Prediction of protein function via graph-theoretic analysis of interaction maps

Elena Nabieva
Department of Computer Science
Princeton University

Determining protein function is one of the most important problems in the post-genomic era. For the typical proteome, there are no functional annotations for one-third or more of its proteins. Traditionally, computational methods to assign protein function have relied largely on sequence homology. Recently, the emergence of high-throughput experimental data sets led to a number of alternative, non-homology based methods for functional annotations. In particular, recent high-throughput experiments have determined proteome-scale protein physical interaction maps for several organisms. These physical interactions are complemented by an abundance of data about other types of functional relationships between proteins, including genetic interactions, knowledge about co-expression and shared evolutionary history. Taken together, these pairwise linkages can be used to build whole-proteome protein interaction maps.

We have developed a network-flow based algorithm, FunctionalFlow, that exploits the underlying structure of protein interaction maps in order to predict protein function. In cross-validation testing on the yeast proteome, we show that FunctionalFlow has improved performance over previous methods in predicting the function of proteins with few (or no) annotated protein neighbors. By comparing several methods that use protein interaction maps to predict protein function, we demonstrate that FunctionalFlow performs well because it takes advantage of both network topology and some measure of locality. Finally, we show that performance can be improved substantially as we consider multiple data sources and use them to create weighted interaction networks.