A3 Report

Name: Krishnasai Addala ,Sahil Deshpande, Kabir Dev Paul Baghel

Q1:

Dataset Description:

Dataset taken from https://snap.stanford.edu/data/ca-GrQc.txt.gz

Arxiv GR-QC (General Relativity and Quantum Cosmology) collaboration network is from the e-print arXiv and covers scientific collaborations between authors papers submitted to General Relativity and Quantum Cosmology category. If an author i co-authored a paper with author j, the graph contains a undirected edge from i to j. If the paper is co-authored by k authors this generates a completely connected (sub)graph on k nodes.

The data covers papers in the period from January 1993 to April 2003 (124 months). It begins within a few months of the inception of the arXiv, and thus represents essentially the complete history of its GR-QC section.

1. Number of nodes: 5242

2. Number of edges: 28980

3. Average In-Degree: 5.528424265547501
   Average In-Degree: The average in-degree of a directed graph is calculated as the sum of the in-degrees of all nodes divided by the number of nodes. The in-degree of a node is the number of edges incoming to the node.

4. Average Out-Degree: 5.528424265547501
   Average Out-Degree: The average out-degree of a directed graph is calculated as the sum of the out-degrees of all nodes divided by the number of nodes. The out-degree of a node is the number of edges outgoing from the node.

5. Node with the highest in-degree: 4233
   Node with Maximum In-Degree: The node with the maximum in-degree is the node with the highest in-degree among all nodes.

6. Node with the highest out-degree: 4233
   Node with Maximum Out-Degree: The node with the maximum out-degree is the node with the highest out-degree among all nodes.

7. Network Density: 0.0010548414931401452

Density of Network: The density of a directed graph is calculated as the number of edges divided by the maximum possible number of edges. For a directed graph with n nodes, the maximum possible number of edges is n * (n - 1).

8. Formula: average_in_degree = sum(in_degrees) / n

9. Formula: average_out_degree = sum(out_degrees) / n

10. Formula: density = m / (n * (n - 1))

11. Local Clustering Coefficient: The local clustering coefficient of a node in a directed graph is calculated as the ratio between the number of edges that exist between its neighbors and the maximum possible number of edges that could exist between its neighbors. For a node with k neighbors, the maximum possible number of edges that could exist between its neighbors is k * (k - 1).
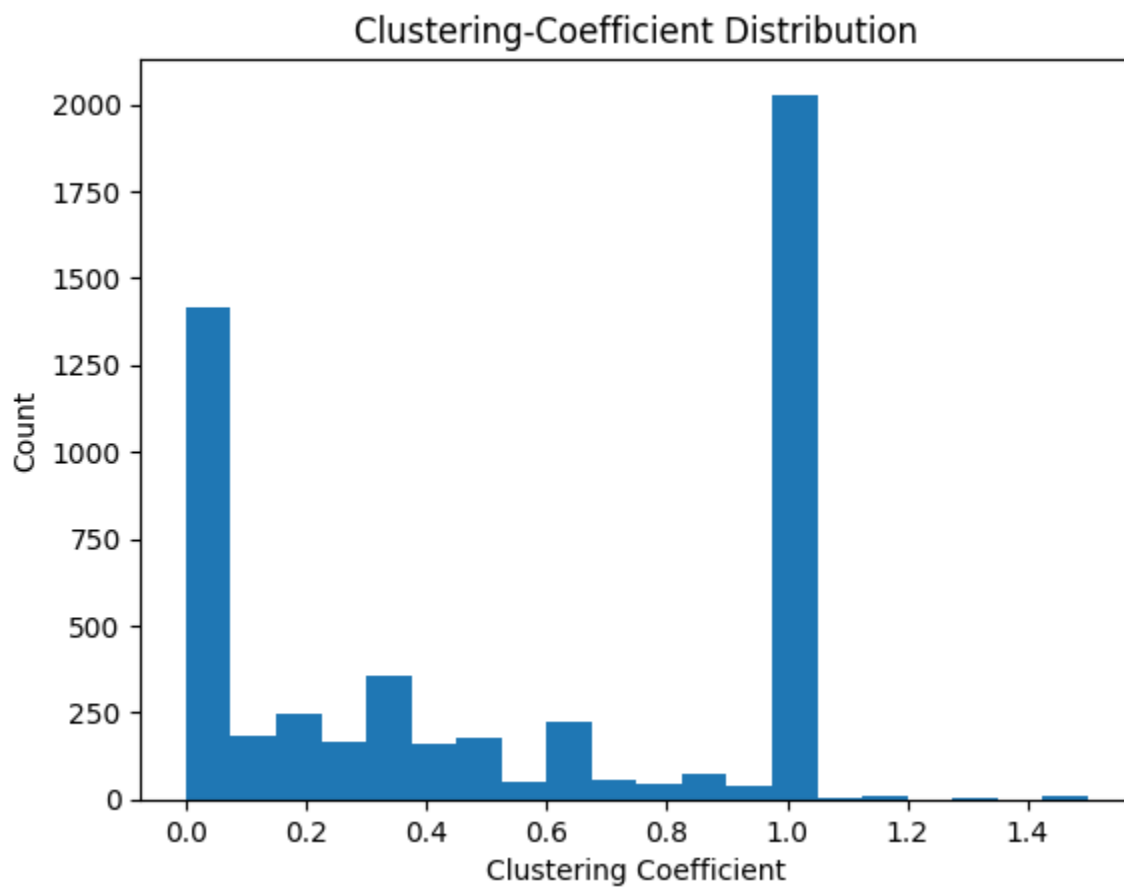
12. Formula: clustering_coefficient = edges_between_neighbors / (k * (k - 1))

1)

Out-Degree Distribution

2)

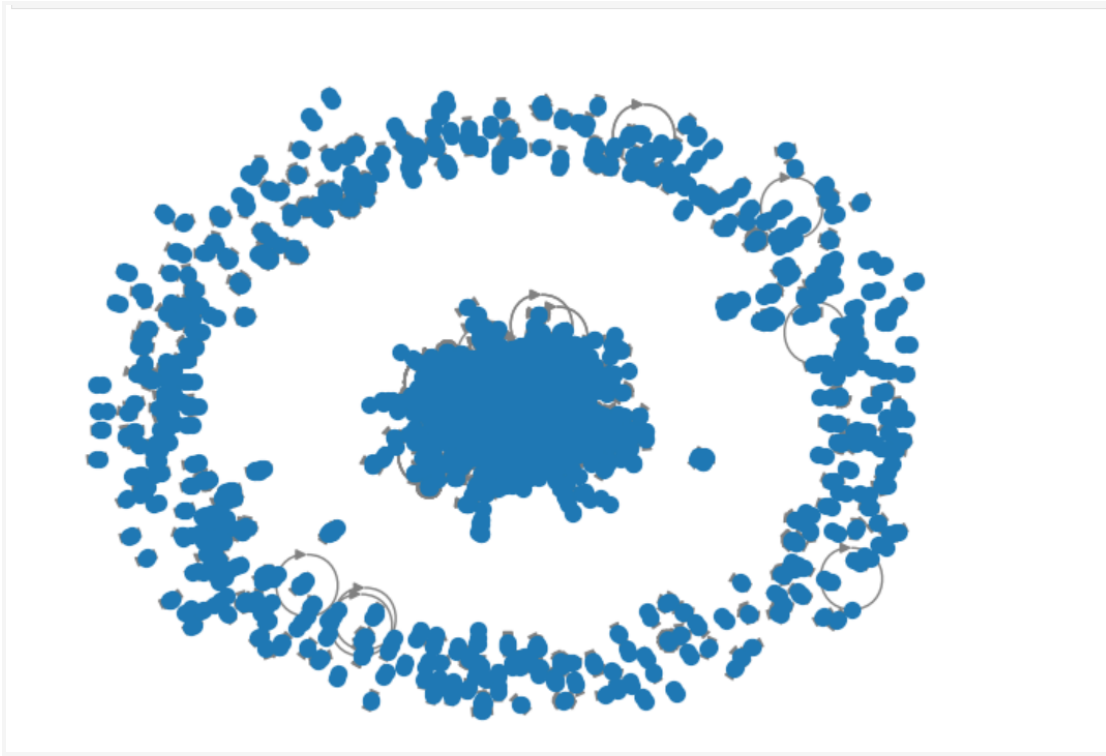

Clustering-Coefficient Distribution
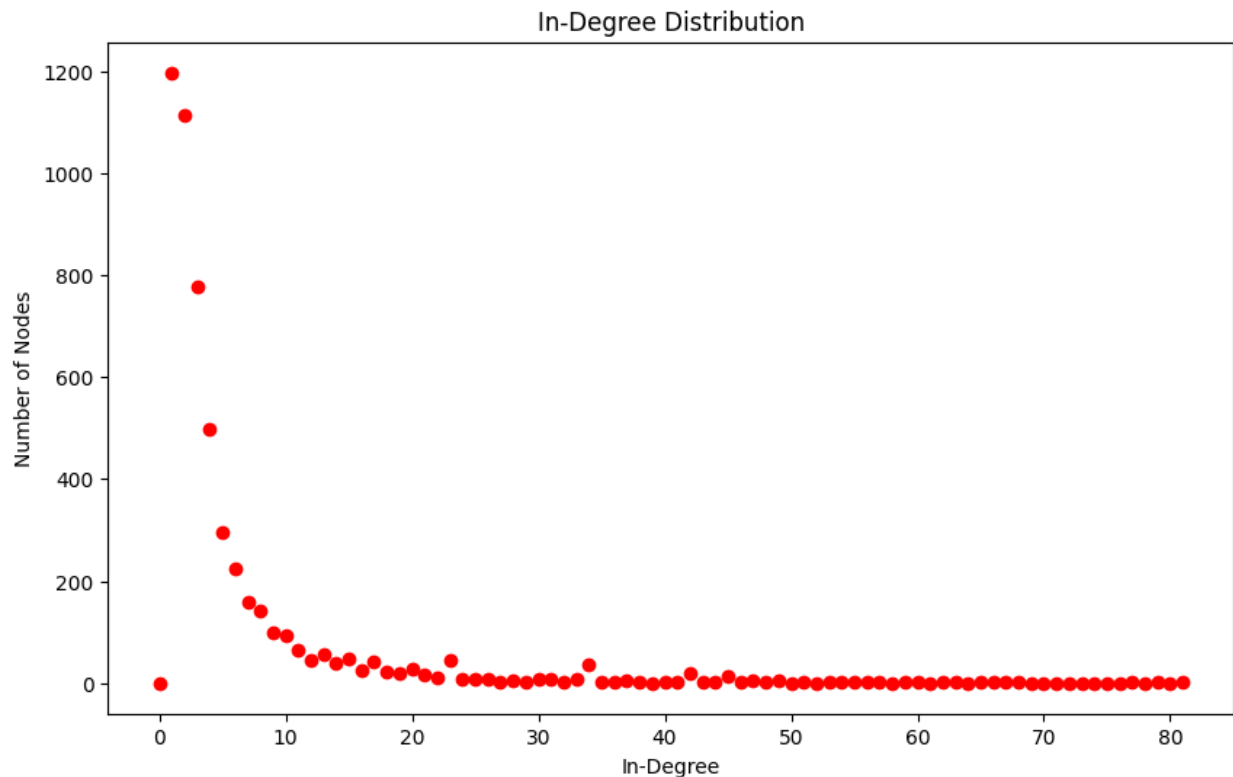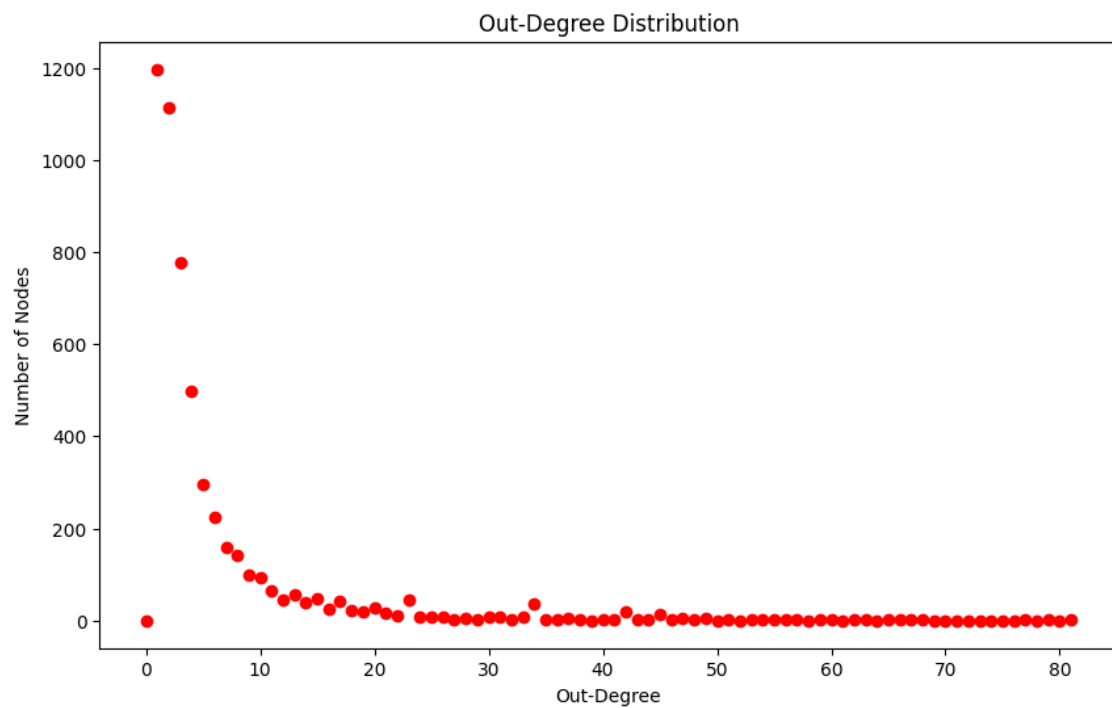
Q2:
Graphical representation of the Dataset:

Sample of the scores obtained from the dataset:

```
Node 3466
 PageRank Score: 0.00028791038987549284
 Authority Score: 4.134765427976227e-06
 Hub Score: 4.134765427988061e-06
_____

Node 937
 PageRank Score: 0.00019009872909030306
 Authority Score: 1.7295705867898783e-06
 Hub Score: 1.7295705867896358e-06
_____
```

```
Spearman's rank correlation coefficient between PageRank and authority scores:  0.05809877657296374
There is a positive correlation between PageRank and hub scores.
```

```
Spearman's rank correlation coefficient between PageRank and hub scores:  0.05809353675717669
There is a positive correlation between PageRank and hub scores.
```

We can see the scores have a weak positive correlation This shows that the two algorithms rank nodes in a similar way based on their importance in the network to some extent, but there are also differences in their ranking methodologies that lead to variations in the rankings.

Scores were obtained using the Networkx library