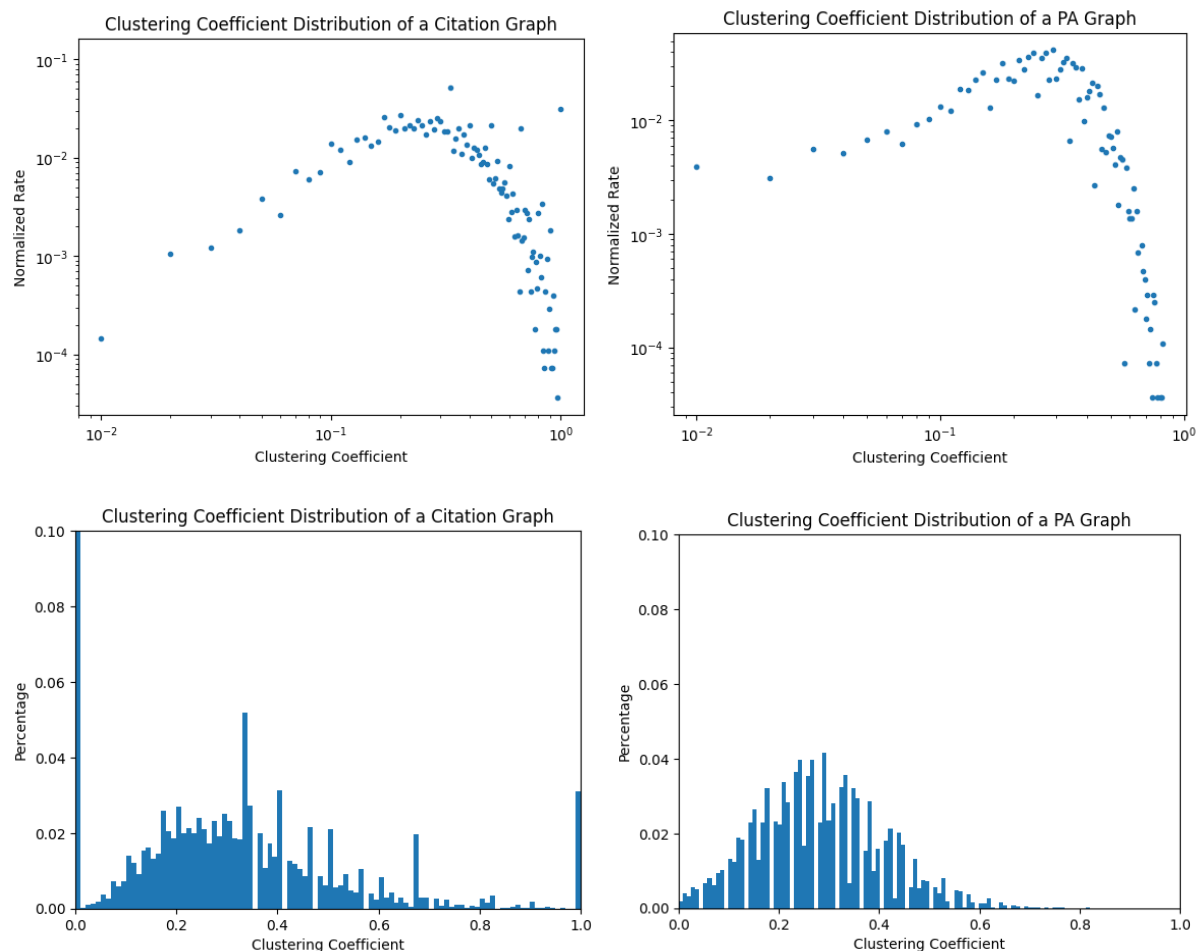
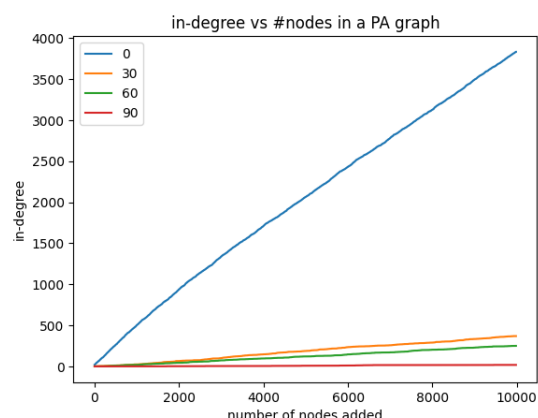


Question 1a



Both graphs have a very similar distribution of clustering coefficients. Hence, it is likely that they are equivalent. Note, the citation graph has more vertices with 1 or less neighbours (0 clustering coefficient) and more vertices that either have perfect clustering or are self-referential (1 clustering coefficient). The spacing artifacts in the histogram are due to the rounding of the clustering coefficient to 2 decimal places.

Question 1b



We see that the in-degree of a vertex increases linearly with respect to the number of nodes added to the network. This is due to the preferential attachment property of the network i.e. the probability a new node is attached to an existing node is proportional to the in-degree of the existing node – previously chosen vertices are more likely to be chosen again

We also see that the in-degree of nodes added earlier to the network increase much quicker than later nodes. This is because of two reasons:

a) vertices added earlier to the network are more likely to be selected as there are less nodes to choose from

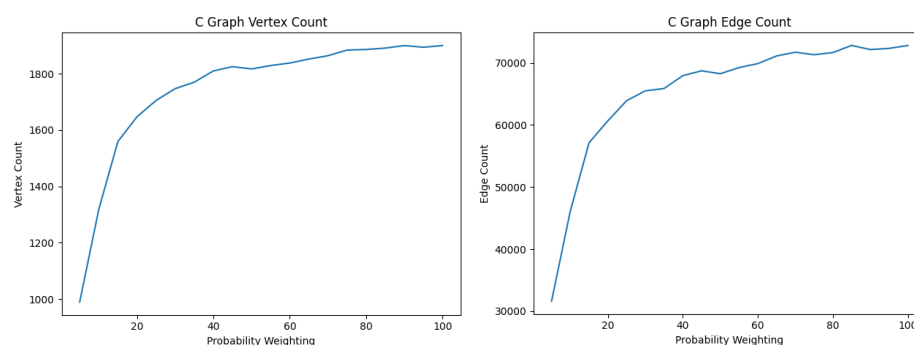
b) new vertices become less likely to be selected as the construction will prefer existing nodes that have already acquired a higher in-degree due to the preferential attachment property.

This is clearly demonstrated by the 0 vertex as its in-degree increases much quicker than the rest. This is because the 0 vertex is in the initial complete network where all nodes have equal in-degree and there are initially only 20 vertices to choose from.

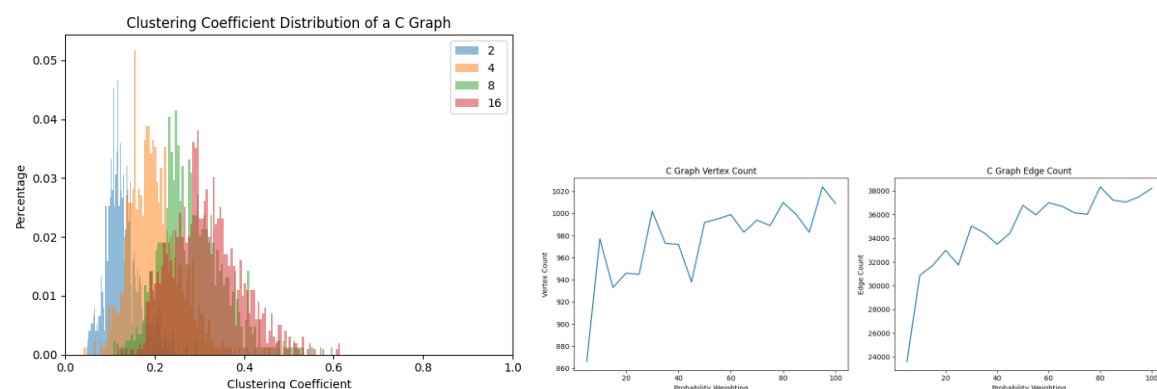
A real graph that has this preferential attachment property is the citation network. This satisfies the preferential attachment property since the more citations a paper has, the higher the chance that it is cited again by another paper. Intuitively, past papers will fade into obscurity unless they are sufficiently cited.

Question 1c

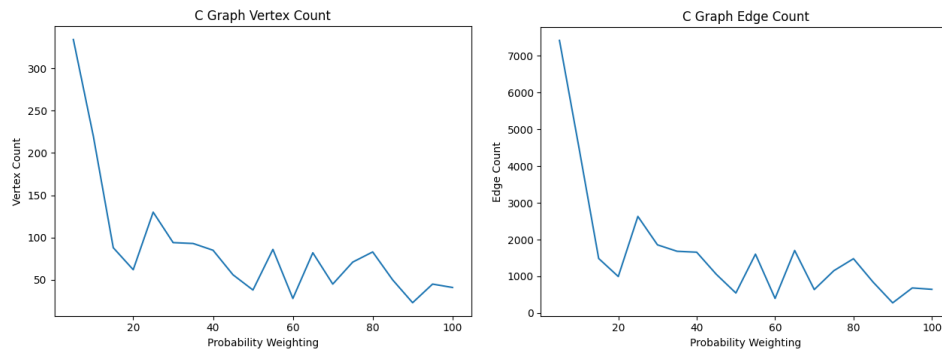
For our analysis of the 4 probability parameters, we alter 1 at a time by weighting it by n times more than the others probabilities where the others are equally weighted e.g. $p_1:p_2:p_3:p_4 \rightarrow n:1:1:1$.



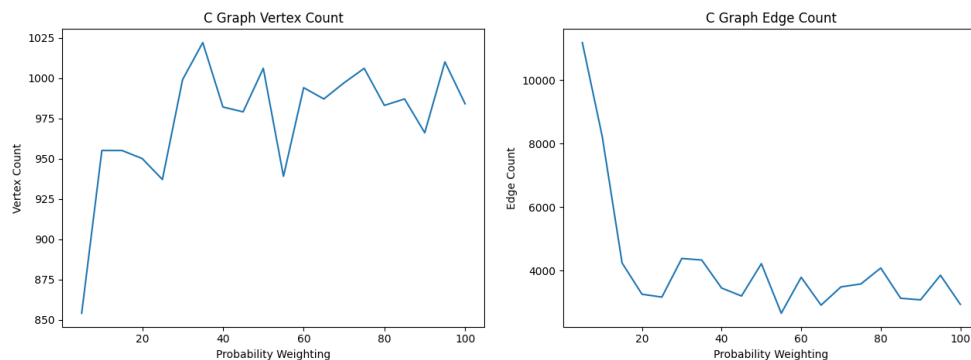
P1 causes the construction to tend towards a bigger graph since it adds new vertices and edges. Since this is the only parameter that increases the size of the network so we need to have a sufficiently high p_1 to ensure that there are sufficient nodes in the graph for analysis. As such, we set $p_1=0.5$ and vary the other probabilities as before for the analysis of the remaining parameters.



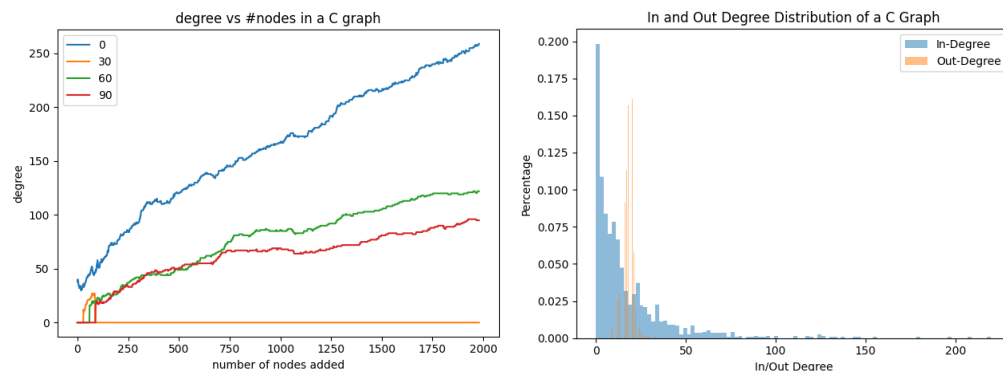
P2 causes the construction to cluster more since it only adds edges to the existing nodes. The reason we don't get a perfect clustering coefficient here is because p_1 is high enough such that nodes are added faster by p_1 than they can be connected by p_2 .



P3 causes the construction to tend towards having 0 nodes since it only removes nodes. Since there are less nodes, there are likely to be less edges.



P4 causes the construction to tend towards having 0 edges since it only removes edges. Unlike above, removing edges does not cause the number of nodes to drop as p4 is not competing with p1.



Here we used $p_1=0.7$ and $p_2, p_3, p_4=0.1$ to ensure there are enough nodes for analysis. From the first diagram, we can see that the C graph has the preferential attachment property as a result of the degree weighting when adding new nodes and edges in p_1 and p_2 . Note, there is a chance that the selected node is deleted from the network – this is what happened to node 30.

From the second diagram, we see that nodes in the C graph have a higher out-degree than in-degree. This is expected since when nodes are added to the graph, they are initialised with 20 out-edges but no in-edges. We can also (just about) see the outliers with a much higher in-degree due to the preferential attachment property.

A real-world example of the C graph is the world wide web where nodes represent webpages and edges represent hyperlinks. As in our model, this is a directed network. This real network also has preferential attachment since bigger/more popular webpages with higher connectivity are more

likely be referenced. When a new webpage is added, it only has out edges (hyperlinks) to other webpages (more likely popular webpages due to preferential attachment), the same as p1. P2 and p4 corresponds to updating a webpage by changing its content such that it has more or less hyperlinks respectively. This even supports the weighted selection of the endpoints of the edges in p2 since bigger webpages are more likely to be update and are more likely to then reference other big pages. Finally, p3 simply corresponds to deleting a webpage. However, if our network would better model the world wide web, we would expect nodes to be deleted inversely proportional to their degree as bigger pages are unlikely to be deleted.