

Tools for analysis of biochemical network topology

Tatjana Rubina^{1*}

¹Department of Computer Systems, Latvia University of Agriculture, Liela iela 2, LV3001, Jelgava, Latvia

*Corresponding author

Tatjana.rubina@llu.lv

Received: 4 November 2012; accepted: 12 November 2012; published online: 13 November 2012.
This paper has no supplementary material.

Abstract: *The biochemical networks can present the relationships between genes and gene products, proteins, metabolites and etc. The exploration of these networks helps to understand cellular processes, functions or properties of biological system. The growing size of interaction models of biological system building elements request determination of the most important topological measurements of given task and powerful automated software tools to perform the analysis.*

The network structure measures and properties are categorized in five groups: topological parameters, topological features, network metrics, network motifs and quantitative parameters of whole network structure. Topology analysis related features of software tools Cytoscape with plug-ins BiNoM and NetworkAnalyzer, VisANT, Biological Networks and CelNetAnalyser are reviewed to simplify the task-dependent choice. The applicability of software tools for calculation of 44 topological features is summarized.

Research resulted in overview on biochemical network structure analysis, on used topological features with the following goals: 1) to accumulate the existing knowledge about the network structure analysis; 2) to provide a list of topological parameters and features; 3) to provide the information of the existing software tools for the structure analysis.

Keywords: Network, structural model, graph theory, computer software.

1. Introduction

The number of biological experiments, corresponding data sets and discoveries in postgenomic era have been very intensive (Gehlenborg et al., 2010; Ideker et al., 2001). One of directions of fast data generation has been systems biology which is concentrating on study of biological interaction networks (Durek and Walther, 2008; Zhu et al., 2007). Cellular proliferation, differentiation, and environmental interactions each requires the production, assembly, operation, and regulation of many thousands of components, and they do so with remarkable fidelity in the face of many environmental cues and challenges (Zhu et al., 2007). Since the end of the 1990s, there has been a flood of interaction data for proteins, carbohydrates, DNA, RNA, lipids and other molecules (Yamada and Bork, 2009) Each completed genome sequencing project generates large data sets of different interactions, specially protein-protein interactions (PPI) (Yamada and Bork, 2009). Adding other types of interactions like metabolic networks, signalling networks, transcription regulatory networks the analysis problem of networks becomes critical due to their size. Still the network representation of mentioned interactions allows application of graph theory (Strogatz, 2001; Watts and Strogatz, 1998) and graph-topological analysis (Assenov et al., 2008) to the biological data to get insight into the global network structure.

Networks have “emergent” properties that are distinct from those of their individual components. Therefore networks have to be studied as systems. Emergent properties are non-linear, aggregated and combinatory effects generated by the interaction of the components of the network. For example, properties such as topology, information flow and the stable

state of a network can only be detected at the network level, not by examining the individual components such as genes or proteins. The structural and dynamic features of genetic networks ultimately contribute to biological functions, robustness and evolvability of the networks (Han et al. 2004). The topological measures can capture the cellular features of cellular networks and provide broad insight into cellular evolution, molecular function, network stability, and dynamic responses (Chen et al., 2009).

The author in this paper reviews and classifies the most popular measures and properties of biochemical network topology and the most relevant freely available software tools for its analysis.

2. Measures and properties of network structure

Examining scientific literature, publications and analyzing software tools, author found many network structure measures and properties, which can be categorized in five groups: topological parameters, topological features, network metrics, network motifs and quantitative parameters of whole network structure that are the global topological parameters (Fig.1). Some of measures and properties will be explained below.

2.1. Topological parameters

Topological parameters can be divided in two groups – local and global parameters (Durek and Walther, 2008), corresponding to the measurable element. Local topological parameters characterize individual network components while global parameters describe the whole network. One of local parameters is the **degree of a network node**. The degree (or connectivity) (Barabási and Oltvai, 2004; Robins et al., 2008; Yamada and Bork, 2009) of an undirected network node, k_i , is

the number of edges (links) that it has with other nodes (Fig. 2) that is incident with i :

$$k_i = \sum_j^n k_{ij} \quad (1)$$

For directed network degree is separated in two types: incoming (in-degree) and out coming (out-degree) degree, depending on the direction of interactions (Hu et al., 2005). A degree is also a feature that distinguishes hubs (highly connected nodes) from leaves or orphans (weakly or non-connected nodes) in the network (Zinovyev et al., 2008). In protein interaction and genetic interaction networks, for example, the degree of a hub (highly connected node) is often hub's importance and essentiality for cell function (Hu et al., 2005), process or whole system.

Degree distribution d_k is the number of nodes with degree k ($k=1,2,\dots,n$) (Chen et al., 2009; Robins et al., 2008). For directed networks the degree distribution is separated into in-degree and out-degree distribution.

Let K be the degree of a network node. Then a statistical model for the degree distribution is represented by:

$$P(K = k) = f(k) = \frac{N_k}{N} \quad (2)$$

,where $f(k)$ is a probability distribution
 N_k – a number of nodes with degree $k=1,2\dots n$
 N – the total number of nodes.

The distribution of degrees $f(k)$ in undirected network, gives the probability that a selected node has degree k (Barabási and Oltvai, 2004). In the case of directed networks one needs to consider two distributions, $P(k_{in})$ and $P(k_{out})$ (Boccaletti et al., 2006).

The degree distribution of many types of real-life networks, such as metabolic or signalling, scientific collaboration networks is called a power law (Robins et al., 2008; Zhang and Shakhnovich, 2008):

$$P(K = k) \sim k^{-\gamma} \quad (3)$$

,where γ - a constant or the degree exponent.

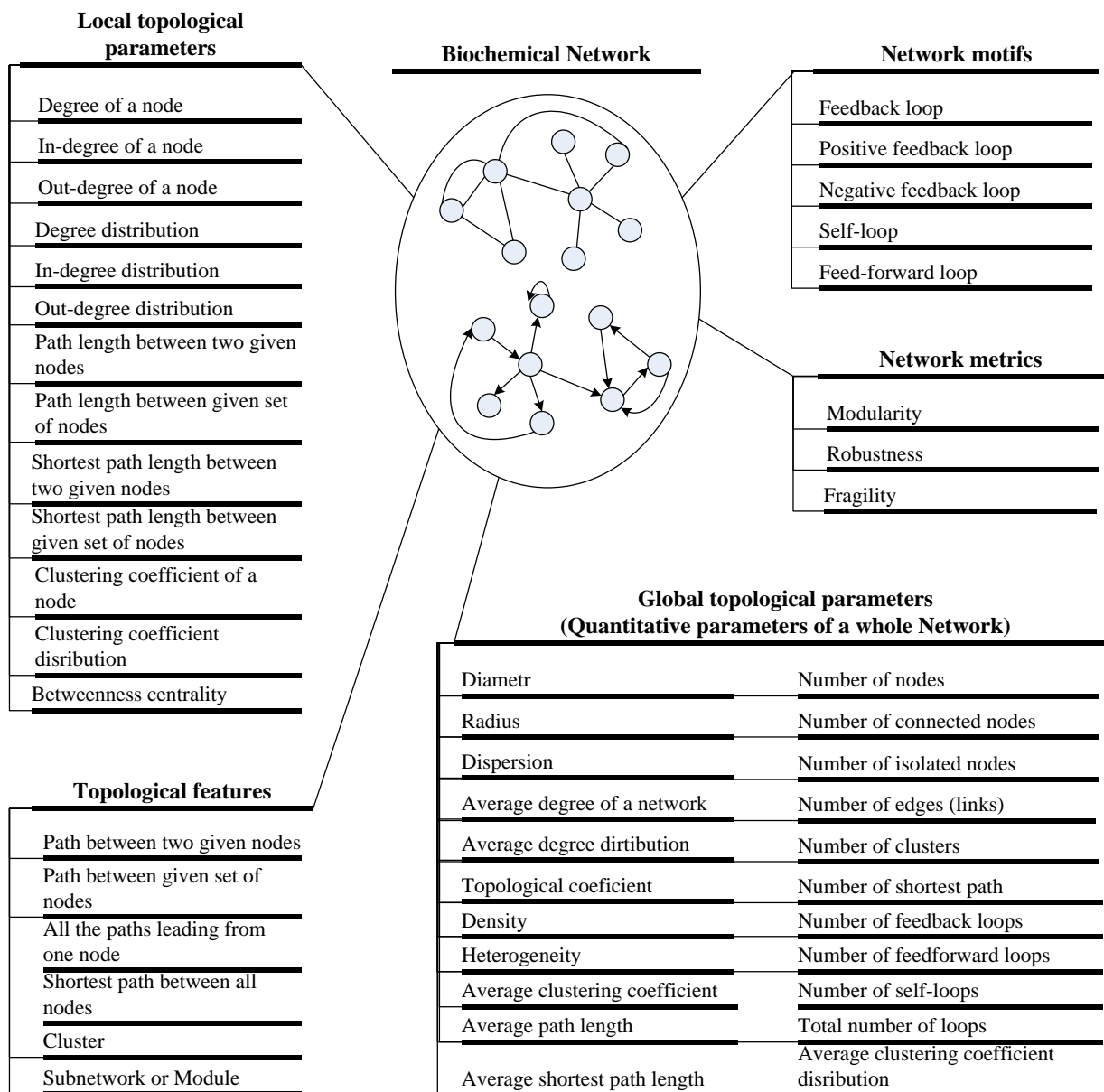


Fig. 1. Biochemical network measures and properties (Rubina and Stalidzans, 2010a).

A power-law degree distribution indicates that a few hubs hold together numerous small components or nodes (Barabási and Oltvai, 2004). In scale-free networks most nodes have only one or two functional links, whereas a small number of nodes, the hubs, have many links (Han, 2008). Nearly all biological networks, including regulatory, interactome and metabolic networks are scale-free (Barabási and Oltvai, 2004; Boccaletti et al., 2006). Still there are other types of networks like random network and hierarchical network (Yamada and Bork, 2009).

2.2. Topological features

Studying the function of pathways, the property of interest is often how a given gene or protein is related to (or responds to) an up- or downstream signal. Given a large data set of interactions, it may be useful in some contexts to find the most direct path between two genes, proteins, complexes or pathways; for example, the overall lengths of such pathways may be related to the immediacy or breadth of signal response (Hu et al., 2005).

According to the graph theory **the path** (Wilson, 1972) is the sequence of nodes from n_0 to n_k . There can be different types of paths: chain (have all different edges), simple chain (have all different nodes), closed chain or cycle (starts and ends with the same node). Cycles compose the separated group of network measures – network motifs.

The path between two given nodes (Fig. 3). In case of signalling networks, the computation of all paths between pair of species helps to recognize all the different ways in which a signal can propagate between two nodes, e.g. all the different ways by which a certain transcription factor (or any other species from the output) can be activated or inhibited by signals riving the input layer.

The path length l_{ij} is the number of edges (or links) in path from node i to j .

The shortest path between two nodes is the path between two nodes in a network with a smallest number of steps compared to alternative paths between the same nodes (Yamada and Bork, 2009).

The shortest path between given set of nodes is the path that connects all the nodes of given set with smallest number of steps.

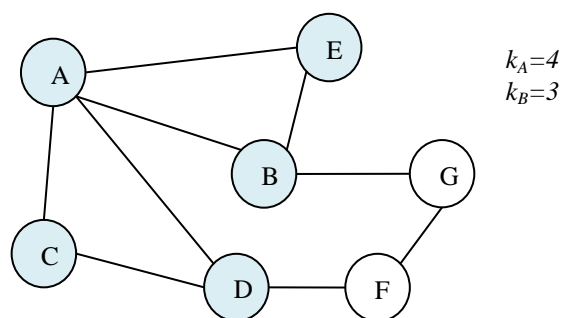


Fig. 2. Undirected network.

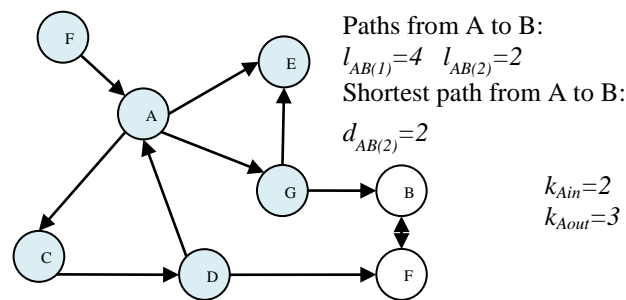


Fig. 3. Directed network.

2.3. Network motifs

Cellular networks are composed of complicated interconnections among nodes, and some subnetworks (in case of graph theory, subgraphs) of particular functioning are often identified as network motifs (Kim et al., 2008). **Network motifs** are the simple building blocks (Milo et al., 2002) of complex biochemical networks and are defined as patterns of interconnections that recur in many different parts of a network. The biochemical networks are composed of three highly significant motifs that repeatedly appear: feedback, feed-forward and self-loops. Each network motif has a specific function and play important dynamical roles in behaviour regulation of biological processes.

Biological systems are known to be considerably robust to environmental changes and genetic perturbations (Barabási and Oltvai, 2004; Kitano, 2004, 2007; Kwon and Cho, 2007a). Robustness is a fundamental feature of complex systems that allows them to maintain its functions despite external and internal perturbations (Kitano, 2004). The main mechanism that ensures the robustness of a system is a *system control* that consists of *negative* and *positive feedback*. Presence of feedback is the important party of control in biological systems (Rubina and Stalidzans, 2012). Negative feedback promotes restoration of an initial condition of system. Positive feedback withdraws system all further from an initial condition and strengthens the processes of ability to live.

From the viewpoint of network structure feedback is organized on *feedback loops*. According to the graph theory *feedback loop* is closed simple cycle (Barabási and Oltvai, 2004; Kwon and Cho, 2007a) of any length (Hallinan and Jackway, 2007) with the set of nodes where the nodes are not revisited except the starting and ending nodes. Exploring dynamic models of biochemical networks, researchers have established that feedback loops are very often found as a coupled structure in cellular circuits. *Coupled feedback loop* is closed cycle with the set of nodes where each node is visited twice (in reverse order) except one node in the middle of loop, for example, coupled feedback loop $A \rightarrow B \rightarrow C \rightarrow D \rightarrow C \rightarrow B \rightarrow A$. Coupled feedback loops can be positive and negative and can form three types of coupled structures (Kim et al., 2007; Kim et al., 2008): positive-positive, positive-negative and negative-negative structures (Fig. 4).

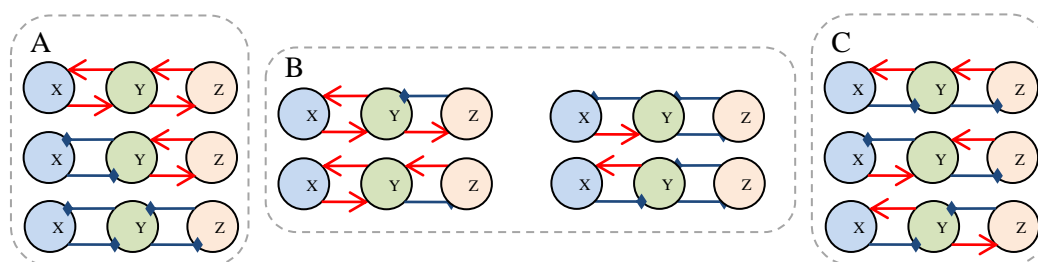


Fig.4. Network motifs of coupled feedback loops (Kim et al., 2008).

(A) Positive-Positive structures. (B) Positive-Negative structures. (C) Negative-Negative structures.

Kwon and colleagues (Kwon and Cho 2007) have verified hypothesis on the relationship between feedback loops and the robustness of a network by employing Boolean network models. They found that three distinct feedback loops are responsible for genetic regulation, mRNA attenuation, and enzyme inhibition that regulate tryptophan concentrations in *Escherichia coli*. The complex regulatory network formed by the feedback loops induces a rapid and stable response, while being robust against uncertainties (Kwon and Cho, 2007a; Venkatesh et al., 2004).

Kim with colleagues (Kim et al., 2007) suggests that coupled positive and negative feedback loops form essential signal transduction motifs in cellular signaling systems or signaling pathways. They performed mathematical simulations and investigations into various experimental evidences, and found that positive and negative coupled feedback circuits can rapidly turn on a reaction to a proper stimulus, robustly maintain its status, and immediately turn off the reaction when the stimulus disappears. In other words, coupled feedback loops enable cellular systems to produce perfect responses to noisy stimuli with respect to signal duration and amplitude (Kim et al. 2007). Likewise Shi with colleagues has proved (Shi et al., 2012) that coupled positive feedback loops can generate reversible and irreversible switch. And coupled positive feedback loops can strengthen bistable, enlarge signal and extend the signal reaction time. It means, that coupled positive feedback loops play an important role in regulation of biological behaviours. Therefore their recognition in biochemical networks is important task.

3. Tools for structure analysis of biochemical networks

According to the earlier performed analysis of existing tools (Rubina and Stalidzans, 2010), the best tools for topology analysis is Cytoscape with plugins BiNoM and NetworkAnalyzer. Tools with good performance are also VisANT, Biological Networks and CelNetAnalyser. Comparative analysis of these tools demonstrates can simplify the choice of appropriate tool for solving of a particular task.

CellNetAnalyser is free for academic use package for MATLAB. CNA provides an environment for structural and functional analysis of biochemical networks such as metabolic, signalling and regulatory networks. It includes metabolic flux

analysis, analysis of basic topological / structural properties, metabolic pathway analysis and for signal flow (signalling, regulatory) networks including analysis of interaction graphs, analysis of logical (boolean) interaction networks.

Cytoscape is an open source tool for visually exploring of biological networks, that support the SBML and BIOPAX standards. Cytoscape specializes in the representation of interaction networks and includes many powerful network display styles (Sudermann and Hallett, 2007). Automatic layout algorithms help to organize massive amounts of interaction data relating to a set of molecules (Chen et al, 2009). **NetworkAnalyzer** is the versatile Cytoscape plug-in (Assenov et al., 2008) that computes a comprehensive list of simple and complex topology parameters (single values and distributions) for directed and undirected networks using efficient graph algorithms. **BiNoM** is a Cytoscape plug-in, developed to analyze a structure of the networks.

VisANT is free and open source integrative web-based software platform for the visualization, mining, analysis and modelling of the biological networks. Visant allows to create multi-scale networks, represent many types of biological data, such as biomolecular interactions, cellular pathways and functional modules and provides a visual interface for combining and annotating network data, supporting function and annotation data for different genomes from the Gene Ontology and KEGG databases. It contains statistical and analytical tools needed for extracting topological properties of the user-defined networks.

Biological networks is free for academic use application for visualization and analysis of biological pathways. It is a graph-based system for creating a combined database of biological pathways, gene regulatory networks and protein interaction maps. After importing expression data, users can apply sorting, normalization and clustering algorithms on the data and then create various tables, heat maps and network views of the data.

Next tables provide summary of topological parameters (Table 1, Table 2) and features of network structure (Table 3) that can be analyzed by selected software tools, dividing topological parameters in two main groups – simple and complex parameters.

Table 1.

Summary of computed simple topological parameters by software tools Visant, Cytoscape with BiNoM, CellNetAnalyzer and Biological Networks.

Parameters \ Tools	Local topological parameters				Global topological parameters											
	Degree of a node	In-degree of a node	Out-degree of a node	Clustering coefficient of node	Average number of neighbours	Network diameter	Network radius	Density	Centralization	Heterogeneity	Number of nodes	Number of edges	Number of self-loops	Number of connected nodes	Number of isolated nodes	Number of shortest paths
Visant	•	•	•													
Cytoscape with BINOM & Network-Analyser	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•
CellNet-Analyser						•					•					
Biological Networks				•												

Table 2.

Summary of computed complex topological parameters by software tools Visant, Cytoscape with BiNoM, CellNetAnalyzer and Biological Networks.

Parameters \ Tools	Local topological parameters						Global topological parameters								
	Node degree distribution	Node in-degree distribution	Node out-degree distribution	Clustering coefficient distribution	Shortest path lengths	Shortest path distribution	Average degree distribution	Neighborhood connectivity's	Neighbours connectivity	Average clustering coefficient	Average clustering coefficient distribution	Topological coefficients	Average path length	Average shortest path length	Shared neighbours of two nodes
Visant	•	•	•	•			•			•	•				
Cytoscape with BINOM & Network-Analyser	•	•	•		•	•		•	•			•		•	•
CellNet-Analyser					•								•		
Biological Networks	•				•										

Table 3.

Summary of computed network motifs and topological features by software tools Visant, Cytoscape with BiNoM, CellNetAnalyzer and Biological Networks.

Parameters	Topological features										Network motifs		
	Shortest path between all nodes	Shortest path between two given nodes	Shortest path between given set of nodes	Optimal and sub-optimal shortest paths	All the paths leading from one node	All non-intersecting paths	Path finding between two given nodes	Path finding between given set of nodes	Cluster finding	Cycle clustering and decomposition	Feedback loops finding	Feed-forward loops finding	Self-loops finding
Tools													
Visant	•	•	•				•	•			•	•	•
Cytoscape with BINOM & NetworkAnalyser	•	•		•	•	•			•	•		•	
CellNet-Analyser	•	•						•			•	•	
Biological Networks	•	•						•	•		•		

4. Conclusion

The topology of biochemical networks can be analysed using tens of measures and parameters. Analysis of software tools Visant, Cytoscape with BINOM and NetworkAnalyser, CellNetAnalyser and Biological Networks gives detailed overview about the functionality of software tools as well as their specialisation on determination of topological measures and parameters. Generally it is concluded that all the mentioned software tools can be involved in analysis of at least some measures and parameters of all five groups.

Software tools Visant and Cytoscape with BINOM and NetworkAnalyser plug-ins has the highest number of calculated parameters among the other compared software tools with relatively high number of simple topological parameters.

Software tools CellNetAnalyser and Biological Networks relatively more concentrate on calculation of complex topological parameters, network motifs and topological features.

All the mentioned software tools can calculate shortest path between all nodes, shortest path between two given nodes and perform finding of feed-forward loops.

Acknowledgments

This work and academic study is funded by a project "Support for doctoral studies in LUA"/2009/0180/1DP/1.1.2.1.2/09/IPIA/VIAA/017" agreement Nr. 04.4-08/EF2.D1.D6".

References

- Assenov, Y., Ramírez, F., Schellhorn, S.-E., Lengauer, T. and Albrecht, M. (2008), "Computing topological parameters of biological networks" *Bioinformatics*, Vol. 24 No. 2, pp. 282–284. <http://dx.doi.org/10.1093/bioinformatics/btm554>
- Barabási, A.-L. and Oltvai, Z.N. (2004), "Network biology: understanding the cell's functional organization" *Nature reviews. Genetics*, Vol. 5 No. 2, pp. 101–113. <http://dx.doi.org/10.1038/nrg1272>
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D. (2006), "Complex networks: Structure and dynamics" *Physics Reports*, Vol. 424 No. 4-5, pp. 175–308. <http://dx.doi.org/10.1016/j.physrep.2005.10.009>

- Chen, L., Wang, R.-S. and Zhang, X.-S. (2009), "Biomolecular networks: Methods and Applications in Systems Biology" *Jersey, John Wiley & Sons; Inc.; Hoboken; New*.
- Durek, P. and Walther, D. (2008), "The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles" *BMC Systems biology*, Vol. 2, p. 100. <http://dx.doi.org/10.1186/1752-0509-2-100>
- Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M., Kitano, H., Kohlbacher, O., et al. (2010), "Visualization of omics data for systems biology" *Nature methods*, Nature Publishing Group, Vol. 7 No. 3, pp. S56–68. <http://dx.doi.org/10.1038/nmeth.1436>
- Hallinan, J.S. and Jackway, P.T. (2007), "Network Motifs, Feedback Loops and the Dynamics of Genetic Regulatory Networks" *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1–7.
- Han, J.-D. (2008), "Understanding biological functions through molecular networks" *Cell research*, Vol. 18 No. 2, pp. 224–37. <http://dx.doi.org/10.1038/cr.2008.16>
- Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., et al. (2004), "Evidence for dynamically organized modularity in the yeast protein-protein interaction network" *Nature*, Vol. Jul 15, 43, pp. 88–93. <http://dx.doi.org/10.1038/nature02555>
- Hu, Z., Mellor, J., Wu, J., Yamada, T., Holloway, D. and Delisi, C. (2005), "VisANT: data-integrating visual framework for biological networks and modules" *Nucleic acids research*, Vol. 33 Web Server issue, pp. W352–357. <http://dx.doi.org/10.1093/nar/gki431>
- Ideker, T., Galitski, T. and Hood, L. (2001), "A new approach to decoding life: systems biology" *Annual review of genomics and human genetics*, Vol. 2, pp. 343–372. <http://dx.doi.org/10.1146/annurev.genom.2.1.343>
- Kim, D., Kwon, Y.-K. and Cho, K.-H. (2007), "Coupled positive and negative feedback circuits form an essential building block of cellular signalling pathways" *BioEssays: news and reviews in molecular, cellular and developmental biology*, Vol. 29 No. 1, pp. 85–90. <http://dx.doi.org/10.1002/bies.20511>
- Kim, J.R., Yoon, Y. and Cho, K.H. (2008), "Coupled feedback loops form dynamic motifs of cellular networks" *Biophysical journal*, Vol. 94 No. 2, pp. 359–365. <http://dx.doi.org/10.1529/biophysj.107.105106>
- Kitano, H. (2004), "Biological robustness" *Nature reviews. Genetics*, Vol. 5 No. 11, pp. 826–837. <http://dx.doi.org/10.1038/nrg1471>
- Kitano, H. (2007), "The theory of biological robustness and its implication in cancer" *Nature Reviews. Genetics*, Vol. 5 No. 11, pp. 826–837. <http://dx.doi.org/10.1038/nrg1471>
- Kwon, Y.-K. and Cho, K.-H. (2007), "Analysis of feedback loops and robustness in network evolution based on Boolean models" *BMC bioinformatics*, Vol. 8, p. 430. <http://dx.doi.org/10.1186/1471-2105-8-430>
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002), "Network motifs: simple building blocks of complex networks,"

- Science* (New York, N.Y.), Vol. 298 No. 5594, pp. 824–827.
<http://dx.doi.org/10.1126/science.298.5594.824>
- Robins, G., Pattison, P. and Koskinen, J. (2008), “Network degree distributions” *Technical report*, pp. 1–9.
- Rubina, T. and Stalidzans, E. (2010a), “Topological features and parameters of biochemical network structure” EUROISIS, *Proceedings of Industrial Simulation Conference*, Budapest, Hungary, pp. 228–236.
- Rubina, T. and Stalidzans, E. (2010b), “Software tools for structure analysis of biochemical networks” *Proceedings of Applied Information and Communication Technologies (AICT)*, Jelgava, Latvia, pp. 33–49.
- Rubina, T. and Stalidzans, E. (2012), “Evolution of alternative control loops of biological systems” *Proceedings of Applied Information and Communication Technologies (AICT)*, Jelgava, Latvia, pp. 317–324.
- Shi, F., Zhou, P. and Wang, R. (2012), “Coupled positive feedback loops regulate the biological behavior” *IEEE 6th International Conference on Systems Biology (ISB)*, Ieee, pp. 169–173.
<http://dx.doi.org/10.1109/ISB.2012.6314131>
- Strogatz, S.H. (2001), “Exploring complex networks” *Nature*, Vol. 410 No. 6825, pp. 268–76. <http://dx.doi.org/10.1038/35065725>
- Suderman M, Hallett M. (2007), “Tools for visually exploring biological networks” *Bioinformatics*, Vol. 23 No 20, pp.2651-2659. <http://dx.doi.org/10.1093/bioinformatics/btm401>
- Venkatesh, K.V., Bhartiya, S. and Ruhela, A. (2004), “Multiple feedback loops are key to a robust dynamic performance of tryptophan regulation in *Escherichia coli*” *FEBS letters*, Vol. 563 No. 1-3, pp. 234–240.
[http://dx.doi.org/10.1016/S0014-5793\(04\)00310-2](http://dx.doi.org/10.1016/S0014-5793(04)00310-2)
- Watts, D.J. and Strogatz, S.H. (1998), “Collective dynamics of ‘small-world’ networks” *Nature*, Vol. 393, pp. 440–442.
<http://dx.doi.org/10.1038/30918>
- Wilson, R.J. (1972), Introduction to graph theory, *New York*, Academic Press.
- Yamada, T. and Bork, P. (2009), “Evolution of biomolecular networks: lessons from metabolic and protein interactions” *Nature reviews. Molecular cell biology*, Vol. 10 No. 11, pp. 791–803. <http://dx.doi.org/10.1038/nrm2787>
- Zhang, J. and Shakhnovich, E.I. (2008), “Sensitivity-dependent model of protein–protein interaction networks” *Physical Biology*, Vol. 5. <http://dx.doi.org/10.1088/1478-3975/5/3/036011>
- Zhu, X., Gerstein, M. and Snyder, M. (2007), “Getting connected: analysis and principles of biological networks” *Genes & development*, Vol. 21 No. 9, pp. 1010–1024. <http://dx.doi.org/10.1101/gad.1528707>
- Zinovyev, A., Viara, E., Calzone, L. and Barillot, E. (2008), “BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks” *Bioinformatics*, Vol. 24 No. 6, pp. 876–877.
<http://dx.doi.org/10.1093/bioinformatics/btm553>