Chapter 33
Modularity of Biochemical Networks

Hiroyuki Kurata
Kyushu Institute of Technology, Japan

ABSTRACT

To reveal the relationships between large-scale, heterogeneous biochemical networks and their associated functions, called design principles in biology, it is critically important to disintegrate the networks into topology- or function-based subnetworks to analyze the mechanism of how each subnetwork generates a specific biological function, and to synthesize them as the whole system in the same manner as engineering, where a variety of parts are assembled into functional machines. This synthesis and analysis approach can be carried out by a computer. In this review, the author describes several methodologies that serve to disintegrate biological systems into biologically meaningful modules, with practical consequences for systems biology studies.

INTRODUCTION

The genome sequences of several prokaryotic and eukaryotic organisms have now been completely been determined(Lander, 2011). Genome biology, which aims to understand living organisms as complete systems at the molecular level, has greatly shifted the paradigms of traditional biological studies that once focused on more simplified and local phenomena. A crucial question exists regarding how scientists can best understand and interpret the massive data sets generated by these studies involving complex networks of thousands of genes. "Systems Biology" is a promising methodology that may aid in the analysis of such molecular networks, revealing the relationships between network structures and their biological functions. In general, it is not feasible to understand a network as a single entity. How then should we define a complete network-based understanding of biological systems? To invoke a common metaphor, a system of living cells can be compared to a computer. An examination of the relationships between the input to a keyboard and the output from a screen hardly reveals the internal structure or mechanism of the computer, although many phenomena are observed. As we disintegrate a computer into its individual modules or elements whose functions are well known based on their architecture, we can then reconstruct a

DOI: 10.4018/978-1-4666-2196-1.ch033

computer using these well-understood modules or elements. Reconstruction may be defined as a perfect understanding of systems. Following this same line of thought, to gain a better understanding of biology, it is important to partition whole biological network into subnetworks or modules responsible for specific biological functions.

A module is characterized as a subsystem that possesses a function that is separable from that of other modules. Biology prefers an approach to modularity that emphasizes connectivity and function over isolation. Indeed, regulatory modules are not necessarily isolated, nor would they preserve their function if isolated or rearranged, except in very structured and organized ways (Barabasi & Oltvai, 2004; Rives & Galitski, 2003). To provide a useful characterization of modularity, it is imperative to classify and identify different types of modularity based on their levels of detail or abstraction.

In this review, I describe several methodologies that serve to disintegrate biological systems into topology- or function-based modules with practical consequences for systems biology studies. A number of different decomposition algorithms have been extensively characterized, the use of which depends not only on the types of biological networks in question, such as protein-protein interaction networks, metabolic networks, and gene regulatory and signal transduction networks, but also on the purpose of what we wish to understand.

LARGE-SCALE PROTEIN-PROTEIN INTERACTION NETWORKS

Many complex networks are naturally divided into modules in which the links within the modules are much more dense than those across the modules. Cellular functions are typically organized in a hierarchical modular architecture, where each module is a discrete object composed of a group of tightly linked components (genes, proteins, and metabolites) and performs a relatively independent task.

Protein-protein interaction (PPI) networks are appealing to biologists wishing to understand a complete and interconnected picture of cellular function. PPI networks can generally be transformed into a graph where a node is specified as a given protein with a given interaction at its edge. A large size and substantial heterogeneity are common features of PPI networks. In typical PPI networks, a few nodes may have a high degree of interconnectedness, while others may have very few interactions. Classical graph-based agglomerative methods employ a variety of similarity measures between nodes to partition PPI networks, but they often result in a poor clustering arrangement that contains only one or a few giant core clusters with many smaller ones (Barabasi & Oltvai, 2004).

While biological modules are not as clearly separated as expected in terms of network topology, a variety of algorithms have been developed to improve clustering results. PPI networks have been weighted based on topological properties, such as shortest path length (Arnau, Mars, & Marin, 2005; Rives & Galitski, 2003), clustering coefficients (Friedel & Zimmer, 2006), node degree, or the degree of experimental validity (Pereira-Leal, Enright, & Ouzounis, 2004). As an alternative and powerful algorithm, the shortest path betweenness (SPB) method has been proposed to define edge-betweenness, which is the number of shortest paths between all pairs of nodes that run through the edge, as a global measure to separate PPI networks into subgraphs in a divisive manner (Dunn, Dudbridge, & Sanderson, 2005; Luo et al., 2007; Newman, 2004).

To explore a biologically meaningful partition of PPI networks, a diffusion model-based spectral clustering method has been developed by introducing a power factor that adjusts the diffusion matrix to the heterogeneity of PPI networks. This method is referred to as adjustable diffusion matrix-based spectral clustering (ADMSC)(Inoue, Li, & Kurata, 2010). The discovery of cluster structures by random walks in this diffusion model is based on spectral graph theory that solves the eigenvectors