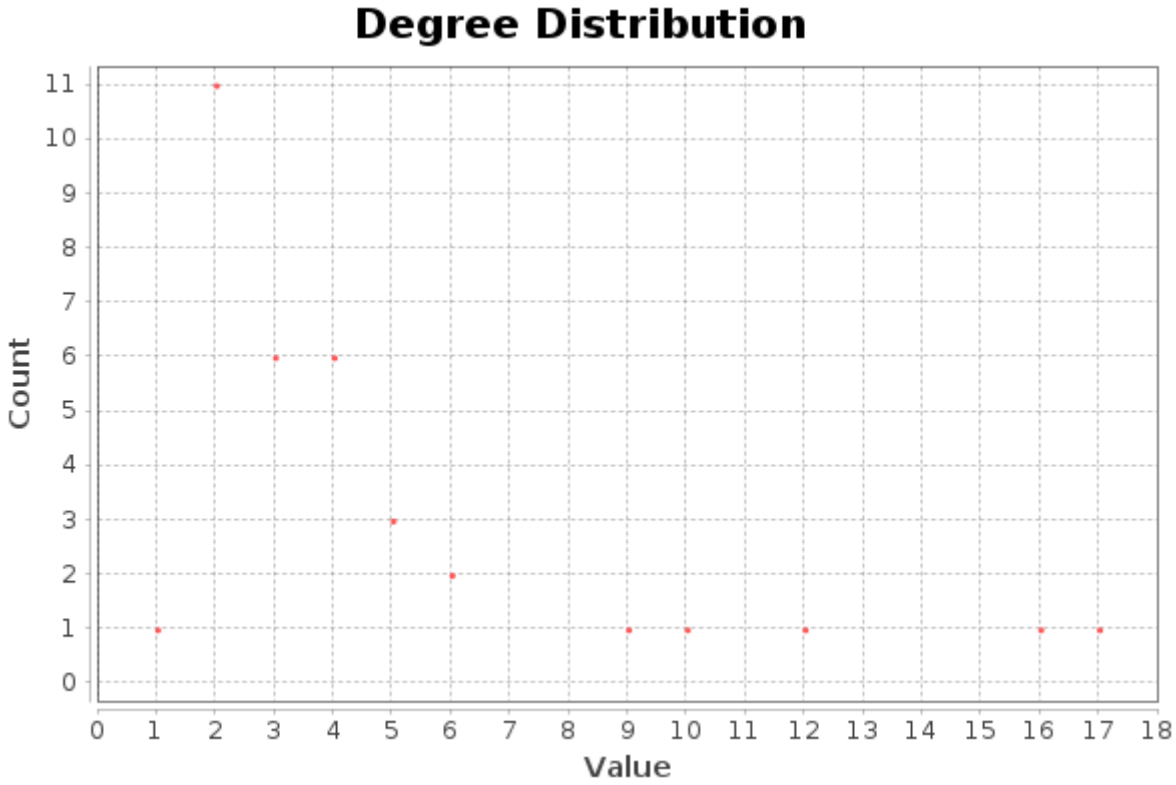
CS529 HW3

Q1 – Network and Node Metrics

Q1-A Degree Distribution



In the figure below, you can see the degree distribution graph of the provided dataset Gephi returns. It suggests that the distribution follows the power law.

Q1-B Network Density

Computing graph density on Gephi statistics panel results in a density value of **0.139.** This suggests the graph is relatively loosely-connected and sparse in terms of its edge content.

Q1-C Degree Centrality

When computing average degree under Gephi statistics panel, this outputs node-level degrees as well. As instructed, the top-10 highest degree values are as follows:

Node ID	Degree
34	17
1	16
33	12
3	10
2	9

4	6
32	6
9	5
14	5
24	5

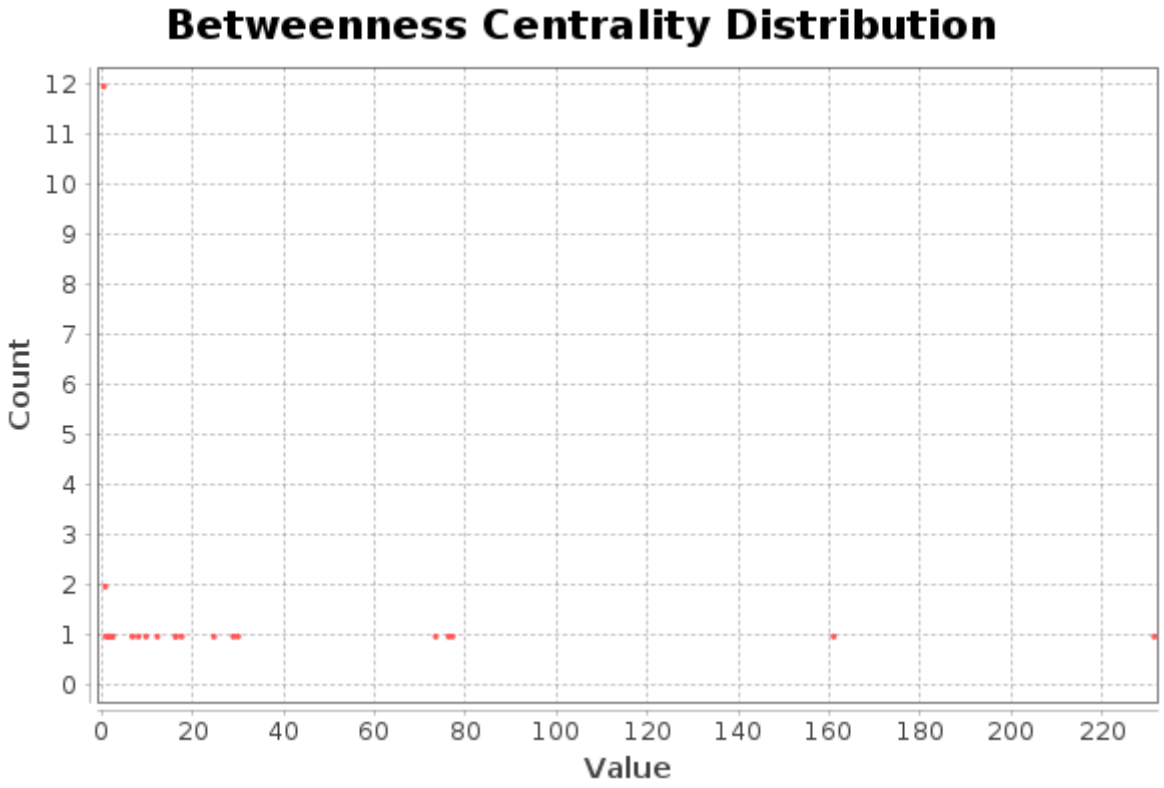
These results are also consistent with the distribution table provided in Q1-A.

Q1-D Betweenness Centrality

When running network diameter under Gephi statistics, betweenness centrality for each node is computed. As instructed, top-10 highest ranking nodes are as follows, rounded for interpretability:

Node ID	Betweenness
1	231.1
34	160.6
33	76.7
3	75.9
32	73.0
9	29.5
2	28.5
14	24.2
20	17.1
6	15.8

Distribution of betweenness centrality is shown below as well:

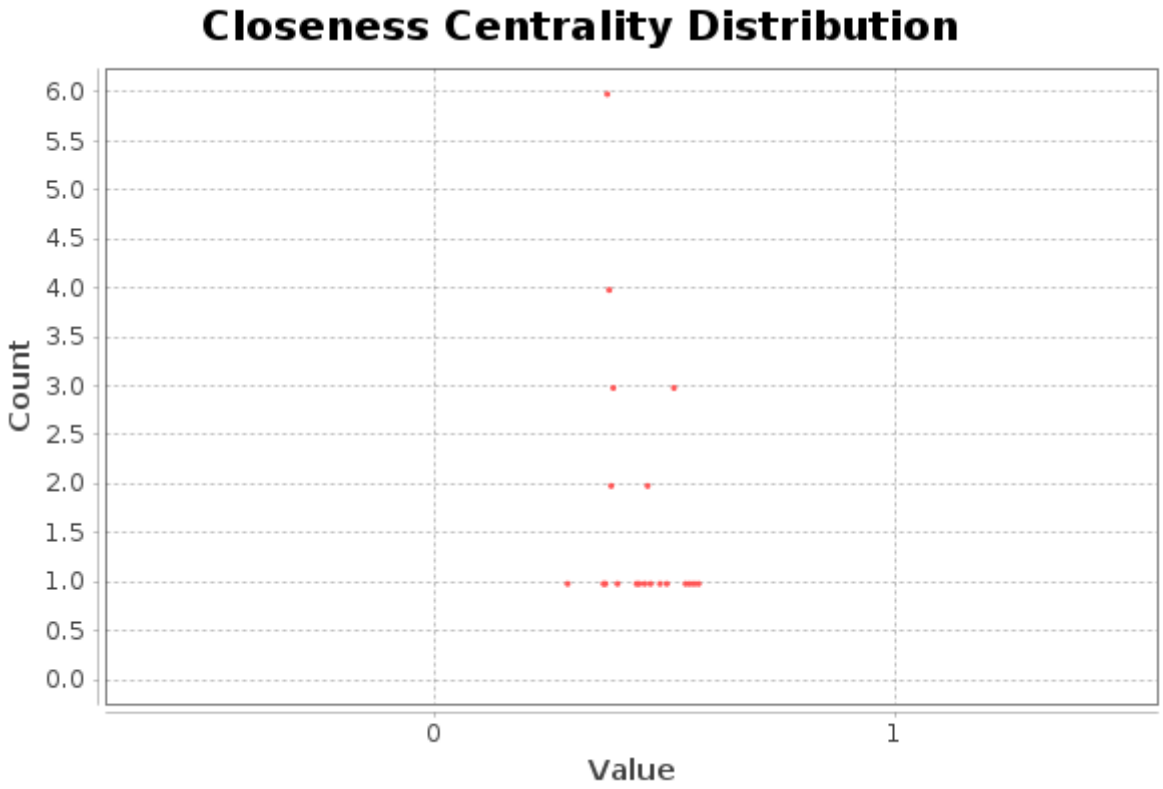


Q1-E Closeness Centrality

Closeness is computed in a similar fashion to betweenness, an the table of top-10 highest ranking nodes are as follows, rounded for interpretability:

Node ID	Closeness
1	0.57
3	0.56
34	0.55
32	0.54
33	0.51
9	0.51
14	0.51
20	0.5
2	0.49
4	0.46

Distribution of closeness centrality is shown below as well:

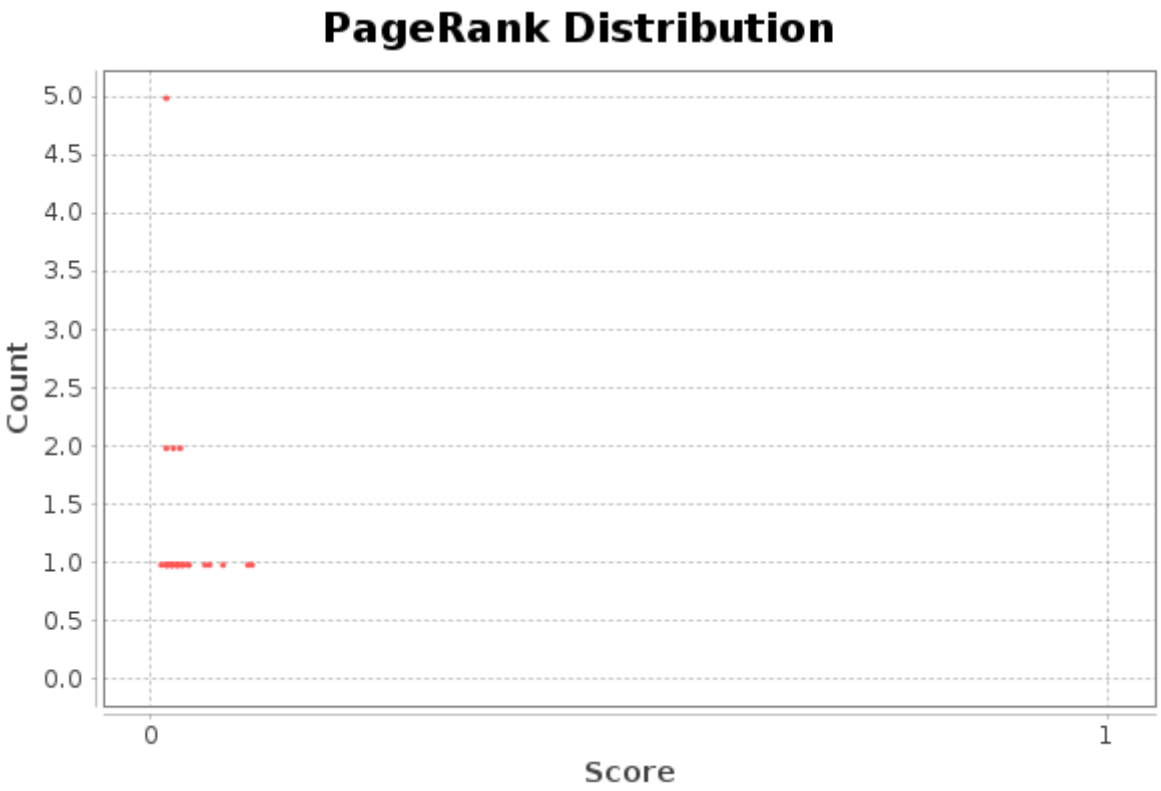


Q1-F Pagerank Centrality

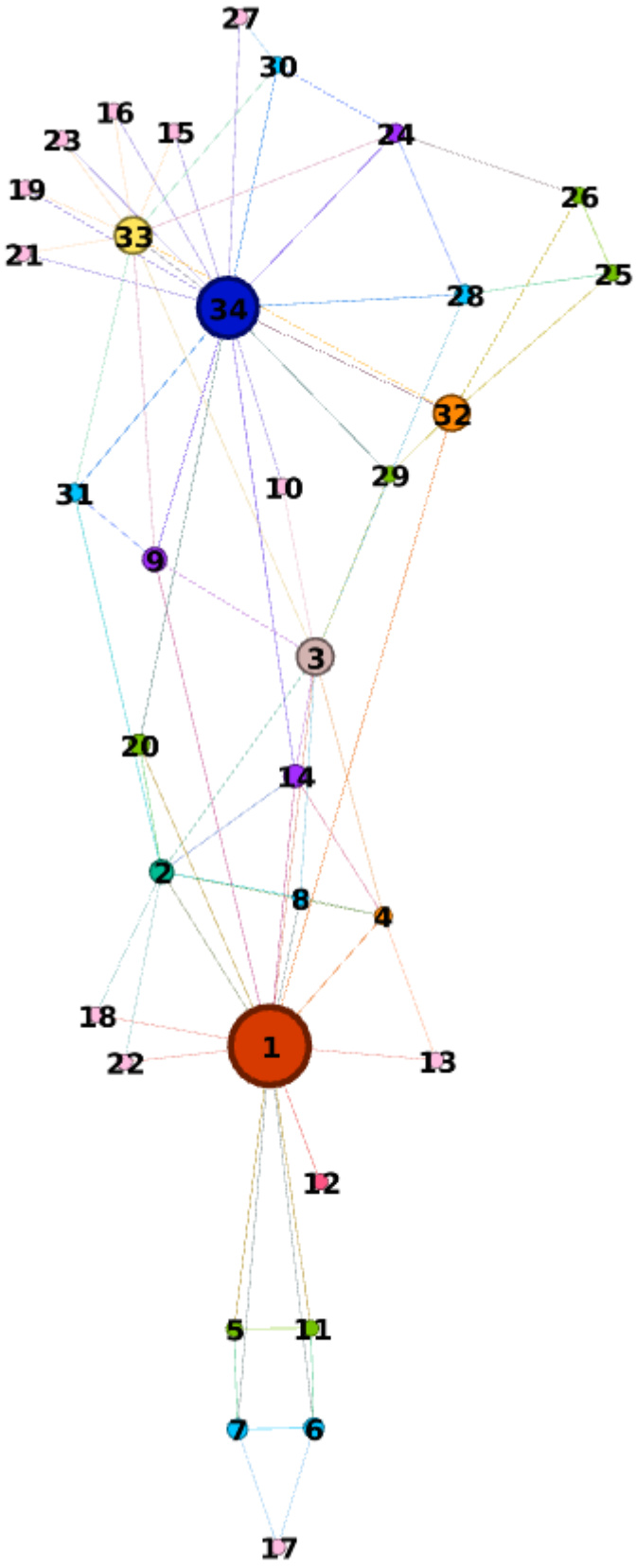
Computed under PageRank segment in Gephi statistics panel, since PageRank is a node-level metric as well, the top-10 ranking nodes are shown below:

Node ID	PageRank
34	0.1034
1	0.0992
33	0.0732
3	0.0591
2	0.0543
32	0.0375
4	0.0366
24	0.0316
9	0.0304
14	0.0302

Also, PageRank distribution is as follows:

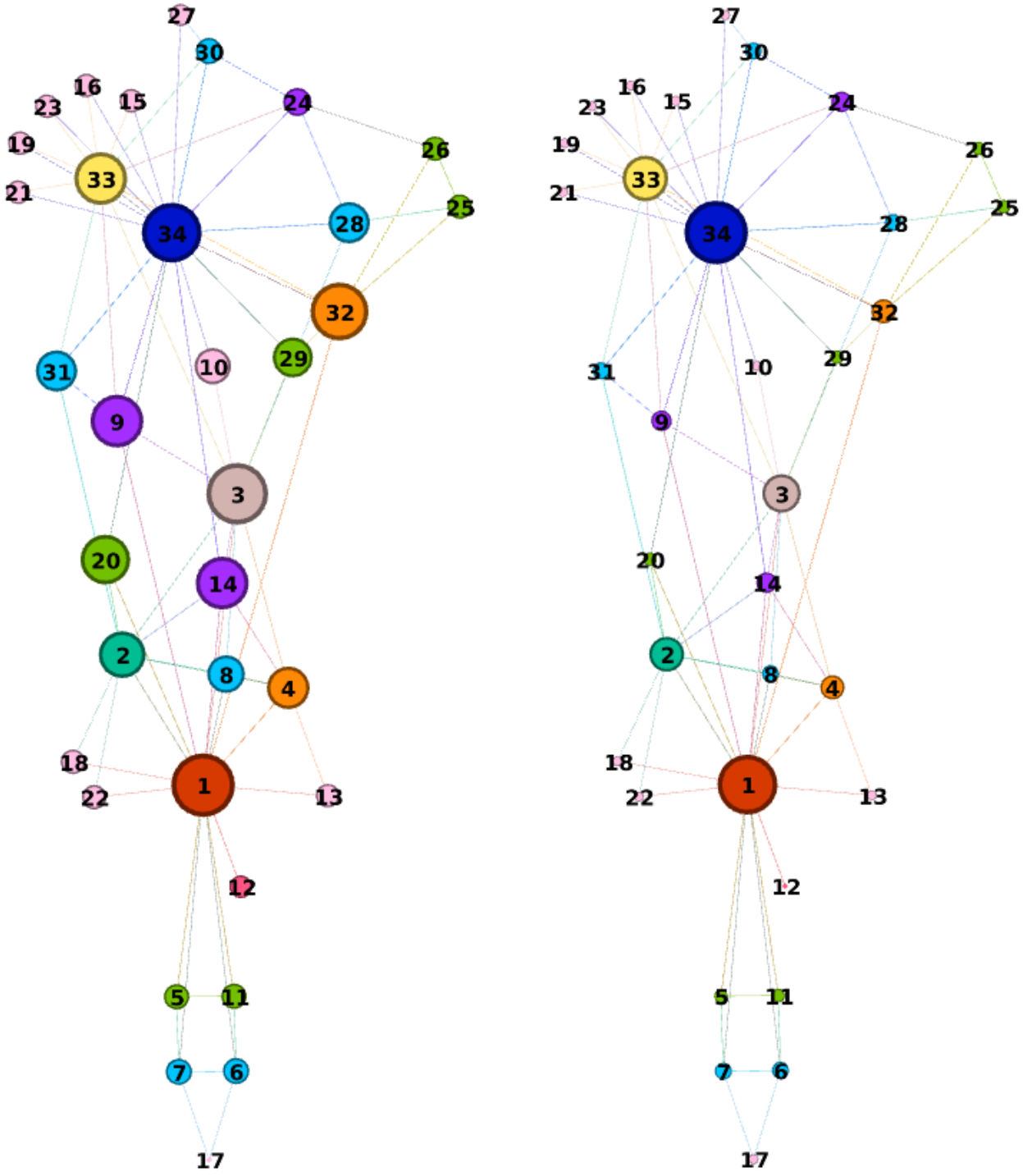


Q2 Visualization Based on Betweenness



In the figure, nodes labeled 1, 34, and 33 are the nodes having the largest size, which is consistent with betweenness ranking shared previously.

Q3 Interpretation and Analysis



On the left, the nodes are ranked based on **closeness** whereas on the right, the ranking is performed based on node **degree**. By definition, if a node has high betweenness it means shortest paths between a pair of nodes likely passes through this node. By extension, if a node has multiple connections, this node likely shortens distance between a node pair. This suggests if a node has high betweenness, it also likely has to have a high degree, which holds true for this case as shown in the figures above.

In terms of closeness, a high closeness value indicates the node in question is located at or around the center. In a sense, we can imagine closeness as water spilled from a certain height: the amount around the impact point contains more water than the edge of the spill. This is why compared to

other centralities closeness distribution is relatively high and uniform around the graph center. While nodes having a high degree or betweeness seem to have a high closeness value, the opposite isn't always true. A node with high degree can 'reach out' to other nodes quickly by using direct connections which is why such nodes tend to lie in the middle and have high closeness value. This is evident for nodes 34,32,3, etc. However, a node located around the center does not imply that it has a strong connection to its neighbourhood, such as nodes 31, 9, 10, 29, etc. Because of this, such nodes have low degree and betweenness values.

In short, a node having a high degree likely has high betweenness, and vice-versa. It's also true that if a node has high betweenness or degree, it tends to have high closeness value. However, a node having a high closeness does not guarantee that this node will have high betweenness or degree. Note that this discussion does not include networks with bridge connections. These graphs contain clusters of weakly connected nodes where the inter-cluster traffic passes through a few bottleneck nodes that tend to have small degrees, yet high betweenness due to the lack of inter-cluster connections. Since the example dataset do not show this characteristic, it is omitted from the discussion.