Analysis of protein-protein interactions has a growing potential of providing a better understanding about collective protein function and cellular machinery. Such interactions can be modeled as a protein interaction network; that is a weighted graph where each node represents a protein and each edge represents an interaction between a pair of proteins, with a weight representing the level of confidence that this interaction truly exists. Such a structure allows easier extraction of hidden information due to the analogy with well-known graph problems.

An important aspect of the analysis of protein interaction networks is detecting signaling pathways. A signaling pathway is a series of proteins in which each protein signals its successor to transmit some biological information through their interaction. Signaling pathways can be viewed as simple paths in protein interaction networks1. Given a confidence value for each interaction, we can assign an overall confidence value for the existence of a pathway simply by multiplying the confidence values of its constituting interactions. An optimal pathway is one with the highest confidence value. To work in an additive framework, edge weights can be assigned as negative logarithmic values of the original interaction confidence.

Problem

Consider a set V of nodes denoting proteins, and a probability value p(u, v) denoting interaction confidence for each u, v E V. A set of undirected weighted edges E can be obtained by adding an edge for each pair (u, v) with p(u, v) > 0, and assigning it a weight w(p, v) = -log p(u, v). Now consider the undirected weighted graph G = (V, E, w) representing the protein interaction network, and a set of start nodes S E V. For each v E V, we want to find a minimum-weight simple path of length m, starting at any node s E S and ending at v. The traveling-salesman problem is polynomial-time reducible to this problem; therefore it is NP-hard.

Scott et al presented a method for detecting pathways2 using a generic technique devised by Alon et al called color-coding3. The basic idea of this method is to randomly assign each node in the graph one of m different colors, and search for an optimal pathway in the restricted domain of colorful pathways. A pathway is considered colorful if and only if all of its nodes are different in color from each other. This process is repeated for several iterations until reaching a given level of confidence that the unknown optimal pathway was among the colorful ones at some instance. This confidence level builds up with each iteration by calculating a probability value that an optimal path is indeed colorful in this iteration. This success probability value depends solely on the pathway length m and doesn’t capitalize on available information like the network topology and color assignment.

Gulsoy et al presented an enhanced color-coding technique called k-hop coloring4. In essence, k-hop coloring makes use of knowing the network topology and the node colors to assign the network a maximal value k such that the network is k-hop colorable. A colored network is k-hop colorable if the shortest path between any pair of same-color nodes is more than k hops in length. This additional piece of information allows for higher success probability in each iteration, with a higher k value resulting in a higher success probability. However, an obvious limitation is that the existence of parts with higher level of connectivity diminishes the k value assigned to the whole network. For example, a network containing a clique of size m cannot be colored with (m-1)-hop coloring using m colors4.

Contribution

Our motivation comes from the need of deeper understanding of the relation between network topology, random color assignment and success probability. We study the possibility of assigning k values to nodes on an individual basis instead of a single k value for the whole network. We also study how this reflects on the resulting success probability for each iteration. We examine the idea that a pathway whose nodes are assigned different k values should result in a higher success probability than if we only consider the minimum of these k values for all nodes. We present a new method for detecting signaling pathways in protein interaction networks using an enhanced k-hop coloring technique based on these findings.

We provide validation experiments to test the biological significance of our results. We use weight p-value and functional enrichment as validation measures. We also compare the performance of our method against the one presented by Scott et al2 with respect to how fast our method reaches a given confidence level as opposed to theirs.

The rest of the paper is organized as follows. Section 2 discusses the background and related work. Section 3 explains how to obtain a tighter bound on success probability and describes our enhanced k-hop coloring method. Section 4 shows the experiments performed and their results. Section 5 is the conclusion of the paper.