Homework 3
BCB 570

1. Analyze the structure of the Union (Y2H_union and CCSB1 datasets to assess the structure of the protein-protein interactions.
   a) Small world, connectivity, radius, etc. Does it follow a power law?
   b) Find hubs, are hubs correlated with phenotype-changing genes (From Yeast deletion project)? compare with null hypothesis (choose set of genes at random and check essentiality, repeat at least 10 times)
   c) Compare your results to the Yu Paper for the Y2H data, do you have similar conclusions?

2. The large file, BG_2018 is a set of the data that is currently in BioGRID for yeast protein-protein interactions that come from physical data. I have removed most of the duplicates. Does this network have a similar structure to the data used in the Yu paper? Is this graph with many more data points self-similar to the smaller data sets? Discuss how this analysis might effect how you use assumptions on the data structure.

3. Do the below for the largest connected component in the network only.
   a) Find the most central genes with respect to betweenness centrality, is there a correlation with essential genes?
   b) Find modules in networks using MCL, compare the modules with 10% and 25% of the connections randomly removed.
   c) Choose 1 cluster and look at the annotation for the genes, is there any common function?

Turn in: Include all your code and write-ups and plots of your results.