

Chapter 34

Modules in Biological Networks: Identification and Application

Bing Zhang

Vanderbilt University School of Medicine, USA

Zhiao Shi

Vanderbilt University, USA

ABSTRACT

One of the most prominent properties of networks representing complex systems is modularity. Network-based module identification has captured the attention of a diverse group of scientists from various domains and a variety of methods have been developed. The ability to decompose complex biological systems into modules allows the use of modules rather than individual genes as units in biological studies. A modular view is shaping research methods in biology. Module-based approaches have found broad applications in protein complex identification, protein function prediction, protein expression prediction, as well as disease studies. Compared to single gene-level analyses, module-level analyses offer higher robustness and sensitivity. More importantly, module-level analyses can lead to a better understanding of the design and organization of complex biological systems.

INTRODUCTION

The twentieth-century biology has been focused on individual cellular components and their functions. Despite the huge success of this approach, a discrete biological function can rarely be attributed to an individual molecule (Hartwell et al., 1999). It is increasingly clear that the cell can be understood as a complex network of interacting

components (Barabasi & Oltvai, 2004). Unraveling the interactions between the components of a cell constitutes a major goal of the post-genomic era.

With recent advances in high-throughput experimental technologies, genomic data are now available for the reconstruction of large-scale biological networks, in which nodes are biological molecules (e.g. proteins, genes, metabolites, microRNAs, etc.) and edges are functional relationships among the molecules (e.g. protein interactions, genetic interactions, transcriptional regu-

DOI: 10.4018/978-1-4666-3604-0.ch034

lations, protein modifications, metabolic reactions, etc.). Biological networks that have been studied include protein interaction networks constructed from protein-protein interaction data (Gavin et al., 2006; Ito et al., 2001; Krogan et al., 2006; Rual et al., 2005; Stelzl et al., 2005; Uetz et al., 2000), gene co-expression networks constructed from gene expression profiling data (Oldham et al., 2006; Stuart et al., 2003; van Noort et al., 2004), transcriptional regulation networks constructed from protein-DNA interaction data (Harbison et al., 2004; Lee et al., 2002), and metabolic networks constructed from bioreaction data (Duarte et al., 2007; Jeong et al., 2000). At a more abstract level, functional association networks have been used to represent integrated information from various types of functional association data (Franke et al., 2006; Jensen et al., 2009).

One of the most important properties of networks representing complex systems is *modularity*, i.e., the organization of nodes in clusters, with many edges connecting nodes of the same cluster and comparatively few edges connecting nodes of different clusters (Girvan & Newman, 2002). Indeed, modularity has been observed in protein interaction (Gavin et al., 2006), transcriptional regulation (Ihmels et al., 2002), and metabolic networks (Ravasz et al., 2002). Network-based module identification has captured the attention of a diverse group of scientists from various domains such as statistical physics, computer science, discrete mathematics, sociology, and computational biology (Fortunato, 2009). Although an ideal solution remains to be reached, the enormous effort of a large interdisciplinary community of scientists has generated a variety of approaches to this problem. In the biology community, besides network-based inference, modules can also be derived from existing knowledge on pathways and biological processes (Wang et al., 2008).

Modules, by definition, are sub-groups of elements (e.g. nodes and edges in the context of networks) that function in a semi-autonomous fashion and serve as building blocks of complex

systems. The ability to decompose complex biological systems into modules allows the use of modules rather than individual genes as units in biological studies. Module-based analyses have several advantages over gene-based methods, including improved robustness against the inherent noise that exists in the sample population and increased sensitivity in identifying patterns that are too subtle to discern when considering individual genes (Chuang et al., 2007; Mootha et al., 2003). More importantly, module-based analyses can achieve a higher-level understanding of the design and organization of biological systems (Gavin et al., 2006; Segal et al., 2004).

In this review, we first summarize module identification algorithms that have been developed in different research communities. Next, we provide examples of module-based applications in biological research, including protein complex identification, protein function prediction, protein expression prediction, and disease studies.

COMPUTATIONAL METHODS FOR MODULE IDENTIFICATION

Module identification in networks is the process of grouping network elements that are similar in some way into different sets, i.e. modules. Depending on different research communities, modules are also called *communities* or *clusters*. Loosely stated, a module is a cohesive set of nodes that are connected “more tightly” to each other than to other nodes in the graph. Despite of various attempts, there is no universally accepted quantitative definition of a module. In many cases, it is induced by the algorithms without a precise *a priori* definition. Not surprisingly, many module identification algorithms have been proposed based on their needs in different scientific domains (Fortunato, 2009). In this section, we present representative module identification methods categorized according to the approaches they take (Table 1).

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/modules-biological-networks/76087

Related Content

Efficient Fault Tolerant Algorithms for Internet Distributed Systems

Swati Mishra and Sanjaya Kumar Panda (2017). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 71-90).

www.irma-international.org/article/efficient-fault-tolerant-algorithms-for-internet-distributed-systems/178608

Lexical Granularity for Automatic Indexing and Means to Achieve It: The Case of Swedish MeSH®

Dimitrios Kokkinakis (2009). *Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration* (pp. 11-37).

www.irma-international.org/chapter/lexical-granularity-automatic-indexing-means/23053

Pharmacy Data Integrity for Optimal Analytics

C. David Butler (2014). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 21-45).

www.irma-international.org/article/pharmacy-data-integrity-for-optimal-analytics/147302

BTISNet: Biotechnology Information Network for Biological Scientific Community

K. N. Kandpal, Mohammad Faheem Khan and S. S. Rawat (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 1564-1569).

www.irma-international.org/chapter/btisnet-biotechnology-information-network-biological/76134

Physiological Cycles and Their Algebraic Models in Matrix Genetics

Sergey Petoukhov and Matthew He (2010). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications* (pp. 222-247).

www.irma-international.org/chapter/physiological-cycles-their-algebraic-models/37904