On Tools for Network Motif Finding: Proposal

2010 NSF Bio-Grid
REU Research Fellows
University of Connecticut
Brittany Baur (Manhattanville 2011) & Elisabeth Wong (Bowdoin 2011)

Networks are integral parts of many real systems and thus it has become a priority in many research fields to analyze them. Graphs have been used to represent networks which allows for graph theory to be used for computational analysis of the networks. It has been found that numerous networks contain network motifs, small sub-graphs that appear more frequently than expected in randomized networks. An ongoing problem in the study of networks is how to develop an algorithm to efficiently identify these network motifs. This problem is made difficult due to the graph isomorphism problem, the fact that the number of network motifs exponentially increases with increases in network size, and the absence of the downward closure property in many networks. These difficulties make it so that full enumeration of sub-graphs can be extremely time consuming and may require large computational power. Multiple algorithms and tools have been developed, each with different advantages and disadvantages, to identify network motifs. Network motif discovery is a crucial problem to solve in order to gain further insights into the important characteristics, functions, and inner workings of systems with networks. Therefore it has been the goal of many researchers to develop ways to efficiently identify network motifs.

We intend to explore and analyze different algorithms and approaches available today for finding network motifs. It is our goal to study as many tools as possible and to look at a wide range of the tools. The myriad of tools that we plan on focusing on include FANMOD, Mfinder, MAVisto, Pajek, Grochow, Kavash, and NeMoFinder. A large part of our survey will deal with classifying these tools. The tools will be classified based on different aspects of their methodology: (1) the algorithm used for full enumeration, (2) strategies of identifying motifs without full enumeration, (3) methods of determining frequencies of motifs, (4) ways of developing randomized networks, and (5) the data sets the tools can be applied to. We will evaluate each tool via literature review and experimental runs. The runs will be carried out on benchmark data sets including PPI networks, social networks, and transcriptional networks. Aspects of the tools that will be considered are runtime, feasible motif size, accuracy (including comparison between tools using the same benchmark data sets), inputs, outputs, and visualization.

With the results obtained from our examination of the tools and networks we will provide further analysis. We will look at real world significance and functions of the motifs found within their respective networks and organisms and suggest possible explanations for these why these motifs are over represented. We will also look for explanations of results based on the methodology of the algorithm used. Comparison of the results of the study will be carried out. From these comparisons we will make suggestions for improvement of the tools and further directions that can be explored in the development of network motif discovery algorithms and tools. These further directions include: (1) the ability to intelligently search respective networks for possible biologically relevant motifs that have been identified as significant sub-graphs from experimental runs and literature review, and (2) the idea of employing modern computing infrastructure to search concurrently for network motifs that are larger than those that presently available tools can search.

The survey of the network motif finding tools existing today will be an important contribution to this area of study. It will provide an explanation of the way each tool works, an analysis of the results of each tool when applied to various networks, and a comparison of the performance and efficiency of the tools. This summary of the preexisting network motif finding tools will greatly benefit future developments and investigations of network motif discovery.