Analysis/Conclusion of results:

Our target node is the yeast homolog of CDC28 - YBR160W:

* Biological function: Involved in the cell division cycle.
* Experimental importance: Protein of interest, key protein in the cell cycle

Node of interest: YLR466W:

* Biological function: Helicase.
* Experimental importance: Phosphorylated by CDC28, therefore CDC28 has possible implications in DNA unwinding. Could monitor DNA unwinding to determine how CDC28 effects this function

I started by computing all the shortest paths between our target node and one interesting node. The shortest paths were computed using the function “all\_shortest\_paths” which used Dijkstra's algorithm (which is the default algorithm used by this function). Then I found the degree of each node along each shortest path. I call these nodes along each shortest path an intermediate node.

Then I found which intermediate node had the largest degree. Then I found which community this intermediate node belonged to. I did this to see if it is a node which is too important (and hence irrelevant for our experiment) by comparing the degree sizes of the intermediate node with its neighbours within its community. If I found an intermediate node which has a large degree compared with its neighbours, then I computed all shortest paths having removed this intermediate node. I did this because I wanted to bypass this too important node. In other words, to find a meaningful path for our experiment. I would then give this node to the biochem students to see if this node is of significance.

Findings:

* The code I’ve provided with the file name: ‘Path Analysis.ipynb’ when opened, contains comments in each cell detailed the steps I’ve taken
* Here is a summary of what I’ve done

The node YFL038C is a node on the shortest path between YBR160W and one node of interest: YLR466W. I computed all the shortest paths and found that the shortest path was of length 3. I found that among all the shortest paths, node YFL038C appeared in 8 such shortest paths and has a degree of 150. I found that within its own community, YFL038C has a degree of 131, making it the largest connected node in its community. I then found all of the next shortest paths between the yeast homolog of CDC28 and the node of interest: YLR466W after removing YFL038C. I found that there were 22 new shortest paths of length 4.

YFL039C lies along 3 out of 22 of the next shortest paths. This node has the 2nd largest degree in its community with degree 98, the 30th largest has degree 32. YCR068W lies along 12 out of the new 22 shortest paths. It appears in the same community that '4932.YFL038C' used to be in prior to being removed. These nodes may be of interest for our biochem colleagues.

The aim here was to find meaningful paths between our target node and an interesting node which contains a biologically significant node. I have posited that a high degree intermediate node relative to the nodes in its community may be important and should be marked for biochemical analysis. Our intention is to use this method for the rest of the interesting nodes, given more time. These interesting nodes were found after pruning the graph and then having the biochem students pick out which nodes are important.