A network, whether it is Protein-Protein Interaction (PPI), social, or electrical, can be defined to be a set of objects and interactions. A network can furthermore be modeled as a graph where the nodes represent objects, and arcs represent interactions. Within a network, there are patterns that occur statistically more often than similar but random networks, which are called “network motifs.” Motif statistical significance suggests that their occurrence satisfies some need within the network. This is the case for the C. Elegans neural network and the Feed Forward Loop (FFL), a three node motif which is exemplary for information processing. The focus of motifs however has been on smaller motifs, usually three to four nodes; this is due to two reasons. One, the smaller motifs appear more often and in various networks, whereas the larger motifs are less sporadically found and are very specialized. The other factor is that motif detection has an exponential time complexity, with respect to the size of motif. This makes studies about large(r) motifs long and expensive. These reasons make researchers wonder if larger motifs are worth further research. We plan on answering that question by first obtaining data using various motif detection tools and then analyzing this data to find the statistical significance of large(r) motifs.

We will first garner our data by utilizing existing motif detection tools to search for motifs up to size six in networks of various disciplines. The tools we plan on using include FanMod, Kavosh, mFinder, and NetMod. They were chosen for their speed in motif detection and accuracy of results. Multiple tools would also help us verify the data that we collect. We chose to collect data on sizes three to six node motifs to have sufficient information to see any trends. The networks we plan on using are of four different categories. The first is biological, which includes E.Coli transcription, Yeast transcription, C.Elegans PPI and Fruit Fly PPI networks. The second is electrical networks which include Forward Logic chip circuit, Digital Fractional multiplier circuit. The third type is social networks which includes World Wide Web and person to person network. The fourth is a hybrid network, the C.Elegans neural network, which has elements of the initial categories. This will provide a broad scope to encounter motifs.

After obtaining all our data, we will compile the various outputs into a database. Using this database, we will examine how motifs of different sizes compare in terms of statistical significance. This will be accomplished through three objectives. We will first plot the average z-score of the top five most significant motifs for each given size as a function of size. Then we will search for which motifs seemed to appear across the various disciplines of networks and then distinguish them from those that appear to a particular discipline. Third, we would like to examine the topology of the most significant five and six node motifs to see if they are just superimposed 3 and 4 motifs or their own structure. These tests would provide us with an answer for our question of the significance of larger motifs.

Through this process, we hope to understand just how significant larger motifs could be. This would be important for whether or not large motifs need to be studied and newer faster motif detection tools should be made. We would also like to provide a list of medium-sized motifs that researchers could then further analyze. The data we acquire from our third objective would be useful for the examination of how motifs cluster; one could use the tool Cytospace to visually examine if certain motifs cluster in the same pattern. This information would significantly affect the interest of researchers on motif detection.