Scale-free networks in cell biology

Réka Albert

Department of Physics and Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA
(e-mail: ralbert@phys.psu.edu)

Summary

A cell's behavior is a consequence of the complex interactions between its numerous constituents, such as DNA, RNA, proteins and small molecules. Cells use signaling pathways and regulatory mechanisms to coordinate multiple processes, allowing them to respond to and adapt to an ever-changing environment. The large number of components, the degree of interconnectivity and the complex control of cellular networks are becoming evident in the integrated genomic and proteomic analyses that are emerging. It is increasingly recognized that the understanding of properties that arise from whole-cell function require integrated, theoretical descriptions of the relationships between different cellular components.

Recent theoretical advances allow us to describe cellular network structure with graph concepts and have revealed organizational features shared with numerous non-biological networks. We now have the opportunity to describe quantitatively a network of hundreds or thousands of interacting components. Moreover, the observed topologies of cellular networks give us clues about their evolution and how their organization influences their function and dynamic responses.

Key words: Protein-protein interactions, Signal transduction, Transcriptional regulatory networks, Metabolic networks, Network modeling, Systems biology

Introduction

Genes and gene products interact on several levels. At the genomic level, transcription factors can activate or inhibit the transcription of genes to give mRNAs. Since these transcription factors are themselves products of genes, the ultimate effect is that genes regulate each other's expression as part of gene regulatory networks. Similarly, proteins can participate in diverse post-translational interactions that lead to modified protein functions or to formation of protein complexes that have new roles; the totality of these processes is called a protein-protein interaction network. The biochemical reactions in cellular metabolism can likewise be integrated into a metabolic network whose fluxes are regulated by enzymes catalyzing the reactions. In many cases these different levels of interaction are integrated – for example, when the presence of an external signal triggers a cascade of interactions that involves both biochemical reactions and transcriptional regulation.

A system of elements that interact or regulate each other can be represented by a mathematical object called a graph (Bollobás, 1979). Here the word 'graph' does not mean a 'diagram of a functional relationship' but 'a collection of nodes and edges', in other words, a network. At the simplest level, the system's elements are reduced to graph nodes (also called vertices) and their interactions are reduced to edges connecting pairs of nodes (Fig. 1). Edges can be either directed, specifying a source (starting point) and a target (endpoint), or non-directed. Directed edges are suitable for representing the flow of material from a substrate to a product in a reaction or the flow of information from a transcription factor to the gene whose transcription it regulates. Non-directed edges are used to represent mutual interactions, such as protein-protein binding. Graphs can be augmented by assigning various attributes to the nodes and edges; multi-partite graphs allow representation of different classes of node, and edges can be characterized by signs (positive for activation, negative for inhibition), confidence levels, strengths, or reaction speeds. Here I aim to show how graph representation and analysis can be used to gain biological insights through an understanding of the structure of cellular interaction networks. For information on other important related topics, such as computational methods of network inference and mathematical modeling of the dynamics of cellular networks, several excellent review articles are available elsewhere (Friedman, 2004; Longabaugh et al., 2005; Ma'ayan et al., 2004; Papin et al., 2005; Tyson et al., 2003).

Graph concepts: from local to long-range

The nodes of a graph can be characterized by the number of edges that they have (the number of other nodes to which they are adjacent). This property is called the **node degree**. In directed networks we distinguish the in-degree, the number of directed edges that point toward the node, and the out-degree, the number of directed edges that start at the node. Whereas node degrees characterize individual nodes, one can define a **degree distribution** to quantify the diversity of the whole network (Fig. 1). The degree distribution $P(k)$ gives the fraction of nodes that have degree k and is obtained by counting the number of nodes $N(k)$ that have $k = 1, 2, 3…$ edges and dividing it by the total number of nodes N. The degree distributions of numerous networks, such as the Internet, human collaboration networks and metabolic networks, follow a well-defined functional form $P(k) = Ak^{-\gamma}$ called a power law. Here, A is a constant that ensures that the $P(k)$ values add up to 1, and the degree exponent γ is usually in the range $2 < \gamma < 3$ (Albert and
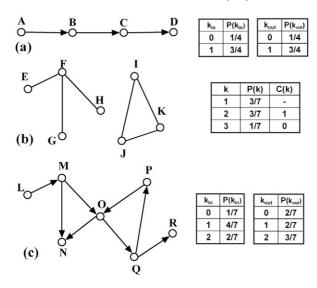
Fig. 1. Graph representation and graph analysis reveals regulatory patterns of cellular networks. The number of interactions a component participates in is quantified by its (in/out) degree, for example node O has both in-degree and out-degree 2. The clustering coefficient characterizes the cohesiveness of the neighborhood of a node – for example the clustering coefficient of I is 1, indicating that it is part of a three-node clique. The graph distance between two nodes is defined as the number of edges in the shortest path between them. For example, the distance between nodes P and O is 1, and the distance between nodes O and P is 2 (along the OQP path). The degree distribution $P(k)$ [$P(k_{in})$ and $P(k_{out})$ in directed networks] quantifies the fraction of nodes with degree k, while the clustering-degree function $C(k)$ gives the average clustering coefficient of nodes with degree k. (a) A linear pathway can be represented as a succession of directed edges connecting adjacent nodes. Because there are no shortcuts or feedbacks in a linear pathway, the distance between the starting node and end node increases linearly with the number of nodes. The in-degree and out-degree distribution indicates the existence of a source ($k_{in}=0$) and a sink ($k_{out}=0$) node. (b) This undirected and disconnected graph is composed of two connected components (EFGH and IJK), has a range of degrees from 1 to 3 and a range of clustering coefficients from 0 (for F) to 1 (for I, J and K). The connected component IJK is also a clique (completely connected subgraph) of three nodes. (c) This directed graph contains a feed-forward loop (MON) and a feedback loop (POQ), which is also the largest strongly connected component of the graph. The in-component of this graph contains L and M, while its out-component consists of the sink nodes N and R. The source node L can reach every other node in the network.

Barabási, 2002). This function indicates that there is a high diversity of node degrees and no typical node in the network that could be used to characterize the rest of the nodes (Fig. 2). The absence of a typical degree (or typical scale) is why these networks are described as '**scale-free**'.

The cohesiveness of the neighborhood of a node i is usually quantified by the **clustering coefficient** C_i, defined as the ratio between the number of edges linking nodes adjacent to i and the total possible number of edges among them (Watts and Strogatz, 1998). In other words, the clustering coefficient quantifies how close the local neighborhood of a node is to being part of a **clique**, a region of the graph (a subgraph) where every node is connected to every other node. Various networks, including protein interaction and metabolic networks (Wagner and Fell, 2001; Yook et al., 2004), display a high average clustering coefficient, which indicates a high level of redundancy and cohesiveness. Averaging the clustering coefficients of nodes that have the same degree k gives the function $C(k)$, which characterizes the diversity of cohesiveness of local neighborhoods (Fig. 1). Several measurements indicate a decreasing $C(k)$ in metabolic networks (Ravasz et al., 2002) and protein interaction networks (Yook et al., 2004), following the relationship $C(k) = B/k^{\beta}$ (where B is a constant and β is between 1 and 2). This suggests that low-degree nodes tend to belong to highly cohesive neighborhoods whereas higher-degree nodes tend to have neighbors that are less connected to each other.

Two nodes of a graph are connected if a sequence of adjacent nodes, a **path**, links them (Bollobás, 1979). A path can thus signify a transformation route from a nutrient to an end-product in a metabolic network, or a chain of post-translational reactions from the sensing of a signal to its intended target in a signal transduction network. The **graph distance** (also called path length) between two nodes is defined as the number of edges along the shortest path connecting them. If edges are characterized by the speed or efficiency of information propagation along them, the concept can be extended to signify, for example, the path with shortest delay (Dijkstra, 1959). In most networks observed, there is a relatively short path between any two nodes, and its length is in the order of the logarithm of the network size (Albert and Barabási, 2002; Newman, 2003b). This **small world** property appears to characterize most complex networks, including metabolic and protein interaction networks. If a path connects each pair of nodes, the graph is said to be **connected**; if this is not the case, one can find connected components, graph regions (subgraphs) that are connected (Fig. 1).

The connectivity structure of directed graphs presents special features, because the path between two nodes i and j can be different when going from i to j or vice versa (Fig. 1). Directed graphs can have one or several strongly connected components, subgraphs whose nodes are connected in both directions; **in-components**, which are connected to the nodes in the strongly connected component but not vice versa; and **out-components**, which can be reached from the strongly connected component but not vice versa. It is important to note that this topological classification reflects functional separation in signal transduction and metabolic networks. For example, the regulatory architecture of a mammalian cell (Ma'ayan et al., 2004) has ligand-receptor binding as the in-component, a central signaling network as the strongly connected component and the transcription of target genes and phenotypic changes as part of the out-component.

The **source nodes** of directed cellular networks (the nodes that only have outgoing edges) can be regarded as their inputs. For example, the substrates consumed from the environment (and not synthesized by the cell) constitute the inputs of a metabolic network, extracellular ligands or their receptors are the sources of signal transduction networks (Ma'ayan et al., 2005), and environmentally (but not transcriptionally) regulated transcription factors constitute the sources of transcriptional networks (Balázsi et al., 2005). Following the paths starting from each source node will reveal a subgraph (termed **origon** in the context of transcriptional networks whose nodes can

potentially be influenced by functional changes in the source node.

Graph models

To understand how the above-defined graph measures reflect the organization of the underlying networks, we should first consider some representative graph families that have had a significant impact on network research (Barabási and Oltvai, 2004; Newman, 2003b).

A **linear pathway** has a well-defined source, a chain of intermediary nodes, and a sink (end) node. The clustering coefficient of each node is zero, because there are no edges among first neighbors. Both the maximum and average path length increase linearly with the number of nodes and are long for pathways that have many nodes (Fig. 1a). This type of graph has been widely used as a model of an isolated signal transduction pathway.

Random graphs, constructed by randomly connecting a given number N of nodes by E edges, reflect the (statistically) expected properties of a network of this size (Bollobás, 1985). They have a bell-shaped degree distribution (Fig. 2), indicating that the majority of nodes have a degree close to the average degree $<k>$. The average clustering coefficient of a random graph equals $<k>/N$ and thus is very small for large N (Albert and Barabási, 2002). Also, the $C(k)$ function is a constant,

Fig. 2. Comparison between the degree distribution of scale-free networks (○) and random graphs (□) having the same number of nodes and edges. For clarity the same two distributions are plotted both on a linear (left) and logarithmic (right) scale. The bell-shaped degree distribution of random graphs peaks at the average degree and decreases fast for both smaller and larger degrees, indicating that these graphs are statistically homogeneous. By contrast, the degree distribution of the scale-free network follows the power law $P(k) = Ak^{-3}$, which appears as a straight line on a logarithmic plot. The continuously decreasing degree distribution indicates that low-degree nodes have the highest frequencies; however, there is a broad degree range with non-zero abundance of very highly connected nodes (hubs) as well. Note that the nodes in a scale-free network do not fall into two separable classes corresponding to low-degree nodes and hubs, but every degree between these two limits appears with a frequency given by $P(k)$.

indicating that the size of a local neighborhood does not influence its chance of being a clique. Thus random graphs are statistically homogeneous, because very small and very large node degrees and clustering coefficients are very rare. The average distance between nodes of a random graph depends logarithmically on the number of nodes, which results in very short characteristic paths (Bollobás, 1985).

Scale-free random graphs are constructed such that they conform to a prescribed scale-free degree distribution but are random in all other aspects. Similar to scrambled but degree-preserving versions of real networks, these graphs serve as a much better suited null model for biological networks than do random graphs, and indeed they have been used to identify the significant interaction motifs of cellular networks (Milo et al., 2002; Shen-Orr et al., 2002). Scale-free random graphs have even smaller path-lengths than random graphs (Cohen et al., 2003), and they are similar to random graphs in terms of their local cohesiveness (Newman, 2003a).

Growing network models strive to arrive at realistic topologies by describing network assembly and evolution. The simplest such model (Barabási and Albert, 1999) incorporates two mechanisms: growth (i.e. an increase in the number of nodes and edges over time) and preferential attachment (i.e. an increased chance of high-degree nodes acquiring new edges). Networks generated in this way have a power-law degree distribution $P(k) = Ak^{-3}$ (Fig. 2); thus they can describe the higher end of the observed degree exponent range. Similarly to random graphs and scale-free random graphs, the average clustering coefficient in this model is small, and the clustering-
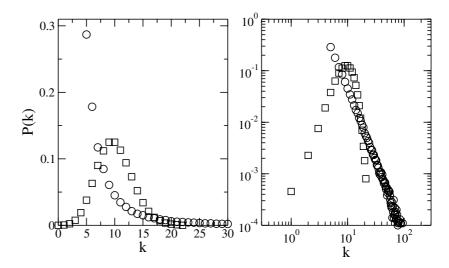
degree function $C(k)$ is constant (Ravasz et al., 2002). The average path length is slightly smaller than that in comparable random graphs (Bollobás and Riordan, 2003). The numerous improvements to this generic model include the incorporation of network evolution constraints and the identification of system-specific mechanisms for preferential attachment (Albert and Barabási, 2002).

Another growing network model, proposed by Ravasz et al., grows by iterative network duplication and integration to its original core (Ravasz et al., 2002). This growth algorithm leads to well-defined values for the node degree (for example, k = 4, 5, 20, 84 when starting from a five-node seed) and clustering coefficient. The degree distribution can be approximated by a power law in which the exponent equals $\gamma = 1 + \log(n) / \log(n-1)$, where n is the size of the seed graph. Thus this model generates degree exponents in the neighborhood of 2, which is closer to the observed values than the degree exponent of the Barabási and Albert model. In contrast to all previous models, and in agreement with protein interaction and metabolic networks, the average clustering coefficient of the Ravasz et al. network does not depend on the number of nodes, and the clustering-degree function is heterogeneous, $C(k) \cong 1/k$, and thus agrees with the lower range of the observed clustering-degree exponent β.

From general to specific: properties of select cellular networks

Protein interaction maps

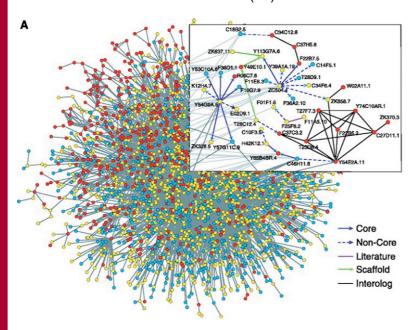During the past decade, genomics, transcriptomics and

Fig. 3. *C. elegans* protein interaction network. The nodes are colored according to their phylogenic class: ancient, red; multicellular, yellow; and worm, blue. The inset highlights a small part of the network. Figure reproduced with permission from the American Association for the Advancement of Science (Li, S. et al., 2004).

proteomics have produced an incredible quantity of molecular interaction data, contributing to maps of specific cellular networks (Burge, 2001; Caron et al., 2001; Pandey and Mann, 2000). In protein interaction graphs, the nodes are proteins, and two nodes are connected by a nondirected edge if the two proteins bind (Fig. 3). Protein-protein interaction maps have been constructed for a variety of organisms, including viruses (McCraith et al., 2000), prokaryotes such as *H. pylori* (Rain et al., 2001) and

eukaryotes such as *S. cerevisiae* (Gavin et al., 2002; Ho et al., 2002; Ito et al., 2001; Uetz et al., 2000), *C. elegans* (Li, S. et al., 2004) and *D. melanogaster* (Giot et al., 2003).

The current versions of protein interaction maps are, by necessity, incomplete and suffer from a high rate of false positives. Despite these drawbacks, there is an emerging consensus in the topological features of the maps of different organisms (Fig. 4). For example, all protein interaction networks have a giant connected component and the distances within this component are close to the small-world limit given by random graphs (Giot et al., 2003; Yook et al., 2004). This finding suggests pleiotropy, since perturbations of a single gene or protein can propagate through the network and have seemingly unrelated effects. The degree distribution of the yeast protein interaction network is approximately scale-free (Fig. 4a). The *Drosophila* protein network exhibits a lower-than-expected fraction of proteins that have >50 interacting partners (Giot et al., 2003); this deviation is suspected to be caused by incomplete coverage and could change as more interactions are discovered – as was the case for the yeast protein interaction network. The heterogeneous clustering-degree function $C(k) = B/k^{\beta}$, where the exponent β is around 2 (Fig. 4b), and the inverse correlation between the degree of two interacting proteins (Maslov and Sneppen, 2002) indicate that the neighborhood of highly connected proteins tends to be sparser than the neighborhood of less connected proteins.

Metabolic networks

Arguably the most detailed representation of a network of reactions such as the metabolic network is a directed and weighted tri-partite graph, whose three types of node are
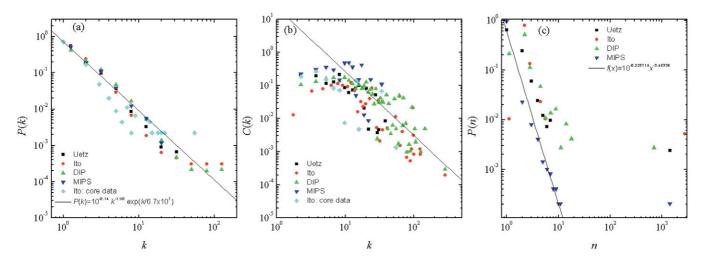

Fig. 4. Topological properties of the yeast protein interaction network constructed from four different databases. (a) Degree distribution. The solid line corresponds to a power law with exponent $\gamma = 2.5$. (b) Clustering coefficient. The solid line corresponds to the function $C(k) = B/k^2$. (c) The size distribution of connected components. All the networks have a giant connected component of >1000 nodes (on the right) and a number of small isolated clusters. Figure reproduced with permission from Wiley-VCH (Yook et al., 2004).
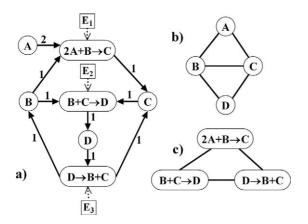
Fig. 5. Three possible representations of a reaction network with three enzyme-catalyzed reactions and four reactants. (a) The most detailed picture includes three types of node – reactants (circles), reactions (ovals) and enzymes (squares) – and two types of edge – mass flow (solid lines) or catalysis (dashed lines). The edges are marked by the stochiometric coefficients of the reactants. (b) In the metabolite network all reactants that participate in the same reaction are connected; thus the network is composed of a set of completely connected subgraphs (triangles in this case). (c) In the reaction network, two reactions are connected if they share a reactant. A similar graph can be constructed for the enzymes as well.

metabolites, reactions and enzymes, and two types of edge represent mass flow and catalytic regulation, respectively (Fig. 5). Mass flow edges connect reactants to reactions and reactions to products, and are marked by the stoichiometric coefficients of the metabolites (Feinberg, 1980; Lemke et al., 2004); enzymes catalyzing the reactions are represented as connected by regulatory edges to the nodes signifying the reaction (Jeong et al., 2000). Several simplified representations have also been studied – for example, the substrate graph, whose nodes are reactants joined by an edge if they occur in the same chemical reaction (Wagner and Fell, 2001), and the reaction graph, whose nodes are reactions that are connected if they share at least one metabolite.

All metabolic network representations indicate an approximately scale-free (Jeong et al., 2000; Tanaka, 2005; Wagner and Fell, 2001) or at least broad-tailed (Arita, 2004) metabolite degree distribution (Fig. 6). The degree distribution of enzymes indicates that enzymes catalyzing several reactions are rare (Jeong et al., 2000). The variability of metabolite degrees can be accounted for if they are functionally separated into high-degree carriers and low-degree metabolites unique to separate reaction modules (such as catabolism or amino acid

biosynthesis) (Tanaka, 2005); however, such a picture does not seem to explain the frequency of intermediate degrees. The clustering-degree function follows the relationship $C(k) \cong 1/k$.

The substrate and reaction graphs indicate a remarkably small and organism-independent average distance between metabolites and reactions (Jeong et al., 2000; Wagner and Fell, 2001). If the preferred directionality of the reactions is known and is taken into account, only the largest strongly connected component (whose nodes can reach each other in both directions) has well-defined average path length. Although this average path length is still small in all the organisms studied, the strongly connected component itself contains fewer than 50% of the nodes (Ma and Zeng, 2003). An alternative representation of the *E. coli* metabolic network defines edges among metabolites as structural changes that convert the source metabolite into the target metabolite (Arita, 2004). Because separate reactions can involve the same structural change in a metabolite, this alternative representation has <50% as many edges as the metabolite graph defined by Jeong et al., and consequently it yields twice as high average metabolite distances.

Transcriptional regulation maps

It is now possible to identify the set of target genes for each transcription factor produced by a cell, and transcription regulation maps have been constructed for *E. coli* (Shen-Orr et al., 2002) and *S. cerevisiae* (Guelzim et al., 2002; Lee et al., 2002; Luscombe et al., 2004). The full representation of such a network has two types of node – transcription factors and the mRNAs of the target genes – and two types of directed edge – transcriptional regulation and translation (Lee et al., 2002). For simplicity, transcription factors are often combined with the genes encoding them; thus all nodes correspond to genes (Fig. 7). The nodes representing target genes that do not encode transcription factors become sinks whereas non-transcriptionally regulated transcription factors correspond to sources.

Both prokaryotic and eukaryotic transcription networks exhibit an approximately scale-free out-degree distribution, signifying the potential of transcription factors to regulate a multitude of target genes. The in-degree distribution is a more restricted exponential function, illustrating that combinatorial regulation by several transcription factors is observed less than regulation of several targets by the same transcription factor (Fig. 8). Neither the *E. coli* nor the yeast transcription network has strongly connected components, which indicates a unidirectional, feed-forward-type regulation mode. The subgraphs found by following paths that start from non-

Fig. 6. Rank (cumulative distribution) of metabolite node degree (left panel) and reaction node degree (right panel) for metabolic networks of *H. pylori*. The straight lines correspond to a power-law degree distribution with exponent $\gamma = |\text{slope}| + 1 = 2.32$. The figure illustrates that functionally different metabolites tend to cover different ranges of the degree spectrum. Reproduced with permission from the American Physical Society (Tanaka, 2005).

Fig. 7. Interactions among 52 genes in the transcriptional regulation network of *S. cerevisiae*. The gene names are arranged in such a way that left to right illustrates c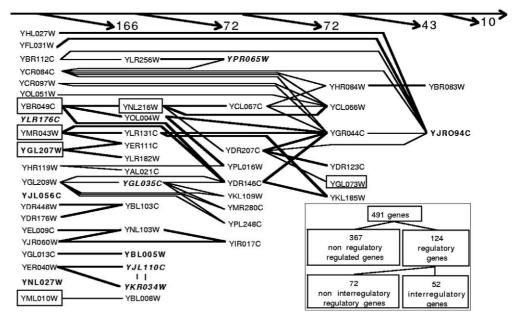ausality. The number of non-regulatory genes regulated by each column of regulatory genes is shown above. Bold type indicates self-activation, bold italics indicates self-inhibition, and borders indicate essential genes. Reproduced with permission from the Nature Publishing Group (http://www.nature.com/naturegenetics) (Guelzim et al., 2002).

transcriptionally regulated genes have relatively little overlap (Balázsi et al., 2005), reflecting the fact that distinct environmental signals tend to initiate distinct transcriptional responses. The source-sink distances are small in both networks, and the longest regulatory chain has only four (in *E. coli*) or five (in *S. cerevisiae*) edges (Fig. 8).

Signal transduction pathways

Elucidation of the mechanisms that connect extracellular signal inputs to the control of transcription factors was until recently restricted to small-scale biochemical, genetic and pharmacological techniques. Signal transduction pathways have traditionally been viewed as linear chains of biochemical reactions and protein-protein interactions, starting from signal-sensing molecules and reaching intracellular targets; however, the increasingly recognized abundance of components shared by several pathways indicates that an interconnected signaling network exists*. The largest reconstructed signal transduction network contains 1259 interactions among 545 cellular components of the hippocampal CA1 neuron (Ma'ayan et al., 2005), based on more than 1200 articles in the experimental literature. This network exhibits impressive interconnectivity: its strongly connected component (the central signaling network) includes 60% of the nodes, and the subgraphs that start from various ligand-occupied receptors reach most of the network within 15 steps. The average input-output path-length is near 4, which suggests that a very rapid response to signaling inputs is possible. Both the in- and out-degree distributions of this network are consistent with a power-law that has an exponent of around 2, the highest degree nodes including four major protein kinases (MAPK, CaMKII, PKA and PKC).

*Note that, despite the separate categories discussed, there is a significant overlap between protein interaction networks, metabolic networks and signal transduction networks.

Functional association networks

In addition to the networks whose edges signify biological interactions, several functional association networks based on gene co-expression (Stuart et al., 2003; Valencia and Pazos, 2002), gene fusion or co-occurrence (von Mering et al., 2002) or genetic interactions have been constructed. For example, synthetic lethal interactions, introduced between pairs of genes whose combined knockout causes cell death, indicate that these genes buffer for one another (Fig. 9). A recent study by Tong et al. shows that the yeast genetic interaction network has small world and scale-free properties, having a small average path length, dense local neighborhoods, and an approximately power-law degree distribution (Tong et al., 2004). The overlap

Fig. 8. Genome-wide distribution of transcriptional regulators in *S. cerevisiae*. (A) Solid symbols represent the number of transcription factors bound per promoter region (corresponding to the in-degree of the regulated gene). Open symbols represent the in-degree distribution of a comparable randomized network. (B) Distribution of the number of promoter regions bound per regulator (i.e. the out-degree distribution of transcription factors). Figure reproduced with permission from the American Association for the Advancement of Science (Lee et al., 2002).

Fig. 9. Connections between pathway redundancy and synthetic lethal interactions. Consider a hypothetical cellular network module (a) that receives exogeneous signals thro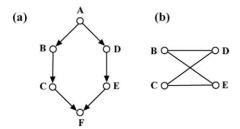ugh node A and whose sink node F determines the response to the signal (or the phenotype). There are two node-independent (redundant) pathways between nodes A and F that can compensate for each other in case of node disruptions. By defining synthetic lethal interactions as pairs of nodes whose loss causes the disconnection of nodes A and F, one would find graph b. The two graphs present complementary and non-overlapping information.

between the yeast protein interaction and genetic interaction network is extremely small, which is expected since genetic interactions reflect a complex functional compensatory relationship and not a physical interaction (Fig. 9). Indeed, the relationships that do overlap with genetic interactions include having the same mutant phenotype, encoding proteins that have the same subcellular localization or encoding proteins within the same complex.

Biological interpretation of graph properties

The architectural features of molecular interaction networks are shared to a large degree by other complex systems ranging from technological to social networks. While this universality is intriguing and allows us to apply graph theory to biological networks, we need to focus on the interpretation of graph properties in light of the functional and evolutionary constraints of these networks.

Hubs

In a scale-free network, small-degree nodes are the most abundant, but the frequency of high-degree nodes decreases relatively slowly. Thus, nodes that have degrees much higher than average, so-called hubs, exist. Because of the heterogeneity of scale-free networks, random node disruptions do not lead to a major loss of connectivity, but the loss of the hubs causes the breakdown of the network into isolated clusters (Albert and Barabási, 2002). The validity of these general conclusions for cellular networks can be verified by correlating the severity of a gene knockout with the number of interactions the gene products participate in. Indeed, as much as 73% of the *S. cerevisiae* genes are non-essential, i.e. their knockout has no phenotypic effects (Giaever et al., 2002). This confirms the cellular networks' robustness in the face of random disruptions. The likelihood that a gene is essential (lethal) or toxicity modulating (toxin sensitive) correlates with the number of interactions its protein product has (Jeong et al., 2001; Said et al., 2004). This indicates the cell is vulnerable to the loss of highly interactive hubs[†]. Among the most well-known

examples of a hub protein is the tumor suppressor protein p53, which has an abundance of incoming edges, interactions regulating its conformational state (and thus its activity) and its rate of proteolytic degradation, and numerous outgoing edges in the genes it activates. p53 is inactivated by mutation in 50% of human tumors, which is in agreement with the vulnerability of cellular networks to their most connected hubs (Vogelstein et al., 2000).

Given the importance of highly connected nodes, one can hypothesize that they are subject to severe selective and evolutionary constraints. Hahn et al. have correlated the rate of evolution of yeast proteins with their degree in the protein interaction network (Hahn et al., 2004), and the rate of evolution of *E. coli* enzymes with their degree in the core metabolic reaction graph constructed by Wagner and Fell (Wagner and Fell, 2001). Although they obtained statistically significant (albeit weak) negative correlation between yeast protein degree and evolution rate, no such correlation was evident in the *E. coli* enzyme network. The latter result has the caveat that the edges linking enzymes do not correspond to interactions; thus further studies are needed to gain a definitive answer.

Modularity

Cellular networks have long been thought to be modular, composed of functionally separable subnetworks corresponding to specific biological functions (Hartwell et al., 1999). Since genome-wide interaction networks are highly connected, modules should not be understood as disconnected components but rather as components that have dense intracomponent connectivity but sparse intercomponent connectivity. Several methods have been proposed to identify functional modules on the basis of the physical location or function of network components (Rives and Galitski, 2003) or the topology of the interaction network (Giot et al., 2003; Girvan and Newman, 2002; Spirin and Mirny, 2003). The challenge is that modularity does not always mean clear-cut subnetworks linked in well-defined ways, but there is a high degree of overlap and crosstalk between modules (Han et al., 2004). As Ravasz et al. recently argued, a heterogeneous degree distribution, inverse correlation between degree and clustering coefficient (as seen in metabolic and protein interaction networks) and modularity taken together suggest hierarchical modularity, in which modules are made up of smaller and more cohesive modules, which themselves are made up of smaller and more cohesive modules, etc.

Motifs and cliques

Growing evidence suggests that cellular networks contain conserved interaction motifs, small subgraphs that have well-defined topology. Interaction motifs such as autoregulation and feed-forward loops have a higher abundance in transcriptional regulatory networks than expected from randomly connected graphs with the same degree distribution (Balázsi et al., 2005; Shen-Orr et al., 2002). Protein interaction

---

[†]Note that different network representations can lead to distinct sets of hubs and there is no rigid boundary between hub and non-hub genes or proteins.

motifs such as short cycles and small completely connected subgraphs are both abundant (Giot et al., 2003) and evolutionarily conserved (Wuchty et al., 2003), partly because of their enrichment in protein complexes. Triangles of scaffolding protein interactions are also abundant in signal transduction networks, which also contain a significant number of feedback loops, both positive and negative (Ma'ayan et al., 2005). Yeger-Lotem et al. have identified frequent composite transcription/protein interaction motifs, such as interacting transcription factors coregulating a gene or interacting proteins being coregulated by the same transcription factor (Yeger-Lotem et al., 2004). As Zhang et al. have pointed out, the abundant motifs of integrated mRNA/protein networks are often signatures of higher-order network structures that correspond to biological phenomena (Zhang et al., 2005) (Fig. 10). Conant and Wagner found that the abundant transcription factor motifs of *E. coli* and *S. cerevisiae* do not show common ancestry but are a result of repeated convergent evolution (Conant and Wagner, 2003). These findings, as well as studies of the dynamical repertoire of interaction motifs, suggest that these common motifs represent elements of optimal circuit design (Csete and Doyle, 2002; Ma'ayan et al., 2005; Mangan and Alon, 2003).
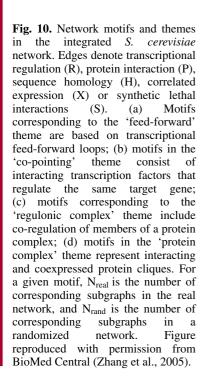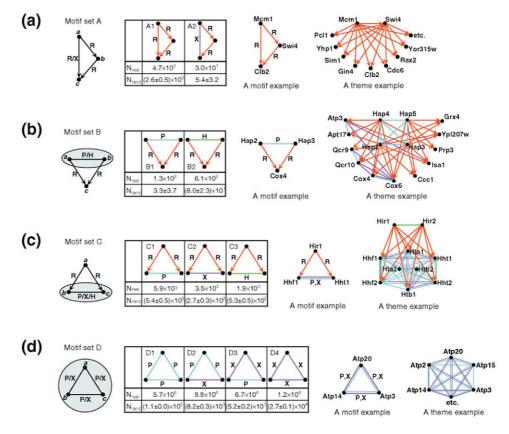
Path redundancy

Any response to a perturbation requires that information about the perturbation spreads within the network. Thus the short path lengths of metabolic, protein interaction and signal transduction networks (their small world property) is a very important feature that ensures fast and efficient reaction to

perturbations. Another very important global property related to paths is path redundancy, or the availability of multiple paths between a pair of nodes (Papin and Palsson, 2004). Either in the case of multiple flows from input to output, or contingencies in the case of perturbations in the preferred pathway, path redundancy enables the robust functioning of cellular networks by relying less on individual pathways and mediators. The frequency of node participation in paths connecting other components can be quantified by their betweenness centrality, first defined in the context of social sciences (Wasserman and Faust, 1994). Node betweenness, adapted to the special conditions of signal transduction networks, can serve as an alternative measure for identifying important network hubs.

Network models specific to biological networks

The topology of cellular networks is shaped by dynamic processes on evolutionary time scales. These processes include gene or genome duplication and gain or loss of interactions owing to mutations. Many researchers have investigated whether the similar topological properties of biological networks and social or technological networks point towards shared growth principles and whether variants of general growing network models apply to cellular networks. The most intriguing question is the degree to which natural selection, specific to biological systems, shapes the evolution of cellular network topologies.

Several growing network models based on random gene duplication and subsequent functional divergence display good agreement with the topology of protein interaction networks

Fig. 10. Network motifs and themes in the integrated *S. cerevisiae* network. Edges denote transcriptional regulation (R), protein interaction (P), sequence homology (H), correlated expression (X) or synthetic lethal interactions (S). (a) Motifs corresponding to the 'feed-forward' theme are based on transcriptional feed-forward loops; (b) motifs in the 'co-pointing' theme consist of interacting transcription factors that regulate the same target gene; (c) motifs corresponding to the 'regulonic complex' theme include co-regulation of members of a protein complex; (d) motifs in the 'protein complex' theme represent interacting and coexpressed protein cliques. For a given motif, N_{real} is the number of corresponding subgraphs in the real network, and N_{rand} is the number of corresponding subgraphs in a randomized network. Figure reproduced with permission from BioMed Central (Zhang et al., 2005).

(Kim et al., 2002; Pastor-Satorras et al., 2003; Vazquez et al., 2003). However, estimates of gene duplication rate and the rate at which point mutations lead to the gain or loss of protein interactions indicate that point mutations are two orders of magnitude more frequent than gene duplications (Berg et al., 2004). Berg et al. have proposed a protein network evolution model based on edge dynamics and, to a lesser extent, gene duplication, and find that it generates a topology similar to that of the yeast protein interaction network. It is interesting to note that both gene duplications and point mutations, specific biological processes, lead to a preferential increase in the degree of highly connected proteins – also confirmed by measurements (Eisenberg and Levanon, 2003; Wagner, 2003). Thus natural selection could affect the balance between interaction gain and loss in such a way that an effective preferential attachment is obtained. The modeling of the evolution of transcriptional, metabolic and signal transduction networks is more challenging owing to their directed nature and to the complexity of the regulatory mechanisms involved, but rapid progress is expected in these fields as well (Light and Kraulis, 2004; Tanay et al., 2005).

Beyond static properties

As illustrated in the specific examples presented in this review, graph representations of cellular networks and quantitative measures characterizing their topology can be extremely useful for gaining systems-level insights into cellular regulation. For example, the interconnected nature of cellular networks indicates that perturbations of a gene or protein could have seemingly unrelated effects (pleiotropy), a result that would seem counterintuitive in a reductionist framework. The graph framework allows us to discuss the cell's molecular makeup as a network of interacting constituents and to shift the definition of gene function from an individual-based attribute to an attribute of the network (or network module) in which the gene participates (Fraser and Marcotte, 2004). Interaction motifs and themes can be exploited to predict individual interactions given sometimes-uncertain experimental evidence or to give a short list of candidates for experimental testing (Albert and Albert, 2004; King et al., 2004; Wong et al., 2004).

It is important to realize that cellular interaction maps represent a network of possibilities, and not all edges are present and active at the same time or in a given cellular location in vivo. Indeed, superposing mRNA expression patterns and protein interaction information in *S. cerevisiae*, Han et al. identified a strong dynamical modularity mediated by two types of highly interactive proteins: party hubs, which interact with most of their partners simultaneously, and date hubs, which bind their different partners at different times or location (Han et al., 2004). Similarly, Luscombe et al. and Balázsi et al. found that only subsets of the yeast and *E. coli* transcriptional networks are active under particular conditions. Exogenous stimuli induce only a few transcription factors with little crosstalk, whereas endogenous responses activate connected clusters of transcription factors and many feed-forward loops (Luscombe et al. 2004; Balázsi et al., 2005).

In addition, the diversity of metabolic fluxes (Almaas et al., 2004) and reaction rates/timescales (Papin et al., 2005) attest that only an integration of interaction and activity information will be able to give a correct dynamic picture of a cellular network (Levchenko, 2003; Ma'ayan et al., 2004). To move significantly beyond our present level of knowledge, new tools for quantifying concentrations, fluxes and interaction strengths, in both space and time, are needed. In the absence of comprehensive time-course datasets, dynamic reconstruction and analysis can usually be carried out only for small networks (Hoffmann et al., 2002; Lee et al., 2003; Tyson et al., 2001). The coupling of experimental data with mathematical modeling enables the identification of previously unknown regulatory mechanisms. For example, the Hoffmann et al. model's prediction regarding the importance of particular IκB isoforms in feedback loops regulating NF-κB (Hoffmann et al., 2002) was experimentally verified, as were the dynamic profiles of β-catenin concentrations in the Lee et al. model of the WNT signaling module (Lee et al., 2003).

Our currently limited knowledge of kinetic parameters makes the construction of detailed kinetic models of complex biological networks next to impossible; however, there is hope that more coarse-grained models will also be successful. Indeed, increasing evidence indicates the crucial role of network topology in determining dynamic behavior and function and robustness to fluctuations in kinetic parameters (Albert and Othmer, 2003; Barkai and Leibler, 1997; Chaves et al., 2005; Li, F. et al., 2004; von Dassow et al., 2000). The topological properties of signal transduction subgraphs (pathways) seem to reflect the dynamics of response to those signals: the subgraphs corresponding to ligands that cause rapid, transient changes – such as glutamate or glycine – exhibit extensive pathway branching, whereas the signaling pathways for responses to FasL or ephrin have many fewer branches (Ma'ayan et al., 2005). Constraint-based modeling of stoichiometrically reconstructed metabolic and signaling networks can lead to verifiable predictions related to their input/output relationships and their changes in the case of gene knockouts (Papin and Palsson, 2004; Papin et al., 2002). Network discovery and network analysis thus have the potential to form a self-reinforcing loop where theory and modeling lead to testable predictions that feed back into experimental discovery. At a minimum, network representations have changed our view of what is functionally 'downstream' (or 'near') a cellular component, and have the potential to lead to predictions of systems-level behavior that will be important for future biochemical and medical research (Cohen, 2002).

References

Albert, I. and Albert, R. (2004). Conserved network motifs allow protein-protein interaction prediction. *Bioinformatics* **20**, 3346-3352.

Albert, R. and Barabási, A. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* **74**, 47-97.

Albert, R. and Othmer, H. G. (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J. Theor. Biol.* **223**, 1-18.

Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z. N. and Barabási, A. L. (2004). Global organization of metabolic fluxes in the bacterium Escherichia coli. *Nature* **427**, 839-843.

Arita, M. (2004). The metabolic world of Escherichia coli is not small. *Proc. Natl. Acad. Sci. USA* **101**, 1543-1547.

Balázsi, G., Barabási, A. L. and Oltvai, Z. N. (2005). Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli. *Proc. Natl. Acad. Sci. USA* **102**, 7841-7846.

Barabási, A. L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286**, 509-512.

Barabási, A. L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101-113.

Barkai, N. and Leibler, S. (1997). Robustness in simple biochemical networks. *Nature* **387**, 913-917.

Berg, J., Lassig, M. and Wagner, A. (2004). Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol. Biol.* **4**, 51.

Bollobás, B. (1979). *Graph Theory: an Introductory Course*. New York: Springer Verlag.

Bollobás, B. (1985). *Random Graphs*. London; Orlando: Academic Press.

Bollobás, B. and Riordan, O. M. (2003). Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks* (ed. S. Bornholdt and H. G. Schuster,), pp. 1-32. Weinheim: Wiley.

Burge, C. B. (2001). Chipping away at the transcriptome. *Nat. Genet.* **27**, 232-234.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M. C., van Asperen, R., Boon, K., Voute, P. A. et al. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289-1292.

Chaves, M., Albert, R. and Sontag, E. D. (2005). Robustness and fragility of Boolean models for genetic regulatory networks. *J. Theor. Biol.* **235**, 431-449.

Cohen, P. (2002). Protein kinases – the major drug targets of the twenty-first century? *Nat. Rev. Drug Discovery* **1**, 309-315.

Cohen, R., Havlin, S. and ben-Avraham, D. (2003). Structural properties of scale-free networks. In *Handbook of Graphs and Networks* (ed. S. Bornholdt and H. G. Schuster), pp. 85-110. Weinheim: Wiley.

Conant, G. C. and Wagner, A. (2003). Convergent evolution of gene circuits. *Nat. Genet.* **34**, 264-266.

Csete, M. E. and Doyle, J. C. (2002). Reverse engineering of biological complexity. *Science* **295**, 1664-1669.

Dijkstra, E. W. (1959). A note on two problems in connection with graphs. *Numerische Math.* **1**, 269-271.

Eisenberg, E. and Levanon, E. Y. (2003). Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* **91**, 138701.

Feinberg, M. (1980). Chemical oscillations, multiple equilibria, and reaction network structure. In *Dynamics of Reactive systems* (ed. W. Stewart, W. Rey and C. Conley), pp. 59-130. New York: Academic Press.

Fraser, A. G. and Marcotte, E. M. (2004). A probabilistic view of gene function. *Nat. Genet.* **36**, 559-564.

Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799-805.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M. et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B. et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387-391.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E. et al. (2003). A protein interaction map of Drosophila melanogaster. *Science* **302**, 1727-1736.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 7821-7826.

Guelzim, N., Bottani, S., Bourgine, P. and Kepes, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* **31**, 60-63.

Hahn, M. W., Conant, G. C. and Wagner, A. (2004). Molecular evolution in large genetic networks: does connectivity equal constraint? *J. Mol. Evol.* **58**, 203-211.

Han, J. D., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J., Cusick, M. E., Roth, F. P. et al. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88-93.

Hartwell, L. H., Hopfield, J. J., Leibler, S. and Murray, A. W. (1999). From molecular to modular cell biology. *Nature* **402**, C47-C52.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K. et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-183.

Hoffmann, A., Levchenko, A., Scott, M. L. and Baltimore, D. (2002). The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241-1245.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569-4574.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. and Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature* **407**, 651-654.

Jeong, H., Mason, S. P., Barabási, A. L. and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41-42.

Kim, J., Krapivsky, P. L., Kahng, B. and Redner, S. (2002). Infinite-order percolation and giant fluctuations in a protein interaction network. *Physical Review E* **66**, 055101.

King, A. D., Przulj, N. and Jurisica, I. (2004). Protein complex prediction via cost-based clustering. *Bioinformatics* **20**, 3013-3020.

Lee, E., Salic, A., Kruger, R., Heinrich, R. and Kirschner, M. W. (2003). The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS Biol* **1**, E10.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I. et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799-804.

Lemke, N., Heredia, F., Barcellos, C. K., Dos Reis, A. N. and Mombach, J. C. (2004). Essentiality and damage in metabolic networks. *Bioinformatics* **20**, 115-119.

Levchenko, A. (2003). Dynamical and integrative cell signaling: challenges for the new biology. *Biotechnol. Bioeng.* **84**, 773-782.

Li, F., Long, T., Lu, Y., Ouyang, Q. and Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. USA* **101**, 4781-4786.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T. et al. (2004). A map of the interactome network of the metazoan C. elegans. *Science* **303**, 540-543.

Light, S. and Kraulis, P. (2004). Network analysis of metabolic enzyme evolution in Escherichia coli. *BMC Bioinformatics* **5**, 15.

Longabaugh, W. J., Davidson, E. H. and Bolouri, H. (2005). Computational representation of developmental genetic regulatory networks. *Dev. Biol.* **283**, 1-16.

Luscombe, N. M., Babu, M. M., Yu, H. Y., Snyder, M., Teichmann, S. A. and Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308-312.

Ma, H. W. and Zeng, A. P. (2003). The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19**, 1423-1430.

Ma'ayan, A., Blitzer, R. D. and Iyengar, R. (2004). Toward predictive models of mammalian cells. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 319-349.

Ma'ayan, A., Jenkins, S. L., Neves, S., Hasseldine, A., Grace, E., Dubin-Thaler, B., Eungdamrong, N. J., Weng, G., Ram, P. T., Rice, J. J. et al. (2005). Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science* **309**, 1078-1083.

Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* **100**, 11980-11985.

Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science* **296**, 910-913.

McCraith, S., Holtzman, T., Moss, B. and Fields, S. (2000). Genome-wide analysis of vaccinia virus protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **97**, 4879-4884.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827.

Newman, M. E. J. (2003a). Random graphs as models of networks. In *Handbook of Graphs and Networks* (ed. S. Bornholdt and H. G. Schuster), pp. 35-65. Weinheim: Wiley.

Newman, M. E. J. (2003b). The structure and function of complex networks. *Siam Review* **45**, 167-256.

Pandey, A. and Mann, M. (2000). Proteomics to study genes and genomes. *Nature* **405**, 837-846.

Papin, J. A. and Palsson, B. O. (2004). Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *J. Theor. Biol.* **227**, 283-297.

Papin, J. A., Price, N. D. and Palsson, B. O. (2002). Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res.* **12**, 1889-1900.

Papin, J. A., Hunter, T., Palsson, B. O. and Subramaniam, S. (2005). Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell. Biol.* **6**, 99-111.

Pastor-Satorras, R., Smith, E. and Sole, R. V. (2003). Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* **222**, 199-210.

Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V. et al. (2001). The protein-protein interaction map of Helicobacter pylori. *Nature* **409**, 211-215.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555.

Rives, A. W. and Galitski, T. (2003). Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA* **100**, 1128-1133.

Said, M. R., Begley, T. J., Oppenheim, A. V., Lauffenburger, D. A. and Samson, L. D. (2004). Global network analysis of phenotypic effects: protein networks and toxicity modulation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA* **101**, 18006-18011.

Shen-Orr, S. S., Milo, R., Mangan, S. and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**, 64-68.

Spirin, V. and Mirny, L. A. (2003). Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA* **100**, 12123-12128.

Stuart, J. M., Segal, E., Koller, D. and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249-255.

Tanaka, R. (2005). Scale-rich metabolic networks. *Phys. Rev. Lett.* **94**, 168101.

Tanay, A., Regev, A. and Shamir, R. (2005). Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl. Acad. Sci. USA* **102**, 7203-7208.

Tong, A. H., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M. et al. (2004). Global mapping of the yeast genetic interaction network. *Science* **303**, 808-813.

Tyson, J. J., Chen, K. and Novak, B. (2001). Network dynamics and cell physiology. *Nat. Rev. Mol. Cell. Biol.* **2**, 908-916.

Tyson, J. J., Chen, K. C. and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221-231.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P. et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-627.

Valencia, A. and Pazos, F. (2002). Computational methods for the prediction of protein interactions. *Curr. Opin. Struct. Biol.* **12**, 368-373.

Vazquez, A., Flammini, A., Maritan, A. and Vespignani, A. (2003). Modeling of protein interaction networks. *ComPlexUs* **1**, 38-44.

Vogelstein, B., Lane, D. and Levine, A. J. (2000). Surfing the p53 network. *Nature* **408**, 307-310.

von Dassow, G., Meir, E., Munro, E. M. and Odell, G. M. (2000). The segment polarity network is a robust developmental module. *Nature* **406**, 188-192.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399-403.

Wagner, A. (2003). How the global structure of protein interaction networks evolves. *Proc. Biol. Sci.* **270**, 457-466.

Wagner, A. and Fell, D. A. (2001). The small world inside large metabolic networks. *Proc. R. Soc. Lond., B, Biol. Sci.* **268**, 1803-1810.

Wasserman, S. and Faust, K. (1994). Social network analysis: methods and applications. Cambridge; New York: Cambridge University Press.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442.

Wong, S. L., Zhang, L. V., Tong, A. H., Li, Z., Goldberg, D. S., King, O. D., Lesage, G., Vidal, M., Andrews, B., Bussey, H. et al. (2004). Combining biological networks to predict genetic interactions. *Proc Natl. Acad. Sci. USA* **101**, 15682-15687.

Wuchty, S., Oltvai, Z. N. and Barabási, A. L. (2003). Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat. Genet.* **35**, 176-179.

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R. Y., Alon, U. and Margalit, H. (2004). Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proc. Natl. Acad. Sci. USA* **101**, 5934-5939.

Yook, S. H., Oltvai, Z. N. and Barabási, A. L. (2004). Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928-942.

Zhang, L. V., King, O. D., Wong, S. L., Goldberg, D. S., Tong, A. H., Lesage, G., Andrews, B., Bussey, H., Boone, C. and Roth, F. P. (2005). Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. *J. Biol.* **4**, 6.