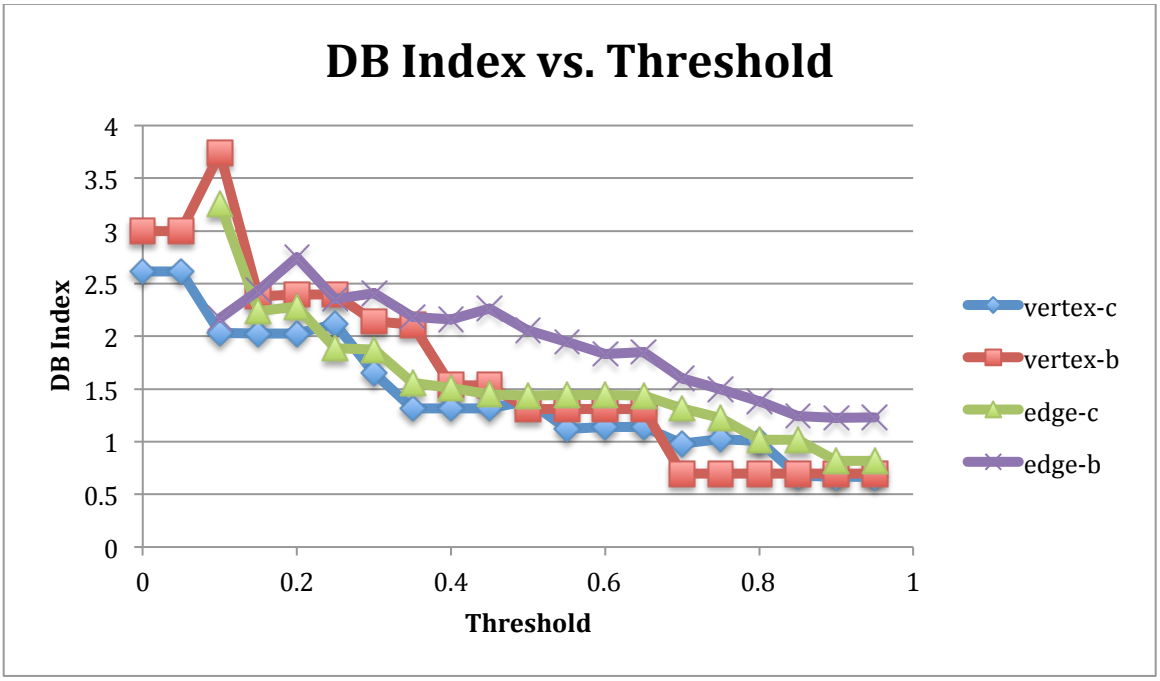
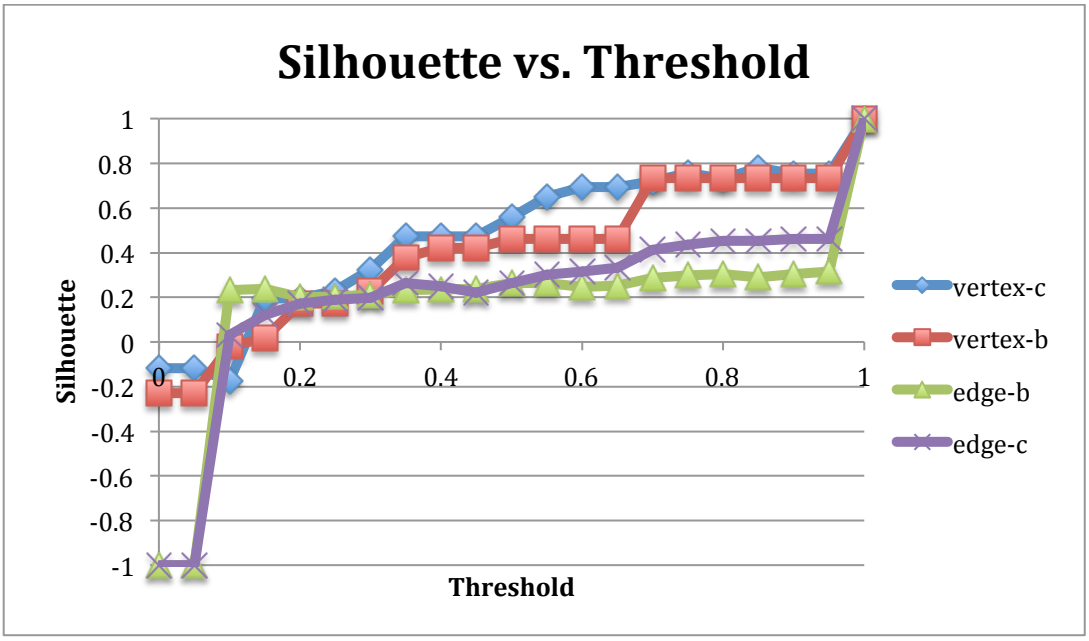
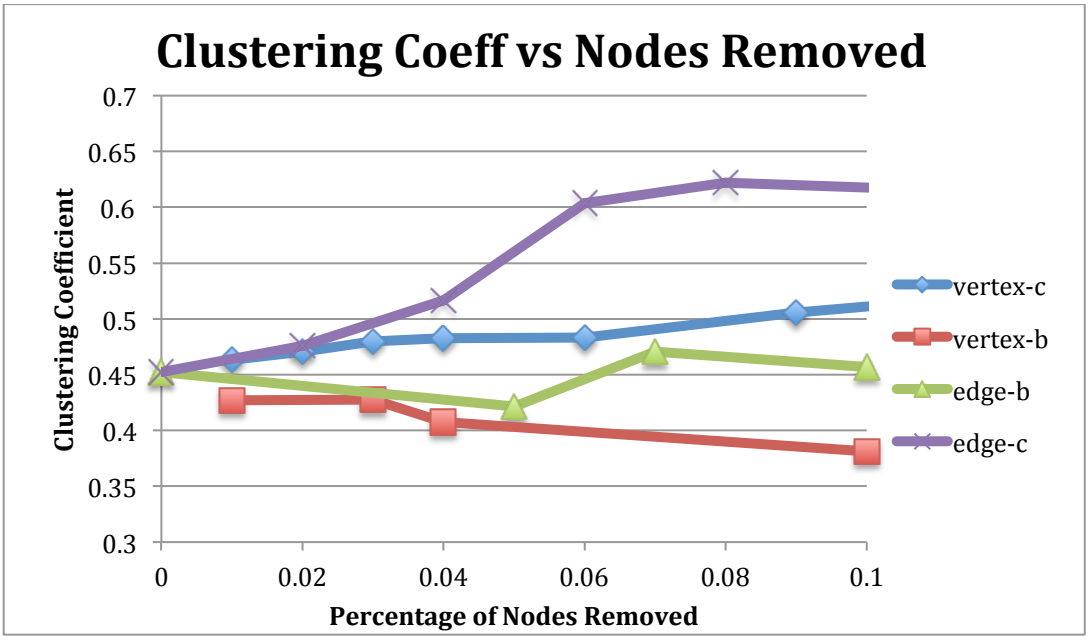
Sensitivity Analysis and Version Comparison



The DB index decreases as the threshold increases. This is because the clusters are becoming more compact and dense as the threshold increases. One important concept to note is that as iteration goes up, there will be more singleton clusters, which is undesirable. Therefore, the end of the graph might look good, but in reality is not desired. If I were to choose a value based on this metric, I would choose somewhere between 0.4 and 0.65.



Silhouette should gradually increase as items in a cluster, on average, become closer and farther away from other clusters. Similar to the problem with DB index, it is best to "ignore" the last few threshold readings. With such a high threshold, there is sure to be a large amount of singletons, which is undesirable. If I were to choose a value based on this metric, I would choose somewhere between 0.4 and 0.6.



The clustering coefficient was an interesting comparison. I expected the vertex removal to be relatively low because we are completely removing vertices from the graph (singletons). My assumption for why edge centrality was scoring so well was that it was truly removing decent clusters from the graph, making the remaining graph more and more cluster-like in itself.