

# Gene function prediction from protein network properties

Shan Zhao | Advisors: Alpan Raval/Animesh Ray  
Keck Graduate Institute, REU Program 2005



In our project we attempt to predict gene essentiality and double lethal gene pairs in *S. cerevisiae* by using only graph-theoretical properties of each protein in the protein interaction network (PIN) of *S. cerevisiae*. Sequence and expression information on these proteins are not used. The graph properties we used to predict essentiality and double lethal pairs include many previously studied properties, such as degree, betweenness, closeness, subgraph, eigenvector, and information centralities. In addition, we used previously unstudied properties, such as closeness flow, and random walk centralities. The new aspect of this work is to integrate these properties into a single prediction system for gene function, and to analyze the predictive value of this system.

The prediction system consists of a supervised, multilayer feed-forward neural network trained on a subset of known essential genes and double lethal pairs. The inputs to our neural network for predicting essentiality were the graph properties of each protein in the PIN and the output is a classification score from 0 to 1, 0 being nonessential and 1 being essential. The inputs to our neural network for predicting double lethal pairs were the graph properties of each protein in the pair along with the shortest path distance between the two proteins in the PIN. The results of our studies demonstrate that essential genes can be predicted at an accuracy of 75% and nonessential ones at an accuracy of 77%. Our results also demonstrate that double lethal pairs can be predicted at an accuracy of 98% and non-double lethal pairs at accuracy of 88%. Our studies demonstrate the predictive power of PIN graph properties for gene essentiality and double lethality in *S. cerevisiae*.

