

# Graph Theory: A Powerful Research Tool for Biological Network Analysis

Chien-Hung Huang\*

*Department of Computer Science and Information Engineering, National Formosa University, Taiwan*

## Article Information

Received date: Jun 21, 2015

Accepted date: Aug 20, 2015

Published date: Sep 16, 2015

## \*Corresponding author

Chien-Hung Huang, Department of Computer Science and Information Engineering, National Formosa University, Taiwan 63205; Email: chhuang@nfu.edu.tw

**Distributed under** Creative Commons CC-BY 4.0

**Keywords** Graph algorithm; Network structure; Motif; Protein complexes; Vascular smooth muscle cell; Lung cancer; Biological network; Network robustness

In the pre-genome era, traditional molecular biology provides very informative knowledge on how individual bio-molecule, i.e. DNA, RNA and protein, perform biological functions. Networks of interactions among bio-molecules are fundamental to all biological processes; for example, the Gene Regulatory Network (GRN) can be described as a complex network of genes regulated by protein binding. Cellular processes are controlled by various types of biochemical networks; such as (i) metabolic networks, (ii) Protein-Protein Interaction (PPI) networks, (iii) GRN, and (iv) signal transduction networks. Biochemical networks are complex in nature; they consist of a large number of bio-molecules, interacting with each other give rise to biological responses and stabilities.

In the post-genome era, it is more productive to investigate how bio-molecules regulate or cooperate on a system level. The graph theory approach is a powerful tool for investigating the underlying topological structures of different molecular networks. A great diversity of graph theoretical notions is discussed to characterize biological networks.

The theory of complex networks plays an important role, ranging from computer science, sociology, engineering and physics, to bioinformatics etc. Within the fields of bioinformatics, potential applications of network analysis include drug target identification, determining bio-molecules' pathways and function, and designing effective strategies for treating various diseases. Molecular networks are the basis of biological processes. Such networks can be decomposed into smaller modules, also known as network motifs. These motifs show interesting dynamical behaviors, in which co-operatively effects between the motif components play a critical role in human diseases. Some of the network motifs are interconnected which can be merged together and form more complex structures, the so-called Coupled Motif Structures (CMS). These structures exhibit mixed dynamical behavior, which may lead biological organisms to perform specific functions.

On the other hand, the ability of a network to perform its intended function depends on how it responds to pressures - both internal and external, Maslov and Sneppen claimed that one could take into account experimental artifacts or mutations by doing an error tolerance study. Therefore, quantifying the robustness of a network is vital in a number of disciplines. Albert et al. proposed four traditional four conventional perturbations: (i) network edges are deleted randomly, during which the process of the total number of nodes are not changed; (ii) network nodes are removed randomly (failure); (iii) the most connected nodes are successively removed (attack) and (iv) nodes are rewired randomly, which is known as the local rewiring algorithm, in which the node degree of each node remains unchanged.

Protein complexes play an essential role in many biological processes. Complexes can interact with other complexes to form Protein Complex Interaction Network (PCIN) that involves in important cellular processes. On the other hand, Cardio Vascular Diseases (CVDs) affect tens of millions of human beings each year. Abnormal proliferation of Vascular Smooth Muscle Cell (VSMC) is a major cause of CVDs. In this study, we examine how VSMC responses when subject to mechanical stress. Time course microarray experiments were used to identify stress-induced Differentially Expressed Genes (DEGs). The gene association network among the DEGs was inferred by using the Gaussian Graphical Model. Besides, Lung cancer is the leading cause of death in Taiwan. By using the microarray data downloaded from GEO, we built the up-regulated and down-regulated PPI (Protein-Protein Interaction) network. By comparing early- and late-stage cancer, we can identify stage-specific lung cancer-associated genes. Although many studies are related to biological networks, but there are relatively few studies on examining the above three kinds of biological networks, and little is known about the stability of biological networks. We will employ graph theoretical approach to reveal hidden properties and features of biological networks. A great diversity of graph theoretical notions is employed to characterize many species' PCINs. Three main issues are addressed; (i) the global and local network topological properties and network structures, (ii) propose 9 new types of network robustness measurements and (iii) derive the key genes and clusters on network topologies.

There are many tasks are undergoing by our research teams to examine the graph theory and robustness analysis of biological networks:

- (1) Construct Protein Complex Interaction Network (including Human, Rat, Mouse, Yeast four species).
- (2) Construct Vascular Smooth Muscle Cell gene association network by Gaussian Graphical Model.
- (3) Construct lung cancer gene association network.
- (4) Employ graph theoretical approach to reveal hidden properties and features of the three biological networks. Twelve topological parameters are addressed, the global network topological properties, including average graph distance, diameter, network efficiency; and local network topological properties, including closeness centrality, degree centrality, eccentricity centrality, betweenness centrality, eigenvector centrality, bridging centrality, clustering coefficient, brokering coefficient and local average connectivity.
- (5) Perform Kolmogorov-Smirnov test on nine local topological parameters for the test of homogeneity.
- (6) Discover key genes based on performing meta-analysis and logistic regression on top topological parameters' scored genes.
- (7) Conduct the gene enrichment analysis to examine whether topological similar regions of PCINs are associated with similar molecular processes or not.
- (8) Examine the role of network motifs (including Auto-regulation loop, Feedback loop, FFL, Bi-fan, Single Input Module (SIM) and Multiple Input Module) of the above biological networks.
- (9) Elucidate the underlying network structures, including random network, scale-free network, hierarchical network, assortative/disassortative network and small world network.
- (10) Compare the network structures of multi-cellular organism with unicellular organism.
- (11) Analyze the stability of the networks under 13 types of perturbations, including nine local topological parameters perturbations and four conventional perturbations (edge deletion, node deletion, deletion of the most connected nodes and nodes are rewired randomly), and the effect of these perturbations.
- (12) Cluster analysis by using CFinder or K-means method.
- (13) Gene set enrichment analysis for three types of biological networks by using DAVID and CPDB tools.
- (14) Potential lung cancer drugs discovery using Connectivity Map (cMap).
- (15) Investigate the key lung cancer genes by MTT™ cell viability test and Clonogenic assay experiments and derive the optimal combination of topological properties.
- (16) Construct bioinformatics open source platform to provide the service of topology parameter calculation and robustness analysis, as well as the extraction and visualization of the key genes and clusters.

In conclusion, graph theory approach is a useful tool for studying biological networks at a system level and provides insights on topology properties usage. It is very likely that graph theory analysis can supply very specific information for further study; it is an indispensable tool for network biology research.