NETWORKS IN CELL BIOLOGY

R. Albert
Departments of Physics and Biology, Pennsylvania State University, USA

Keywords: networks, protein-protein interaction, cell biology

Contents

Summary

Cells use signaling and regulatory pathways connecting numerous constituents such as DNA, RNA, proteins and small molecules to coordinate multiple functions, allowing them to adapt to changing environments. High-throughput experimental methods enable the measurement of expression levels for thousands of genes and the determination of thousands of protein-protein or protein-DNA interactions. This abundance of information serves as a fertile ground for the application of network theory, and it is increasingly recognized that theoretical methods such as network inference, graph analysis and dynamic modeling are needed to make sense of the data. This chapter reviews three topics of special relevance for molecular and cellular interaction networks: graph inference (i.e. reconstructing the network of interactions among a set of biological entities), graph analysis (i.e. mining the information content of the network), and dynamic network modeling (i.e. connecting the interaction network to the dynamic behavior of the system).

1. Introduction

In order to understand the function of a cell or of higher units of biological organization, often it is beneficial to conceptualize them as systems of interacting elements. For such a systems-level description (which represents the main goal of systems biology), one needs to know (i) the identity of the components that constitute the biological system; (ii) the dynamic behavior of these components, i.e. how their abundance or activity changes over time in various conditions and (iii) the interactions among these components. Ultimately, this information can be combined into a biological network model that is not only consistent with current knowledge, but provides new insights and predictions such as the behavior of the system in conditions that were previously unexplored.

In some cases, the organization of the network of interactions underlying a biological system is straightforward (e.g. a linear chain of interactions), while in other cases a more formal representation, offered by graph theory, is required. The simplest possible graph representation reduces the system's elements to graph nodes (also called vertices) and reduces their pairwise relationships to edges (also called links) connecting pairs of nodes (see Figure 1). The nodes of (sub)cellular systems may be genes or mRNA-, protein- or other molecules. Directed edges (also called arcs) have a specified source (starting) node and target (end) node, and are most suited to represent chemical transformations and regulatory relationships. Non-directed edges are most appropriate for mutual interactions such as protein-protein binding or for relationships whose source and target are not yet distinguishable. Depending on the availability of information, edges are characterized by signs (positive for activation, negative for inhibition) or weights quantifying confidence levels, strengths, or reaction speeds. As the abundance of cellular constituents spans a large range and varies in time, nodes also need to be characterized by quantitative information describing the concentration of the corresponding molecules or the copy number of the corresponding mRNAs; this information is usually denoted as node state (or status).

This paper focuses on three topics related to molecular networks: graph inference, graph analysis, and dynamic network modeling. Graph inference refers to the problem in which the information on the identity and the state of a system's elements is used to infer interactions or functional relationships among these elements and to construct the interaction graph underlying the system. Graph analysis means the use of graph theory to analyze a known (complete or incomplete) interaction graph, and to extract new biological insights and predictions from the results. Dynamic network modeling aims to describe how known interactions among defined elements determine the time-course of the state of the elements, and of the whole system, under different conditions. A dynamic model that correctly captures experimentally observed normal behavior allows researchers to track the changes in the system's behavior due to perturbations. These three lines of inquiry are often combined in the literature since they provide three facets of the same objective: to understand, predict, and if possible control (tune toward a desired feature) the dynamic behavior of biological interacting systems.

2. Inference of Interaction Networks from Expression Information

The most prevalent use of graph inference is using gene/protein expression information to predict network structure, i.e. to predict which gene/protein influences which other genes/proteins through transcriptional or (post)translational regulation. A predicted regulatory relationship among two genes can be verified by experimental testing of the interactions and regulatory relationships among the two genes/proteins.

Genes with statistically similar (highly correlated) expression profiles in time or across several experimental conditions can be grouped using clustering algorithms. These methods give insight into groups of genes that respond in a similar manner to varying conditions, and that might therefore be co-regulated , however, that two nodes belong to the same group does not imply a causal relationship among them. The ability to extract meaning from clustering depends on the user's prior biological understanding of the objects that are organized. Most applications derive biological insight through "guilt by

association", that is, they predict the function of unknown gene products by their association with recognized clusters.

Data analysis methods such as principal component analysis aim to highlight the global patterns in the expression of a large number of genes/proteins by condensing the multivariate data into just two or three composite variables that capture the maximal co-variation between all the individual patterns. This method was used to link the level of 19 proteins involved in apoptotic signaling in human colon adenocarcinoma cells to four quantitative measures of apoptosis, leading to the prediction of cell-death responses to molecular perturbations and of the roles of key signaling intermediaries.

Bayesian methods aim to find a directed, acyclic (i.e. with no feedback loop) graph describing the causal dependency relationships among components of a system, and a set of local joint probability distributions that statistically convey these relationships. The starting edges are established heuristically based on an initial assessment of the experimental data, and are refined by an iterative search-and-score algorithm until the causal network and posterior probability distribution best describing the observed state of each node are found. Bayesian inference was recently used to infer the signaling network responsible for embryonic stem cell fate responses to external cues based on measurements of 28 signaling protein phosphorylation states across 16 different factorial combinations of stimuli. The inferred network predicted novel influences between ERK phosphorylation and differentiation as well as between RAF phosphorylation and differentiated cell proliferation.

Model-based methods of regulatory network inference from time-course expression data seek to relate the rate of change in the expression level of a given gene with the levels of other genes. Continuous methods postulate a system of differential equations, while discrete methods assume a logical (Boolean) relationship. Experimental data on gene expression levels is substituted into the relational equations, and the ensuing system of equations is then solved for the regulatory relationships between two or more components (see Figure 1). Because often there are far more biochemical components in the network than there are experimental time points, multiple networks will be possible solutions; these are filtered by making plausible assumptions on the objectives of the underlying system such as economy of regulation (reflected by having the fewest edges that satisfy the conditions) or maximal biomass production flux.

Several types of experimental results are best interpreted as indirect causal evidence that indicates the involvement of a protein or molecule in a certain process or pathway. Differential responses to a stimulus in wild-type organisms versus an organism where the respective protein's expression or activity is disrupted is an example of such indirect causal evidence connecting the stimulus, protein and response. These observations can be represented by two intersecting paths (successions of adjacent edges, see below) in the underlying interaction network: one connecting stimulus to response and the other connecting the protein to response. Graph-based inference algorithms integrate indirect causal relationships and direct interactions to find the most parsimonious network consistent with all available experimental observations. This method was used to reconstruct the signal transduction network corresponding to stomatal closure in plants in response to the stress hormone abscisic acid, and is implemented in the software NET-SYNTHESIS.

-
-
-

Bibliography

Albert, R. and A. L. Barabási (2002). Statistical mechanics of complex networks. Reviews of Modern Physics 74(1): 47-97.[A comprehensive review of network analysis and modeling at the time.]

Albert, R., B. DasGupta, R. Dondi, S. Kachalo, E. D. Sontag, A.Zelikovsky, K.Westbrooks . (2007). A novel method for signal transduction network inference from indirect experimental evidence. Journal of Computational Biology 14(7): 927-949.[Description of an algorithm and software for inference of signal transduction networks.]

Almaas, E., Z. N. Oltvai, A. L. Barabási (2005). The activity reaction core and plasticity of metabolic networks. PLoS Comput Biol 1(7): e68.[Identifies a connected metabolic core catalyzed by essential enzymes.]

Alon, U. (2006). An Introduction to Systems Biology: Design Principles of Biological Circuits.[Presents the design principles of biological systems, and highlights the recurring circuit elements that make up biological networks.]

Andrews, S. S. and A. P. Arkin (2006). Simulating cell biology. Curr Biol 16(14): R523-7.[Reviews modeling approaches with focus on spatial and stochastic models.]

Balázsi, G., A. L. Barabási, Z. N. Oltvai (2005). Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli. Proc Natl Acad Sci U S A 102(22): 7841-6.[Identifies regulatory subnetworks that originate at sensor transcription factors]

Bandyopadhyay, S., R. Sharan, T. Ideker (2006). Systematic identification of functional orthologs based on protein network comparison. Genome Res 16(3): 428-35.[Proposes a strategy for identifying functionally related proteins that supplements sequence-based comparisons with information on conserved protein-protein interactions.]

Barabási, A. L. and Z. N. Oltvai (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2): 101-13. [A comprehensive review of network analysis of molecular interaction networks.]

Bollobás, B. (1979). Graph theory: an introductory course. New York, Springer-Verlag. [An insightful introduction to graph theory with examples from electrical networks, flows, connectivity and matchings, random graphs.]

Chen, T., H. L. He, G. M. Church (1999). Modeling gene expression with differential equations. Pac Symp Biocomput: 29-40.[Proposes a differential equation framework for inferring gene regulatory networks from time-course expression data.]

Friedman, N., M. Linial, I. Nachman, D. Pe'er (2000). Using Bayesian networks to analyze expression data. J Comput Biol 7(3-4): 601-20.[Proposes a Bayesian probabilistic method to infer gene regulatory networks from expression data.]

Giot, L., J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L.Hao, C. E. Ooi, B.Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, Jr., K. P. White, M.Braverman, T. Jarvie,

S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant, J. M. Rothberg (2003). A protein interaction map of Drosophila melanogaster. Science 302(5651): 1727-36.[The first genome-wide protein interaction network of the model organism Drosophila melanogaster (fruit fly).]

Herrgard, M. J., B. S. Lee, V. Portnoy, B. O. Palsson. (2006). Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. Genome Res 16(5): 627-35.[Integrates a genome-scale metabolic network model with the transcriptional regulatory network of enzymes.]

Janes, K. A., J. G. Albeck,S. Gaudet, P. K. Sorger, D. A.Lauffenburger, M. B.Yaffe (2005). A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. Science 310(5754): 1646-53.[Uses principal component analysis to link the level of 19 proteins involved in apoptotic signaling in human colon adenocarcinoma cells to four quantitative measures of apoptosis.]

Jeong, H., S. P. Mason, A. L. Barabási, Z. N. Oltvai (2001). Lethality and centrality in protein networks. Nature 411(6833): 41-2.[Connects the centrality of proteins with the essentiality of their genes.]

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabási (2000). The large-scale organization of metabolic networks. Nature 407(6804): 651-4.[An early, comprehensive study of the topological properties of metabolic networks.]

Kauffman, S. A. (1993). The origins of order: self organization and selection in evolution. New York, Oxford University Press.[Proposes a conceptual model of gene regulatory networks through random Boolean networks.]

Kelley, B. P., B. Yuan, F. Lewitter, R.Sharan, B. R.Stockwell, T Ideker, T (2004). PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res 32(Web Server issue): W83-8.[Describes the PathBLAST software.]

Lee, I., S. V. Date, A. T. Adai, E. M. Marcotte . (2004). A probabilistic functional network of yeast genes. Science 306(5701): 1555-8.[Proposes a probabilistic conceptual framework for integrating diverse data to find functional linkages among genes.]

Lee, T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K.Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, R. A.Young (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298(5594): 799-804.[The first reconstruction of transcriptional regulatory networks in the model organism Saccharomyces cerevisiae (baker's yeast).]

Li, S., S. M. Assmann, R. Albert (2006). Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS Biol 4(10): e312.[A comprehensive study that integrates network inference and Boolean modeling into a predictive model.]

Ma'ayan, A., S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. J. Rice, A. Kershenbaum, G. A. Stolovitzky, R. D. Blitzer, R. Iyengar (2005). Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science 309(5737): 1078-83.[Demonstrates how network analysis can give insights into signal propagation.]

Palsson, B. O. (2006). Systems Biology: Properties of Reconstructed Networks, Cambridge University Press.[Describes how to reconstruct and analyze biochemical reaction networks.]

Prill, R. J., P. A. Iglesias, A. Levchenko (2005). Dynamic properties of network motifs contribute to biological network organization. PLoS Biol 3(11): e343.[Demonstrates that robustness to small perturbations is highly correlated with the relative abundance of network motifs.]

Shmulevich, I., E. R. Dougherty, eS. Kim, W. Zhang (2002). Probabilistic Boolean Networks: a rule based uncertainty model for gene regulatory networks. Bioinformatics 18(2): 261 74.[Proposes a probabilistic Boolean framework for the inference of gene regulatory networks.]

Tavazoie, S., J. D. Hughes, M. J. Campbell, R. J. Cho, G. M. Church . (1999). Systematic determination of genetic network architecture. Nat Genet 22(3): 281-5.[Uses clustering methods to uncovered co-regulated genes from gene expression data.]

Tyson, J. J., K. C. Chen, . B. Novak (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol 15(2): 221-31.[A clear and insightful

introduction to mathematical modeling of molecular interaction networks.]

Vazquez, A., A. Flammini, A. Maritan, A. Vespignani (2003). Global protein function prediction from protein-protein interaction networks. Nat Biotechnol 21(6): 697-700.[Proposes a method to assign proteins to functional classes on the basis of their interaction network.]

Wagner, A. and D. A. Fell (2001). The small world inside large metabolic networks. Proceedings of the Royal Society of London Series B-Biological Sciences 268(1478): 1803-1810.[An early study of the topology of metabolic networks.]

Woolf, P. J., W. Prudhomme, L. Daheron, G. Q. Daley, D. A. Lauffenburger (2005). Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. Bioinformatics 21(6): 741-53.[Uses a Bayesian framework to infer the signaling network responsible for embryonic stem cell fate responses.]

Yeger-Lotem, E., S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Y. Pinter, U. Alon, H. Margalit (2004). Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc Natl Acad Sci U S A 101(16): 5934-9.[Determines the over-represented network motifs integrating transcriptional and protein-level interactions.]

Yook, S. H., Z. N. Oltvai, A. L. Barabási (2004). Functional and topological characterization of protein interaction networks. Proteomics 4(4): 928-42.[A comparative analysis of protein interaction networks available at the time.]

Biographical Sketch

Réka Albert is Professor of Physics and Adjunct Professor of Biology at the Pennsylvania State University in the USA. She received her PhD in Physics from the University of Notre Dame, working on modeling complex networks with Prof. Albert-László Barabási. Her current interests focus on biological networks at the molecular, cellular and physiological level.