

# Chapter VIII

## Discovering Network Motifs in Protein Interaction Networks

**Raymond Wan**  
*Kyoto University, Japan*

**Hiroshi Mamitsuka**  
*Kyoto University, Japan*

### ABSTRACT

*This chapter examines some of the available techniques for analyzing a protein interaction network (PIN) when depicted as an undirected graph. Within this graph, algorithms have been developed which identify “notable” smaller building blocks called network motifs. The authors examine these algorithms by dividing them into two broad categories based on two definitions of “notable”: (a) statistically-based methods and (b) frequency-based methods. They describe how these two classes of algorithms differ not only in terms of efficiency, but also in terms of the type of results that they report. Some publicly-available programs are demonstrated as part of their comparison. While most of the techniques are generic and were originally proposed for other types of networks, the focus of this chapter is on the application of these methods and software tools to PINs.*

### INTRODUCTION

Data from high-throughput experimental methods are currently being used to construct complex biological networks which include transcription regulatory networks, protein interaction networks (PINs), and metabolic networks. As more and more of these networks become available, an increasing level of attention is being paid to characterize these large, complex structures.

One technique originally proposed by Milo et al. (2002) is to break down these networks into smaller components called *network motifs*. These network motifs can be seen as the networks’ basic building blocks. Determining the number and structure of these network motifs allows researchers to provide a general overview of the network and even assign each to a so-called “superfamily” (Milo, Itzkovitz, Kashtan, Levitt, Shen-Orr, et al., 2004). While the initial work with biological network motifs focused on transcription regulatory networks, Milo, Itzkovitz, Kashtan, Levitt, Shen-Orr, et al. (2004) and others since have extended this work to protein interaction networks.

The focus of this chapter is to examine the algorithms and software tools that are available to users for network motif discovery. Since Milo et al. (2002) highlighted the importance of network motifs, various researchers have sought various methods of finding them. The underlying theme in this chapter is to examine two definitions of “network motifs” that stem from their seminal work. Our aim is to compare and contrast the methods within the context of their corresponding algorithms and software systems. A recent review paper on network motifs for PINs provides a broader range of methods (Ciriello & Guerra, 2008); this chapter focusses on several selected algorithm within the context of this theme. Readers are encouraged to consult this review to gain another perspective on the motif discovery problem.

## MOTIVATION

The analysis of protein interaction networks (hereafter abbreviated as PIN) is based on the work on many types of networks, including other biological networks and even the structure between web pages on the World Wide Web. Because of this, we explore the motivation for analyzing a PIN by occasionally digressing to other types of networks.

A PIN can be represented as an undirected graph  $G$  such that each node represents a protein and unweighted edges are placed between two nodes if an interaction exists between them. By depicting proteins in this manner, we can examine the patterns of interacting proteins.

Starting at the node-level, one such analysis examines the number of immediate neighbors of each node, which is defined as the *degree*. Using tools such as the power law, it has been shown that PINs and other biological networks are scale-free based on the distribution of the node degrees (Barabási & Albert, 1999). That is, many nodes have few edges and few nodes are *hub nodes* and have many edges. In the context of PINs, this means that a random loss of a protein would have a small effect on the overall biological system. However, a focussed attack on a specific protein could have severe consequences if it was a hub node.

Motifs represent the next level in biological network analysis. According to one definition, they are small subgraphs which are repeated and conserved and used to form the building blocks of larger modules (Wolf & Arkin, 2003). However, as suggested by Wolf and Arkin, the separation between motifs and modules is unclear and remains a topic for future work. In their work and the subsequent review by Alon (2007), functionality was the focus of the definition of a motif, and in both cases, the discussion was centered on transcription regulation networks.

An alternate view of motifs is to focus on the topology of the network by giving priority to the configuration of each motif. If motifs are found through their structure, they can then be used to describe the structure and functionality of the PIN. Wuchty, Oltvai, and Barabási (2003) showed that certain topological motifs are important since they are conserved through evolution. To arrive at this conclusion, protein interaction data of yeast (*S. cerevisiae*) was shown to have certain motifs of two to five nodes which also existed in 5 higher eukaryotes. In a more recent study, the evolution of the yeast PIN was modeled from a “preduplication network” by inferring the duplication and divergence of genes using motifs from a PIN (Presser, Elowitz, Kellis, & Kishony, 2008).

Thus, once motifs of a PIN have been found, additional insights can be obtained from the data. In this chapter, our focus is on discovering such motifs using the topology of the PIN. Following the work of Milo et al. (2002), these motifs are subgraphs which occur with frequencies that are higher than one would expect by chance. This definition resembles the usage of “motifs” in other areas of theoretical biology. DNA sequence motifs are short DNA sequences which occur more frequently than some background distribution of nucleotide frequencies. For the networks examined by Milo et al. (2002), the motifs were dubbed *network motifs* to differentiate them from the more general usage of “motifs” above. An alternative definition of network motifs is based on the frequency of the subgraph. We explore both definitions in this chapter. As a matter of convenience, since we concentrate on this topological definition for PINs for the remainder of this chapter, we use “motifs” and “network motifs” interchangeably.

25 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/discovering-network-motifs-protein-interaction/5562](http://www.igi-global.com/chapter/discovering-network-motifs-protein-interaction/5562)

## Related Content

---

### Drug Optimization for Cystic Fibrosis Patients Based on Disease Pathways Crosstalk

Shuting Lin and Yifei Wang (2021). *International Journal of Applied Research in Bioinformatics* (pp. 1-11).

[www.irma-international.org/article/drug-optimization-for-cystic-fibrosis-patients-based-on-disease-pathways-crosstalk/267820](http://www.irma-international.org/article/drug-optimization-for-cystic-fibrosis-patients-based-on-disease-pathways-crosstalk/267820)

### Predictive Toxicity of Conventional Triazole Pesticides by Simulating Inhibitory Effect on Human Aromatase CYP19 Enzyme

Tamar Chachibaia and Joy Harris Hoskeri (2016). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 44-56).

[www.irma-international.org/article/predictive-toxicity-of-conventional-triazole-pesticides-by-simulating-inhibitory-effect-on-human-aromatase-cyp19-enzyme/172005](http://www.irma-international.org/article/predictive-toxicity-of-conventional-triazole-pesticides-by-simulating-inhibitory-effect-on-human-aromatase-cyp19-enzyme/172005)

### A Review of Kernel Methods Based Approaches to Classification and Clustering of Sequential Patterns: Sequences of Discrete Symbols

T. Veena, A. D. Dileep and C. Chandra Sekhar (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 72-92).

[www.irma-international.org/chapter/review-kernel-methods-based-approaches/76057](http://www.irma-international.org/chapter/review-kernel-methods-based-approaches/76057)

### Computational Systems Biology Perspective on Tuberculosis in Big Data Era: Challenges and Future Goals

Amandeep Kaur Kahlon and Ashok Sharma (2015). *Big Data Analytics in Bioinformatics and Healthcare* (pp. 240-264).

[www.irma-international.org/chapter/computational-systems-biology-perspective-on-tuberculosis-in-big-data-era/121461](http://www.irma-international.org/chapter/computational-systems-biology-perspective-on-tuberculosis-in-big-data-era/121461)

### Computational Methods for Identification of Novel Secondary Metabolite Biosynthetic Pathways by Genome Analysis

Swadha Anand and Debasisa Mohanty (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 1642-1666).

[www.irma-international.org/chapter/computational-methods-identification-novel-secondary/76139](http://www.irma-international.org/chapter/computational-methods-identification-novel-secondary/76139)