

Graph Theoretic Concepts in the Study of Biological Networks

M. Indhumathy, S. Arumugam, Veeky Baths and Tarkeshwar Singh

Abstract The theory of complex networks has a wide range of applications in a variety of disciplines such as communications and power system engineering, the internet and worldwide web (www), food webs, human social networks, molecular biology, population biology and biological networks. The focus of this paper is on biological applications of the theory of graphs and networks. Graph theory and several graph theoretic properties serve as an ideal mathematical tool in the analysis of complex networks. We present the basic concepts and notations from graph theory which is widely used in the study of biological networks. Various biological networks such as Protein interaction networks, Metabolome based reaction network, Gene regulatory network, Gene coexpression network, Protein structure network, Structural brain network, Phylogenetic networks, Ecological networks and Food web networks are described. We also deal with various centrality measures which provide deep insight in the study of biological networks. Applications of biological network analysis in several areas are also discussed.

M. Indhumathy · S. Arumugam (✉)
National Centre for Advanced Research in Discrete Mathematics,
Kalasalingam University, Anand Nagar, Krishnankoil 626126, Tamil Nadu, India
e-mail: s.arumugam.klu@gmail.com

M. Indhumathy
e-mail: indhumathy.bio@gmail.com

S. Arumugam
Department of Computer Science, Liverpool Hope University, Liverpool, UK

S. Arumugam
Department of Computer Science, Ball State University, Muncie, IN, USA

V. Baths
Department of Biological Sciences, Birla Institute of Technology
and Science Pilani, K K Birla Goa Campus, NH-17B, Zuarinagar, Goa, India
e-mail: veeky@goa.bits-pilani.ac.in

T. Singh
Department of Mathematics, Birla Institute of Technology and Science Pilani,
K K Birla Goa Campus, NH-17B, Zuarinagar, Goa, India
e-mail: tksingh@goa.bits-pilani.ac.in

Keywords Biological networks · Centrality measures · Graph · Motifs

1 Introduction

The theory of complex networks has a wide range of applications in a variety of disciplines such as communications and power system engineering, the internet and worldwide web (www), food webs, human social networks, molecular biology, population biology and biological networks. The focus of this paper is on biological applications of the theory of graphs and networks. Network analysis leads to a better understanding of the critical role of these networks in many key questions. For instance the complex interplay between the structure of social networks and the spread of disease is a topic of critical importance and hence in recent years the researches in ecology and epidemiology have focused attention in network analysis.

Graph theory and several graph theoretic properties serve as an ideal mathematical tool in the analysis of complex networks. In Sect. 2, we present the basic concepts and notations from graph theory which is widely used in the study of biological networks. Various biological networks such as Protein interaction networks, Metabolome based reaction network, Gene regulatory network, Gene coexpression network, Protein structure network, Structural brain network, Phylogenetic networks, Ecological networks and Food web networks are described in Sect. 3. Section 4 deals with the various centrality measures which provide deep insight in the study of biological networks. Topology of biological networks and network motifs have been stated in Sects. 5 and 6 respectively. Sections 7 and 8 deal with network databases and network visualizing tools. Applications of biological network analysis in several areas and concluding remarks are given in Sects. 9 and 10 respectively.

2 Graph Theoretic Concepts

In this section we present a few basic concepts in graph theory which are essential for the study of biological networks. For graph theoretic terminology we refer to Chartrand and Lesniak [8].

A graph G is a finite nonempty set of objects called *vertices* or *nodes* together with a set of unordered pairs of distinct vertices of G called *edges* or *links*. The vertex set and the edge set of G are denoted by $V(G)$ and $E(G)$ respectively. The edge $e = \{u, v\}$ is said to join the vertices u and v . We write $e = uv$ and say that u and v are adjacent vertices; u and e are incident, as are v and e . If e_1 and e_2 are distinct edges of G incident with a common vertex, then e_1 and e_2 are adjacent edges.

The number of vertices in G is called the *order* of G and the number of edges in G is called the *size* of G . A graph of order n and size m is called a (n, m) -graph. A graph is *trivial* if its vertex set is a singleton. A graph $G = (V, E)$ is called a weighted graph if for every edge e of G a weight $w(e)$ is assigned. The weight is

usually a positive number. The graph G is called a signed graph if every edge e of G is assigned a positive or a negative sign.

A vertex u is called a *neighbor* of a vertex v in G , if uv is an edge of G . The set of all neighbors of v is the *open neighborhood* of v and is denoted by $N(v)$; the set $N[v] = N(v) \cup \{v\}$ is the *closed neighborhood* of v in G .

A graph H is called a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A subgraph H of a graph G is a *proper subgraph* of G if either $V(H) \neq V(G)$ or $E(H) \neq E(G)$. A *spanning subgraph* of G is a subgraph H of G with $V(H) = V(G)$.

For a set S of vertices of G , the *induced subgraph* is the maximal subgraph of G with vertex set S and is denoted by $\langle S \rangle$. Thus two vertices of S are adjacent in $\langle S \rangle$ if and only if they are adjacent in G . The induced subgraph $\langle S \rangle$ is also denoted by $G[S]$.

Let v be a vertex of a graph G and $|V(G)| \geq 2$. Then the induced subgraph $\langle V(G) \setminus \{v\} \rangle$ is denoted by $G - v$ and it is the subgraph of G obtained by the removal of v and the edges incident with v . If $e \in E(G)$, the spanning subgraph with edge set $E(G) \setminus \{e\}$ is denoted by $G - e$ and it is the subgraph of G obtained by the removal of the edge e . The graph obtained from G by adding an edge e is denoted by $G + e$.

The *degree* of a vertex v in a graph G is defined to be the number of edges incident with v and is denoted by $\deg(v)$.

The minimum of $\{\deg(v) : v \in V(G)\}$ is denoted by $\delta(G)$ or simply δ and the maximum of $\{\deg(v) : v \in V(G)\}$ is denoted by $\Delta(G)$ or simply Δ .

A graph G is *complete* if every pair of distinct vertices of G are adjacent in G . A complete graph on n vertices is denoted by K_n .

A *clique* in G is a complete subgraph of G . The maximum order of a clique in G is called the *clique number* of G and is denoted by $\omega(G)$ or simply ω . A clique H in G with $|V(H)| = \omega$ is called a *maximum clique* in G .

A *bipartite* graph is a graph G whose vertex set $V(G)$ can be partitioned into two nonempty subsets X and Y such that each edge of G has one end in X and the other end in Y . The pair (X, Y) is called a *bipartition* of G . If further, every vertex in X is adjacent to all the vertices of Y , then G is called a *complete bipartite graph*. The complete bipartite graph with bipartition (X, Y) such that $|X| = r$ and $|Y| = s$ is denoted by $K_{r,s}$. In particular, the graph $K_{1,n-1}$ is called a *star* and the graph $K_{1,3}$ is called a *claw*.

A *walk* in a graph G is an alternating sequence $u_0, e_1, u_1, \dots, u_{n-1}, e_n, u_n$ of vertices and edges of G , beginning and ending with vertices such that $e_i = u_{i-1}u_i$, for $1 \leq i \leq n$. This walk joins u_0 and u_n and may also be denoted $(u_0, u_1, u_2, \dots, u_{n-1}, u_n)$; it is sometimes called a $u_0 - u_n$ walk. It is *closed* if $u_0 = u_n$ and is *open* otherwise.

A *path* P of length n in a graph G is a sequence $(u_0, u_1, u_2, \dots, u_{n-1}, u_n)$ of distinct vertices such that for $0 \leq i \leq n - 1$, the vertices u_i and u_{i+1} are adjacent. We say that P is an $u_0 - u_n$ path. The vertices u_0 and u_n are called the *origin* and *terminus* of P respectively. The vertices u_1, u_2, \dots, u_{n-1} are called the *internal vertices* of P . A path on n vertices is denoted by P_n .

A cycle of length $n \geq 3$ in a graph G is a sequence $(u_0, u_1, u_2, \dots, u_{n-1}, u_0)$ of vertices of G such that for $0 \leq i \leq n - 2$, the vertices u_i and u_{i+1} are adjacent, u_{n-1} and u_0 are adjacent and $u_0, u_1, u_2, \dots, u_{n-1}$ are distinct. A cycle on n vertices is denoted by C_n . A cycle C_n of length n is called *even* or *odd* according as n is even or odd.

A graph G is said to be *connected* if every pair of vertices of G are joined by a path. A maximal connected subgraph of G is called a *component* of G .

A graph G having more than one component is called a *disconnected* graph. An edge e of a connected graph G is called a *cut-edge* if $G - e$ is disconnected. A vertex v of a connected graph G is called a *cut-vertex* if $G - v$ is disconnected.

The *distance* $d(u, v)$ between two vertices u and v of a connected graph G is defined to be the length of any shortest path joining u and v . A shortest $u-v$ path is often called a *geodesic*.

The *diameter* of a connected graph G is the length of any longest geodesic and is denoted by $\text{diam}(G)$. We call an $u - v$ path in G for which $d(u, v) = \text{diam}(G)$ as a *diametrical path*.

3 Biological Networks

In this section we present some of the popular biological networks which have been investigated by several authors.

Protein-Protein Interaction network (PPI-Network) is a graph $G = (V, E)$ where V is a set of proteins and two proteins are joined by an edge if they interact physically. The interaction between viral proteins and human proteins can be represented as a bipartite graph G . The vertex set of G is $V_1 \cup V_2$, where V_1 is the set of viral proteins and V_2 is the set of all human proteins. A viral protein $v \in V_1$ is joined to a human protein $w \in V_2$ if v interacts with w . This bipartite graph is called viral-human protein interaction network and this network has been investigated by Mukhopadhyay and Maulik [26].

Human protein and disease association network is a bipartite graph G whose vertex is $V_1 \cup V_2$, where V_1 is the set of human proteins and V_2 is the set of diseases and $v_1 \in V_1$ is joined by an edge to $v_2 \in V_2$, if the human protein v_1 is associated with the disease v_2 . This network has been investigated by Mukhopadhyay and Maulik [26].

Metabolome based reaction network is a directed graph $D = (V, A)$ where V is a set of metabolites and a vertex v is joined to a vertex w by an arc (v, w) if there is a reaction or interaction which transforms the metabolite v to the metabolite w . This network has been investigated by Veeky Baths et al. [5].

Gene regulation is a general term for cellular control of the synthesis of protein at the transcription step. Often one gene is regulated by another gene via the corresponding protein. Thus gene regulation leads to the concept of gene regulatory network, which has been investigated by Yue and Chunmei [36]. Gene regulatory network is a directed graph $D = (V, A)$ where V is the set of genes and two genes $g_1, g_2 \in V$ are joined by an arc if there is a regulatory relationship between g_1 and

g_2 , or more precisely g_1 regulates g_2 . The regulatory relationship between two genes may be either positive direct regulatory influence or inverse causality or no correlation. Hence gene regulatory network can also be represented as a directed weighted graph, where the weight of an arc is an estimate of the probability of relationship between the genes in the network. This network has been investigated by Raza and Jaiswal [29]. Positive regulatory relationship represents activation and negative regulatory relationship represents inhibition. This leads to the representation of a gene regulatory network as a signed directed graph where an arc (g_1, g_2) is assigned a positive sign if the corresponding regulatory relationship is activation and is assigned a negative sign if the corresponding relationship is inhibition. A study of gene regulatory network leads to a better understanding of the regularity mechanism of the genes and prediction of the behavior of some unknown genes. This network has been studied in Christensen et al. [9].

A gene coexpression network is a graph $G = (V, E)$ where V is a set of genes and two genes are connected by an edge if there is a significant coexpression relationship between them. There are several methods for constructing the gene coexpression network and this network has been studied in Perkins et al. [28]. A coexpression measure is selected and a similarity score is calculated for each pair of genes using this measure. Two genes which have a similarity score higher than a selected threshold value are joined by an edge (Azuaje [2]).

To understand the protein structures, a graph representation of protein structure has recently been introduced. Protein structure is modeled as residue interaction graph (RIG) in which nodes represent the amino acid residuals and an edge represents a pairwise contact between residuals. A contact between two residuals is defined if the distance between any pair of their heavy atoms is within a specified distance cut-off and the cut-off is normally taken in the range (4,5). Understanding RIGs may provide deeper insights into protein structures binding and folding mechanisms as well as inter protein stability and function. Properties of this network are given in [24].

Structural brain network can be represented as a graph whose vertex set is the set of neural elements (neurons or brain regions) and edges representing physical connections (Synapses or axonal projections). This is used to understand the complex structure of the brain and brain associated diseases such as Alzheimers disease, brain tumor and epilepsy. This network has been studied in [7].

Consider an ecological community V consisting of predators and preys where a predator eats a prey. The food web network is a directed graph D whose vertex set is V and if $u, v \in V$, then (u, v) is an arc in D if u is a predator, v is a prey and u eats v . The competition graph is a graph whose vertex set is the set of predators and there is an edge between two predators if they have a common prey. If further we associate a weight with edge in the competition graph, where the weight is the number of common preys then we obtain a weighted competition graph. These concepts have been investigated by Dunne et al. [10].

A Phylogenetic network is a graph which is used to visualize evolutionary relationships between nucleotide sequences, genes, chromosomes or genomes. They are used when reticulated events such as hybridization, horizontal gene transfer, recom-

bination or gene duplication and loss are believed to be involved. For further details we refer to [17].

An Ecological network is a representation of the biotic interactions in an ecosystem. This is an undirected graph whose vertex set is a set of species and two species are joined by an edge if they interact. These interactions can be trophic or symbiotic. Applications of ecological networks include exploration of how the community context affects pairwise interactions. Other related studies include Metapopulations, epidemiology and evolution of cooperation. For some basic results in Ecological network we refer to Sole and Montoya [32].

4 Centrality Measures

Graph theoretic concepts are being extensively used in the analysis of biological networks. In this section we present several centrality measures which are used to rank the nodes of a network in the order of their performance.

4.1 Stress

The stress is a node centrality index, which has been studied by Shannon et al. [31]. The stress of a node v is the number of shortest paths passing through v . A node with a high stress is traversed by a high number of shortest paths. However, a node has high stress value does not imply that it is critical to maintain communication. Indeed two nodes may be connected by other shortest paths not passing through v . Hence “high” and “low” stress are more meaningful when stress of a node v is compared with the average stress value of the graph G . In biological terms, the stress of a node indicates its relevance in holding together communicating nodes. Hence, if the stress is higher, then the relevance of the node connecting other node is higher. Due to the nature of this centrality it may also be possible that the stress indicates how a molecule is heavily involved in cellular processes but not necessarily in maintaining the communication between the other nodes.

4.2 Betweenness

Betweenness is another node centrality index which is similar to stress, but provides more information. Let v_1, v_2 and v be three distinct nodes. Let $\sigma_{v_1 v_2}(v)$ denote the number of shortest v_1-v_2 paths which pass through v . Let $\sigma_{v_1 v_2}$ denote the total number of shortest v_1-v_2 paths. The betweenness centrality index of v is defined by

$$C_B(v) = \sum_{\substack{v_1, v_2 \in V \\ v_1 \neq v_2}} \left(\frac{\sigma_{v_1 v_2}(v)}{\sigma_{v_1 v_2}} \right).$$

Since for computing the betweenness centrality index, the summation is taken overall pairs of nodes, we divide it by $\frac{(n-1)(n-2)}{2}$ for graphs, where n is the total number of nodes; so that the betweenness index of v lies in the range $[0, 1]$. A high betweenness index indicates that the node for certain paths is crucial to maintain node connections (Scardoni and Laudana [30]). The betweenness index of a node in a protein-signaling network indicates the relevance of a protein as functionally capable of holding together communicating proteins. The higher the value the higher the relevance of the protein as organizing regularity molecule.

4.3 Edge Betweenness

Edge betweenness centrality is the edge version of the node betweenness centrality. Let v_1, v_2 be two distinct nodes and let e be an edge. Let $\sigma_{v_1 v_2}(e)$ denote the number of shortest v_1-v_2 paths which pass through the edge e . Let $\sigma_{v_1 v_2}$ denote the total number of shortest v_1-v_2 paths. Then the edge betweenness centrality index of e is defined by

$$C_B(e) = \sum_{\substack{v_1, v_2 \in V \\ v_1 \neq v_2}} \left(\frac{\sigma_{v_1 v_2}(e)}{\sigma_{v_1 v_2}} \right).$$

The edge betweenness index is normalized as in the case of node betweenness index. In the context of a protein-signaling network, edge betweenness centrality indicates that a specific biochemical reaction has a central role in the network functional organizations.

4.4 Diameter

The diameter of a graph G is defined by $diam(G) = \max \{d(u, v) : u, v \in V\}$, where $d(u, v)$ is the distance between the vertices u and v . It is a simple general parameter which indicates the compactness of the network. If G has high diameter, then G has two vertices whose distance is high. However, a graph with high diameter may have subgraphs which are compact. If a graph has low diameter, then it surely indicates that all the nodes are close to each other and the graph is compact.

For example, the diameter of a protein-signaling network can be interpreted as the overall easiness of the proteins to communicate or influence their reciprocal function.

4.5 Average Distance

The average distance of a graph G is defined by

$$Ad(G) = \frac{\left(\sum_{\substack{u, v \in V \\ u \neq v}} d(u, v) \right)}{\frac{n(n-1)}{2}}$$

where n is the number of nodes in G .

In general $Ad(G)$ is not an integer. In most cases, $Ad(G)$ is more informative than the diameter. High average distance indicates that the nodes are distant, implying that the network is not compact. A low average distance indicates that the nodes are close to each other and the network is compact. For example if a big protein signaling network has low average distance, then the proteins within the network have the tendency to generate functional complexes or modules.

4.6 Closeness

Closeness is another node centrality index. The closeness centrality of a node v is defined by

$$CC(v) = \frac{n - 1}{\sum_{u \in V - \{v\}} d(u, v)},$$

where n is the number of nodes in the network. Here high and low values are more meaningful when compared to the average closeness of G . High value of closeness of v indicates that all the nodes are in proximity to v and low value of closeness of v indicates that all other nodes are distant from v .

For example, the closeness of a node in a protein-signaling network can be interpreted as the probability of a protein to be functionally relevant for several other proteins. Thus a protein with high closeness will be easily central to the regularity of other proteins. This concept has been used in [1] for the study of protein structures.

4.7 Eigenvector Centrality

The adjacency matrix A of a graph G with vertex set $V(G) = \{v_1, v_2, \dots, v_n\}$ is the $n \times n$ matrix defined by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases}$$

A number λ is called an eigen value of A , if there exists a vector e such that $Ae = \lambda e$ and e is called the eigen vector corresponding to the eigen value λ . Since the matrix A is symmetric, all its eigen values are real. Let e_1 be the eigen vector corresponding to the largest eigen value λ_1 of A . Then the i^{th} component of the vector e_1 is the eigen vector centrality of the node v_i . In biological terms, a node with high eigen vector centrality value is adjacent to the other nodes that themselves have high eigenvector centrality value. This concept is given in [6].

4.8 Eccentricity

The eccentricity of a node v is defined by $e(v) = \max\{d(u, v) : u \in V - \{v\}\}$. Thus eccentricity of v is the distance between v and a node which is farthest from v . Let $f(v) = \frac{1}{e(v)}$, the reciprocal of the eccentricity. If $f(v)$ has a high value, then all other nodes are in proximity with v . On the other hand, if $f(v)$ is low, then there is at least one node which is far from v . For example in a protein-signaling network a protein with high value of f will be more easily influenced by activity of other proteins.

4.9 Subgraph Centrality

Subgraph centrality was introduced by Estrada in 2005. This centrality measure helps in finding hidden subgraph within a network. Here, smaller subgraph has given more weightage than the larger one as smaller subgraph can reveal the network motifs [13].

5 Topology of Biological Network

5.1 Watts and Strogatz Small World Network

Small world network was proposed by Watts and Strogatz [34]. This network is a random network with high clustering coefficient value indicating that most pairs of nodes contain at least one shortest path of small length between them. Therefore, in this type of network, mean length of shortest path is always small and in a given time, it is possible to reach from one node to another with few steps. Internet connectivity, social network, gene network are examples of small world network. This type of topology is only applied to the network where single nodes have few neighbors. If links between nodes are grown in huge number, then this theory fails as there may not be a shortest path of small length between two distant nodes.

5.2 *Erdos Renyi Random Network*

Paul Erdos and Alfred Renyi proposed this random graph network for non regular complex networks. Here edges are randomly added between pairs of randomly selected nodes with an initial condition of N nodes without any edges between them. This network follows poissonian distribution assuming added edges $\ll N^2$ [11]. Real world network like internet, social network, biological network etc. do not follow this network.

5.3 *Barabasi–Albert Scalefree Network*

Scalefree network was proposed by Barabasi–Albert [3]. This model opposes the idea that all complex networks are random in nature. According to this network, there are some special kind of mechanisms which shape this randomness of complex network. In scalefree network, structure and evolution are closely related and it is constantly changing by addition of new node or link to the existing network. When a new node comes in the existing network, it will tend to link with the node having maximum number of connections in a given network. This type of attachment is known as preferential attachment, and the node with maximum number of connections is known as hub. Here, degrees are distributed following the power law distribution resulting a few nodes with maximum links(hubs) while many nodes with a very few links.

6 Network Motifs

In biological network, it has been observed that a particular group of nodes with a fixed structural pattern, are involved in specific functions and these are known as motifs. Motifs are often called simple building blocks of complex network [25]. In graphs, motifs are basically repeating units of small subgraphs within a single network or among many networks. As motifs are involved in the particular functions, it is possible to predict the function of unknown proteins by comparing with the known motifs and then with its function. There are many motifs finder algorithm like Mavisto, FANMOD used to identify motifs within a network.

7 Network Databases

To date, vast amount of biological data has been created with the help of high throughput techniques like yeast two hybrid screening systems, DNA microarray and next generation sequencing. These data can be accessed through databases. There are

many such online databases housing these data and are freely accessible. Following are some important databases of biological network.

- **KEGG**. Kyoto Encyclopedia of Genes and Genomes [21].
- **STRING**. Search Tool for the Retrieval of Interacting Genes/Proteins [18].
- **HPRD**. Human Protein Reference Database [23].
- **MINT**. Molecular Interaction Database [37].
- **DIP**. Database of Interacting Proteins [35].
- **Reactome**. It is database for reaction pathways and biological processes. The pathways represented here are species specific [20].
- **BioGRID**. Biological General Repository for Interaction Datasets [33].
- **SPIKE**. Signaling Pathway Integrated Knowledge Engine [27].
- **IntAct**. InAct Molecular interaction database [22].

8 Network Visualizing and Analyzing Tools

There are many open source tools routinely used for network visualization and also for calculating different centrality values along with many network parameters like diameter, degree, shortest pathlength, clustering coefficient etc. Following three tools are widely used in biological network visualizing and analysis.

- **Cytoscape** [31]
- **Pajek** [4]
- **Visant** [16]

9 Applications

Applications of graph theory in the fields of biology and medicine include identification of drug targets, determination of the role of proteins or genes of unknown function [12, 14], design of effective containment strategies for infectious diseases and early diagnosis of neurological disorders by detecting abnormal patterns of neural synchronization. The knowledge of the topologies of biological networks and their impact on biological processes is needed to develop more sophisticated treatment strategies for complex diseases such as cancer [19]. Protein-Protein interaction networks have recently been combined with the networks describing the relationships between the diseases and disease gene causing them as well as between drugs and their protein targets, thus giving insights into pharmacology. Another application is the study of genetic disorders. A disease is represented as a bipartite graph, whose vertex set is the set of genetic disorders and disease genes. A genetic disorder X is joined to the disease gene Y if there is a mutation in the gene Y which gives genetic disorder X [15]. Two other graphs in this connection are constructed as follows. The human disorder network has its vertex set the set of genetic disorders and two

disorders are joined by an edge if they are both caused by at least one common gene. The disease gene network has its vertex set the set of all disease genes and two genes are joined by a link if they are associated with at least one common disorder. These networks are used to examine and understand human disease gene and phenotype associations. Measures of centrality are used to identify structurally important genes or proteins in interaction networks. Network motifs help in finding structural pattern along with unknown protein functions.

10 Conclusion

In this survey article a detailed account of various biological networks, basic graph theoretic concepts, various centrality measures which play a crucial role in the analysis of biological networks, biological databases which are essential for the construction of biological networks and software tools required for the network visualizing and analyzing have been presented. Analysis of the biological networks using graph theoretic tools lead to the identification of influential proteins or genes, which can be confirmed experimentally.

References

1. Amitai, G., Shemesh, A., Sitbon, E., Shklar, M., Netanel, D., Venger, I., Pietrokovski, S.: Network analysis of protein structures identifies functional residues. *J. Mol. Biol.* **344**(4), 1135–1146 (2004)
2. Azuaje, F.J.: Selecting biologically informative genes in coexpression networks with a centrality score. *Biol. Direct* **9**(12), 1–23 (2014)
3. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
4. Batagelj, V., Mrvar, A.: Pajek - program for large network analysis. *Connections* **21**, 47–57 (1998)
5. Baths, V., Roy, U., Singh, T.: Disruption of cell wall fatty acid biosynthesis in *Mycobacterium tuberculosis* using a graph theoretic approach. *Theor. Biol. Med. Model.* **8**(5), 1–13 (2011)
6. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**, 113–120 (1972)
7. Bullmore, E.D., Sporns, O.: Complex brain networks graph theoretical analysis of structural and functional systems. *Nature Rev. Neurosci.* **10**, 186–198 (2009)
8. Chartrand, G., Lesniak, L.: *Graphs & Digraphs*, 4th edn. Chapman and Hall, CRC, Boca Raton (2005)
9. Christensen, C., Gupta, A., Maranas, C.D., Albert, R.: Large scale inference and graph theoretical analysis of gene-regulatory networks in *B. Subtilis*. *Physica A* **373**, 796–810 (2007)
10. Dunne, J.A., Williams, R.J., Martinez, N.D.: Food-web structure and network theory: the role of connectance and size. *PNAS* **99**(20), 12917–12922 (2002)
11. Erdős, P., Rényi, A.: On the strength of connectedness of a random graph. *Acta Mathematica Academiae Scientiarum Hungarica* **12**, 261–267 (1964)
12. Estrada, E.: Virtual identification of essential protein within the protein interaction network of yeast. *Proteomics* **6**(1), 35–40 (2006)

13. Estrada, E., Rodríguez-Velázquez, J.A.: Subgraph centrality in complex networks. *Phys. Rev.* **71** (2005)
14. Fuller, T.F., Ghazalpour, A., Aten, J.E., Drake, T.A., Lusic, A.J., Horrath, S.: Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome* **18**(6–7), 463–472 (2007)
15. Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabási, A.: The human disease network. *Proc. National Acad. Sci.* **104**(21), 8685–8690 (2007)
16. Hu, Z., Mellor, J., Wu, J., Yamada, T., Holloway, D., DeLisi, C.: VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res.* **33**, 352–357 (2005)
17. Huson, D.H., Rupp, R., Scornavacca, C.: *Phylogenetic Networks*. Cambridge University Press, Cambridge (2010)
18. Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C.: STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, 412–416 (2009)
19. Jiang, W., Li, X., Rao, S., Wang, L., Du, L., Li, C., Wu, C., Wang, H., Wang, Y., Yang, B.: Constructing disease-specific gene networks using pair-wise relevance metric: application to colon cancer identifies interleukin 8, desmin and enolase 1 as the central elements. *BMC Syst. Biol.* **2**(1), 72 (2008)
20. Joshi-Tope, G., Gillespie, M., Vastrik, I., D’Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L., Lewis, S., Birney, E., Stein, L.: Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, 428–432 (2005)
21. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M.: KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, 355–360 (2010)
22. Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Liefink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R., Hermjakob, H.: IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.* **35**, 561–565 (2007)
23. Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Abdul Rahiman, B., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R., Pandey, A.: Human protein reference database. *Nucleic Acids Res.* **37**, 767–772 (2009)
24. Memisevic, V., Milenkovic, T., Przulj, N.: An integrative approach to modeling biological networks. *J. Integr. Bioinform.* **7**(3), 1–22 (2010)
25. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. *Science* **298**(5594), 824–827 (2002)
26. Mukhopadhyay, A., Maulik, U.: Network-based study reveals potential infection pathways of Hepatitis-C leading to various diseases. *PLOS One* **9**(4), 1–12 (2014)
27. Paz, A., Brownstein, Z., Ber, Y., Bialik, S., David, E., Sagir, D., Ulitsky, I., Elkon, R., Kimchi, A., Avraham, K.B., Shiloh, Y., Shamir, R.: SPIKE: a database of highly curated human signaling pathways. *Nucleic Acids Res.* **39**, 793–799 (2011)
28. Perkins, A.D., Langston, M.A.: Threshold selection in gene coexpression networks using spectral graph theory techniques. *BMC Bioinform.* **10**(54), 1–11 (2008)
29. Raza, K., Jaiswal, R.: Reconstruction and analysis of cancer-specific gene regulatory networks from gene expression profiles. *Int. J. Bioinform. Biosci.* **3**(2), 25–34 (2013)
30. Scardoni, G., Laudana, C.: Centralities based analysis of complex networks. In: Zhang, Y. (ed.) *New Frontiers in Graph Theory*, InTech (2012)
31. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Idekar, T.: Cytoscape: a software environment for integrated models of bio-molecular interaction networks. *Genome Res.* **13**(11), 2498–2504 (2003)

32. Sole, R.V., Montoya, J.M.: Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B* **268**, 2039–2045 (2001)
33. Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M.: BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, 535–539 (2006)
34. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998)
35. Xenarios, I., Rice, D.W., Salwinski, L., Baron, M.K., Marcotte, E.M., Eisenberg, D.: DIP: the database of interacting proteins. *Nucleic Acids Res.* **28**(1), 289–291 (2000)
36. Yue, H., Chunmei, L.: Study of Gene regulatory network based on graph. In: 4th International Conference on Biomedical Engineering and Informatics, pp. 2236–2240. IEEE (2011)
37. Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., Cesareni, G.: MINT: a Molecular INTeraction database. *FEBS Lett.* **513**(1), 135–140 (2002)