Dr. Khalid Raza

Assistant Professor, Department of Computer Science, Jamia Millia Islamia (Central University), New Delhi, kraza@jmi.ac.in

Network Biology: A New Way to Understand Complex Disease and Future of Medicine

We have various kinds of networks around us ranging from computer networks to social networks, food web, disease transmission networks and terrorist networks. Today, the concept and the reality of networks are playing an important role in modern society. From the last several years, scientists from different fields: including computer science, mathematics, physics, chemistry, sociology, and of course biology, deals with some kind of network and more recently coined a new terminology "network science" (Newman, et.al., 2006). In this article, biological networks are discussed with special focus on understanding complex molecular interactions for better insight about a disease and how biological networks help in the development of future medicines.

Network biology is an emerging area of research where researchers try to understand functional organization of cell and complex interactions between various constituents of cells (such as proteins, RNAs, DNAs, metabolites and other molecules). Even most cell performs their functions through interaction across cells and even across organs. Network biology approach may help to identify the root-cause of various human diseases, including cancers. Due to rapid advancement in high-throughput data collection technologies, such as Microarrays, it is possible to know the states of a cell's constituents at any given time. Similarly, other modern technologies, such as protein chips or yeast two hybrid screens, reveal how and when these cell's constituents interact (Barabasi & Oltavai, 2004). There are interactions between different constituents of cell that lead to different kinds of biological networks, such as protein-protein interaction network, gene regulatory network and metabolic pathways. These networks are not independent to each other, rather they form network of networks which is responsible for cell's behaviour and functions. These biological networks can be analysed using graph theory for inferring biologically relevant knowledge. These networks can also be exploited for developing templates for modelling and simulation of biological systems that may be further revised in consultation with concerned domain experts. The field of network biology has integration with other fields including text mining, data analysis, biological databases, modelling & simulation, sequence analysis, and so on, as shown in Fig. 1.

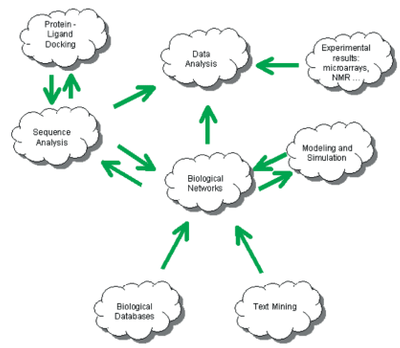

Fig. 1. Depicts an overview how different fields are integrated with biological networks (Koehler, et.al., 2004)

After Human Genome Project, the main aim of biomedical research is to develop systematic catalogue of all cell molecules and model their interactions. The major challenge is how to integrate theoretical and experimental techniques to understand and model the topological and dynamic properties of different kinds of networks in quantifiable terms.

The Human Interactome

The complexity of biological networks in human, also known as "human interactome", is daunting. The human interactome contains around 1,00,000 interacting molecules that includes ~25,000 protein coding genes, ~1,000 metabolites and an undefined number of proteins and functional RNAs (Barabasi, et.al., 2011). Mathematically, any network can be represented in the form of a graph G = {V, E}, where V denotes the set of N vertices (i.e., biological molecules) and E denotes set of edges that connect two elements in V. The biological networks can be of various types such as gene regulatory networks, protein-protein interaction networks, metabolic networks and RNA networks.

Gene regulatory networks: As described, the interactions between various cellular molecules form very complex "biological circuits". The gene regulatory network (GRN) is a kind of such "biological circuit". A typical GRN may consists of input signalling pathways, regulatory proteins which assimilate input signals, target genes, RNAs, and produces proteins from these target genes. These networks may also include dynamic feedback loops. The GRNs are the systematic biological networks that describe regulatory interactions among genes in the form of a graph, where nodes represents genes and edges their regulatory interactions. In general, there can be two types of interactions: *activation* or *inhibition*. Generally, edges are directed

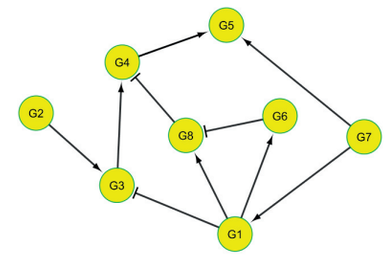

Fig. 2. A typical gene regulatory network having eight nodes and 10 edges, where sharp arrows depict activation and blunt arrows inhibition

but if transfer of information between nodes is bidirectional, then the edge is considered as undirected. Figure 2 shows an example of a gene regulatory network. Gene regulatory network is a very useful mechanism used to demonstrate complex dependencies between developmental transcription factors, target genes and their regulators.

Protein-protein interaction networks: It is also a network where nodes are proteins and are linked to each other by a physical interaction. Protein-protein interactions play a vital role in living cells that control most of the biological processes. Proteins mostly perform their functions with the help of interactions with other proteins. Mutations occurred due to disease affects protein interactions that disrupt protein-DNA interactions, adds new unwanted

interactions and lead to protein misfold. Aberrant protein-protein interactions have implicated in a number of neurological disorders, for example, Alzheimer's disease. With appropriate knowledge of interaction scientist can easily predict pathways in the cell, potential novel therapeutic target, and protein functions. This has motivated to map interactions on the proteome-wide scale.

Metabolic networks: As we know, metabolism is one of the complex cellular processes. The connectivity between biochemical and metabolites form very complex metabolic networks which can be analysed using graph theory for inferring lots of biologically meaningful information.

RNA networks: Last few years brought a better understanding about RNA world and its key role in biological networks. RNA networks refer to those networks that contain RNA-RNA or RNA-DNA interactions. It is found that non-coding RNAs regulate in various aspects of cell differentiation and development and it also works as a key player in regulating flow of genetic information. Several microRNA-gene interaction networks have been reported in the literature and are available in various databases such as microRNA, PicTar, miRBase, miRecords and TargetScan.

Basic Properties of Disease Networks

Today, the advancement in graph and network theory helped to get insights about the properties of biological networks, especially disease networks. From the studies of networks, it is revealed that networks in biological, social and technological systems are not functioning randomly, but are organized by set of principles, such as motifs, modules, hubs, centrality measures, etc. These network principles help us to extract some of the basic properties of genes involved in disease. Some of the basic properties of a disease network are describe as follows:

Hub genes: One of the distinguishing properties of biological networks is a limited number of highly connected genes, called hub genes. These hub genes performs special biological role. It is revealed from the model organism that hub genes are older and evolve more slowly, and also deletion of hub genes lead to a larger number of phenotypic outcomes, than non-hub genes. It is a hypothesis that hub genes are typically associated with disease genes in human interactome. Many studies support this hypothesis; however, evidence of some of these effects is still poor and debatable.

Disease modules: A module is a sub-network having high degree of clustering and represents highly interlinked local regions within a network. It is commonly accepted that biological system is modular. A biological functional module is an entity, consists of different interacting molecules and their function is separable from other modules (Cho, et.al., 2012). It is suspected that the direct interacting molecules participating in a specific disease, also play some role in the same disease. Over the last few years, many network-clustering tools have developed but still identification of network modules is computationally challenging. Given a large network, we may require to discover meaningful functional modules from such biological networks. The most commonly used method is to identify densely connected subgraphs or clusters. Alternatively, gene expression data or (and) protein interaction data can be used to extract modules by grouping co-expressed genes into one module. Cho & his colleagues described some of the disease modules as: (i) disease modules are enriched with genetic alterations, so, genes or genomic regions which are altered in a disease of interest are identified first, and then it is mapped to an interaction network; (ii) differentially expressed network modules – a subnetworks enriched with genes whose expression values are significantly changed in disease sample profile; and (iii) uncovering information propagation modules – clusters of genes related to either phenotypic or genotypic information, or combining both that may be an effective approach to understand disease mechanisms. There are several applications of network modules in medical science, such as (i) disease classification, (ii) finding disease similarity, and (iii) response to treatment.

Motifs: A motif is a sub-graph in which a group of nodes linking to each other to form a small subnetwork within a biological network. Motifs are considered as elementary units of cellular networks. Motifs are more likely to be associated with some biological functions, such as negative feedback loops, positive feedforward loops and oscillators.

Predicting Disease Genes and Their Interaction Networks

Several sophisticated network-based tools have been proposed to identify disease associated genes but most of these methods identify a large number of disease associated genes. Hence, identification of particular genes and casual mutation remains a challenge. These tools can be grouped into three categories (Barabasi, et.al., 2011):

Linkage methods: It hypothesise that the direct interacting disease-proteins are likely to be involved in the same disease phenotype.

Module-based methods: It hypothesise that all cell components belonging to the same topological, functional or disease module have a high chance for being associated in the same disease. In these methods, disease modules are identified first then examines their members as potential disease genes. Several variants of these methods have been used to a large number of diseases including cancers, systemic inflammation, obesity, cardiovascular disease, neurological diseases, asthma and diabetes.

Diffusion-based methods: These methods identify the pathways closest to some known disease. Here, 'random walk' algorithm is applied and a walker is released from a known disease genes, and they are allowed to diffuse along the links of the network, moving to any neighbouring nodes with equal probability. By this way, the nodes and links closest to the known disease genes are identified because these genes will be most often visited by random walkers. Hence, proteins interacting with large number of disease proteins are likely to gain a high probability weight. Several variants of these methods have also been applied to detect diseases genes including cancers and Alzheimer.

Future of Medicines: Network-based & Personalized Medicines

The current and established diagnosis of disease relies on humble correlation between clinical syndromes and pathological analysis. This traditional method is largely based on Oslarian clinico-pathological correlation, where disease is defined on the basis of organ in which symptoms are identified. Although, this conventional approach to disease diagnosis has been serving the society well for over a century but it has

several severe flaws in the post-genomic modern era of genomic medicine. Some of the limitations of conventional approach to disease diagnosis and treatment are as follows:

i. The focus is not on specific genetic of the disease and hence therapy mechanisms do not focus on targeted disease source.
ii. Diagnosis is based on late-appearing symptoms in the organ.
iii. Neglects underlying pathobiological mechanism of the disease and molecular & environmental factors are not taken into account which governs the evolution of disease.
iv. In fact, disease is not based on single effector gene product, but on interactions in the complex biological networks.

To solve shortcomings & limitations of conventional disease definition, we need to reconsider and redefine the factors of the disease and shifting the entire healthcare system to modern genome-based approach.

Personalized medicines: The beginning of a disease is triggered by a breakdown in a biological network system. The diagnostic and medical system based on our understanding of interacting network system is called "network medicine". It explores the interaction mechanism at the system-level and translates discoveries to better diagnosis of disease. Personalized medicine is the emerging practice that uses an individual's unique molecular characteristics and their genetic profile to diagnose more finely and helps the doctors to give proper medication and therapy according to diseases which also decreases its risk. Network-based technique is applied to gain better insight in disease phenotyping and developing novel therapeutics which address system-wide molecular disturbance occurred due to disease processes. The main objective of personalized medicine is to develop individualized medicine which is more effective and specific to a particular patient. The novel personalized medicine is likely to reduce the adverse effects of drug and a significant influence on healthcare. This approach optimizes the drug discovery and drug development process to get improved understanding of disease processes, drug safety profile and drug efficacy. It also stratifies patient's molecular profiles which is main root toward treatment. The personalized medicine allows development of "targeted" diagnostics and therapeutics. Genomic testing allows physicians to find out an individual's susceptibility to disease, predict how a particular patient would respond to a specific drug molecule, reduce adverse effects of drugs, increase the effectiveness of treatments and finally improve overall health system. Some of the benefits of such approach to healthcare are:

i. Moving towards 'evidence-based' system from 'trial-and-error' based system
ii. Anticipating health problems and focusing on wellness, not on disease
iii. A new medical model that creates a customizable healthcare according to the genetic information of an individual
iv. Remove trial-and-error inadequacies; decrease healthcare cost, time and clinical trial failure and lead to better patient care.

Recently, a personalized medicine paradigm, called "P4 medicine", has been proposed by Leroy Hood, President of Institute for Systems Biology, Washington, USA. The "P4 medicine" focuses on four key attributes: Personalized, Preventive, Predictive and Participatory. Personalized means considering a person's genetic information; preventive – handing health problems by focusing on wellness of a person rather than disease; predictive – prescribing treatment that avoid drug reaction; and participatory – empowering patients to be attentive and responsible for their healthcare. According to Leroy Hood "the most unique about P4 medicine is that it will represent a network of networks – molecular networks, genetic networks, cellular networks, tissue networks, individual networks, population networks, social networks".

Conclusions

Network biology lets us study about the complex diseases. Complex disease can be better understood by network-based approaches. Modularity is broadly recognised concept in molecular biological networks and module-based methods offer a number of benefits including robustness in the identification of pathways and better disease classification. Furthermore, network-based formalism lets us apply a plenty of methods already developed in graph theory, such as bipartite graph, shortest paths, module and motif identification, hubs, network flow, and Steiner trees. Network-based approach has some limitations including lack of mechanistic explanations. Despite of limitations, network biology has been successfully applied in several diseases that suggest testable hypotheses. The personalized medicines, that targets individualized diagnosis based on individual's genetic variations is an innovation which is expected to create both opportunities and challenges for conventional healthcare and emerging markets.

References

[1] Barabasi, A-L & Oltvai, ZN (2004). Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5.
[2] Barabasi, A-L, Gulbahce, N & Loscalzo, J (2011). Network medicine: a network-based approach to human disease. Nature Reviews, 12.
[3] Cho, D-Y, Kim, Y-A, & Przytycka, TM. (2012). Chapter 5: Network Biology Approach to Complex Diseases. PLoS Comput Biol 8(12): e1002820.
[4] Koehler, J, et.al. (2004). Linking experimental results, biological networks and sequence analysis methods using Ontologies and Generalised Data Structures. In Silico Biology 5, 0005.
[5] Newman, M, Barabási, A-L & Watts, DJ (2006). The Structure and Dynamics of Networks. Princeton Studies in Complexity, Princeton University Press.

About the Author

Dr. Khalid Raza is an active researcher and currently working as an Assistant Professor at the Department of Computer Science, Jamia Millia Islamia (Central University), New Delhi, India. He has completed his Bachelor's & Master's Degree in Computer Science & Applications and also cleared National Eligibility Test (NET) & Junior Research Fellow (JRF) in the same subject conducted by University Grants Commission (UGC), India. He has completed his Ph.D in Computational Biology from Jamia Millia Islamia, New Delhi. He has published more than 10 research articles in refereed international journals, conference proceedings and as book chapters. His research area of interest is soft computing techniques, artificial intelligence, systems biology and bioinformatics. He has received research grants for two Govt. funded research projects and working as Principal Investigator for both. He is reviewer of several international journals, member of several conference review committee and delivered several invited talk in national/international conferences/workshops.