5.521278520917273e-05
Node: 15 Authority: 1.1497913183781055e-05 Hub: 0.0
Node: 16 Authority: 0.00014632827374524336 Hub: 0.0
Node: 17 Authority: 6.457460767129105e-05 Hub: 3.0088616612213085e-05
Node: 18 Authority: 2.164947149799312e-05 Hub: 0.0
Node: 19 Authority: 1.3930575351370015e-05 Hub: 1.8276447242537122e-05
Node: 20 Authority: 0.0001171430445086412 Hub: 5.3844302980788876e-05
Node: 21 Authority: 2.1000700253293717e-06 Hub: 3.32760450609083e-05
Node: 22 Authority: 7.905916887068662e-05 Hub: 0.0
Node: 23 Authority: 5.388828331658944e-05 Hub: 6.84993499458857e-06
Node: 24 Authority: 2.1000700253293717e-06 Hub: 0.0
...
Node: 10877 Authority: -4.723418436092913e-21 Hub: 0.0
Node: 10873 Authority: 1.553061751775325e-06 Hub: 0.0
Node: 10874 Authority: 0.0 Hub: 2.1421087293819075e-05
Node: 10875 Authority: 2.137185791790055e-06 Hub: 0.0
Node: 10876 Authority: 2.137185791790055e-06 Hub: 0.0
```

Comparison of the two algorithms:
Comparing the results obtained from both algorithms, we can see that there are some similarities and differences in the node scores. In general, nodes with higher PageRank scores tend to have higher Authority and Hub scores as well, indicating that they are important nodes in the graph.

However, there are some nodes where the PageRank score and Authority/Hubs score differ significantly.

Overall, while both PageRank and Authority/Hubs algorithms provide valuable insights into the importance and influence of nodes in a graph, they may emphasize different aspects of the node's role and function in the network, leading to some variation in the scores obtained.