Question 4a

Calculating r(N,C) for N the Zachary karate club network and C both degree and closeness centrality, we get a relatedness index of -0.476 and -0.082 with p values 0.000 and 0.312 respectively (to 3 d.p.). This means that there is a negative correlation for the degree centrality between adjacent nodes and no correlation for the closeness centrality between related nodes. Note that the p value for the relatedness index of the closeness centrality is fairly high, so the confidence in the given value is lower. Also note that the Pearson Product-Moment Correlation Coefficient only identifies linear relationships.

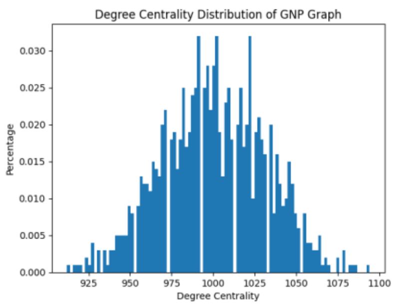
Question 4b

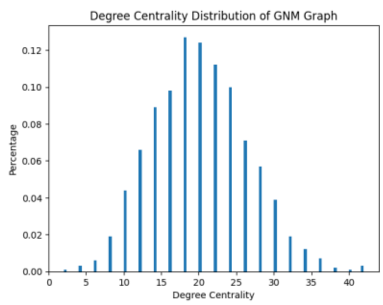
We use the Erdős-Rényi GNP/GNM models and Watts-Strogatz model as our two network models studied in the course. For the real networks we use the karate, football, and the words/ladder networks from networkx. Note, all of these networks are undirected as required in the question.

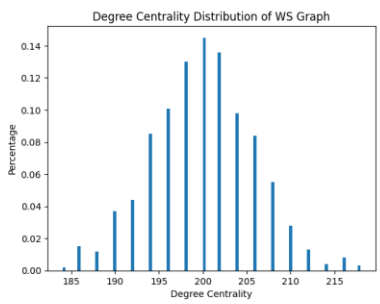
Graph	r(N,Degree)	r(N,Closeness)	Degree P Value	Closeness P Value
GNP Graph	-0.003	-0.003	0.016	0.016
GNM Graph	0.001	0.175	0.900	0.000
WS Graph	0.011	0.014	0.001	0.000
Karate Graph	-0.476	-0.082	0.000	0.312
Football Graph	0.162	0.306	0.000	0.000
Words Graph	0.649	0.970	0.000	0.000

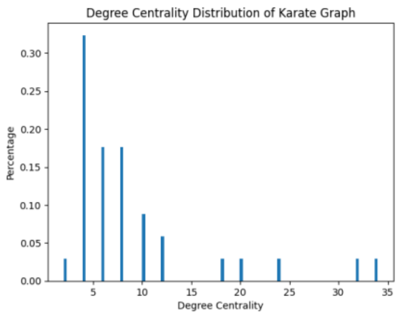
We see that there is almost no correlation for either centrality measures between adjacent nodes for the first 3 networks – although, there is little confidence about the degree centrality relatedness index for the GNM graph due to the very high p value. Note, there is actually a very slight positive correlation in the relatedness index for the closeness centrality of the GNM graph.

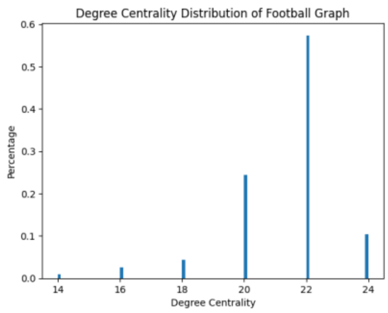
We see that the karate graph is the only network with a strong negative correlation. Note this is only for the degree centrality and interestingly there is no correlation for the closeness centrality – however, this may again be due to the relatively high p value. The football graph has a slight positive correlation, and the words graph has a very strong positive correlation for both centrality measures.

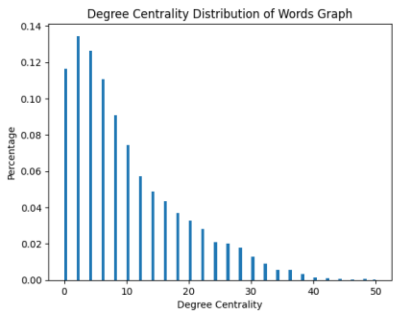




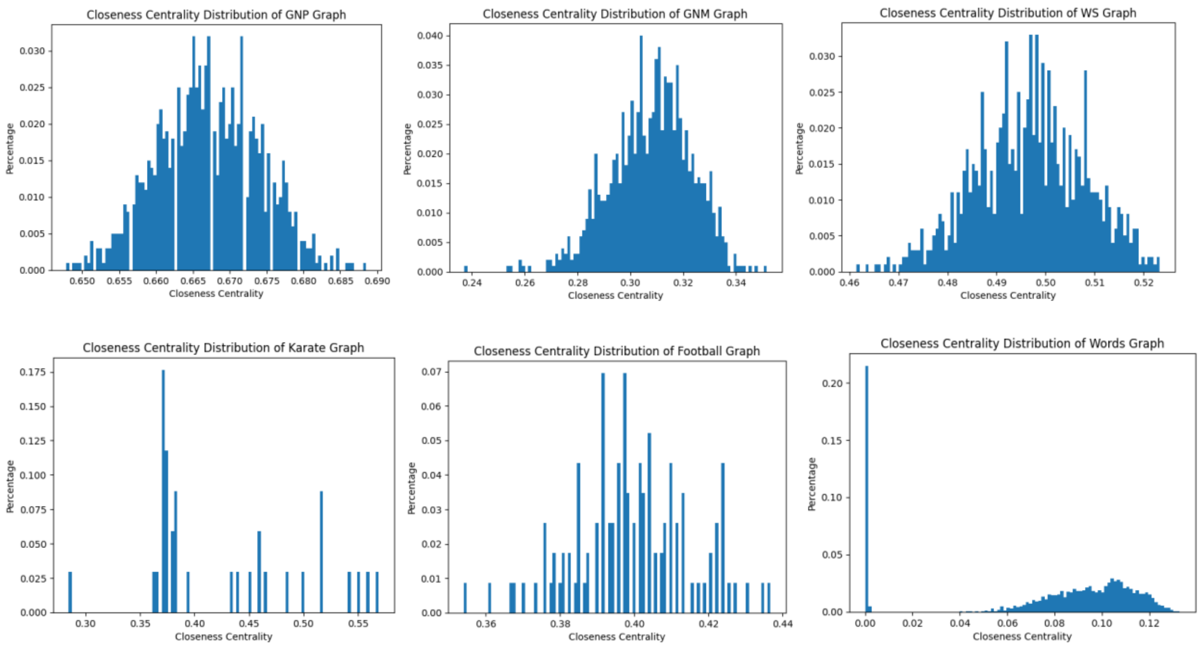




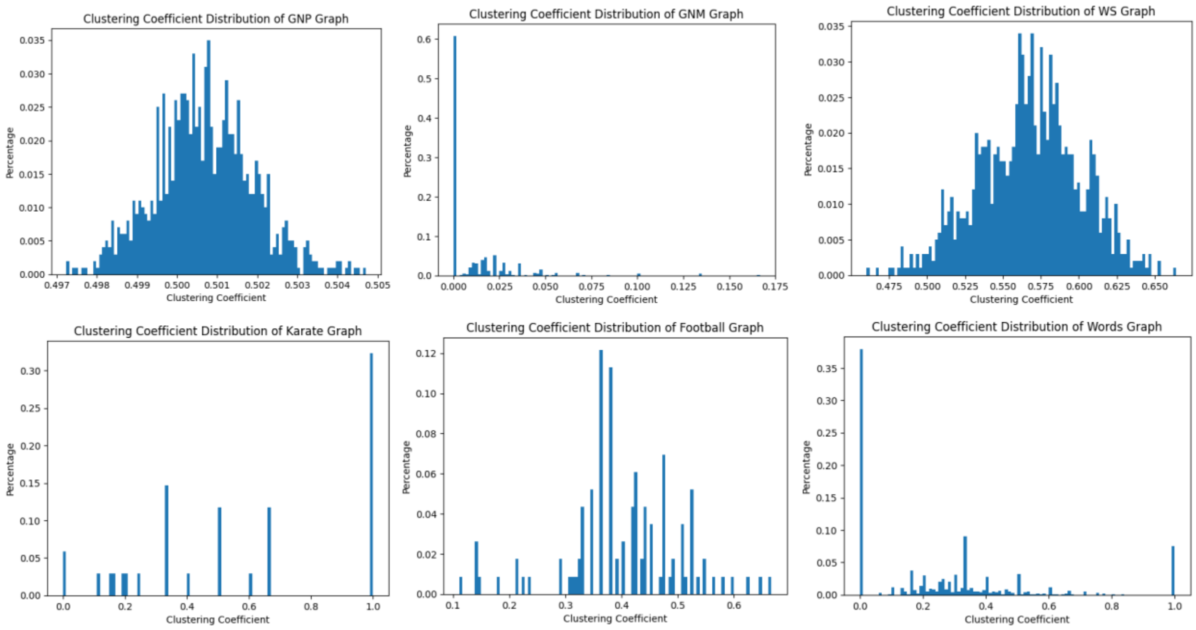




We first look at the degree centrality distributions of each graph to try and explain our results. We can clearly see that the networks with an almost zero degree centrality relatedness index follow a normal distribution, while the networks with a non-zero degree centrality relatedness index follow a weighted normal distribution – the more significant the weighting, the higher the relatedness index. It also seems that a positive or negative weighting reflects a positive or negative relatedness index (see karate and football graphs). In contradiction, the words graph has a high positive relatedness index but a strong negative weighting.



We see the same pattern if we look at the closeness centrality distributions of each graph – a normal distribution with the size of the positive/ negative weighting reflecting the size of the positive/ negative relatedness index. For example, we can see the slightly positive weighting of the closeness centrality distribution of the GNM graph giving the slightly positive value of its relatedness index. Notably, the words graph does not follow a normal distribution – a large set of nodes all have 0 closeness centrality with the remaining nodes having a very similar closeness centrality.



If we consider the local clustering coefficients, we see that this can have little impact on the relatedness coefficients. For example, the GNP and GNM graphs have similar degree/closeness centrality relatedness indices and degree/closeness centrality distributions but have extremely different clustering coefficient distributions.

If we instead consider the karate graph, we see that a large proportion of nodes have a clustering coefficient of 1, which may explain the fairly strong negative correlation for the degree centrality. However, we don't get a fairly large negative relatedness index for the closeness centrality of the karate graph, although this may be due to the lack of confidence in the relatedness index value.

If we finally consider the words graph, we can also see that the nodes with 0 closeness centrality also have 0 clustering whereas the remaining nodes are all highly clustered with very similar closeness centrality. This splits the nodes into 2 distinct groups – those with very low/high closeness centrality and clustering – which creates a very large positive correlation giving the very high relatedness index for the closeness centrality.

Interestingly, it is also worth noting that the relatedness index of the closeness centrality is always higher than the relatedness index of the degree centrality. This is possibly because the degree of a node has no influence on the degree of a neighbouring node. On the contrary, shortest paths between nodes have common edges that are also used in the calculations of the closeness centrality for neighbouring nodes so are more likely to have similar centrality.

Note, it is hard to perform analysis / see any obvious correlations on some of the graphs due to their limited size e.g. the karate graph only has 34 nodes.