Adding Color-Coding algorithm as an extension to Cytoscape

G Sanjay(CS19B019), Sriram Goutam P(CS19B073)

1 Introduction

A cell cannot perform its basic function, or even decide whether to survive/grow/die, without knowing what is happening outside it! Signaling pathways are essential for a cell to sense its external environment. Research on protein-protein interactions is crucial for understanding which pathways are used by a cell for sensing. A popular open-source software platform for viewing and studying biological networks is called Cytoscape. Nevertheless, Cytoscape currently has only a few tools accessible for studying protein-protein interactions.

Color-coding is one such algorithm for finding signalling pathways in a protein-protein interaction network. We suggest creating a plugin for this algoirthm in Cytoscape that would allow users to examine intricate protein interaction networks and recognise significant interactions and pathways.

2 Problem Statement and Objectives

Problem Statement: To implement and add the color-coding algorithm for analyzing protein-protein interactions and finding signalling pathways as a plugin to Cytoscape.

Objectives:

1. To implement color-coding algorithm for analyzing protein-protein interactions.
2. To integrate the algorithm into the existing Cytoscape platform.
3. To visually demonstrate the utility of the algorithm by applying it to examples.

3 Approach to solve the problem

1. Implement color-coding algorithm and evaluate its performance on an example protein network(like Yeast protein network)
2. Go through the documentation and create a sample app to run on Cytoscape. Then integrate the algorithm as a plugin to Cytoscape

4 Expected Outcomes

1. An efficient implementation of the color-coding algorithm
2. To correctly establish communication between our algorithm and Cytoscape, then to integrate the algorithm to Cytoscape as a plugin
3. To visually demonstrate the signalling pathways of a protein-protein interaction network to the users.

5 Time Line

Week 1: Go through the documentation of Cytoscape to create a custom plugin.

Week 2: Implementation of the color-coding algorithm and establishing a communication between Cytoscape and the algorithm.

Week 3: Integrating the algorithm to the Cytoscape and testing the plugin.

Closely related research paper

Efficient Algorithms for Detecting Signaling Pathways in Protein Interaction Networks

1 Summary

This paper presented efficient algorithms for finding simple paths and rooted trees in graphs (like protein interaction networks) based on the color-coding technique and several biologically motivated extensions of this technique. Linear time Algorithms were proposed for locating Signaling pathways and trees in networks under various biologically inspired restrictions. A signalling route is originally described by the authors as a subgraph of the protein interaction network, where the nodes correspond to proteins and the edges to interactions between them.

The authors presented two algorithms for identifying signaling paths of length k from a given set of source vertices (representing receptor proteins) to a set of terminating vertices (representing regulator proteins). The first algorithm is an exact algorithm which uses dynamic programming to efficiently enumerate all possible signaling pathways in the network. The time complexity of this algorithm is found to be $O(kn^k)$ and the space complexity of this algorithm is $O(n^k)$.

They then introduced the color coding technique, which involves assigning colors randomly. The idea behind this algorithm is to search paths with different colors rather than searching paths with different vertices. This reduces the time complexity of the algorithm. But this has a drawback of assignment of same colors to two different vertices in a path. To overcome this many random colorings need to be tried to ensure that no paths are missed. This algorithm has a time complexity of $O(2^k km)$ and a space complexity of $O(2^k n)$.

Then this paper extended the color coding method to other biologically motivated problems, such as finding more general structures like rooted trees, Two-terminal series-parallel graphs. Finally this algorithm was applied to Yeast protein network and significant speedups were observed as compared to standard algorithm.

2 Strengths

1. The time complexity of the proposed algorithm is significantly lower than the standard algorithm (especially for a small k and large n)
2. It was discovered that 68% of the identified pathways and 63% of the identified trees were considerably functionally enriched.

3 Weaknesses

1. Even though the coloring can be repeated many times some paths may still be missed because of the randomization.