

COMPUTATIONAL BIOLOGY

Introduction to Network Analysis in Systems Biology

Avi Ma'ayan*

This Teaching Resource provides lecture notes, slides, and a problem set for a set of three lectures from a course entitled “Systems Biology: Biomedical Modeling.” The materials are from three separate lectures introducing applications of graph theory and network analysis in systems biology. The first lecture describes different types of intracellular networks, methods for constructing biological networks, and different types of graphs used to represent regulatory intracellular networks. The second lecture surveys milestones and key concepts in network analysis by introducing topological measures, random networks, growing network models, and topological observations from molecular biological systems abstracted to networks. The third lecture discusses methods for analyzing lists of genes and experimental data in the context of prior knowledge networks to make predictions.

Introduction

This Teaching Resource is intended for use by instructors and students who have some knowledge of network analysis and modeling. This advanced material is not appropriate for those with limited experience in network analysis. Familiarity with the following programs is also required for the problem set: Cytoscape, Pajek, Genes2Networks, and FANMOD. Network analysis of biological systems is increasingly gaining acceptance as a useful method for data integration and analysis. However, it is recognized that network representation of the complexity of biological systems is just the beginning, given that it only provides an overview of the system under investigation. Biological intracellular systems are dynamic, and the quantitative balance between components often determines biological outcome. Network models that include more details about the kinetics, localization, and quantities of the molecular components should aid in understanding cellular behavior and cellular regulation at the molecular level from a systems perspective.

Lecture Notes

Lecture 1: Representation of biological systems as networks

Network representation of intracellular biological networks typically considers molec-

ular components within a cell as nodes and their direct or indirect interactions as links (1) (Slide 2). Different types of intracellular molecular biological networks can be represented by different types of mathematical structures called graphs (2) (Slides 3 and 4). Some examples are (i) metabolic networks, for instance, the EcoCyc database provides access to metabolic networks across many organisms (3) (Slide 5); (ii) cell signaling networks (Slides 6 to 8), such as one representing pathways in mammalian neurons (4); (iii) kinase-substrate networks, such as the one we constructed for developing the tool kinase enrichment analysis (KEA) (5) (Slide 9); (iv) gene regulatory networks (Slides 10 and 11), such as the one described in the RegulonDB database for *Escherichia coli* (6); (v) protein-protein interaction networks (Slides 13 to 15), for instance, the human protein reference database (HPRD), which contains a large manually extracted data set describing mammalian protein-protein interactions (7); (vi) epistasis interaction networks, which are networks created by connecting genes if they exhibit a genetic interaction when knocked out or down-regulated (8) (Slides 16 and 17); (vii) disease gene interaction networks, which connect diseases to genes that when mutated cause or contribute to the disease (9) (Slide 20); and (viii) drug interaction networks in which drugs are linked to their targets (10, 11) (Slide 21).

There are many different types of graphs, including directed or undirected graphs (Slide 4), directed acyclic graphs (DAGs), trees, forests, minimum spanning trees, Boolean networks (Slide 8), and

Steiner trees (12, 13). For in-depth coverage of these data structures and the algorithms used to traverse and manipulate such structures, the classic computer science text book by Cormen *et al.* is recommended (14). Network representation enables integration of data from many different studies into a single framework. For example, Tanay *et al.* (15) showed how many different types of experimental data can be integrated using genes as anchors (Slide 22). Networks can be generated directly from time-series data or from perturbation data (Slides 18 and 19). The topology of regulatory networks can be “reverse engineered” directly from data tables of the changing quantities of mRNA expression or protein abundance over time or under different perturbations using Bayesian networks (16, 17) derived from advanced statistical learning techniques, or using tools, such as ARACNE, that use the concept of mutual information from information theory initially developed by Claude Shannon (18).

One of the challenges in network analysis is data visualization. One common method to visualize networks is the ball-and-stick diagram, and several useful software tools exist for creating these: Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>) (Slide 23), GraphViz (<http://www.graphviz.org/>), Cytoscape (19) (Slide 24), VisANT (20), SNAVI (21), AVIS (22), and yEd (<http://www.yworks.com/products/yed/>) are a few examples.

From this lecture, students should have an understanding that there are different type of representations for biological networks, that networks can be based on integration of multiple sources of published information or reconstructed directly from the data, that some biological networks can be connected to diseases or drugs, and that different data sets can be integrated through the abstraction to a network representation (Slide 25).

Lecture 2: Milestones and key concepts in network analysis

The field of graph theory, a subfield in mathematics, was first introduced by Euler, who used this approach to prove the famous Seven Bridges of Königsberg problem in 1736 (Slide 27). Major breakthroughs in the field of graph theory are attributed to Paul Erdos and Alfred Renyi, who studied the properties of random graphs (Slide 28). They simply asked: What are the consequences of throwing on a table a bunch of buttons and then randomly connecting them

Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, Mount Sinai School of Medicine, New York, NY 10029, USA.

*Contact information: E-mail, avi.maayan@mssm.edu

with links (Slide 28)? Erdos and Renyi are also famous for proving the Map Coloring problem, which answers the question, what is the minimum number of colors needed to distinctly color states on a map.

The seminal publications by Watts and Strogatz in 1998 (23) (Slides 29 to 31) and Barabási and Albert in 1999 (24) (Slides 32 to 34) popularized the notion that complex systems can be viewed as networks wherein components within a complex system can be represented as nodes and are linked through their interactions, which are called edges. This approach has been applied in many scientific disciplines, including systems biology and cell signaling research. Representing the complexity of biological regulatory systems as networks enables analysis of the network's topology, which provides insight into the organizational principles of the cell, achieved through evolution. Network topology includes information about the general and specific properties of nodes, properties of edges, properties of the entire network (global topological properties), and modules within the network.

Properties of nodes include connectivity degree, which is the number of links for each node; node betweenness centrality (25), which is the number of shortest paths that go through a node among all shortest paths between all possible pairs of nodes; closeness centrality, which is the average shortest path from one node to all other nodes; eigenvector centrality (26), which is a more sophisticated centrality measure that assesses the closeness to highly connected nodes; and bioinformatic analyses of the molecules represented by the nodes, for example, their Gene Ontology annotations (27), which describe the nodes' function, location in the cell, and involvement biological processes. Properties of edges include edge betweenness centrality (28), which is the number of number of shortest paths that go through an edge among all possible shortest paths between all the pairs of nodes; the types of relationship, for example, edges may represent activating or inhibiting relationships between a pair of nodes; and edge directionality, which indicates the upstream and downstream nodes connected by a particular link. Analyses based on the type of interactions, for example, phosphorylation, binding, gene regulation, or other types as specified in the *Science Signaling* Connections Maps in the Database of Cell Signaling (29), or those proposed by the Edge Ontology (30) can be used to further

characterize the network's topology. Global topological characteristics of networks include connectivity distribution (24), which is represented by a histogram showing how many nodes have a certain amount of links; characteristic path length (23), which can be computed using Floyd-Warshall's or Dijkstra's algorithms and represents the average shortest path between all pairs of nodes; clustering coefficient (23), which represents the local density of interactions by measuring the connectivity of neighbors for each node averaged over the entire network; grid coefficient, which extends the clustering coefficient from only looking at first neighbors to also examining second neighbors; network diameter, which represents the longest shortest path; and assortativity, which assesses whether nodes prefer to attach to other nodes on the basis of common nodal properties (31). Networks can also have recurring circuits of a few nodes and their edges, which are called network motifs. Network motifs are recurring circuits composed of a few nodes and their edges (Slide 38), and these types of circuits appear in the topology of biological regulatory networks much more frequently than they are observed in random or shuffled networks (4, 32, 33). Some motifs are particularly important because they directly influence a system's overall dynamics; examples include feedback loops (34), feedforward loops (35), bifans (36, 37), and other types of cycles (38). Network motifs can be identified within directed or oriented graphs or in undirected networks, which are called graphlets (39) (Slide 39), and a biological example of graphlets includes motifs present in protein-protein interaction networks.

Another characteristic of networks is their modularity, which represents the modules, or network clusters, which are dense areas of connectivity separated by regions of low connectivity. Modules can be found using unsupervised clustering algorithms, such as nearest neighbors clustering, Markov clustering, and betweenness centrality-based clustering, which uses nodes with high betweenness centrality and low connectivity to separate clusters (25).

Understanding the structural organization of biological networks using these and other topology analysis measures gives clues to the evolutionary processes that may produce the observed topology of biological regulatory networks. Network evolution algorithms attempt to recreate realistic topologies on the basis of a few simple rules gov-

erning network growth. Network evolution models can help to intuitively understand the design principles embedded within complex biological regulatory networks. Some of these algorithms include rich-get-richer (24) (Slide 34), growth by duplication-divergence (40) (Slide 36), exponential growth (24), and geometric growth (41) (Slide 37). Other models start with a random network, or a very regular structured network where the network gradually evolves to a realistic topology based on rules that reposition links (23, 42) (Slides 31 and 47).

To determine whether the topological properties observed in real networks deviate from random connectivity and are not observed by chance, topological observations can be compared to Erdos-Renyi random networks (43) (Slide 28) or other types of shuffled networks (44), which serve as statistical controls.

By applying various topological analyses, general properties of biological regulatory networks emerge. Most biological molecular regulatory networks are scale-free (Slide 32), meaning that the connectivity distribution, the distribution of edges per node, fits a power law (24). The scale-free architecture makes the networks robust to random failures (45). Biological networks are also small-world, meaning that they are highly clustered, with shortcuts that connect the clusters (23) (Slide 29). Biological networks have "party" hubs and "date" hubs (46): Party hubs are nodes that interact with many proteins at one cellular compartment at a specific time, whereas date hubs are proteins that can be found in many places inside the cell and interact with various partners at different times. Hubs are also categorized as multisite or single-site (47) (Slide 40). Some cell signaling networks have a bow-like structure in which many receptors (one loop of the bow) share few downstream adaptor proteins (the knot in the bow) (48), and these central adaptor nodes integrate information from the many receptors and then distribute the information to many effectors in multiple cell signaling pathways (the other loop of the bow) (Slide 41). The topology of regulatory biochemical networks contains fewer feedback loops than observed in random Erdos-Renyi networks. In addition, the feedback loops tend to be nested (38). In the topology of gene and cell signaling networks, bifan motifs are the most abundant (32, 36), probably due to evolution through duplication and divergence (40) (Slide 43).

By examining the structure of nine directed networks from the real world (six biological and three human-made complex systems (38)), we identified, counted, and characterized the cycles (the closed loop pathways) in these networks. Most cycles in real networks are made of source and sink nodes, which deplete feedback loops (Slide 44). Such topology contributes to optimal design for dynamical stability (Slide 45). We also developed a network evolution model that starts with a random directed Erdos-Renyi-type network (42). The network evolution process reassigns links if they contribute to a local increase of sources and sinks. After several evolutionary steps, the random network becomes dynamically stable and displays a power-law, scale-free, connectivity distribution. By adding a simple rule, the evolution can continue forever, maintaining the network scale-free structure, as well as displaying dynamics that are at the cusp between stability and chaos. In addition, the hubs, the highly connected nodes in the network, rise and fall in their connectivity while the global properties of the network remain (Slide 46). Hence, the simple theoretical network model captures dynamical evolutionary features observed in real complex systems.

From this lecture, students should gain a basic understanding of the history of graph theory and how graph theory was used to characterize the topological properties of biological networks. Biological networks have distinct topologies, and this influences the system's dynamical behavior (Slide 48).

Lecture 3: Making predictions using network analysis

The networks described in the first lecture can be used to analyze new experimental data in the context of known biology. Different algorithms can be applied to predict or find “connections” between lists of “seed” nodes using prior knowledge of molecular interactions. For example, genes or proteins that are differentially regulated under control versus treatment conditions could be used as seed nodes for building functional networks using information from prior publications (49, 50). The algorithms that can be used to find connections between genes, proteins, or both, using information from background networks to form subnetworks, include mean-first-pass-time (MFPT) (51), nearest neighbor expansion, Steiner trees (12, 13, 52), and the shortest path search algorithm (53). We used the shortest path approach to predict key regulators of neurite

outgrowth in Neuro2A cells stimulated with a cannabinoid agonist (50) (Slides 53 to 61) and to identify a previously unknown Noonan syndrome gene (54) (Slides 62 and 63). We also applied the Steiner tree approach to minimally connect up-regulated genes in breast cancer, as well as a Steiner tree to connect up-regulated genes in colorectal cancer (13) (Slide 64). The Steiner tree approach was also used by Huang and Fraenkel to connect signaling pathways to transcription factors (52) (Slide 65), and others have used a similar approach to connect commonly expressed genes in stem cells (55) (Slide 66). Such methods are useful for finding functional connections between nodes and predicting additional nodes that had not been detected experimentally. Similarly, network analyses that utilize background knowledge interaction networks can be used to predict interactions. For example, protein-protein interactions can be predicted by completing circuits that appear to have missing links in clusters within the network structure (56) or by combining clusters in the network structure with function of nodes in clusters (57) (Slides 51 and 52). Not only can networks or portions of networks be predicted, but information about the molecules within the network can be inferred or predicted from various properties of the network and its constituents. For example, protein function can be predicted on the basis of known protein-protein interactions and the assumption that proteins close in network space are likely to share functions (58) (Slide 50).

Making predictions using network analysis and using Gene Set Enrichment Analysis (GSEA) (59) [see also the Teaching Resource by Clark and Ma'ayan on GSEA that is part of this course (60)] are conceptually related. Gene sets can be converted to networks, and networks can be converted to gene sets. For example, one can create a network by connecting genes if they are targeted by the same microRNAs, share many Gene Ontology terms, or are found in many pathways together. One can also create gene sets from networks. For instance, we have created a gene set of all genes encoding proteins in the human interactome that interact with at least 50 other proteins, which serves as a resource gene set library for the gene set enrichment analysis tool Lists2Networks (61). We also created enrichment analysis tools for kinase enrichment analysis (KEA) (62) (Slide 67) and for chip-X enrichment analysis (ChEA) (63) (Slides 68 and 69) by integrating net-

works of kinase-substrate interactions for KEA or transcription factor-target interactions from chip-seq and chip-chip studies for ChEA. The resulting directed graphs, representing kinase-substrate interactions and transcription factor-target interactions, respectively, were converted to gene set libraries for the purpose of performing gene set enrichment analysis.

From this lecture, students should have an understanding of the algorithms that can be used to seed genes and proteins within biological networks and how network analysis can drive hypothesis generation (Slide 70).

Problem Set

The following problem sets provide reinforcement of the concepts presented in the lecture and do not require sophisticated computational skills. The software systems Cytoscape, Pajek, Genes2Networks, and FANMOD required to complete the assignments are freely available from the Web and are cross-platform capable to run on Windows, Linux, or Mac systems (except Pajek, which only runs on Windows).

1. Write, using your own words, definitions for the following terms: epistasis interaction networks, minimum spanning tree, Steiner tree, mean-first-pass-time (MFPT), betweenness centrality, closeness centrality, characteristic path length, clustering coefficient, and grid coefficient. For each term, describe how it is relevant to understanding intracellular regulatory networks.

2. Using the supporting table S3 from de Chassey *et al.* (64), create a network of human-HCV protein-protein interaction visualized using Cytoscape or Pajek. Make the nodes appear in two colors, viral proteins in one color, and human proteins in another color. Answer the following two questions:

Which host proteins interact with many HCV proteins?

Which HCV proteins show the most interactions with the host proteins?

Note: User manuals for using Cytoscape or Pajek are available at http://www.cytoscape.org/manual/Cytoscape2_2Manual.pdf and <http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf>.

3. Cordeddu *et al.* (54) identified and characterized a novel gene that, when mutated, can cause Noonan-like syndrome. *SHOC2* was discovered using the program Genes2Networks (49) by entering into the program the previously known Noonan syndrome disease genes *KRAS*, *RAF1*, *SOS1*, and *PTPN11*. Using the Genes2Networks program from

<http://actin.pharm.mssm.edu/genes2networks/>, enter these four genes and reproduce the ranked list of predicted disease genes that was reported in supporting table S1 from Cordeddu *et al.* (54). You will need to deselect the ncbi_hprd and mammalian_kegg databases from the list of databases included in the databases option on the left-side menu.

Using a similar approach, select a disease from the OMIM Morbid Map database <http://www.ncbi.nlm.nih.gov/Omim/getmorbidity.cgi> and enter the list of disease genes for that disease into Genes2Networks. Report the ranked list of predicted disease genes for the disease you chose. Using the results, answer the following questions:

For what type of diseases should this method work well in identifying disease-associated genes?

For which types of diseases is this method likely to fail?

4. Zaidel-Bar *et al.* published a signaling network made of the molecular components of focal adhesions and their interactions (65). Find the network motif profile of this network by downloading the adhesome network from <http://www.adhesome.org/download/fa.txt> and loading it into FANMOD (<http://theinfl.informatik.uni-jena.de/~wernicke/motifs/index.html>) for analysis. Explain why some of the motifs are overrepresented in this network. Are the results consistent with other networks analyzed and reported in Milo *et al.* (66)?

5. Fill in the blanks in the table shown in Fig. 1.

Educational Details

Learning resource type: Lecture, assignment, digital presentation

Context: Graduate

Intended users: Teacher, learner

Intended educational use: Learn, plan, teach

Discipline: Biochemistry, biocomplexity; bioinformatics, genomics and proteomics, biostatistics, biotechnology, cell biology, molecular biology, pharmacology, proteomics,

systems biology, theoretical biology

Keywords: Cell signaling, computational biology, gene regulatory networks, graph theory, interactome, networks, pathway analysis

Technical Details

Software: Cytoscape

Requirements: Platform-independent, open-source Java application that is free for noncommercial use

Download: http://www.cytoscape.org/download.php?file=cyto2_6_3

Software: Pajek

Requirements: Runs on Windows and is free for noncommercial use

Download: <http://pajek.imfm.si/doku.php?id=download>

Software: Genes2Networks

Requirements: Platform-independent, Web-based software that is free for noncommercial use

Download: <http://actin.pharm.mssm.edu/genes2networks/>

Software: FANMOD

Requirements: Platform-independent, Web-based software that is free for noncommercial use


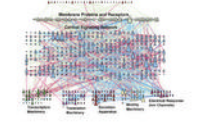
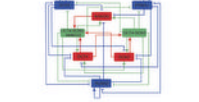
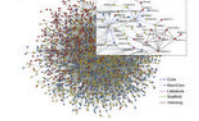



Study	Type of network	Type of graph	Nodes	Links	Method used to build network
 PMID: 17608928					
 PMID: 16099987					
 PMID: 18769478					
 PMID: 14704431					
 PMID: 11743205					
 PMID: 17516560					
 PMID: 17502601					

Fig. 1. Sample table for completing question 5 of the problem set.

Download: <http://theinfl.informatik.uni-jena.de/~wernicke/motifs/index.html>

Supplementary Materials

<http://stke.sciencemag.org/cgi/content/full/sigtrans;4/190/tr5/DC1>

Slides: Introduction to Network Analysis in Systems Biology

Problem set key is available upon request

References and Notes

- D. Eisenberg, E. M. Marcotte, I. Xenarios, T. O. Yeates, Protein function in the post-genomic era. *Nature* **405**, 823–826 (2000).
- A. Ma'ayan, R. D. Blitzer, R. Iyengar, Toward predictive models of mammalian cells. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 319–349 (2005).
- P. D. Karp, M. Riley, S. M. Paley, A. Pelligrini-Toole, EcoCyc: An encyclopedia of Escherichia coli genes and metabolism. *Nucleic Acids Res.* **24**, 32–39 (1996).
- A. Ma'ayan, 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, Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science* **309**, 1078–1083 (2005).
- A. Lachmann, A. Ma'ayan, KEA: kinase enrichment analysis. *Bioinformatics* **25**, 684–686 (2009).
- H. Salgado, S. Gama-Castro, M. Peralta-Gil, E. Díaz-Peredo, F. Sánchez-Solano, A. Santos-Zavaleta, I. Martínez-Flores, V. Jiménez-Jacinto, C. Bonavides-Martínez, J. Segura-Salazar, A. Martínez-Antonio, J. Collado-Vides, RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.* **34** (Database issue), D394–D397 (2006).
- T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahman, S. Mohan, P. Rangathan, S. Ramabadran, R. Chaerkady, A. Pandey, Human Protein Reference Database—2009 update. *Nucleic Acids Res.* **37** (Database issue), D767–D772 (2009).
- D. Segrè, A. Deluna, G. M. Church, R. Kishony, Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2005).
- K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, A. L. Barabási, The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8685–8690 (2007).
- M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabási, M. Vidal, Drug-target network. *Nat. Biotechnol.* **25**, 1119–1126 (2007).
- A. Ma'ayan, S. L. Jenkins, J. Goldfarb, R. Iyengar, Network analysis of FDA approved drugs and their targets. *Mt. Sinai J. Med.* **74**, 27–32 (2007).
- M. T. Dittrich, G. W. Klau, A. Rosenwald, T. Danker, T. Müller, Identifying functional modules in protein-protein interaction networks: An integrated exact approach. *Bioinformatics* **24**, i223–i231 (2008).
- A. G. White, A. Ma'ayan, Connecting seed lists of mammalian proteins using Steiner Trees. In: *Signals, Systems and Computers*, 2007 ACSSC Conference Record of the Forty-First Asilomar Conference, 4–7 Nov 2007 (IEEE, 2007), 155–159.
- T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to Algorithms, Second Edition* (MIT Press, Cambridge, MA, 2001).
- A. Tanay, R. Sharan, R. Kupiec, R. Shamir, Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2981–2986 (2004).
- J. Pearl, *Causality* (Cambridge Univ. Press, Cambridge, 2000).
- K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, G. P. Nolan, Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308**, 523–529 (2005).
- A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, A. Califano, ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7** (suppl. 1), S7 (2006).
- P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Z. Hu, E. S. Snitkin, C. DeLisi, VisANT: An integrative framework for networks in systems biology. *Brief. Bioinform.* **9**, 317–325 (2008).
- A. Ma'ayan, S. L. Jenkins, R. L. Webb, S. I. Berger, S. P. Purushothaman, N. S. Abul-Husn, J. M. Posner, T. Flores, R. Iyengar, SNAVI: Desktop application for analysis and visualization of large-scale signaling networks. *BMC Syst. Biol.* **3**, 10 (2009).
- S. I. Berger, R. Iyengar, A. Ma'ayan, AVIS: AJAX viewer of interactive signaling networks. *Bioinformatics* **23**, 2803–2805 (2007).
- D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- A. L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- M. E. Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **64**, 016132 (2001).
- J. L. Morrison, R. Breitling, D. J. Higham, D. R. Gilbert, GeneRank: Using search engine technology for the analysis of microarray experiments. *BMC Bioinformatics* **6**, 233 (2005).
- M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. WhiteGene Ontology Consortium, The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32** (Database issue), D258–D261 (2004).
- J. Yoon, A. Blumer, K. Lee, An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics* **22**, 3106–3108 (2006).
- N. R. Gough, Science's signal transduction knowledge environment: The connections maps database. *Ann. N.Y. Acad. Sci.* **971**, 585–587 (2002).
- L. J. Lu, A. Sboner, Y. J. Huang, H. X. Lu, T. A. Gianoulis, K. Y. Yip, P. M. Kim, G. T. Montelione, M. B. Gerstein, Comparing classical pathways and modern networks: towards the development of an edge ontology. *Trends Biochem. Sci.* **32**, 320–331 (2007).
- M. E. J. Newman, Mixing patterns in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **67**, 026126 (2003).
- R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).
- S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68 (2002).
- D. Angeli, J. E. Ferrell Jr, E. D. Sontag, Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1822–1827 (2004).
- S. Mangan, A. Zaslaver, U. Alon, The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.* **334**, 197–204 (2003).
- A. Lipshtat, S. P. Purushothaman, R. Iyengar, A. Ma'ayan, Functions of bifans in context of multiple regulatory motifs in signaling networks. *Biophys. J.* **94**, 2566–2579 (2008).
- P. J. Ingram, M. P. Stumpf, J. Stark, Network motifs: structure does not determine function. *BMC Genomics* **7**, 108 (2006).
- A. Ma'ayan, G. A. Cecchi, J. Wagner, A. R. Rao, R. Iyengar, G. Stolovitzky, Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19235–19240 (2008).
- N. Przulj, D. G. Corneli, I. Jurisica, Efficient estimation of graphlet frequency distributions in protein-protein interaction networks. *Bioinformatics* **22**, 974–980 (2006).
- S. A. Teichmann, M. M. Babu, Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492–496 (2004).
- D. J. Higham, M. Rasajski, N. Przulj, Fitting a geometric graph to a protein-protein interaction network. *Bioinformatics* **24**, 1093–1099 (2008).
- B. D. MacArthur, R. J. Sánchez-García, A. Ma'ayan, Microdynamics and criticality of adaptive regulatory networks. *Phys. Rev. Lett.* **104**, 168701 (2010).
- P. Erdos, A. Rényi, On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kutatóintez. Közl.* **5**, 17–61 (1960).
- S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
- R. Cohen, K. Erez, D. Ben-Avraham, S. Havlin, Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).
- J. D. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. Walhout, M. E. Cusick, F. P. Roth, M. Vidal, Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
- P. M. Kim, L. J. Lu, Y. Xia, M. B. Gerstein, Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* **314**, 1938–1941 (2006).
- K. Oda, H. Kitano, A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol.* **2**, 2006, 0015 (2006).

49. S. I. Berger, J. M. Posner, A. Ma'ayan, Genes2Networks: Connecting lists of gene symbols using mammalian protein interactions databases. *BMC Bioinformatics* **8**, 372 (2007).
50. K. D. Bromberg, A. Ma'ayan, S. R. Neves, R. Iyengar, Design logic of a cannabinoid receptor signaling network that triggers neurite outgrowth. *Science* **320**, 903–909 (2008).
51. J. D. Noh, H. Rieger, Random walks on complex networks. *Phys. Rev. Lett.* **92**, 118701 (2004).
52. S. S. Huang, E. Fraenkel, Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci. Signal.* **2**, ra40 (2009).
53. E. W. Dijkstra, A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
54. V. Cordeddu, E. Di Schiavi, L. A. Pennacchio, A. Ma'ayan, A. Sarkozy, V. Fodale, S. Cecchetti, A. Cardinale, J. Martin, W. Schackwitz, A. Lipzen, G. Zampino, L. Mazzanti, M. C. Digilio, S. Martinnelli, E. Flex, F. Lepri, D. Bartholdi, K. Kutsche, G. B. Ferrero, C. Anichini, A. Selicorni, C. Rossi, R. Tenconi, M. Zenker, D. Merlo, B. Dallapiccola, R. Iyengar, P. Bazzicalupo, B. D. Gelb, M. Tartaglia, Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. *Nat. Genet.* **41**, 1022–1026 (2009).
55. F. J. Müller, L. C. Laurent, D. Kostka, I. Ulitsky, R. Williams, C. Lu, I. H. Park, M. S. Rao, R. Shamir, P. H. Schwartz, N. O. Schmidt, J. F. Loring, Regulatory networks define phenotypic classes of human stem cell lines. *Nature* **455**, 401–405 (2008).
56. H. Yu, A. Paccanaro, V. Trifonov, M. Gerstein, Predicting interactions in protein networks by completing defective cliques. *Bioinformatics* **22**, 823–829 (2006).
57. I. Albert, R. Albert, Conserved network motifs allow protein-protein interaction prediction. *Bioinformatics* **20**, 3346–3352 (2004).
58. R. Sharan, I. Ulitsky, R. Shamir, Network-based prediction of protein function. *Mol. Syst. Biol.* **3**, 88 (2007).
59. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
60. N. R. Clark, A. Ma'ayan, Introduction to statistical methods for analyzing large data sets: Gene Set Enrichment Analysis (GSEA). *Sci. Signal.* **4**, tr3 (2011).
61. A. Lachmann, A. Ma'ayan, Lists2Networks: Integrated analysis of gene/protein lists. *BMC Bioinformatics* **11**, 87 (2010).
62. A. Lachmann, A. Ma'ayan, KEA: Kinase enrichment analysis. *Bioinformatics* **25**, 684–686 (2009).
63. A. Lachmann, H. Xu, J. Krishnan, S. I. Berger, A. R. Mazloom, A. Ma'ayan, ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
64. B. de Chasse, V. Navratil, L. Tafforeau, M. S. Hiet, A. Aublin-Gex, S. Agaugué, G. Meiffren, F. Pradezynski, B. F. Faria, T. Chantier, M. Le Breton, J. Pellet, N. Davoust, P. E. Mangeot, A. Chaboud, F. Penin, Y. Jacob, P. O. Vidalain, M. Vidal, P. André, C. Rabourdin-Combe, V. Lotteau, Hepatitis C virus infection protein network. *Mol. Syst. Biol.* **4**, 230 (2008).
65. R. Zaidel-Bar, S. Itzkovitz, A. Ma'ayan, R. Iyengar, B. Geiger, Functional atlas of the integrin adhesion. *Nat. Cell Biol.* **9**, 858–867 (2007).
66. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
67. **Acknowledgments:** A.M. thanks S. L. Jenkins for comments and suggestions and Y.-S. Lee for validating the problem sets. **Funding:** This work was supported by NIH grants 5P50GM071558-03, 1R01DK088541-01A1, KL2RR029885-0109, and RC2OD006536-01.

10.1126/scisignal.2001965

Citation: A. Ma'ayan, Introduction to network analysis in systems biology. *Sci. Signal.* **4**, tr5 (2011).

Introduction to Network Analysis in Systems Biology

Avi Ma'ayan

Sci. Signal. **4** (190), tr5.

DOI: 10.1126/scisignal.2001965 originally published online September 6, 2011

ARTICLE TOOLS

<http://stke.sciencemag.org/content/4/190/tr5>

SUPPLEMENTARY MATERIALS

<http://stke.sciencemag.org/content/suppl/2011/09/12/4.190.tr5.DC1>

RELATED CONTENT

<http://stke.sciencemag.org/content/sigtrans/4/190/tr2.full>

REFERENCES

This article cites 63 articles, 18 of which you can access for free
<http://stke.sciencemag.org/content/4/190/tr5#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Signaling (ISSN 1937-9145) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Signaling* is a registered trademark of AAAS.