In this section, we start by properly formulating the problem and defining common terms that we use in our methods. We then present new thoughts about pathway detection using color coding. We study the network topology and how we can involve it in our calculation to obtain a better success probability, and hence needing less number of iterations and improving performance. Last, we present an enhanced color-coding method for detecting pathways in protein interaction networks.

Problem formulation

Given a graph G = (V, E, w), where V is its set of nodes, E is its set of edges and w is the edge-weight function, and given a set S \subset V and a path length m. Assume Pi is the set of all simple paths of length m starting at any node s \in S and ending at node i. Our goal is to find, for each node i, the path p \in Pi whose sum of edge weights is minimum.

The following are the definition of some commonly used terms:

* k neighborhood. For a given node v \in V and an integer k, the k neighborhood of v is a set of nodes U \subset V where a node u \in U if and only if u can be reached from v in k hops or less.
* max-k. For a given node v \in V, max-k of v is the maximal value of k such that \forall u \in k-neighborhood of v, the color assigned to u is not equal to the color assigned to v.
* max-k configuration. For a given path P, the max-k configuration of P is the sequence of max-k values corresponding to the sequence of nodes in P.

Based on the work presented by Scott et al.\cite{scott}, a generic color-coding approach to solving the problem consists of three main steps. The first step is coloring the network; each node v \in V is independently assigned a color drawn uniformly at random from a set of m different colors. The second step is finding optimal colorful paths; \forall v \in V we want to find the minimum-weight colorful path of length m starting in S and ending at v.