Methods

1. Yeast

For the yeast protein interaction network, **Louvain Community Detection and various kinds of centrality measures** are obtained to discover the 'hub' proteins.

After reading in the PPI network downloaded from string database, we try to edit on the large graph by deleting some of the edges which have low confidence. We have tried two of the threshold scores for deleting edges, so we get two networks to work on. The first graph is the network with more edges preserved which have weight more than 400 threshold score (graph400), which is the medium value of the confidence. The second one contains edges have weight more than a high medium value of confidence 700 (graph700).

The second procedure is to discard some of the housekeeping proteins, which as defined as proteins which can be detected in many different tissues. This category of housekeeping proteins has very high degrees in the PPI network. In order to better conduct community detection algorithms, we remove the top 5% high degree nodes in the networks and proceed to next-step analysis.

The Louvain community detection method available in the workspace are applied to both networks. The advantage of Louvain community detection method in PPI network is fast, which is appropriate to use especially in large networks. The graph400 is found to have 16 communities and 6 of them have less than 5 nodes. The graph700 has 45 communities, while 23 of them are very small groups. Then we conduct different centrality measures to identify the central nodes in each community which could have a significant function in the PPI network.

Regardless of the centrality measures listed in the workspace, two more centrality measures are adapted, which are closeness centrality and harmonic centrality measures. Harmonic centrality is another distance-based centrality and new compared to others. It has been regarded as the best 'centrality measure' in a scientific paper. Closeness centrality is the best centrality discovered to best find central nodes in another literature. After getting all the high centrality nodes in each measure, we consider the frequency of nodes in the table with high ranks and make outputs of them. These nodes are the central nodes in each of their communities.

Then we can do further analysis on the target protein ALD2 with these central proteins that we have found, including degrees, paths, and within community behaviours. And further research can be conducted by editing more on the graphs, such as deleting the central nodes in each cluster and redo a community detection. More important proteins

might be discovered. And other community detection methods can be

2. Human

We do similar network analysis on the human network. However, the human protein interaction network is far more larger than yeast PPI so that we can hardly get results using too many centrality measure, so we just choose betweenness and closeness centrality, which are two classic ways for centrality measures and are proved to be efficient. Then we have some of the hub proteins and we can see the paths and relations to ALDH2, which is homologue compared to ALD2 in yeast.