

Topological parameters, patterns, and motifs in biological networks

Arvind Kumar Yadav^{1,*}, Rohit Shukla^{1,*} and Tiratha Raj Singh²

¹Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology (JUIT), Solan, India, ²Centre of Excellence in Healthcare Technologies and Informatics (CHETI), Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology (JUIT), Solan, India

22.1 Introduction

Since the advancement of high-throughput biotechnology, a large amount of biological and biomedical data has been generated. These large datasets provide new research avenues. Due to the advancement of technologies, such as next-generation sequencing (NGS), the rapid decline of cost has provided the momentum for the generation of several extensive multigenomics biomedical datasets that describe multiple phenotypes. These datasets include exome and whole-genome sequencing, proteomics, transcriptomics, lipidomics, and microbiomics (Schadt & Björkegren, 2012; Shukla, Yadav, & Singh, 2021; Singh et al., 2018). In the field of computational biology, the networks have raised from a curious sideshow into a major approach of analysis in the last 20 years. In the process of this advancement, the network provides a conceptual framework for computing. With the help of network-based bioinformatics methods, networks are used to reveal new insights into large-scale biological data from multiple biological system types.

Biological systems are composed of different molecular entities, such as genes, proteins, and other biomolecules, and the interactions among these components. The interaction between diverse components at diverse levels can be expressed in the type of biological networks, such as protein–protein interactions (PPIs) and gene regulatory networks (Sonawane, Weiss, Glass, & Sharma, 2019). These biological networks help to depict how molecules interrelate to perform a variety of cellular functions. Different biological networks confine the complex interactions among biological components, such as genes, RNA molecules, proteins, genetic variants, and metabolites in the cells of the organism. These complex networks are composed of nodes and edges (Vidal, Cusick, & Barabási, 2011). The suitable depiction of the biological components is a graph, where the nodes symbolize the interacting molecules and edges symbolize the interaction between these molecules (Wang, Lü, & Yu, 2014). Therefore a discipline that combines graph theory, system biology, and statistical analysis to study the overall relationships between a variety of biological components is called network biology (Lindfors, 2011). Network biology allows the use of tools derived from graph theory to represent and analyze biological systems. The idea of network motifs was introduced by Milo, Alon colleagues in system biology in 2002 (Milo et al., 2002). The dynamic properties of networks depends on the motifs of the network (Stone, Simberloff, & Artzy-Randrup, 2019).

A motif is a common topological pattern in a given network. The motif is a basic element to understand the function of biological networks. To reveal the basic structure and design principles of a network, a motif can be used. Motifs are also often regarded as the fundamental building blocks of a network (Masoudi-Nejad, Schreiber, & Kashani, 2012). The motifs can be used to categorize networks into functional subunits. Although studying motifs are crucial to network analysis (Elhessa & Kahveci, 2016). In this chapter, we have provided an outline of the biological networks, network motifs, and topological parameters of biological networks.

**These authors contributed equally to this chapter.

22.2 Biological networks

Network-based methods are much accepted for analyzing “social networks,” where nodes correspond to people and the edges correspond to social connection types, such as friendship. In biology, this method is often used to express ecological relationships, such as the relationship between predators and prey (“food webs”) because there is a sufficient amount of data to correlate them. Recently molecular networks came into consideration because the huge datasets required to accumulate them are also recent. Such datasets typically come from “omics” techniques, which can retrieve huge molecular information on a large scale, some of which can be expressed as networks (Chagoyen, Ranea, & Pazos, 2019).

A network provides a mathematical construction to interpret the association between biological molecules (Vidal et al., 2011). Among them, nodes represent the biomolecules, and edges represent the relationship among them. Genes and their products interact with each other in the form of networks to perform their normal or dysfunctional roles (Parikshak, Gandal, & Geschwind, 2015). Due to the recent mounting data in biomolecular interactions, the research on biomolecular networks has been flourishing (Kang, Moore, Li, Sontag, & Bleris, 2015). Networks can be used to model many types of biological data (Fig. 22.1).

22.2.1 Construction of biological networks

Capturing and presenting scientific knowledge in a structured format is crucial for improving the understanding of biological mechanisms. Known interactions identified from earlier experiments are often used to construct biological networks. Diverse information can be represented in the form of a network to model the cell. The network is composed of nodes representing proteins and genes. The ranking of nodes is defined by the topological features that help to identify biologically important proteins or genes (Winterbach, Mieghem, Reinders, Wang, & Ridder, 2013). The most common types of biological networks are protein and gene networks, as described below:

22.2.1.1 Network construction based on protein interaction

The PPI network represents the physical relationship of proteins. The study of PPI is crucial for understanding cell physiology under normal or disease conditions. In this network, nodes contain proteins connected by nondirectional edges. The connection of PPIs is specific and occurs among defined binding regions with specific biological significance. PPI can help to assign putative roles for uncharacterized proteins. It also helps to characterize the relationship between proteins (such as proteasome) that form multimolecular complexes. Various databases provide the physical PPI of human proteins (Table 22.1). The physical PPIs have been used to construct large-scale reference biological networks. The resources for tissue-specific PPI networks are also available (Table 22.1). These resources offer improved biological enrichment and are more helpful for the study on disease progression (Kitsak et al., 2016; Magger, Waldman, Ruppim, & Sharan, 2012).

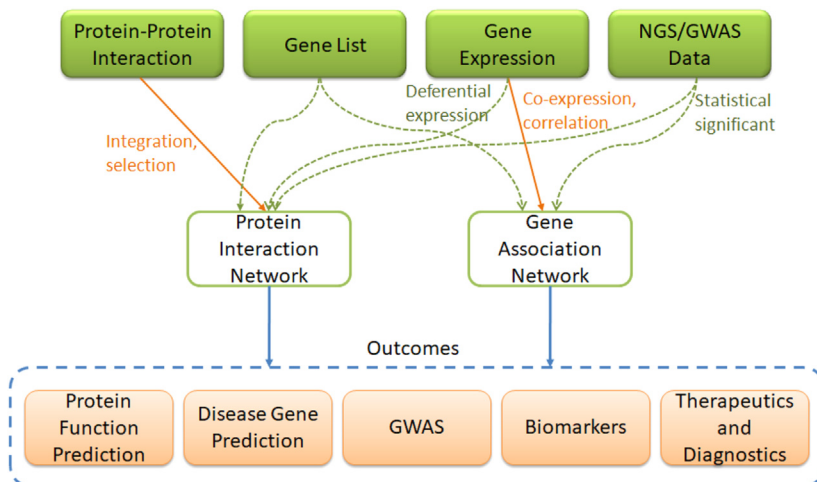


FIGURE 22.1 Overview of general biological networks and their outcomes. *Red* lines represent the biological network construction and *blue* lines represent the mapping of data onto the network.

TABLE 22.1 Resources of protein–protein interaction data for humans.

Databases for global PPI					
Name of database	Interaction type	Number of interactions	Number of proteins	Link	
STRING	Physical, association, text mining	3,123,056,667	24,584,628	https://string-db.org/	
BioGRID	Physical, genetic	514,501	26,126	https://thebiogrid.org/	
IntAct	Physical	1,118,895	118,345	https://www.ebi.ac.uk/intact/	
HIPPE	Physical	273,900	17,000	http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/	
HPRD	Physical	41,327	30,047	https://www.hprd.org/	
Databases for cell/tissue-specific PPI					
Name of database	Type of interaction	Number of tissues	Number of interactions	Number of genes/proteins	Link
GIANT	Physical, coexpression	144	Na	na	http://giant.princeton.edu/
IID	Physical, predicted	26	975,877	19,250	http://iid.ophid.utoronto.ca/
TISPIN	Physical	53	128,579	13,123	http://bidd2.nus.edu.sg/%20TISPIN/
TissueNet	Physical	47	243,706	17,283	http://netbio.bgu.ac.il/tissuenet/

PPI, protein–protein interaction.

22.2.1.2 Based network construction based on gene association/correlation

Genes can also be associated with each other based on the correlation or statistical significance of their expression levels. A gene network can be constructed based on the gene expression dataset generated by the user's experiment, which implies the information of gene regulation in particular conditions (such as tissue specific, disease specific, or drug specific). Lots of statistical methods have been developed to understand networks constructed from gene expression datasets. We briefly introduce here about four methods of reconstructing gene–gene correlation/association network: Gaussian graphical model (Ni, Müller, Wei, & Ji, 2018), Bayesian network (Needham, Bradford, Bulpitt, & Westhead, 2007), correlation network (Langfelder & Horvath, 2008), and information theory (Mousavian, Kavousi, & Masoudi-Nejad, 2016). Table 22.2 sums up the methods introduced for gene-based networks, as well as implementation languages and URLs for access.

22.3 Network motifs and patterns

Since the network quantity enlarges more frequent and statistically important related patterns will emerge in the actual network. Such type of pattern is described as the network motif. These motifs appear more frequently in actual networks in comparison to random networks (Yu et al., 2019). It has been found that such motifs are associated with desired (or undesired) biological functions (or dysfunction). It is now generally understood that the motifs constitute the essential building blocks of the cellular network. The design pattern study of regulatory circuits of these networks may help to understand the whole system. Numerous motifs have been revealed to be functionally significant in biological networks (Fig. 22.2). Certain motifs are mainly vital because they directly affect the overall dynamics of the system, such examples include feedforward loops, feedback loops, bifans, and other cycle types. Motifs also play an important role in network development and optimization (Yu et al., 2019). Therefore mining of motifs in the network helps to understand network patterns in depth.

22.3.1 Motif discovery and counting

The scale of the biological network is very large because it contains thousands of vertices. However, identification of motif is a lengthy task, particularly in large networks (such as PPIs networks, gene regulatory networks, and metabolic networks)

TABLE 22.2 Gene regulatory network constriction methods.

Approach	Description	Method	Implementation	Link
Bayesian network	It is a probabilistic framework for representing a direct acyclic graph structure. To obtain the network structure, this method selects the graph structure	B-Course (Myllymäki, Silander, Tirri, & Uronen, 2002)	Java	http://b-course.hiit.fi/obc/
		BNT (Murphy, 2001)	Matlab	http://code.google.com/p/bnt/
		Werhli's BN (Werhli, Grzegorzcyk, & Husmeier, 2006)	Matlab	http://www.bioss.ac.uk/people/adriano/comparison/
Correlation network	This method represents pairwise correlations between two nodes into edges	WGCNA (Langfelder & Horvath, 2008)	C, R	http://cran.r-project.org/web/packages/WGCNA
Gaussian graphical model	It is a probabilistic framework that considers undirected graph structure.	SPACE (J. Peng, Wang, Zhou, & Zhu, 2009)	C, R	http://cran.r-project.org/web/packages/space
		Graphical Lasso (Banerjee, Ghaoui, & d'Aspremont, 2008; d'Aspremont, Banerjee, & Ghaoui, 2006)	Fortran, R	http://cran.r-project.org/web/packages/glasso
		CLIME (Cai, Liu, & Luo, 2011)	R	http://cran.r-project.org/web/packages/clime
		GeneNet (Schäfer & Strimmer, 2005)	R	http://cran.r-project.org/web/packages/GeneNet
Information theory	To measure the dependencies between variables, this method constructs the GRN by using information-theoretic scores, for example, mutual information.	Relevance network (Butte & Kohane, 2000)	Java	http://www.newatlantictech.com/products.html
		ARACNE (Basso et al., 2005; Margolin et al., 2006)	C++, Java	http://wiki.c2b2.columbia.edu/califanolab/index.php/
		CLR (Faith et al., 2007)	C, Matlab	http://m3d.mssm.edu/network_inference.html
		GTRNetwork (Fu, Jarboe, & Dickerson, 2011)	Matlab	http://www.biomedcentral.com/1471-2105/12/233
		NARROMI (Zhang et al., 2013)	Matlab	https://pubmed.ncbi.nlm.nih.gov/23080116/

GRN, gene regulatory network.

(Kashani et al., 2009). A range of algorithms has been developed to reduce the computational complexity for effective motif discovery (Khakabimamaghani, Sharafuddin, Dichter, Koch, & Masoudi-Nejad, 2013; Peng et al., 2018). These algorithms are divided into three categories, such as serial algorithms, parallel algorithms, and sampling algorithms (Table 22.3). In many cases, motifs must be discovered first, and then they will be counted. In the motif properties analysis, the frequency of the motif is very important. To calculate the correct frequency of network motif, we must first calculate motifs number in the network (Yu et al., 2019). Thus in the following two tables (Tables 22.3 and 22.4), we selectively introduce the motif discovery algorithms and motif counting algorithms according to the application scope of the motif.

22.4 Analysis of biological network

The history of biological network analysis originated from the tools and concepts for the analysis of social networks in social sciences. On behalf of network complexity, the network topology can be analyzed. Graph theory is used to

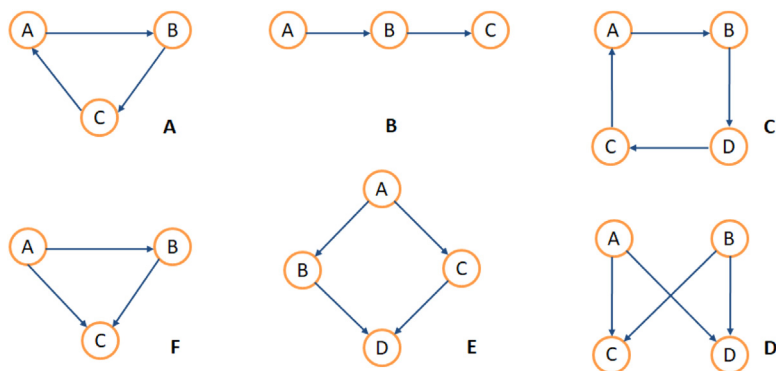


FIGURE 22.2 Functionally important motifs found in biological networks: (A) feed-forward loop, (B) three chain, (C) four node feedback, (D) bifan, (E) biparallel, and (F) three-node feedback.

TABLE 22.3 Algorithms for motif discovery with a short description.

Algorithm name	Description
Serial algorithm	
MotifCut (Fratkin, Naughton, Brutlag, & Batzoglou, 2006)	It is based on the graph theory. The purpose of MotifCut is to perform high-order motif detection. The input sequence of this algorithm is divided into several subcolumns, which are used as vertices to create the graph. Edges are weighted by using a function.
MotifNet (Monti, Otness, & Bronstein, 2018)	This is an open-source website based on FANMOD. It takes a network diagram with node numbers or edge numbers as input, and a user can find motifs that consist of up to eight nodes. It generates a graphical output of predicted motifs.
G-trie (Ribeiro & Silva, 2014)	This is the most efficient algorithm for discovering motifs. G-trie structure uses the common prefix of the subgraph. All trie nodes to be in contact with a subgraph node. It can be used for higher-order motif discovery.
Parallel algorithms	
Parallel G-trie Algorithm (Ribeiro, Silva, & Lopes, 2012)	It can improve calculation efficiency. Among them, every node in the recursive search tree is considered as a work unit, and every unit is self-regulating.
Iterative MapReduce Algorithm (Diwakar Reddy & Geetha Reddy, 2018)	It used three independent iterations of MapReduce, such as enumeration of subgraphs, calculation of subgraph labels, and result aggregation.
GPU-based Parallel Motif Discovery (Lin, Xiao, Xie, & Li, 2015)	It uses matching subgraphs for motif discovery. It reduces the calculation time.
Sampling algorithms	
Efficient Sampling Algorithm (Kashtan, Itzkovitz, Milo, & Alon, 2004)	This is an edge sampling-based method.
G-Trie Sampling Algorithm (Luo, Ding, Liang, & Tu, 2018)	It is used as a recurrent search process. A recurrent tree can be used to represent the entire search process. The idea behind this algorithm is to get back the recurrent tree branches through a certain probability.

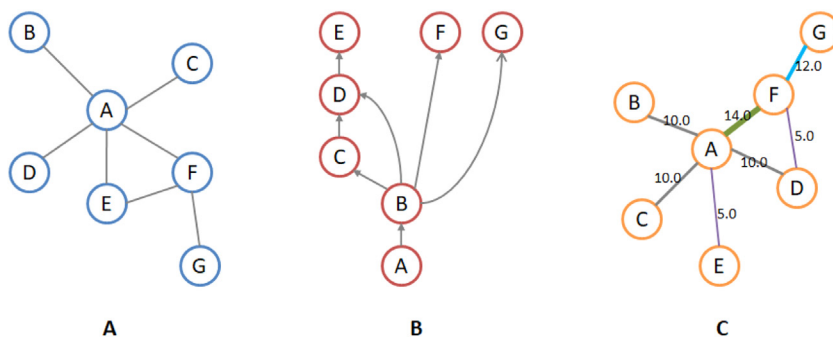
describe the properties of topology in biological networks. The biological network has a unique topological structure, which affects the dynamic behavior of the system.

22.4.1 Graph theory

A graph is a set of abstract concepts and methods that can be used to visualize and analyze networks. Swiss mathematician Leonard Euler described the graph theory and topology and applied them to the problem of the seven bridges of Konigsberg (Ahmed, 2019). Generally, networks or graphs are used to form associations among entities or objects. In a usual representation, the graph consists of node sets that connected with edges. In biology, networks are made up of

TABLE 22.4 A method for motif counting and frequency estimation.

Method name	Description
Accurate counting method	
Flexible Pattern Finder (FPF) (Schreiber & Schwöbbermeyer, 2005)	This algorithm is supported tree-based pattern, It is a pattern algorithm based on a tree, where the tree is composed of different nodes on behalf of patterns. The child node graph of the tree is achieved by adding edges with a graph signify by the current node.
Enumerate Subgraphs (ESU) (Wernicke, 2006)	It is based on Mfinder and enumerates algorithms. The ESU algorithm can count both high- and low-order motifs.
Grochow-Kellis (Grochow & Kellis, 2007)	It is similar to enumerate a subgraph algorithm based on a specified motif that is represented as a query graph. It is a motif-centric approach that performs all feasible mappings in the entire network.
RAGE (Marcus & Shavitt, 2010)	This algorithm used three-order and four-order structural feature motifs to calculate the low-order subgraphs in a large network very proficiently.
Frequency estimation method	
Mfinder (Kashtan et al., 2004)	This is an edge random sampling-based method that aim is to estimate the big subgraphs concentration in a huge network. It extensively reduces the calculation time for motif counting.
Rand-ESU (Wernicke, 2006)	It is ESU-based algorithm to investigate only the part of the ESU tree. It can be used for both high- and low-order motif counting.
MODA (Omid, Schreiber, & Masoudi-Nejad, 2009)	It used “extension tree,” in combination with frequency characteristics of the FPF algorithm and method of motif growth. It can identify high-order motifs with more than eight nodes, effectively.

**FIGURE 22.3** A representation of (A) directed; (B) undirected; and (C) weighted graphs found in the network.

molecules, such as DNA, RNA, proteins, and metabolites, and graphs can be used to confine the connections among these molecules (Huber, Carey, Long, Falcon, & Gentleman, 2007). Thus it is crucial to know the different types of networks that can be capable to communicate and visualize such types of interactions. There are various types of graphs but directed, undirected, and weighted graphs are the main type of graphs in the network (Koutrouli, Karatzas, Paez-Espino, & Pavlopoulos, 2020).

The edges of *undirected graphs* have no direction. In these graphs, the edges represent a *two-way* relationship, because each edge can go across in two directions (Fig. 22.3A). The edges of *directed graphs* have direction. The edges represent a *one-way* relationship, and each edge can simply be represented in one direction (Fig. 22.3B). This type of connection is usually found in gene regulation or metabolism and signal transduction networks. This type of graph usually appears in PPI networks. In a *weighted graph*, directed or undirected edges have certain weights or quantitative values (Fig. 22.3C). This quantitative value is important to study the relationship between nodes. In biological networks, this value indicates how close two genes are in terms of sequence similarity. Edges can also be weighted by other topological parameters, such as centrality and betweenness (Pavlopoulos et al., 2011). A weighted graph is currently the most widely used network in the entire bioinformatics field.

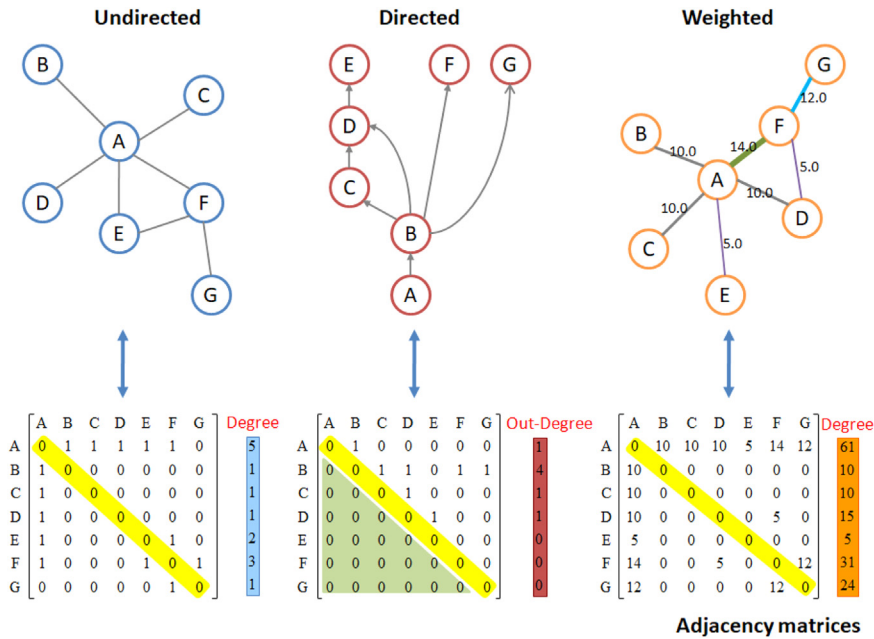


FIGURE 22.4 Different types of graphs and their adjacency matrices.

Any network can be mathematically represented and stored in a form of a matrix called “adjacency matrix” (Fig. 22.4). In these matrices, columns and rows are allocated to the nodes of the network, and the occurrence of the edge is represented numerically. By using the representation of matrix in the network, we can calculate the properties of the network (Pavlopoulos et al., 2011).

22.4.2 Adjacency matrices

In graph theory, the adjacency matrix is a square matrix whose size is $N \times N$ (N = number of vertices present in the network) used to characterize a graph. In a simple graph, the diagonal of the adjacency matrix is zero (Fig. 22.4). In both undirected and unweighted networks, the edges are represented by a symmetric matrix. The matrix only contains a value of 1 indicating that there is a connection and 0 indicating that there is no connection (Yue et al., 2020). In the case of directed and weighted networks, different values in the matrix are used to represent this more complex relationship.

Graphs have some important properties which are very helpful when unscrambling the information that they contain. The purpose of network analysis is to deal with the network complexity to extract important information, and if you examine each component individually, there will be no such information. By representing the complexity of the biological regulatory system as a network, the topology of the network can be analyzed. Network properties, especially *topological properties*, can help to identify significant substructures within the network (Ma’ayan, 2011). The network topology consists of information about some common and specific node properties, edges properties, and the properties of the whole network (global topological properties), and the information of the modules within the network.

22.5 Topological parameters

The topological parameters of the network describe the metrics and quantitative patterns of the edges and nodes (Zhao & Liu, 2019). For example, the node degree refers to the number of edges of node occurrence. The properties of topology (degree and its distribution) help to provide functional descriptions of networks in detail. Therefore the localization and organizational structure of the network are reflecting in its parameters of topology. Topological parameters can also characterize the patterns of genes in a network system (Zhao & Liu, 2019). Therefore in this section, we have focused on the topological parameters of the network and its properties.

Properties of nodes include the degree of connectivity, that is, the number of links per node; the node betweenness centrality (BC), that is, the number of shortest paths, pass from first to the last node between all shortest paths among all pairs of possible nodes; the closeness centrality that characterizes the average shortest path starting from one node to

all other nodes; and eigenvector centrality, which is a complex centrality measurement used to evaluate the closeness to highly connected nodes. The properties of the edge include the BC of the edge, that is, the number of shortest paths through an edge between possible shortest paths among all pairs of nodes (Ma'ayan, 2011). To explain the location properties of network structure, several topological parameters are defined that help to quantify its centrality or functionality (Pradhan, C.U., & Jalan, 2020). Various sets of topological properties have been effectively used for the study of biological networks. Various local and global properties of network topology that are commonly used to reveal biological significance are tabulated (Table 22.5). Some of the most commonly used network topological properties are described here (Ramadan, Alinsaif, & Hassan, 2016).

22.5.1 Node degree

The node degree in a network is described as the summation of the entire connected edges to it (Sporns, 2013). The highly connected node is called hubs. If the degree of a node is n , it means n neighbor nodes connected to the node. Generally, the probability distribution of the degree of all nodes in the network is called degree distribution. In a complex biological network, a power-law distribution is proven (Chattoadhyay, Das, & Ghosh, 2020; Rivkind, Schreier, Brenner, & Barak, 2020).

22.5.2 The average of shortest path length

The shortest path in an unweighted network is used to model the information flows. Between the two nodes, i and j shortest path means the path that has the smallest edges count between nodes i and j . A distance of connecting two

TABLE 22.5 Local and global topological properties are used in biological studies with their graph theory explanation.

Level	Topological property	Explanation of graph theory	Biological implementation
Global	Connectivity centralization	Distinguish between well connected networks or distributed networks.	It is used to study the structural differences present in metabolic Networks.
Global	Heterogeneity	It measures the changes in connection distribution.	Reflects the trend that the network has a hub gene.
Global	Global efficiency	Indicates the efficiency of information exchange on the entire network or defined subnetworks.	It is used to describe the neural connectivity of the brain.
Local	Clustering coefficient	Measures the tendency of a group of nodes and neighboring nodes.	It is used to analyze the organizational characteristics of the human protein network and verify the association of drugs with accessible proteins present in the network of drug–target.
Local	Closeness centrality	Measures the speed at which information can extend from a given particular gene to further accessible genes.	The central role is used to prioritize the disease of the candidate genes. The identification of important genes in the discovery of the drug, and clarify the relationship between disease–disease associations.
Local	Betweenness centrality	It designates the number of times. The given node has worked as a link on the smallest path among any other two nodes.	These centralities have been utilized to prioritize the candidate disease genes. Identification of significant genes in the process of drug discovery, and provide insight into the disease–disease relationships.
Local	PageRank centrality	It is used to measure the significance of the node by taking into consideration both the number of node connections and the magnitude of the connected nodes.	This centrality is used for the identification of protein targets present in metabolic networks. The prediction of potential marker genes helps to identify the patients suffering from pancreatic cancer.

nodes (d_{ij}) represents the smallest path between them. In the whole network, the average shortest path represents the path length average of every feasible pair of nodes (Mao & Zhang, 2013; Zhao & Liu, 2019), that is,

$$L = \frac{2}{N(N+1)} \sum_{i \leq j, i, j \in G} d_{ij}$$

where N is the total number of nodes present in network G and i and j represent the nodes of network G . The self node distance is defined to zero.

22.5.3 Clustering coefficient

The clustering coefficient (CC) of a network reveals the aggregation properties behind the nodes (Zhao & Liu, 2019) and refers to the propensity of these nodes to cluster together. The CC essentially describes the average ratio of the real edge of a node in the large network to the entire promising edges,

$$CC = \frac{n}{\binom{k}{2}} = \frac{2n}{k(k-1)}$$

where n signifies the edges connected number with node and their first-order neighbors. The $\binom{k}{2}$ denotes the edges number of the neighbors. The purpose of calculating CC is to gain its values. The node neighbor subgraph algorithm is used to pull off it from the network adjacency matrix. All neighbors of a position element are found in the adjacency matrix and got the edges number. Due to adjacency matrix symmetry, the neighbor number is twice the number of edges in the case of an undirected graph. Thus the real value should be divided by 2 (Zhao & Liu, 2019).

22.5.4 Betweenness centrality

BC is a metric of centrality that evaluates the significance of each node in the network based on the shortest (Zhao & Liu, 2019). Essentially, the BC value of a node represents the ratio of all straight paths among other nodes.

$$BC(v) = \sum_{i \neq v \neq j} \frac{\partial_{ij}(v)}{\partial_{ij}}$$

∂_{ij} denote the number of the shortest paths between node i to node j . $\partial_{ij}(v)$ is the shortest paths from node i to node j , through node v .

22.5.5 Statistical comparison

Statistical comparison of biological networks provides important insight to understand the biological mechanism of disease genes. For the analysis of statistical comparison, topological parameters need to be characterized. Various methods have been used for network topological information (Ji, Yuan, Zhang, & Xue, 2016). Most of the methods focus on the differences of the network topology but unable to report the chances of vertices. Such methods are usually used non-parametric permutation procedures and it takes much time in the case of big data. Weighted statistical tests for differences in network topology can capture the chances of nodes and edges. The network comparison methods can be divided into two categories. One is an alignment-based method, whose purpose is to find a mapping between the nodes of two or more networks. The other is the alignment-free method, which aims to quantify the overall similarity between networks. Currently, the Graphlet correlation distance method is the best alignment-free method (Yuan et al., 2016). Many disciplines are in great need of statistical comparison between biological networks.

22.6 Biological significance of network motifs

Network biology has promoted the advancement of various fields of biomedical. Its simple and powerful concept allows us to understand the relationship between genes and proteins (Pace-Schott et al., 2019). Biological networks offer a starting point for many analyses and help to advance our understating of biological systems. It can be generated by complex experimental data that observe the interaction between proteins or the relationship along with different genes (Camacho, Collins, Powers, Costello, & Collins, 2018; Gupta, Singh, Shukla, & Misra, 2013). The network represents various biological

systems. A typical example of a biological network is the network of metabolic pathways that represent the metabolic substrates and products. Many authors have studied various biological networks (Bansal, Singh, & Chauhan, 2017; Kumar & Singh, 2017; Rezola et al., 2015; Souza, Alseekh, Brotman, & Fernie, 2020). A biological network illustrates the functional interaction between genes or proteins in organisms or between organisms. It can include an internal communication flow that does not necessarily need physical binding (Mulder, Akinola, Mazandu, & Rapanoel, 2014). The activities of molecular interaction networks appear by the interaction of various small subgraphs or motifs. Motifs can be used to describe the networks more globally. It is found that the global motif signature is unique to different types of networks. The motifs can also be considered an important feature, which can sometimes be explained biologically.

Network motifs help to describe and reveal the local properties of the network (Sehgal & Singh, 2017). The network motif is the basic building block of transcription networks. These motifs are also used to study the integrated transcriptional regulation network and PPIs. The recognition of network motifs has led to many important applications, for example, understanding the modularity and large-scale structure of biological networks, network classification into superfamily, and annotation of protein functions (Pratap, Taliyan, & Singh, 2014; Takes, Kusters, Witte, & Heemskerk, 2018). Biological networks greatly help to improve the understanding of various diseases, such as cancer (Sehgal, Gupta, Moussa, & Singh, 2015) and Alzheimer's disease (Panigrahi & Singh, 2013). It also helps to identify potential drug targets and biomarkers for various diseases by highlighting important genes or proteins (Bansal, Srivastava, & Singh, 2018; Jia, Nie, Li, Liang, & Zhang, 2016; Sehgal & Singh, 2012). While biological networks have many uses and applications, the field is still mounting.

22.7 Applications of network biology

- Provide powerful tools for the study of biological processes and complex diseases (Gupta, Singh, Rath, & Misra, 2012).
- Help to identify and prioritize candidate genes that cause disease.
- Identify subnetworks related to diseases and logical perturbation of diseases.
- Capture curative responses to make it possible to the identification of targets and drug development.
- Identification of influential nodes in the network can help to understand the information related to diseases.
- Identify the interaction between drugs and targets, and assist in the drug development process.
- Accelerate the development of personalized medicine by using the integration of multiomics data and network-based methods.

22.8 Limitations and challenges

The complexity of biological networks increases with the accumulation of data. Therefore improved technology to integrate data from different sources and to visualize is essential for understanding the function of complex networks (Pavlopoulos et al., 2011). More high-quality biological data are needed to extract more reliable biological insights. However, data quality is a key limitation in computational network biology applications. An in-depth understanding of the biological network structure should be introduced. The methods in computational network biology need further development (Bebek, Koyutürk, Price, & Chance, 2012).

There are some key challenges in network biology:

- shared multicellular data standards;
- shared multicellular observation representation;
- standard support for computational tools;
- sharing tools for configuring models and explore data;
- high-quality multiscale benchmark dataset;
- community-curated public database;
- quality and curation standards; and
- link data to the model.

22.9 Conclusion

With the development of experimental technology, network-based methods can be improved and extended widely. Biological networks provide a deep understanding of complex biological systems and are described as network graphs.

Graphs are used as a convenient mathematical representation of the interactive network. The network graph allows us to determine the topological properties of nodes. These properties are node degree, betweenness, centrality, etc., which tell us how the node is associated with other nodes in the network. The topology of networks has been gaining much attention from researchers in the last few decades because the network topology in molecular interaction is associated with biological functions. There is too much to do in the future development of network biology. Therefore the study of network biology is still in its infancy.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Ahmed, H. (2019). Graph routing problem using Euler's theorem and its applications. *Engineering Mathematics*, 3(1), 1. Available from <https://doi.org/10.11648/j.engmath.20190301.11>.
- Banerjee, O., Ghaoui, L. E., & d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine Learning Research*, 9(15), 485–516.
- Bansal, A., Singh, T. R., & Chauhan, R. S. (2017). A novel miRNA analysis framework to analyze differential biological networks. *Scientific Reports*, 7(1), 14604. Available from <https://doi.org/10.1038/s41598-017-14973-x>.
- Bansal, A., Srivastava, P. A., & Singh, T. R. (2018). An integrative approach to develop computational pipeline for drug-target interaction network analysis. *Scientific Reports*, 8(1), 10238. Available from <https://doi.org/10.1038/s41598-018-28577-6>.
- Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., & Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nature Genetics*, 37(4), 382–390. Available from <https://doi.org/10.1038/ng1532>.
- Bebek, G., Koyutürk, M., Price, N. D., & Chance, M. R. (2012). Network biology methods integrating biological data for translational science. *Briefings in Bioinformatics*, 13(4), 446–459. Available from <https://doi.org/10.1093/bib/bbr075>.
- Butte, A. J., & Kohane, I. S. (2000). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Pacific Symposium on Biocomputing*, 418–429. Available from https://doi.org/10.1142/9789814447331_0040.
- Cai, T., Liu, W., & Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494), 594–607.
- Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., & Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell*, 173(7), 1581–1592. Available from <https://doi.org/10.1016/j.cell.2018.05.015>.
- Chagoyen, M., Ranea, J. A. G., & Pazos, F. (2019). Applications of molecular networks in biomedicine. *Biology Methods and Protocols*, 4(1). Available from <https://doi.org/10.1093/biomethods/bpz012>.
- Chattopadhyay, S., Das, A. K., & Ghosh, K. (2020). Finding patterns in the degree distribution of real-world complex networks: Going beyond power law. *Pattern Analysis and Applications*, 23(2), 913–932. Available from <https://doi.org/10.1007/s10044-019-00820-4>.
- d'Aspremont, A., Banerjee, O., & Ghaoui, L. E. (2006). First-order methods for sparse covariance selection. ArXiv:Math/0609812. <http://arxiv.org/abs/math/0609812>
- Diwakar Reddy, A., & Geetha Reddy, J. (2018). Algorithms for iterative applications in MapReduce framework. In M. S. Reddy, K. Viswanath, & S. P. K.M (Eds.), *International proceedings on advances in soft computing, intelligent systems and applications* (pp. 51–61). Springer, <https://doi.org/10.1007/978-981-10-5272-9-5>.
- Elhesha, R., & Kahveci, T. (2016). Identification of large disjoint motifs in biological networks. *BMC Bioinformatics*, 17(1), 408. Available from <https://doi.org/10.1186/s12859-016-1271-7>.
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., . . . Gardner, T. S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 5(1), e8. Available from <https://doi.org/10.1371/journal.pbio.0050008>.
- Fratkin, E., Naughton, B. T., Brutlag, D. L., & Batzoglou, S. (2006). MotifCut: Regulatory motifs finding with maximum density subgraphs. *Bioinformatics (Oxford, England)*, 22(14), e150–e157. Available from <https://doi.org/10.1093/bioinformatics/btl243>.
- Fu, Y., Jarboe, L. R., & Dickerson, J. A. (2011). Reconstructing genome-wide regulatory network of E. coli using transcriptome data and predicted transcription factor activities. *BMC Bioinformatics*, 12, 233. Available from <https://doi.org/10.1186/1471-2105-12-233>.
- Grochow, J. A., & Kellis, M. (2007). Network Motif discovery using subgraph enumeration and symmetry-breaking. In T. Speed, & H. Huang (Eds.), *Research in Computational Molecular Biology* (4453, pp. 92–106). Springer Berlin Heidelberg. Available from <https://doi.org/10.1007/978-3-540-71681-5-7>.
- Gupta, M. K., Singh, D. B., Rath, S. K., & Misra, K. (2012). Metabolic modeling and simulation analysis of thyroid disorder pathway. *Journal of Computer Science and Systems Biology*, 5, 052–061. Available from <https://doi.org/10.4172/jcsb.1000090>.
- Gupta, M. K., Singh, D. B., Shukla, R., & Misra, K. (2013). A comprehensive metabolic modeling of thyroid pathway in relation to thyroid pathophysiology and therapeutics. *OMICS*, 17(11), 584–593. Available from <https://doi.org/10.1089/omi.2013.0007>.

- Huber, W., Carey, V. J., Long, L., Falcon, S., & Gentleman, R. (2007). Graphs in molecular biology. *BMC Bioinformatics*, 8(6), S8. Available from <https://doi.org/10.1186/1471-2105-8-S6-S8>.
- Ji, J., Yuan, Z., Zhang, X., & Xue, F. (2016). A powerful score-based statistical test for group difference in weighted biological networks. *BMC Bioinformatics*, 17. Available from <https://doi.org/10.1186/s12859-016-0916-x>.
- Jia, Y., Nie, K., Li, J., Liang, X., & Zhang, X. (2016). Identification of therapeutic targets for Alzheimer's disease via differentially expressed gene and weighted gene co-expression network analyses. *Molecular Medicine Reports*, 14(5), 4844–4848. Available from <https://doi.org/10.3892/mmr.2016.5828>.
- Kang, T., Moore, R., Li, Y., Sontag, E., & Bleris, L. (2015). Discriminating direct and indirect connectivities in biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 112(41), 12893–12898. Available from <https://doi.org/10.1073/pnas.1507168112>.
- Kashani, Z. R. M., Ahrabian, H., Elahi, E., Nowzari-Dalini, A., Ansari, E. S., Asadi, S., . . . Masoudi-Nejad, A. (2009). Kavosh: A new algorithm for finding network motifs. *BMC Bioinformatics*, 10, 318. Available from <https://doi.org/10.1186/1471-2105-10-318>.
- Kashtan, N., Itzkovitz, S., Milo, R., & Alon, U. (2004). Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics (Oxford, England)*, 20(11), 1746–1758. Available from <https://doi.org/10.1093/bioinformatics/bth163>.
- Khakabimamaghani, S., Sharafuddin, I., Dichter, N., Koch, I., & Masoudi-Nejad, A. (2013). QuateXelero: An accelerated exact network motif detection algorithm. *PLoS One*, 8(7). Available from <https://doi.org/10.1371/journal.pone.0068073>.
- Kitsak, M., Sharma, A., Menche, J., Guney, E., Ghiassian, S. D., Loscalzo, J., & Barabási, A.-L. (2016). Tissue specificity of human disease module. *Scientific Reports*, 6(1), 35241. Available from <https://doi.org/10.1038/srep35241>.
- Koutrouli, M., Karatzas, E., Paez-Espino, D., & Pavlopoulos, G. A. (2020). A guide to conquer the biological network era using graph theory. *Frontiers in Bioengineering and Biotechnology*, 8. Available from <https://doi.org/10.3389/fbioe.2020.00034>.
- Kumar, A., & Singh, T. R. (2017). Analysis for biological network properties of Alzheimer's disease associated gene set by enrichment and topological examinations. *International Journal of Bioinformatics Research and Applications*, 13(3), 214. Available from <https://doi.org/10.1504/IJBRA.2017.085856>.
- Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559. Available from <https://doi.org/10.1186/1471-2105-9-559>.
- Lin, W., Xiao, X., Xie, X., & Li, X.-L. (2015). Network motif discovery: A GPU approach. In *2015 IEEE 31st international conference on data engineering* (pp. 831–842). <https://doi.org/10.1109/ICDE.2015.7113337>.
- Lindfors, E. (2011). *Network biology: Applications in medicine and biotechnology* (Dissertation). <https://cris.vtt.fi/en/publications/network-biology-applications-in-medicine-and-biotechnology-disser>.
- Luo, J., Ding, L., Liang, C., & Tu, N. H. (2018). An efficient network motif discovery approach for co-regulatory networks. *IEEE Access*, 6, 14151–14158. Available from <https://doi.org/10.1109/ACCESS.2018.2796565>.
- Ma'ayan, A. (2011). Introduction to network analysis in systems biology. *Science Signaling*, 4(190), tr5. Available from <https://doi.org/10.1126/scisignal.2001965>.
- Magger, O., Waldman, Y. Y., Ruppín, E., & Sharan, R. (2012). Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Computational Biology*, 8(9), e1002690. Available from <https://doi.org/10.1371/journal.pcbi.1002690>.
- Mao, G., & Zhang, N. (2013). Analysis of average shortest-path length of scale-free network. *Journal of Applied Mathematics*. Available from <https://doi.org/10.1155/2013/865643>.
- Marcus, D., & Shavitt, Y. (2010). Efficient counting of network motifs. In *2010 IEEE 30th international conference on distributed computing systems workshops* (pp. 92–98). <https://doi.org/10.1109/ICDCSW.2010.41>.
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., & Califano, A. (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(Suppl 1), S7. Available from <https://doi.org/10.1186/1471-2105-7-S1-S7>.
- Masoudi-Nejad, A., Schreiber, F., & Kashani, Z. R. M. (2012). Building blocks of biological networks: A review on major network motif discovery algorithms. *IET Systems Biology*, 6(5), 164–174. Available from <https://doi.org/10.1049/iet-syb.2011.0011>.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science (New York, N.Y.)*, 298(5594), 824–827. Available from <https://doi.org/10.1126/science.298.5594.824>.
- Monti, F., Otness, K., & Bronstein, M. M. (2018). MotifNet: A motif-based graph convolutional network for directed graphs. ArXiv:1802.01572 [Cs]. <http://arxiv.org/abs/1802.01572>.
- Mousavian, Z., Kavousi, K., & Masoudi-Nejad, A. (2016). Information theory in systems biology. Part I: Gene regulatory and metabolic networks. *Seminars in Cell and Developmental Biology*, 51, 3–13. Available from <https://doi.org/10.1016/j.semcdb.2015.12.007>.
- Mulder, N. J., Akinola, R. O., Mazandu, G. K., & Rapanoel, H. (2014). Using biological networks to improve our understanding of infectious diseases. *Computational and Structural Biotechnology Journal*, 11(18), 1–10. Available from <https://doi.org/10.1016/j.csbj.2014.08.006>.
- Murphy, K. P. (2001). *The Bayes net toolbox for MATLAB*. *Computing Science and Statistics*, 33, 2001.
- Myllymäki, P., Silander, T., Tirri, H., & Uronen, P. (2002). B-course: A web-based tool for bayesian and causal data analysis. *International Journal on Artificial Intelligence Tools*, 11(03), 369–387. Available from <https://doi.org/10.1142/S0218213002000940>.
- Needham, C. J., Bradford, J. R., Bulpitt, A. J., & Westhead, D. R. (2007). A primer on learning in Bayesian networks for computational biology. *PLoS Computational Biology*, 3(8), e129. Available from <https://doi.org/10.1371/journal.pcbi.0030129>.
- Ni, Y., Müller, P., Wei, L., & Ji, Y. (2018). Bayesian graphical models for computational network biology. *BMC Bioinformatics*, 19(3), 63. Available from <https://doi.org/10.1186/s12859-018-2063-z>.

- Omidi, S., Schreiber, F., & Masoudi-Nejad, A. (2009). MODA: An efficient algorithm for network motif discovery in biological networks. *Genes and Genetic Systems*, 84(5), 385–395. Available from <https://doi.org/10.1266/ggs.84.385>.
- Pace-Schott, E. F., Amole, M. C., Aue, T., Balconi, M., Bylsma, L. M., Critchley, H., . . . VanElzakker, M. B. (2019). Physiological feelings. *Neuroscience and Biobehavioral Reviews*, 103, 267–304. Available from <https://doi.org/10.1016/j.neubiorev.2019.05.002>.
- Panigrahi, P. P., & Singh, T. R. (2013). Computational studies on Alzheimer’s disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases. *Journal of Theoretical Biology*, 334, 109–121. Available from <https://doi.org/10.1016/j.jtbi.2013.06.013>.
- Parikhshak, N. N., Gandal, M. J., & Geschwind, D. H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nature Reviews. Genetics*, 16(8), 441–458. Available from <https://doi.org/10.1038/nrg3934>.
- Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., . . . Bagos, P. G. (2011). Using graph theory to analyze biological networks. *BioData Mining*, 4(1), 10. Available from <https://doi.org/10.1186/1756-0381-4-10>.
- Peng, J., Wang, P., Zhou, N., & Zhu, J. (2009). Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486), 735–746. Available from <https://doi.org/10.1198/jasa.2009.0126>.
- Peng, S., Cheng, M., Huang, K., Cui, Y., Zhang, Z., Guo, R., . . . Shi, B. (2018). Efficient computation of motif discovery on Intel Many Integrated Core (MIC) architecture. *BMC Bioinformatics*, 19(9), 282. Available from <https://doi.org/10.1186/s12859-018-2276-1>.
- Pradhan, P., C.U., A., & Jalan, S. (2020). Principal eigenvector localization and centrality in networks: Revisited. *Physica A: Statistical Mechanics and Its Applications*, 554, 124169. Available from <https://doi.org/10.1016/j.physa.2020.124169>.
- Pratap, A., Taliyan, S., & Singh, T. R. (2014). Nmdb: Network motif database envisaged and explicated from human disease specific pathways. *Journal of Biological Systems*, 22(01), 89–100. Available from <https://doi.org/10.1142/S0218339014500053>.
- Ramadan, E., Alinsaf, S., & Hassan, M. R. (2016). Network topology measures for identifying disease-gene association in breast cancer. *BMC Bioinformatics*, 17(Suppl 7). Available from <https://doi.org/10.1186/s12859-016-1095-5>.
- Rezola, A., Pey, J., Tobalina, L., Rubio, Á., Beasley, J. E., & Planes, F. J. (2015). Advances in network-based metabolic pathway analysis and gene expression data integration. *Briefings in Bioinformatics*, 16(2), 265–279. Available from <https://doi.org/10.1093/bib/bbu009>.
- Ribeiro, P., & Silva, F. (2014). G-Tries: A data structure for storing and finding subgraphs. *Data Mining and Knowledge Discovery*, 28(2), 337–377. Available from <https://doi.org/10.1007/s10618-013-0303-4>.
- Ribeiro, P., Silva, F., & Lopes, L. (2012). Parallel discovery of network motifs. *Journal of Parallel and Distributed Computing*, 72(2), 144–154. Available from <https://doi.org/10.1016/j.jpdc.2011.08.007>.
- Rivkind, A., Schreier, H., Brenner, N., & Barak, O. (2020). Scale free topology as an effective feedback system. *PLoS Computational Biology*, 16(5), e1007825. Available from <https://doi.org/10.1371/journal.pcbi.1007825>.
- Schadt, E. E., & Björkegren, J. L. M. (2012). NEW: Network-enabled wisdom in biology, medicine, and health care. *Science Translational Medicine*, 4(115), 115rv1. Available from <https://doi.org/10.1126/scitranslmed.3002132>.
- Schäfer, J., & Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics (Oxford, England)*, 21(6), 754–764. Available from <https://doi.org/10.1093/bioinformatics/bti062>.
- Schreiber, F., & Schwöbbermeyer, H. (2005). Frequency concepts and pattern detection for the analysis of motifs in networks. In C. Priami, E. Merelli, P. Gonzalez, & A. Omicini (Eds.), *Transactions on computational systems biology III* (pp. 89–104). Springer. Available from https://doi.org/10.1007/11599128_7.
- Sehgal, M., Gupta, R., Moussa, A., & Singh, T. R. (2015). An integrative approach for mapping differentially expressed genes and network components using novel parameters to elucidate key regulatory genes in colorectal cancer. *PLoS One*, 10(7), e0133901. Available from <https://doi.org/10.1371/journal.pone.0133901>.
- Sehgal, M., & Singh, T. (2017). *Principles and analysis of biological networks: Biological pathways and network motifs. Library and information services for bioinformatics education and research*. IGI Global. Available from <https://doi.org/10.4018/978-1-5225-1871-6.ch006>.
- Sehgal, M., & Singh, T. R. (2012). Identification and analysis of biomarkers for mismatch repair proteins: A bioinformatic approach. *Journal of Natural Science, Biology, and Medicine*, 3(2), 139–146. Available from <https://doi.org/10.4103/0976-9668.101887>.
- Shukla, R., Yadav, A. K., & Singh, T. R. (2021). Application of deep learning in biological big data analysis. In H. Saini, G. Rathee, & D. K. Saini (Eds.), *Large-scale data streaming, processing, and blockchain security*. IGI Global. Available from <https://doi.org/10.4018/978-1-7998-3444-1.ch006>.
- Singh, S., Rao, A., Mishra, P., Yadav, A. K., Maurya, R., Kaur, S., & Tandon, G. (2018). Bioinformatics in next-generation genome sequencing. In G. Wadhwa, P. Shanmughavel, A. K. Singh, & J. R. Bellare (Eds.), *Current trends in bioinformatics: An insight* (pp. 27–38). Springer. Available from https://doi.org/10.1007/978-981-10-7483-7_2.
- Sonawane, A. R., Weiss, S. T., Glass, K., & Sharma, A. (2019). Network medicine in the age of biomedical big data. *Frontiers in Genetics*, 10. Available from <https://doi.org/10.3389/fgene.2019.00294>.
- Souza, L. P. D., Alseekh, S., Brotman, Y., & Femie, A. R. (2020). Network-based strategies in metabolomics data analysis and interpretation: From molecular networking to biological interpretation. *Expert Review of Proteomics*, 17(4), 243–255. Available from <https://doi.org/10.1080/14789450.2020.1766975>.
- Sporns, O. (2013). Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, 15(3), 247–262.
- Stone, L., Simberloff, D., & Artyz-Randrup, Y. (2019). Network motifs and their origins. *PLoS Computational Biology*, 15(4), e1006749. Available from <https://doi.org/10.1371/journal.pcbi.1006749>.
- Takes, F. W., Kusters, W. A., Witte, B., & Heemskerk, E. M. (2018). Multiplex network motifs as building blocks of corporate networks. *Applied Network Science*, 3(1), 1–22. Available from <https://doi.org/10.1007/s41109-018-0094-z>.

- Vidal, M., Cusick, M. E., & Barabási, A.-L. (2011). Interactome networks and human disease. *Cell*, *144*(6), 986–998. Available from <https://doi.org/10.1016/j.cell.2011.02.016>.
- Wang, P., Lü, J., & Yu, X. (2014). Identification of important nodes in directed biological networks: A network motif approach. *PLoS One*, *9*(8), e106132. Available from <https://doi.org/10.1371/journal.pone.0106132>.
- Werhli, A. V., Grzegorzczak, M., & Husmeier, D. (2006). Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics (Oxford, England)*, *22*(20), 2523–2531. Available from <https://doi.org/10.1093/bioinformatics/btl391>.
- Wernicke, S. (2006). Efficient detection of network motifs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *3*(4), 347–359. Available from <https://doi.org/10.1109/TCBB.2006.51>.
- Winterbach, W., Mieghem, P. V., Reinders, M., Wang, H., & Ridder, D. de (2013). Topology of molecular interaction networks. *BMC Systems Biology*, *7*(1), 90. Available from <https://doi.org/10.1186/1752-0509-7-90>.
- Yu, S., Xu, J., Zhang, C., Xia, F., Almakhadmeh, Z., & Tolba, A. (2019). Motifs in big networks: Methods and applications. *IEEE Access*, *7*, 183322–183338. Available from <https://doi.org/10.1109/ACCESS.2019.2960044>.
- Yuan, Z., Ji, J., Zhang, X., Xu, J., Ma, D., & Xue, F. (2016). A powerful weighted statistic for detecting group differences of directed biological networks. *Scientific Reports*, *6*(1), 34159. Available from <https://doi.org/10.1038/srep34159>.
- Yue, X., Wang, Z., Huang, J., Parthasarathy, S., Moosavinasab, S., Huang, Y., ... Sun, H. (2020). Graph embedding on biomedical networks: Methods, applications and evaluations. *Bioinformatics (Oxford, England)*, *36*(4), 1241–1251. Available from <https://doi.org/10.1093/bioinformatics/btz718>.
- Zhang, X., Liu, K., Liu, Z.-P., Duval, B., Richer, J.-M., Zhao, X.-M., ... Chen, L. (2013). NARROMI: A noise and redundancy reduction technique improves accuracy of gene regulatory network inference. *Bioinformatics (Oxford, England)*, *29*(1), 106–113. Available from <https://doi.org/10.1093/bioinformatics/bts619>.
- Zhao, X., & Liu, Z.-P. (2019). Analysis of topological parameters of complex disease genes reveals the importance of location in a biomolecular network. *Genes*, *10*(2). Available from <https://doi.org/10.3390/genes10020143>.