1 Zache Karate Network

1.1 Visualization

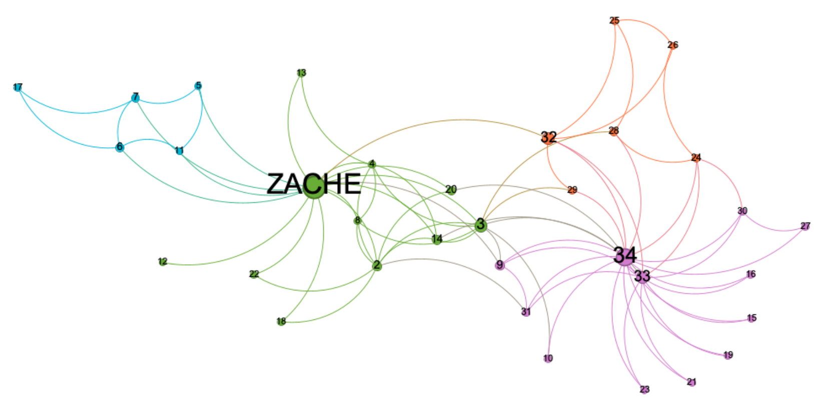

Figure 1: Visualization of Zache's karate network. Larger nodes have a larger betweenness centrality. Nodes are colored by their inferred class. From this visualization, we can see there appear to be a few friend groups and a clear two factions within the network. The blue cliche on the left seemed to follow Zache and his green cliche, while the orange cliche followed 34 and his cliche. Both Zache and 34 have a very high betweenness centrality and degree, seeming to suggest their importance to the network.

1.2 B-E

Connected	Yes
Max Degree	17
Average Degree	4.5
Diameter	5
Clustering Coefficient	0.5706

Table 1: Various measures of the network. The network has low diameter, and a high clustering coefficient which suggests that it takes only a few connections to get from a given node to any other node. The max degree is also much larger than the average degree. This suggests that there is at least one hub, very central node to the network or a lot of nodes that have very few edges to other nodes.

1.3 Shortest Paths

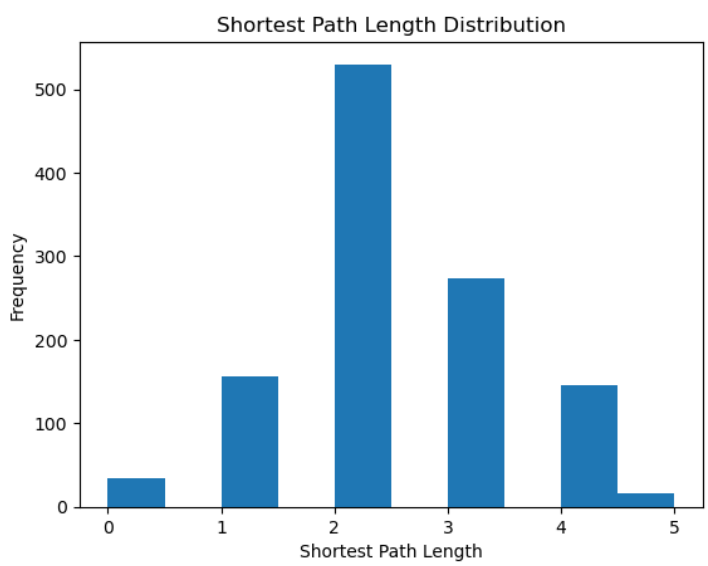

Figure 2: Histogram of the shortest paths between any two nodes in the graph, computed as $ShortestPath(s, d) \ \ \forall s \in V, \ \forall d \in V$. From this data, we can see that the average shortest path is around 2, meaning that on average, it takes only 2 connections to get from one node to another node.

1.4 Degree Distribution

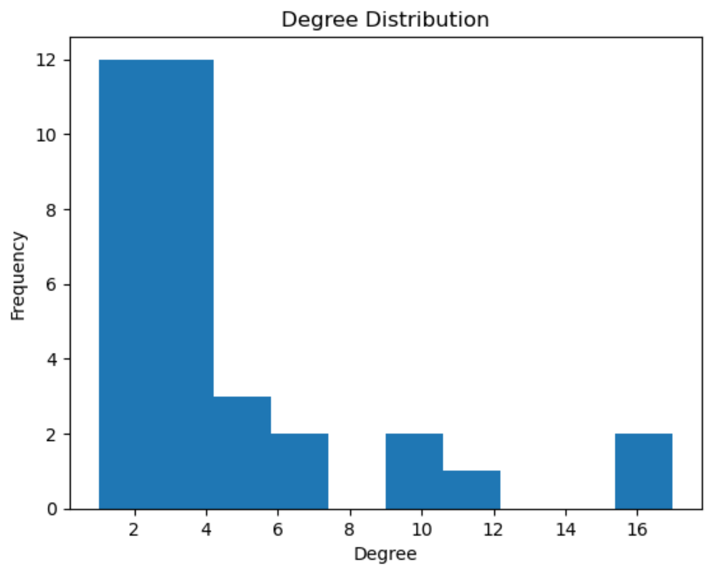

Figure 3: Degree distribution for the karate network. There are many nodes that have less than 6 degree, while a few that have greater and 2 that have over 16. These two nodes are node 34 and Zache

1.5 Betweenness Centrality

Node	Betweenness Centrality
Zache	0.437635
34	0.304075
33	0.145247
3	0.143657
32	0.138276

Table 2: Top 5 most central nodes to the karate network by betweenness centrality

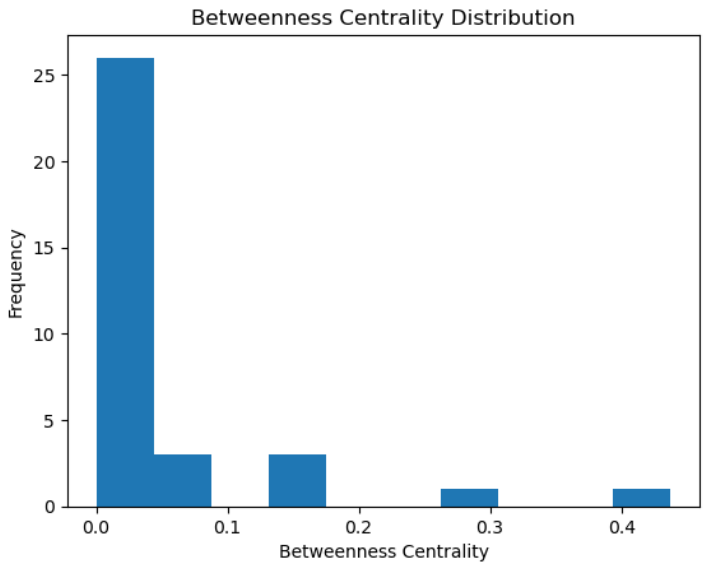

Figure 4: Histogram of the betweenness centrality for each node in the network. A lot of nodes have a very low betweenness centrality, with 2 nodes having a standout betweenness centrality. In total there are 4 nodes that have an above 0.1 betweenness centrality which seems to suggest that there are 4 nodes that are crucial to connecting the network, which is partly why we see 4 cliches in Figure 1

1.6 Clustering Coefficient

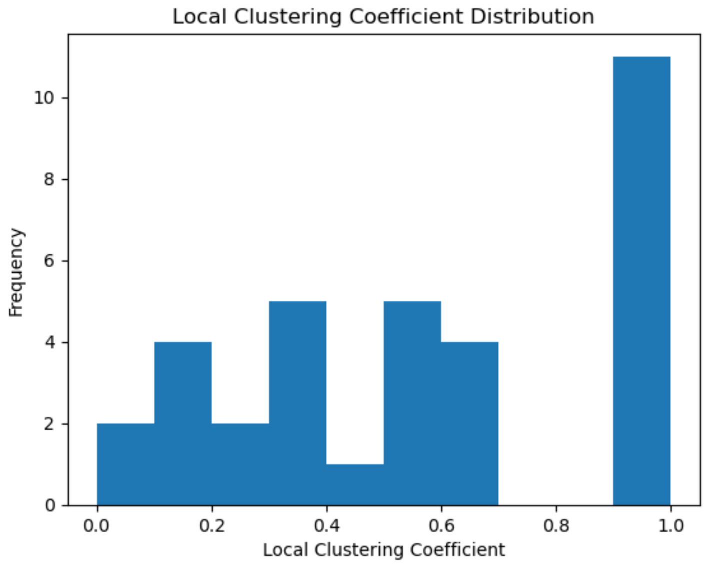

Figure 5: Histogram of the clustering coefficient of each node in the graph.

2 Question 2:

2.1 Visualization

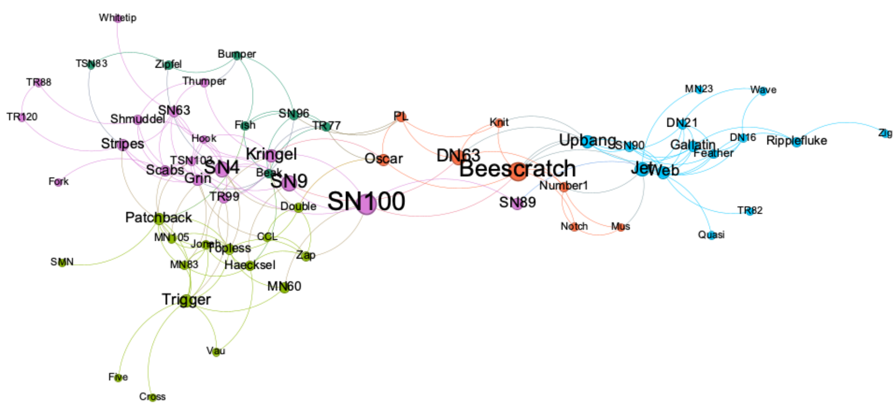

Figure 6: Visualization of the dolphin communication network. The size of the node corresponds to its betweenness centrality. The nodes and edges are colored by their inferred class. We can see that there are 4 or 5 communities in this network, each being very interconnected. Interestingly, there is still a lot of communication across communities but usually, 2-3 dolphins do the majority of the intra-community communication.

2.2 Dolphin Network Questions B-E

Connected	Yes
Max Degree	12
Average Degree	5.129
Diameter	8
Clustering Coefficient	0.259

Table 3: Various measures of the network. Compared to the karate network data in Table 1, the dolphin network has a smaller max degree, larger diameter, and lower clustering coefficient. All of these suggest that it is overall less connected than the karate network. We can see that this appears to be the case when we compare the two visualizations.

2.3 Shortest Paths

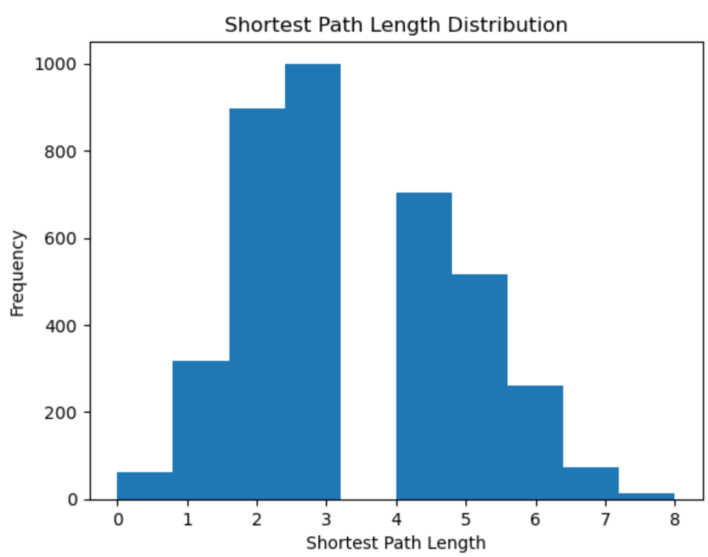

Figure 7: Histogram of the shortest paths between any two dolphins in the network, computed in the same was as Figure 2. We can see that the average shortest path is around 3.5 and that the distribution of paths looks mostly Gaussian, concentrated around the average and getting sparser towards the extremes.

2.4 Degree Distribution

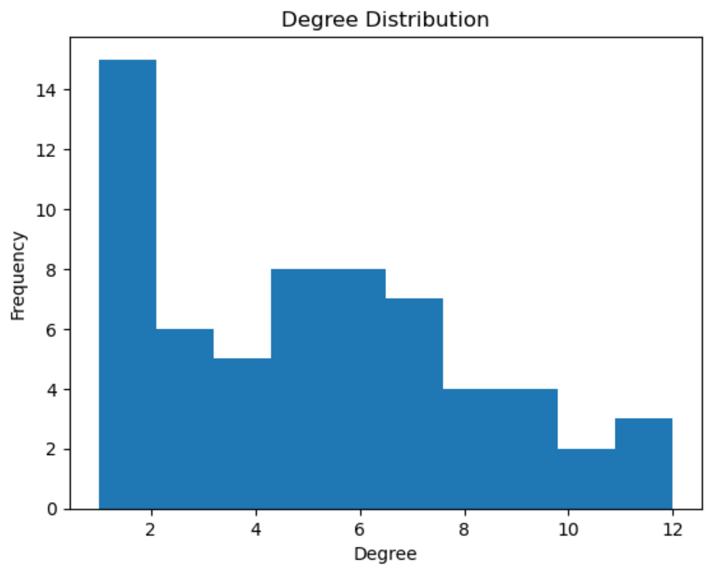

Figure 8: Histogram of the degrees of each dolphin node in the network. The degrees are mostly uniform with a spike around the lower end of the distribution, suggesting there are a number of loner dolphins.

2.5 Betweenness Centrality

Node	Betweenness Centrality
SN100	0.248237
Beescratch	0.213324
SN9	0.143150
SN4	0.138570
DN63	0.118239

Table 4: Top 5 most central nodes to the dolphin network by betweenness centrality

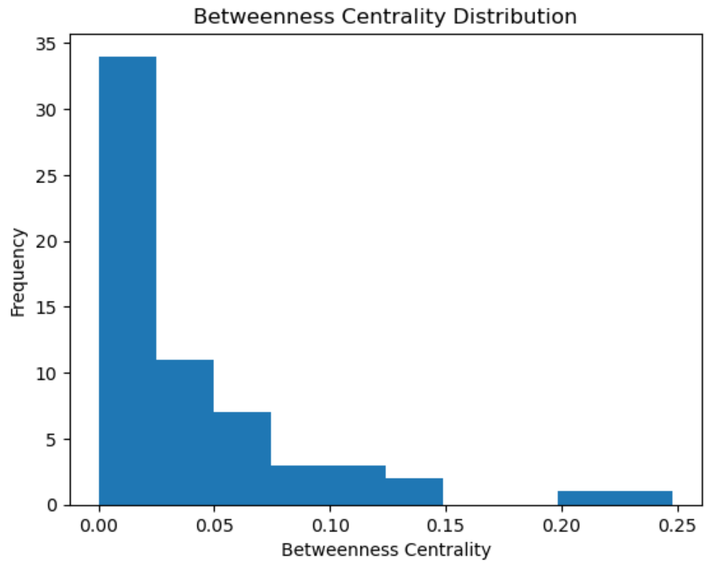

Figure 9: Histogram of the betweenness centrality for each node in the dolphin network. A lot of nodes have very low betweenness centrality with a gradual drop off in the number of nodes that have betweenness centrality from 0 to 0.15. There are two nodes that have a standout betweenness centrality: SN100 and Beescratch. This suggests that these two nodes are crucial to communicating between the multiple groups of dolphins. We can see that in Figure 6, this appears to be the case; both of these dolphins are positioned in the center of the graph connecting two groups of otherwise mostly unconnected dolphins.

2.6 Clustering Coefficient

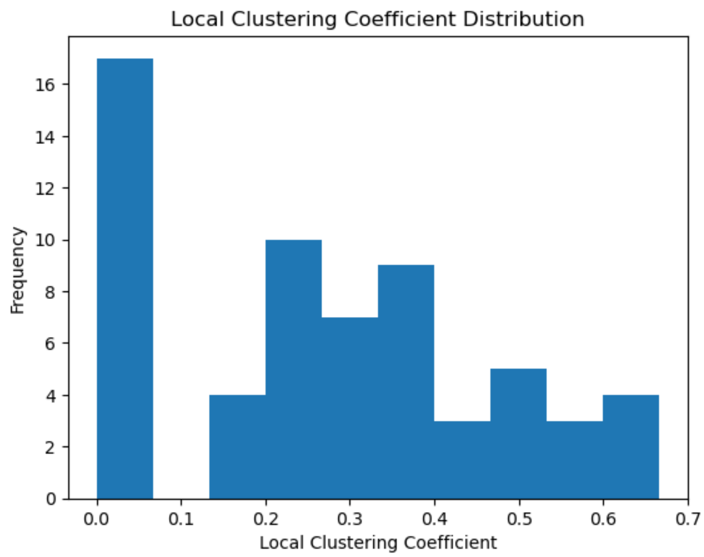

Figure 10: Histogram of the clustering coefficient of each dolphin node in the network. The clustering coefficient graph is similar to the degree distribution, with some nodes having very low clustering coefficient, and then a mostly uniform distribution over medium to high clustering coefficient values. This suggests that apart from around 17 "loner" dolphins, the network is tightly connected, with a lot of dolphins belonging to moderately strong to strong cliches.

3 Asymmetric Prison Network

3.1 Visualization

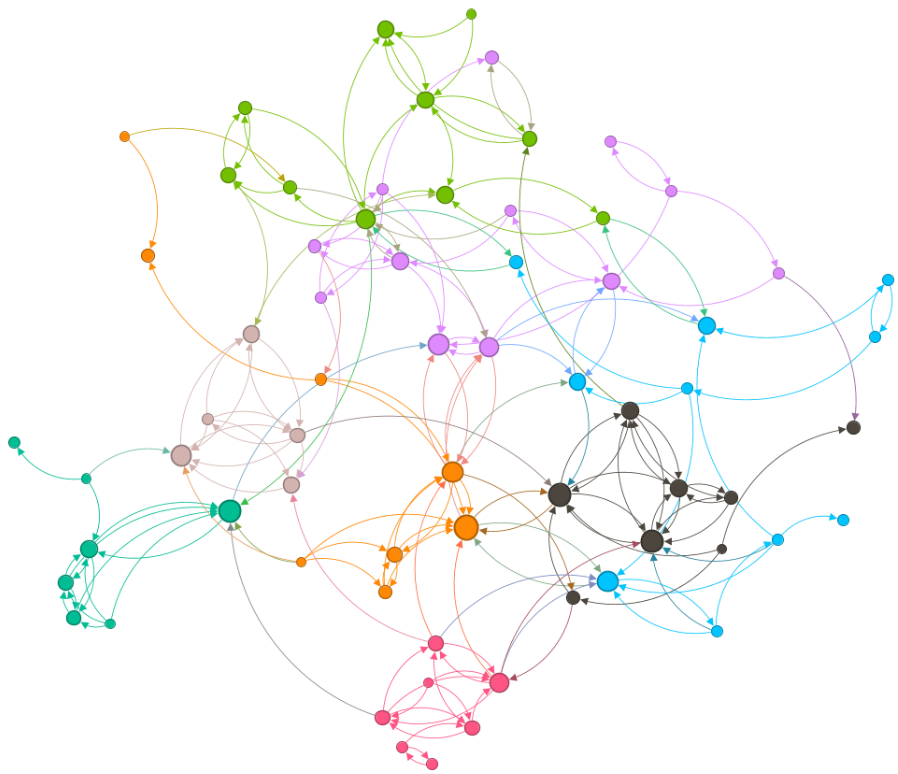

Figure 11: Prison network visualization.

3.2 Connected

While this measure is still relevant to a non-symmetric network, we need to specify what we mean by connected. The graph is weakly connected in the sense that if we ignore edge directions, there is a path from any node to any other node, i.e. there is no node that does not have an edge to or out of it. The graph however, is not strongly connected. The graph does not have a directed path between all nodes. For example, the orange node in the top left of Figure 11 has directed edges going into it meaning that it is inaccessible from any other node in the graph.

So the graph is connected in a weak sense but is not strongly connected.

3.3 Maximum Degree

This is still a valid measure in a non-symmetric graph, but again, we need to qualify what we are talking about when we mean degree. In directed graphs, nodes have both an in-degree – the number of edges going into the node – and an out degree – the number of edges that originate at this node and go out to other nodes.

If we are talking about the maximum in-degree, that is **8** and the maximum out-degree is also **8** but these values could be different.

3.4 Diameter

The diameter of a network is still a useful measure for a directed graph, but if the graph is not strongly connected, if we compute it the same way as we did for undirected graphs, this could result in an infinite value and a not useful metric.

There are generally two approaches for applying diameter to a directed graph. With one approach, we consider the graph with only undirected edges, similar to the definition for weakly connected. In this case, we have a diameter of **7** for this network. In the other case, we consider the diameter of the largest strongly connected component of the graph. Basically, we remove nodes that make the graph not strongly connected. When we define the diameter like this we get a value of **11**.

3.5 Clustering Coefficient

Similar to diameter, the clustering coefficient for a network can be computed for directed graphs by considering directed triangles or ignoring edge directions. It remains a relevant measure. The clustering coefficient for this graph is **0.218** when we consider directed triangles and **0.33** when we ignore edge directions.

3.6 Shortest Paths

Shortest paths is still a valuable metric for a directed graph, but we only consider paths that can be reached. If there is no path from one node to another, we won't consider it in the shortest paths.

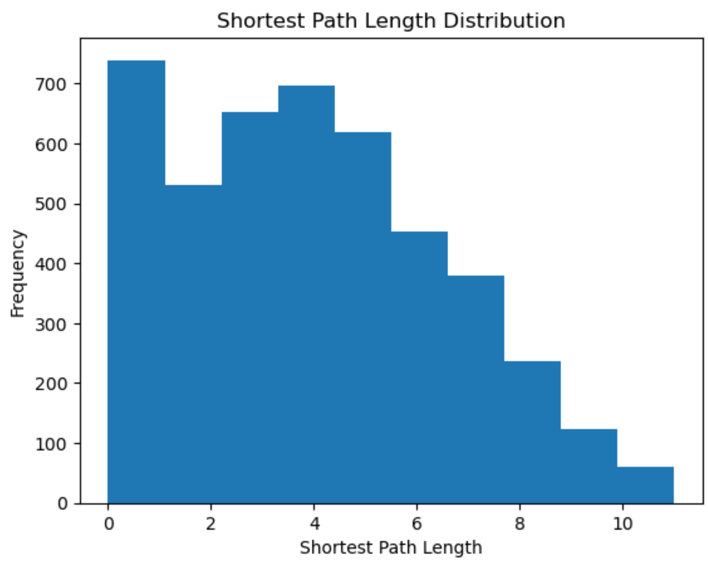

Figure 12: Histogram of the length shortest paths that exist between nodes. There are roughly an equal amount of 1-5 length paths between nodes, which suggests the small world effect, that people are at most a few connections away from each other. There are many less 6-11 length shortest paths.

3.7 Degree Distribution

This is still applicable to directed graphs. I created two plots to show in-degree and out degree. You could also combine these two by adding the in-degrees and out degrees, or some other operation.

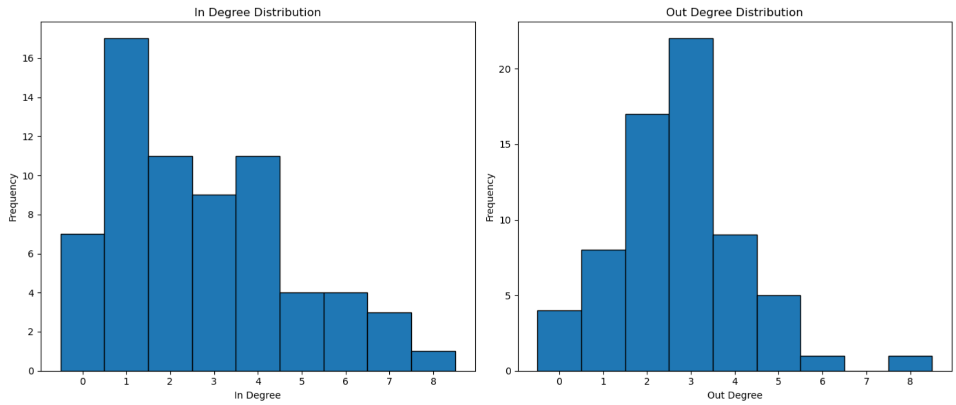

Figure 13: Histogram of both in-degree and out-degree for the prison network. The distributions look mostly similar, although the out degrees look more normal. This makes sense because the data set was collected by out-degree – prisoners were asked to name their closest friends without specifying an exact number of friends to say.

3.8 Betweenness Centrality

Betweenness centrality considers directed shortest paths, making it relevant for non-symmetric graphs.

Node	Betweenness Centrality
29	0.207862
51	0.180780
7	0.149196
32	0.148570
54	0.146465

Table 5: Top 5 most central nodes to the prison network by betweenness centrality

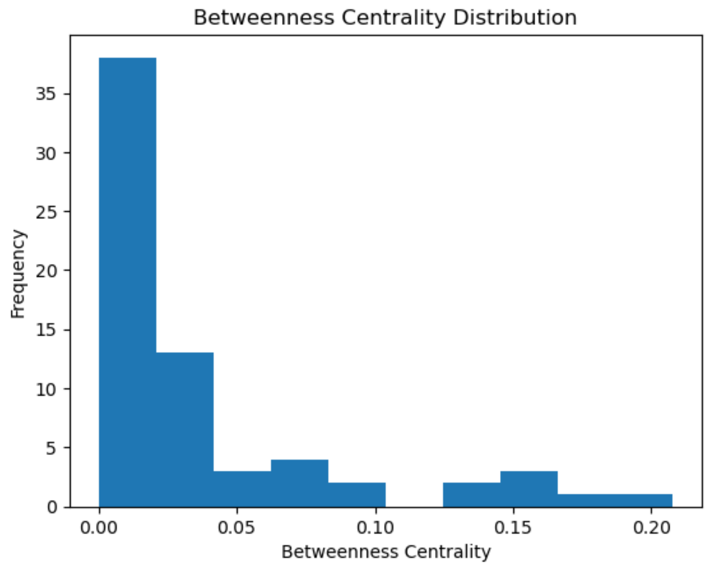

Figure 14: Histogram of the betweenness centrality of nodes in the prison graph. Notice that a lot of nodes have low betweenness centrality with a few central ones that have a high betweenness centrality. This implies that there are a few popular prisoners, who a lot more of the shortest paths between nodes go though.

3.9 Clustering Coefficient

Similar to other metrics, local clustering coefficients can be computed for non-symmetric graphs, either by ignoring directions or considering directed triangles.

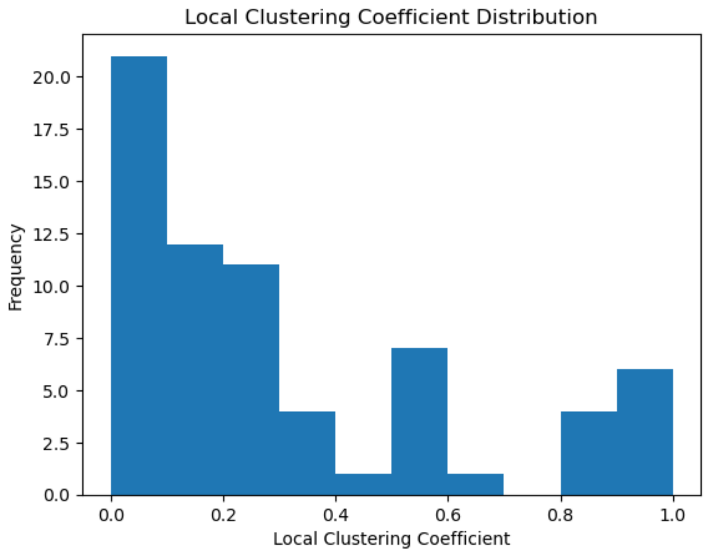

Figure 15: Histogram of the local clustering coefficient distribution. From this, we can see that there are many nodes with low clustering and a few nodes with very high clusters. This suggests that there are some very strong cliches in the network, which we can see in Figure 11.

4 Problem 2

Intuitively, no.

In this formulation of the six degrees of separation question, we only allow people to create directed edges in the graph based on their 10 closest friends. This means that each node in the graph has an out degree of exactly 10. With this construction, we can easily measure how many distinct nodes we could possibly see after 6 steps. If we start at a given node, there are 10 nodes to go to. We then go to one of those 10, and from there, there are exactly 10 nodes to go to. In the best case, one of these 10 is not the node that we just came from. We can continue to do this until we step 6 times. In the best case, it is clear that we observe up to 10^6, 1,000,000, nodes. This is substantially less than the 8 billion people in the world, so the probability that we find a random person within 6 steps is, in the best case,

$$\frac{10^6}{8 * 10^9} = \frac{1}{8000}$$

So, we expect that most people are not connected by 6 or less connections, which violates the six degrees of separation idea. However, if we changed this to undirected and say that a person knows 100 people on a first name basis the calculation changes dramatically. After 6 steps, in the best case we observe 10^{12} nodes. This is much bigger than the 10^9 population of the world, so we would expect that the diameter of the network is at most 6, meaning a random two people have a path of connections between them that is less than 6.