

Evolutionary Resilience of Protein Interaction Networks

Jesse Hautala, Shawn Houser, Angyalka Valcsics
Research Proposal
STA 596

October 6, 2021

Proteins control all biological systems in a cell and, through various interactions with each other, enable cells to complete tasks such as: enzyme activation, gene regulation, and intercellular communication. Protein interactions can be modeled by an undirected network of protein-protein interactions (a PPI network, or PPIN), with nodes representing proteins and edges representing interactions. The complete set of such interactions for a species is called the protein interactome. When interaction relationships between proteins break, possibly due to environmental factors or random mutations, such breakage can cause disease and death, of the cell and of the organism. Since death and disease interfere with an organism's ability to survive and reproduce, we expect selective pressure to increase PPI network resilience. We intend to use statistical methods and network analysis to identify evidence that evolutionary processes increase resilience of protein interactomes against network failure.

Advances in proteomics allow researchers to study the protein interactome, but limitations of experimental methods in practice prevent PPI networks from being comprehensive and free of noise. Therefore we regard extant PPIN data, including the data used in this project, as a noisy sample from an unknown population (the true protein interactome). For example, the yeast-two-hybrid method for mapping protein interactomes was first developed in 1989 by Fields and Song using *Saccharomyces cerevisiae* as a biological model. The accuracy of this experimental method is estimated to be less than 10 percent. Since we plan to model and analyze the PPI networks of just a few species, which share some evolutionary relationship, any conclusion we reach should be considered carefully if generalizing to all species. Furthermore, there are many possible confounders which should be considered: investigative biases towards modeling common or popular organisms, network size, genome size, and more.

Understanding how protein interactomes evolve and developing methods for analyzing these PPI networks is a central goal of evolutionary systems biology (Maddamsetti (2021)). In a paper by Rohan Maddamsetti they provided evidence that protein interactomes in E-Coli appear to show a generational increase in network resilience. Marinka Zitnik (Zitnik et al. (2019)) defined network resilience as the measure of how quickly a network breaks down as edges between nodes are randomly removed. Connectivity in a PPIN is essential to the survival of the cell. As a fraction of edges f are removed at random, proteins can no longer communicate with each other and the cell cannot function properly. The current research uses the modified Shannon diversity index to compare resilience between species, this index measures how the PPI network fragments into isolated components at a given network failure rate f . The resilience measure integrates the modified Shannon diversity index over all possible failure rates. A resilience rating of 1 implies that the network is most resilient while a rating of 0 implies a complete loss of connectivity in the PPI network.

Our dataset comes from the Stanford Network Analysis Platform (or SNAP). SNAP produced their PPI network data by combining datasets containing “regulatory interactions, binary interactions derived from yeast-two-hybrid high-throughput datasets, metabolic enzyme-coupled interactions, protein complexes, kinase-substrate pairs, and signaling interactions”. This data was collected using the Search Tool for the Retrieval of Interacting Genes/Proteins (or STRING), from the European Molecular Biology Laboratory, EMBL. The data is organized in multiple text files, joined by species ID. It comprises taxonomy information, an edge set for each interactome, and a numerical variable for evolution of the species. Evolution of a species is represented by the total branch length from the root to the corresponding leaf in the tree of life.

In addition to assessing the resilience of the PPI networks by randomly removing a fraction of edges, we also intend to find various network statistics for each PPI network. Some of the characteristics we will explore are degree centrality, eigencentrality, modularity, and the fraction of closed triangles. We expect to see a relationship between these statistics and the resilience of the PPI networks, mainly that networks which are strongly connected in terms of topological stability will be more resilient. Our intuition is that there will be a linear relationship between these network characteristics, the network’s resilience to failure, and the given evolutionary number. We will inspect this by fitting linear models with evolution, resilience, and the network characteristics. The degree distribution for each protein in the PPI network may provide some insight into which nodes are strongly

interconnected and which are at risk of easily being separated from the giant component. Moreover, we can discover critical edges in networks with high modularity and expect that when these critical edges are removed the effect will ripple through the PPI network—causing major disconnections.

In conclusion, we hope that our systematic analysis of the network data provides evidence of evolutionary resilience in protein interactomes. However, we hope to add to the existing research by evaluating the relationship between commonly used network characteristics, evolution, and network resilience.

References

- [1] Evolution of resilience in protein interactomes across the tree of life Marinka Zitnik, Rok Sosič, Marcus W. Feldman, Jure Leskovec bioRxiv 454033; doi: <https://doi.org/10.1101/454033>
- [2] Rohan Maddamsetti, Selection Maintains Protein Interactome Resilience in the Long-Term Evolution Experiment with *Escherichia coli*, Genome Biology and Evolution, Volume 13, Issue 6, June 2021, evab074, <https://doi.org/10.1093/gbe/evab074>