# Analysis and Simulation of Biological Complex Networks

B. Ristevski and S. Savoska
Department of Software Engineering and Information Systems
Faculty of Information and Communication Technologies – Bitola
"St. Kliment Ohridski" University - Bitola
Ul. Partizanska bb. 7000 Bitola, Republic of North Macedonia
blagoj.ristevski@fikt.edu.mk, snezana.savoska@fikt.edu.mk

**Abstract - Modeling, simulation and discovering of interactions and regulatory mechanisms among networks' nodes, is still a challenging task in the analysis of complex networks in many domains, such as biology, computer networks, social networks, physics and power systems analysis. Particularly, complex networks' analysis in biology can provide a very significant insight to the relevant information for the biological processes, such as diseases, interactions and regulatory mechanisms. The aim of this paper is to describe and survey the application of complex network analysis in biology to comprehend the relationships in the complex dynamic processes and interactions in biological networks and their properties.**

**Keywords: Network Science, Applied Informatics and Simulation, Complex Networks, Bioinformatics, Big Data.**

## I. Introduction

Nowadays in the era of big data, to understand molecular basis of various disease, huge amounts of high-throughput data are created. Understanding the underlying diseases' processes requires integration of numerous heterogeneous data and then to study the complex interactions and relationships among entities (i.e., genes, proteins, non-coding RNAs, metabolites). These relationships might belong to different types and they depict complex biological networks.

To analyze, describe, control and monitor many real systems in biology, network theory, power systems, engineering, physics, social science, computer science, network analysis is a very promising and powerful method. The organization and behavior of complex systems can be represented by graphs and collection of automata [3]. Graphs are suitable to deal with the static properties of the complex systems. To study nonlinear dynamics of the systems that are represented with graphs, networks node should have discrete states such as Boolean networks.

Networks provides very suitable and efficiently representation of large amounts of data, particularly those who have graph-based structures. The structure of many real systems is a graph that contains modules, often named as communities, consisted of cluster of densely interconnected nodes. The nodes within a module are densely linked, while nodes that belong in different communities are sparsely linked. The entities (nodes) in the same community represents functional units of the network that share similar behavior, common characteristics, interests, or are involved in the similar activities.

Study of the network structure enables discovering of several organization fundaments of complex systems, the community structure and the network node degree distribution, whether the networks are scale-free, random or small-world networks.

The edges in a complex network depict the interaction between nodes. As a result of these interactions, perturbations of one node can trigger off changes in the state of the other nodes [7]. These state alterations are important to control a network when it transits from an initial state to a desired state by handling the state variables of a subset of nodes.

Such control of the complex networks plays a key role in the biological networks, such as protein-protein interaction networks, gene regulatory networks, cellular networks, brain networks, microRNA-mediated regulatory networks, metabolic networks etc.

The remainder of the paper is structured as follows. Section II describe the common network properties, whereas the link prediction and community detection are described in the subsequent section. The models and software tools for modeling and analysis of biological networks are depicted in Section IV The last section provides concluding remarks and direction for further work in analysis of complex networks that are based on big omics data.

## II. Network properties

To study complex networks, network properties such as diameter of the graph, nodes' degree distribution, centrality measures, clustering coefficient, network motifs and graphlets should be taken into account.

Let $G(V, E)$ be a graph, where $V$ is the set of vertices and $E$ is the set of undirected edges and $u,v \in V(G)$ are adjacent. The diameter of a graph is the maximal distance

$$d(u,v) | \forall u,v \in V(G)$$

The clustering coefficient *C* of a network is the average of $C_v$ for all $v \in V(G)$ in the network, and $C_v$ is the clustering coefficient for node *v:*

$$C_v = \frac{2E_v}{n_v(n_v - 1)}$$

where $E_v$ is the number of edges between all the neighbors of *v*, and $n_v$ is the number of neighbors of *v*

The clustering coefficient measures how connected are the neighbors of any node. The degree centrality measures the importance of the role vertex u plays in a graph by measuring the number of interactions u is involved in. Let the degree of the node (vertex) *u* be denoted as *d*(*u*) and given by:

$$d(u) = \sum_{i \in V(G)} e_{ui}$$

where $e_{ui} \in E(G)$. The degree centrality of vertex *u* is defined as *Cd*(*u*)=*d*(*u*)

The betweenness centrality measures the importance of vertex *u* in a graph by measuring the proportion of paths between other vertices in *G [6]*. The betweenness centrality of *w* is given by the following equation:

$$BC(w) = \sum_{u,v \in V} \frac{S_{uv}(w)}{S_{uv}}$$

where $S_{uv}$ is the number of shortest paths between *u* and *v*.

Network motifs are small sub-graphs in a network such that when compared to randomized networks, their structures appear significantly more. Different motifs are found in different complex networks (the feed-forward loop, a 3-node motif etc.)

Graphlets are all non-isomorphic connected induced graphs on a certain number of vertices, and by definition, they have the ability to capture all the local structures on a certain number of vertices.

For instance, when gene regulatory networks are studied, several properties should be considered such as sparseness, scale-free topology, modularity and structurality of the network. The gene regulatory networks should be sparse, which means a limited number of genes regulates genes. Some genes in the network called "hubs" can regulate many genes, i.e. the out-degree of the nodes is not limited. Another important feature is the scale-free topology. Scale-free networks have the power distribution function of the connectivity degree. This property provides the robustness of the networks regarding the random topology changes. Structures with small connectivity follow the regulatory hierarchy. The structurality allows network decomposition into basic modular elements composed of several genes, called network motifs. The network modularity refers to the existence of clusters of highly co-expressed genes and genes with similar function.

Centrality measures for complex networks are used to identify important elements of the networks through their structural topological properties [6]. Each centrality measures cover a different aspect of vertex local or global importance in a given network. Complex network studies have shown that real complex networks have several important properties such as small-word effect, scale-free topology and community organization [6].

Real networks have additional properties that are not associated with their node degree distributions, such as degree correlations, local clustering and community structure [7]. Many real networks have nodes with one degree, so-called dead ends that can erode the complex system stability.

## III. Link prediction and community detection in complex networks

The high-order structures are small network subgraphs that are referred as network motifs. Network motifs such as feed-forward loops, two-hop paths, open bidirectional wedges and triangular motifs are network building blocks that are crucial to understand fundamental network functions, patterns and properties [11].

Motif discovery is one of the essential research fields in complex networks in biology. These networks subgraphs that are usually considered as buildings blocks, Because of the algorithms' complexity, heuristic approaches are commonly used to discover networks motifs.

Nodes belonging to the same community in a network have a higher probability to share functional properties and hence detection of community can discover new functional relationships or characteristics of that network. Community detection algorithms search for the optimal community structure that represents network characteristics as better as possible [1]. To solve community detection problem, many heuristic algorithms are proposed such as those based on simulated annealing, swarm intelligence, genetic and evolutionary algorithms, generational genetic algorithm (GGA+).

For community detection, two main types of algorithms are usually used: edge betweenness-based algorithm and modularity-based algorithm. The edge-betweenness value of an edge *e* in a network/graph *G* is the fraction of all pairwise shortest paths that pass through *e*. When an edge e that bridges a graph is removed and graph is disconnected, that edge *e* has a very high edge-betweenness value. Let a graph *G* is divided into *k* clusters and $e_{ii}$ is the ratio of edges in cluster *i* and $a_i$ is the fraction of edges with at least one end in the *i*-th cluster. The graph modularity is calculated by the following equation:

$$Q = \sum_{i=1}^{k} \left(e_{ii} - a_i^2\right)$$

When the percentage of the edges within the clusters are higher than ones with one node in a cluster, a higher value of Q is expected. The value of Q approaches to 1 when there are only few inter-cluster edges [8].

The graphs used to model complex networks are large and hence many problems in these networks are NP-hard, which make them to use suitable heuristic algorithms [8].

Connectivity in a graph denotes that any two of its nodes are connected by a path. A graph is *k*-connected if there are at least *k* disjoint paths between any two nodes in the graph. When the value of k is higher, the graph becomes more strongly connected. That means that networks with higher connectivity values *k* are more reliable and tolerant of node failures, compared to the networks with lower *k* values.

To understand organization of a network, prediction of the missing links between networks nodes is needed. The link prediction problem is very challenging in contemporary computer science. The main goal of link prediction is to estimate the existence likelihood of unobservable links based on the known network topology and node attributes [5]. In complex network analysis, link prediction is used to reveal missing parts (e.g., of biological and social networks).

To compare network properties, several commonly used network models should be considered, such as Erdős–Rényi random graphs, Barabási–Albert scale free networks, scale free networks that model gene duplication and mutations, geometric random graphs, geometric graphs that model gene duplication and mutations, stickiness-index based networks and generalized random graphs with the same degree distribution as the data [12].

Comparing of biological networks as a whole or partly is used to find their similarities. A comparison between two networks is made in pairwise alignment, whereas comparison of many networks in multiple alignment [8]. Entire networks are compared for similar species in global alignment, while similar subnetworks for diverse species are compared in local alignment [8]. When assessing similarity of two biological networks, topological similarity or node similarity can be calculated.

## IV. Models and Tools for modelling and analysis of biological networks

Biological complex networks are consisted of nodes that represent biological entities, while edges represent the interactions among them. In protein-protein interaction networks, proteins are represented by nodes, as shown on Fig.1. Genes and interacting proteins are nodes in gene regulatory networks. Metabolic networks represent biochemical reactions in the cell that generate metabolism [8]. Besides these networks within the cells, brain networks, neural networks, phylogenetic networks are another type of biological networks outside the cell.

Protein-protein interaction networks are present outside cells' nucleus and detecting clusters and network motifs and aligning two networks are challenging tasks in biological complex network analysis. Network motifs are assumed to have basic functional and building blocks of an organism. Finding similar networks motifs in two or more organisms may be a clue for having common ancestry [8]. While network alignment between two or more networks is usually used to compare different networks and shows the similarities between networks.

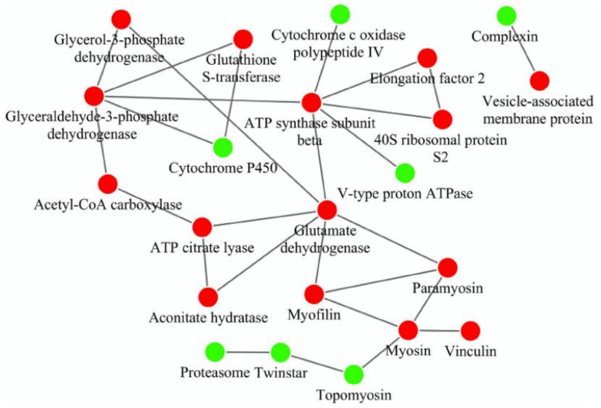

**Figure 1:** Protein-protein interaction network [14].

Clustering in biological networks is employed to find dense regions of significant biological activity that show sometimes diseases states of a particular organism [8]. The clustering in biological networks can be categorized as density-based clustering, hierarchical clustering, spectral clustering and flow-based clustering. A survey of computational approaches for reverse engineering of micro-RNAS-mediated and gene regulatory networks are given in [13], whereas computation approaches for detection of protein complexes from protein-protein interaction networks are given in [19].

Many tools are developed to model, simulate and analyze complex networks with highlight to their application in biology. Since the beginning of the 1990s, Petri nets became a powerful tool to model and simulate complex biological and biochemical networks [2] [4].

Petri nets formalism is suitable to model, simulate and analyze processes that occur in the complex systems. Their application for modeling of biological networks started in the early 1990s. A Petri net PN = (P, T, F, W, $m_0$) is consisted of a finite set of places (P) and finite set of transitions (T) that are connected by directed arcs (F), while $F \subseteq (P \times T) \cup (T \times P)$ is a finite set of arcs [4]. The tuple (P, T, F) is called a net and W is the weight function of the Petri net. The places, transitions and arcs are represented by circles, rectangles and directed arrows, respectively. Places can contain tokens drawn as dots. When tokens are assigned to places, a Petri net is configured and $m_0$ is Petri net's start configuration. When all pre-places represent more or equal token than the input arrow announced, then a transition has concession. If any transition has concession the transition can fire [4]. When the elements of a Petri net have more properties, then more types of Petri nets can be defined.

Functional Petri net are those Petri nets whose arcs can represent functions and tokens can be represented by nonnegative integer numbers. When a transition can be controlled by a timer, timed Petri nets are introduced. Stochastic Petri nets are timed Petri net where each transition is provided by a random delay instead a fixed value. Continuous Petri nets are those nets that use real positive numbers, instead integer numbers [4]. To model biological processes, a hybrid Petri nets are introduced, which have discrete and continuous transitions and places, as well as inhibitory and test arcs, as shown on Fig. 2. Such transitions, places and arcs provide interpretation of concentration of different biological entities, reactions, pathways and regulations. Another extension of the hybrid Petri nets by introducing hybrid functional Petri nets [2].

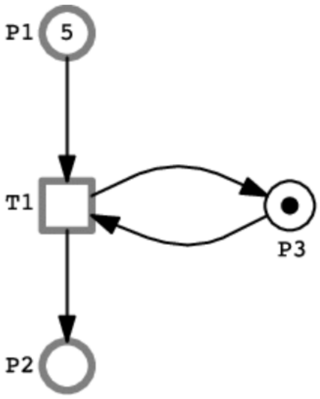

**Figure 2:** A Hybrid Petri net consisted of two continuous places, one discrete place and one continuous transition [15].

Very suitable software packages for modeling, analysis and simulation of complex networks in biology are Matlab with its toolboxes [17], R (programming language) with its R packages [18] and Cytoscape with its plugins [16] [9].

## V. Conclusion and future work

To expedite the study of various biological processes, the analysis of network structural and graph theoretical properties are very important. To study properties of biological complex networks, as a further work development of suitable heuristic, parallel and distributed algorithms is needed.

With recent advances in high-throughput technologies, a huge amount of biological omics data is generated. To store, analyze and query these biological data sets, graph databases are very promising data model. As a further work, more software applications and tools based on graph databases should be developed. Using graph databases can improve the biological network analysis, particularly the understanding the interactions among biological entities (nodes).

## Acknowledgement

This paper was supported by the Ministry of Education and Science of the Republic of Macedonia and the Ministry of Science and Technology (MOST) of the Government of the People’s Republic of China within the bilateral project “Modeling, Simulation and Analysis of Complex Networks in Biology”.

## References

[1] Guerrero, Manuel, Francisco G. Montoya, Raúl Baños, Alfredo Alcayde, and Consolación Gil. "Adaptive community detection in complex networks using genetic algorithms." *Neurocomputing* 266 (2017): 101-113.

[2] Hofestädt, Ralf. "Advantages of Petri-Net Modeling and Simulation for Biological Networks." *Journal of Bioscience, Biochemstry and Bioinformatics* 7, no. 4 (2017): 221-229.

[3] Gates, Alexander J., and Luis M. Rocha. "Control of complex networks requires both structure and dynamics." *Scientific reports* 6 (2016): 24456.

[4] Hofestädt, Ralf, Christoph Brinkrolf, and Philo Reipke. "OMPetri: A New Petri Net Simulation Environment Based on OpenModelica." In *Proceedings of the 2018 10th International Conference on Bioinformatics and Biomedical Technology*, pp. 57-61. ACM, 2018.

[5] Lü, Linyuan, Liming Pan, Tao Zhou, Yi-Cheng Zhang, and H. Eugene Stanley. "Toward link predictability of complex networks." *Proceedings of the National Academy of Sciences* 112, no. 8 (2015): 2325-2330.

[6] Grando, Felipe, Diego Noble, and Luis C. Lamb. "An analysis of centrality measures for complex and social networks." In *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6. IEEE, 2016.

[7] Yan, Gang, Georgios Tsekenis, Baruch Barzel, Jean-Jacques Slotine, Yang-Yu Liu, and Albert-László Barabási. "Spectrum of controlling and observing complex networks." *Nature Physics* 11, no. 9 (2015): 779.

[8] Erciyes, K. *Guide to Graph Algorithms*. Springer International Publishing, 2018.

[9] Kucera, Mike, Ruth Isserlin, Arkady Arkhangorodsky, and Gary D. Bader. "AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations." *F1000Research* 5 (2016).

[10] Iyer, Swami, Timothy Killingback, Bala Sundaram, and Zhen Wang. "Attack robustness and centrality of complex networks." *PloS one* 8, no. 4 (2013): e59613.

[11] Benson, Austin R., David F. Gleich, and Jure Leskovec. "Higher-order organization of complex networks." *Science* 353, no. 6295 (2016): 163-166.

[12] Yaveroğlu, Ömer Nebil, Noël Malod-Dognin, Darren Davis, Zoran Levnajic, Vuk Janjic, Rasa Karapandza, Aleksandar Stojmirovic, and Nataša Pržulj. "Revealing the hidden language of complex networks." *Scientific reports* 4 (2014): 4547.

[13] Ristevski, Blagoj. "Overview of computational approaches for inference of microRNA-mediated and gene regulatory networks." In *Advances in Computers*, vol. 97, pp. 111-145. Elsevier, 2015.

[14] Wang, Hui, Keke Wu, Yan Liu, Yunfeng Wu, and Xifeng Wang. "Integrative proteomics to understand the transmission mechanism of Barley yellow dwarf virus-GPV by its insect vector Rhopalosiphum padi." *Scientific reports* 5 (2015): 10971.

[15] Herajy, Mostafa, Fei Liu, and Monika Heiner. "Efficient modelling of yeast cell cycles based on multisite phosphorylation using coloured hybrid Petri nets with marking-dependent arc weights." *Nonlinear Analysis: Hybrid Systems* 27 (2018): 191-212.

[16] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Research 2003 Nov; 13(11):2498-504

[17] MATLAB Release 2019b, The MathWorks, Inc., Natick, Massachusetts, United States.

[18] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[19] Li, Xiaoli, Min Wu, Chee-Keong Kwoh, and See-Kiong Ng. "Computational approaches for detecting protein complexes from protein interaction networks: a survey." *BMC genomics* 11, no. 1 (2010): S3.