Networks in Biology

Handling Biological Complexity Requires Novel Inputs into Network Theory

PETER SCHUSTER

Peter Schuster, Editor-in-Chief of Complexity, is at the Institut für Theoretische Chemie der Universität Wien, Wien 1090, Austria (pks@tbi.univie.ac.at)

The year 2009 was the tenth anniversary of the first publication on scale-free networks [1] and the fiftieth anniversary of the invention of random graphs [2]. *Science* magazine devoted a special section to review the present status of network theory, complexity research and its application to different disciplines [3]. Understanding and modeling complex systems without consideration of network topology and network evolution became out-of-date and practically impossible. The Erdös-Rényi model initiated a real break-through in the sense that statistical properties of graphs and networks became accessible without knowledge of the connection details. Among many other applications random networks became a useful reference in the biology of macromolecules for mapping polynucleotide sequences into structures [4]. Real world social networks, in particular communication networks, were found to have substantial shorter mean distances between agents than those predicted by the theory of random networks. Watts and Strogatz [5] invented networks that were found to match the available empirical data. These networks are called small-world networks as they have the small-world property[1] [6], and they can be understood as intermediates between a regular lattice and a random graph. Watts and Strogatz characterize small-world networks by a degree of randomness (p), which varies between $p = 0$ for the regular and $p = 1$ for the random network. Only 1 year later, Barabási and Albert conceived small-world networks, which are created by the construction principle of preferential attachment [1]: Starting with a fully connected network of three nodes, further nodes are attached one by one with a larger probability for the incoming node to connect to a node that has already more neighbors. The resulting networks are scale-free—besides having the small-world property—and accordingly many of their properties fulfill power laws.[2] Comparison with real world data from a great

[1] *The small-world property is demonstrated best by communication between humans: The degree of separation between two persons A and B is the number of people in a shortest chain from A to B via people who know each other, e.g., A↔C↔D↔B is a chain with a degree of separation of 2. The average degree of separation of two persons on the Earth is between 6 and 7, a fact that is often called the six degrees of separation phenomenon.*

[2] *Scale-free or self-similar objects exhibit (roughly) the same appearance at all scales and quantitative properties fulfill power laws. Examples are scale-free networks, where the distribution of the degree of nodes fulfills a power law. The degree of a node is the number of nodes to which it is connected by an edge.*

 Vol. 16, No. 4
DOI 10.1002/cplx.20375
Published online 1 March 2011 in Wiley Online Library
(wileyonlinelibrary.com)

variety of different sources ranging from physics and biology to climate modeling, economics, communication networks, and other social interactions revealed fairly good agreement with the predictions that focused, in essence, on power laws describing the distribution of node degrees being the numbers of connections of individual nodes. The degree distribution was found to be sensitive to the class of random networks. On the transition from random and small-world networks to scale-free networks, the degree distribution changes from a Poisson distribution for Erdös-Rényi random graphs to a power law for Barabási-Albert scale-free networks.

Unanswered still is the question what the problems are that wait to be solved by a comprehensive theory of networks in biology. The degree distribution of nodes and other statistical quantities may be interesting for a physicist but they are commonly of little relevance for biology. The real importance of the degree distribution is rather of diagnostic nature in the sense that it allows for the decision whether or not a given network is scale-free. Another general problem is related to self-similarity in biology: Biological networks hardly cover more than two orders of magnitude—corresponding, for example, to one to one hundred connections—and this is often not sufficient to demonstrate self-similarity.[3] Moreover, biologists are interested in the biochemical nature of the hubs, physiologists study the molecular approach to cellular behavior because they are aiming at explanations of holistic properties like robustness and homeostasis, and pharmacologists investigate protein–protein interactions to find specific targets for novel drugs. Knowledge of network topology is an indispensable requirement for modeling cellular dynamics, but it is only a first step. Network statistics provides relevant information, but for answering most of the open questions, the specific details are essential. In particular, the mechanism establishing the connections between the nodes—genetic regulation, chemistry of metabolism or signaling cascades—is crucial for the construction of biologically relevant models.

One month after the special section on networks had appeared in *Science* magazine, *Molecular BioSystems* published a whole issue on Computational and Systems Biology, which contained among other articles an article dealing with critical issues of the application of idealized networks to problems in biology [7]. The authors, Lima-Mendez and van Helden, name five points of conventional and uncritical belief, which they call myths in the sense of '*a traditional story, especially one concerning the early history of people or explaining some natural or social phenomenon*' and which they show to be of very limited validity when tested on real data:

1. the degree distribution of biological networks follows a power law,
2. biological networks are scale-free,
3. the metabolic network is a small-world,
4. small-world networks are tolerant to random deletions, but vulnerable to targeted attacks, and
5. biological networks grow by preferential attachment.

Dispensing from details and making a long story short, Lima-Mendez and van Helden don't find the arguments for the occurrence of scale-free networks in biology convincing. In particular, they make the point that Erdös-Rényi random networks and scale-free networks differ much less when the mean degree of connectivity (λ) is small. The evolutionary growth of biological networks does not obey preferential attachment and the metabolic network is not small-world. Reported small-world property of metabolism [8, 9], they claim, is an artifact of the method used to determine connections between metabolites, because it does not exclude reaction steps that violate mass conservation [10]. From biochemical experience, it is indeed hard to believe that the average distance for the interconversion of any pair of biochemical compounds is only three reactions with a very narrow distribution that contains almost all shortest paths between one and four.

Are statistical properties of networks then completely irrelevant for biology? The answer certainly is: No! Depending on the generic properties, a class of networks and its generation, one of the random networks—Erdös-Rényi, small-world, scale-free or another—will save best as a reference state for the whole class and deviation from the reference can be explained by the specific structure of the special case. An impressive example for the usefulness of a random network as reference was presented in case of RNA sequence-structure relations [11]. The preimage of an RNA (secondary) structure is a graph called neutral network in sequence space. Considering neutral networks as Erdös-Rényi random graphs provides a suitable reference state and allows, for example, for accurate predictions of the connectedness of the graph—a single fully connected component or several separated components without connections between them. Understanding the details of the partitioning into different components—one giant component (Erdös-Rényi prediction) or deviations from the reference in the form of two, three, or four large components—requires specific knowledge of the RNA structure and can be readily explained by it. Other examples can be found with metabolic networks, where the reaction graph is at the beginning and flux balance analysis [12] as described below is the next logical step.

Systems biology is heading for a comprehensive description of the functional dynamics of cells, organs, and organisms. The ultimate goal is modeling biological systems by means of a

[3] *Exceptions are, for example, branching patterns of transport systems in higher organisms, plants and animals, which often cover several orders of magnitude.*

DOI 10.1002/cplx

bottom-up approach from bio-molecules and their interactions that allows for reproduction, analysis, understanding, and prediction of properties and functions. Genetic, metabolic as well as signaling networks are in the core of understanding biological function and hence of primary importance. Biochemical kinetics developed in quantitative terms since the beginning of the twentieth century sets the stage for network dynamics. In principle, it is possible to write down all kinetic differential equations of a cell and assuming the deterministic approach is sufficient[4] the solution would provide the desired information on the biological unit. The actual problem is the dimension of the dynamical system, which is a direct result of the enormously large number of molecular players: Several thousand to thirty thousand genes produce protein molecules—regulators and metabolic enzymes—and form a huge and complex network of dynamic interaction. In addition, there are other players like RNA molecules also encoded by the DNA of the organism and epigenetically controlled molecules, proteins or RNA, which were encoded in the DNA of some ancestor, and which complicate the holistic picture. Many thousands of differential equations require at least as many parameters, but these parameters are not known with sufficient accuracy yet. The program of systems biology thus looks completely unrealistic at a first glance. Simplifications of the gigantic problem, however, were successful. We mention here only one, the most popular example, flux-balance analysis [12, 13], which represents a mathematical and computational method for the analysis of metabolism without explicit knowledge of metabolite concentrations or details of enzyme kinetics. The flux distribution over a biological network is constrained by the stoichiometry of chemical reactions and optimized by means of some objective function, for example, maximal growth of the cell or the organism. The system is assumed to be homeostatic or stationary. Flux balance analysis starts from the reconstruction of the metabolic network of an organism from empirical data. A typical task for flux balance analysis is the computation of the metabolic fluxes, which maximizes the growth rate of an organism for a set of known nutrients. Examples of prokaryotes that were fully analyzed by flux balance analysis are: *Escherichia coli* [14], *Staphylococcus aureus* [15], and *Salmonella typhimurium* [16]. An impressively large number of projects are dealing with systems biology of eukaryotes including man. Representative for others, we mention only one impressive recent example of a systems biology study of a whole organ that is based on a preceding extensive investigation of the corresponding cell: the German liver project, which is the follow-up of a previous program studying the liver cell, the hepatocyte [17]. It is thought to serve as an example that concerted systems biology programs can be carried out, in principle, on all levels of biological complexity spanning eight orders of magnitude in diameter from the individual protein molecule in the nanometer range to the organ or organism in the meter range. Further development of network theory and biochemical dynamics, both merged in systems biology are needed in the process of understanding biology in its overwhelming complexity.

[4]*This point is definitely questionable as many molecular species in biology are present only in very small numbers. A stochastic treatment is possible and shows in most cases that the expectation values of the stochastic description coincide with the deterministic solution. Some data are available also for measurements inside cells and the results are—interestingly—in remarkable good agreement with the deterministic solutions of ordinary (ODE) or partial differential equations (PDE).*

REFERENCES

1. Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. Science 1999, 286, 509–512.
2. Erdös, P.; Rényi, A. On random graphs I. Publ Math Debrecen 1959, 6, 290–297.
3. Jasni, B.R.; Zahn, L.M.; Marshall, E. Connections. Science 2009, 325, 405.
4. Reidys, C.; Stadler, P.F.; Schuster, P. Generic properties of combinatory maps. Bull Math Biol 1997, 59, 339–397.
5. Watts, D.J.; Strogatz, S.H. Collective dynamics of "small-world" networks. Nature 1998, 393, 440–442.
6. Milgram, S. The small-world problem. Psychol Today 1967, 2, 60–67.
7. Lima-Mendez, G.; van Helden, J. The powerful law of the power law and other myths in network biology. Mol Biosyst 2009, 5, 1482–1493.
8. Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabási, A.-L. The large-scale organization of metabolic networks. Nature 2000, 407, 651–654.
9. Fell, D.A.; Wagner, A. The small-world of metabolism. Nat Biotech 2000, 18, 1121–1122.
10. Croes, D.; Couche, F.; Wodak, S.J.; van Helden, J. Inferring meaningful pathways in weighted metabolic networks. J Mol Biol 2006, 356, 222–236.
11. (a) Grüner, W.; Giegerich, R.; Strothmann, D.; Reidys, C.; Weber, J.; Hofacker, I.L.; Schuster, P. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. Monatsh Chem 1996, 127, 355–374; (b) Grüner, W.; Giegerich, R.; Strothmann, D.; Reidys, C.; Weber, J.; Hofacker, I.L.; Schuster, P. Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structures of neutral networks and shape space covering. Monatsh Chem 1996, 127, 375–389.

DOI 10.1002/cplx

12. Palsson, B.Ø. Systems biology. Properties of reconstructed networks; Cambridge University Press: New York, 2006.
13. Orth, J.D.; Thiele, I.; Palsson, B.Ø. What is flux balance analysis? Nat Biotechnol 2010, 28, 245–248.
14. Edwards, J.S.; Ibarra, R.U.; Palsson, B.Ø. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 2001, 19, 125–130.
15. Lee, D.-S.; Burd, H.; Liu, J.; Almaas, E.; Wiest, O.; Barabási, A.-L.; Oltvai, Z.N.; Kapatral, V. Comparative genome-scale metabolic reconstruction nd flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. J Bacteriol 2009, 191, 4015–4024.
16. Raghunathan, A.; Reed, J.; Shin, S.; Palsson, B.; Daefler, S. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. BMC Syst Biol 2009, 3, 38.
17. Abbott, A. Germans cook up liver project. Biologists join physicists in a bid to map the workings of the human organ at all scales. Nature 2010, 468, 879.

DOI 10.1002/cplx