Identifying Bipartite Subgraphs within Protein-Protein Interaction Networks

Julie L. Morrison¹² Rainer Breitling¹³ Desmond J. Higham² David R. Gilbert¹
1. Bioinformatics Research Centre, University of Glasgow, Davidson Building, Glasgow G12 8QQ
2. Dept Mathematics, University of Strathclyde, 26 Richmond Street, Glasgow, G1 1XH
3. Groningen Bioinformatics Centre, University of Groningen, Netherlands

A protein-protein interaction network model is defined which suggests the existence of complete or near complete bipartite subgraphs within the network. We propose an algorithm to find these bipartite subgraphs and uncover some unknown functional modules within a number of species.

Lock-and-Key Model

Model: We define interactions to occur in a protein-protein interaction network between proteins which contain complementary aspects of some domain. We call these complementary domains the lock and key, and interactions occur with a probability < 1. This results in complete or near complete bipartite subgraphs. This is based on an idea proposed by Thomas et al. [1].

Analysis: By making some simplifying modeling assumptions, we conducted analysis on our model definition. From this we were able to predict the existence of pairs of eigenvalues of the same size but opposite sign. The sum and difference of the corresponding eigenvectors can be used to predict domain information.

Synthetic Data Testing

We tested our analysis on a network generated using the model definition (Figure 1). All predicted links are present. As predicted we were able to identify pairs of eigenvalues of equal sign and opposite magnitude. The sum and difference of the eigenvectors for the first pair are shown in Figure 2(a). The non-zero values in each case correspond to proteins labelled 17,18,19,20. To identify the proteins containing the remaining two domains we look at the sum and difference vectors for another eigenvector pair (Figure 2(b)). Again we are able to gain domain information about the remaining proteins.

A similar test was carried out using the same network as above with some links removed. We no longer see equal non-zero values in the sum and difference vectors. By choosing a threshold and assigning proteins above the threshold to a domain, we are able to correctly assign all proteins (Figures 2(c) and 2(d)).

Results

Based on the idea of defining a threshold, we were able to define a heuristic algorithm to identify bipartite subgraphs. Since the model is based on pairwise interactions we test the algorithm using Y2H data. We were able to identify the bipartite subgraphs shown in Figures 3(a) and (b) from the yeast interaction network found by the Uetz et al. genome-wide scan [2]. For another 6 species we were able to find other bipartite subgraphs. All were biologically verified.