Graph Theory Properties of Cellular Networks

Baruch Barzel[1,2], Amitabh Sharma[1,2], Albert-László Barabási[1,2]

[1]Center for Complex Network Research, Department of Physics, Northeastern University, 360 Huntington avenue, Boston, Massachusetts 02115, USA

[2]Center for Cancer System Biology (CCSB) and Department of Cancer Biology, the Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, 44 Binney street, Boston, Massachusetts, USA

Telephone: 617-373-2355

Fax: 617-373-4385

Mailing Address:

Center for Complex Network Research

Northeastern University Physics Department

111 DA/Physics Dept., 110 Forsyth Street

Boston, MA 02115

Email addresses:

Baruch Barzel: baruchbarzel@gmail.com

Amitabh Sharma: amitabhsharma13@gmail.com

Albert-László Barabási: barabasi@gmail.com (Corresponding author)

Summary

The functionality of the living cell is enabled due to an intricate network of biochemical, metabolic and information transporting processes. These processes are carried out by the different network systems that the cell's activity is comprised of, among which are the transcriptional regulatory network, the protein-protein interaction network and the metabolic network. To understand the functional design of these complex systems, it is worthwhile to refer to their abstract representation as graphs, where the interacting components, be them proteins, metabolites or genes, are designated as nodes, and the interactions between them – as edges. Once the graphical description has been established, the tools of graph theory can be utilized in order to analyze the networks and obtain a better understanding of their overall construction. This approach has led to several groundbreaking discoveries on the nature of networks, crossing fields of research from biology, to social science and technology. In this chapter we present the basic tools and concepts brought forth by the graph theoretic approach, and show their application to biological networks. We especially focus on the universal appearance of various features, such as small-world topologies, scale-free degree distributions and hierarchical and modular structures. These recurring patterns in the structure of the cellular networks are key to understanding their evolution, their design principles, and, most importantly, the way they function.

Keywords:

Cellular networks; Network topology; Scale-free networks; Power-law; Degree-distribution; Metabolic networks; Protein-protein interaction networks; Regulatory networks; Disease networks.

1 Overview

From a conceptual point of view, the rise of systems biology can be described as the adoption of a broad-based perspective on biological systems. In that sense, the classical detailed biological analysis is complemented by a macroscopic description of the cell as a holistic unit [1-4]. This approach, aiming at a system level understanding of biology, mandates a crude simplification of biological processes. In this light, the graph theoretic approach to biological systems, focuses on the structural aspects of the interaction patterns, where the interacting species, be them genes, proteins or other biological components, are signified by nodes, and their interactions by the edges drawn between them. These network systems express the underlying architecture which enables the cellular functions to be carried out [5-9].

Undoubtedly, the functionality of the cell cannot be attributed to just one network, but rather to a set of interdependent networks ranging from the level of transcription to the processes of metabolism. It is common to divide the cellular functions into three distinct networks - the transcriptional network, the protein interaction network and the metabolic network [2]. While we follow this division throughout this chapter, it should be acknowledged that the true functionality of the cellular unit is a result of the interdependence between these networks, and not merely the interactions in each of them alone. At the current state, the topology of these three fundamental cellular networks has been thoroughly mapped using high throughput techniques. As a result, we now have reliable data on the interaction maps of many organisms. Some examples are protein-protein interaction networks, which have been constructed for organisms such as *Homo sapiens*, *Saccharomyces cerevisiae* (*S. cerevisiae*) *Helicobacter Pylori* and others [10-20]. Regulatory and metabolic networks have also been successfully mapped for yeast, *Escherichia coli* (*E. coli*) and various other organisms [21-23]. However, we still lack data regarding other networks taking part in the cellular processes, such as sRNA and RNAi mediated networks, of which we know little about at this point.

2 Biological Systems as Graphs

While the foundations of graph theory were laid out of a purely mathematical curiosity, its applicability as a tool for the characterization of complex systems has been appreciated since. The notion is that the behavior of complex systems arises from the coordinated actions of many interacting components. The network abstraction can then be used in order to reveal the underlying structure of these interactions. The interacting components are signified by a series of nodes, and the interactions between selected pairs of these nodes are represented by the links (or edges) drawn between them. This abstract description eliminates some of the details associated with the specific nature of the system at hand. However it allows one to utilize the well-established formalisms of graph theory, thus providing a powerful tool for the analysis and understanding of these complex systems. Moreover, this categorical representation applied to various systems provides the grounds for comparison between seemingly distinct networks. This process has proven highly beneficial, as one of the most important discoveries of recent years was that despite the diversity of cellular networks, several important universal properties are shared by them all [5].

In some cases the network description of a cellular system is straightforward and natural. Consider, for instance, the set of physical binding reactions between pairs of proteins or between proteins and other molecules, such as nucleic acids or metabolites. Here it seems natural to use a node for the representation of each molecular type, and an edge to denote each potential binding reaction. However, in other cases, the network description is not unique, and may differ according to the motivation of the study. A simple example regards the transcriptional regulatory network. In this network the edges link between transcription factors and the genes that they regulate. Here the information flows from the regulating gene to the regulated one, so that the links are not symmetrical. The network is thus a *directed* network. Moreover, the relationship between a pair of interacting genes can be of an activating nature, or of an inhibitory nature. Thus two different types of directed edges exist, which can be denoted by positive versus negative, or graphically by → versus —┤. Nevertheless, in many contexts, the directed nature of the interactions, or their sign, is not important, and it is sufficient to model the system using a regular undirected network. As a

more complicated system we refer to metabolic networks. These systems can be conceptualized as networks in many levels of abstraction. For instance, one can visualize the molecular substrates as nodes, and the reactions transforming substrates to products as links. In this case the links are attributed by the enzyme catalyzing the reaction, and the graph is directed. However, there are contexts in which it is sufficient to use a simpler description, where the enzymes are ignored, and the links are undirected. In this case the graph simply describes the interconnections between metabolites leaving out the detailed chemistry which underlies these connections.

3 The Tools of Graph Theory

In the basis of graph theory lays the insight that a complex system could be reduced into a series of abstract components tied together by a set of connections. The spark of this idea is commonly attributed to the eighteenth century mathematician, Leonhard Euler, who in 1735 used it to solve the problem of the Seven Bridges of Königsberg – a problem which back then boggled the minds of the Prussian town residents. To show that one cannot visit all of the city's islands, without crossing at least one of the city's seven bridges twice, Euler constructed an abstract map of the city, in which the islands are represented by nodes, and the bridges – by edges. In doing so, Euler mapped a realistic problem, in all of its complexities, into a clean abstract mathematical representation, which allowed him to focus strictly on the structural crux of the problem. However, this spark will remain dim, and only reemerge as an elaborate, formalized mathematical theory some two hundred years later, in the twentieth century, following the work of Paul Erdős and Alfréd Rényi.

Erdős-Rényi – The Benchmark Network

The most elementary network considered in graph theory, is the Erdős-Rényi random network, where each pair of nodes is connected with equal probability [24-26]. The properties of this prototypic network serve as a benchmark, to which we later compare the more realistic networks of cellular biology. To construct an Erdős-Rényi network, we consider a set of N nodes. For each of the $N(N-1)/2$ pairs of nodes in the network we assign an edge with probability p, typically

chosen such that $p \ll 1$. Simple as it may be, the Erdős-Rényi network features some surprising characteristics, commonly observed in many real world biological networks.

Degrees and Degree Distribution

To analyze the components of the network we introduce some elementary network measures. For concreteness, we use the Erdős-Rényi network to exemplify them. The most basic characteristic of a node is its degree, k, defined as the number of links it has to other nodes in the network. In the Erdős-Rényi graph, every node can potentially be linked with any of its $N-1$ counterparts with independent probability p. The average degree will thus be $\langle k \rangle = p(N-1) \approx pN$. The random nature of this network invokes some variability in the degree of the nodes, so that several nodes will have more links than the average, while other nodes will have less. This variability can be described by the *degree distribution* of the graph. Denoted by $P(k)$, it is defined as the probability that a randomly selected node will have exactly k links. As we will see later in this chapter, the degree distribution is one of the most fundamental characteristics of a network, carrying crucial information about its evolution and formation process. In the Erdős-Rényi network $P(k)$ follows a Poisson distribution, which indicates that most nodes are characterized by roughly the same degree, the probability to encounter a node with a degree significantly different than $\langle k \rangle$ being vanishingly small. The average degree is thus the *characteristic scale* of the degree distribution.

Location of Fig. 1

In directed networks, we distinguish between the *in-degree* of a node, denoting the number of incoming links, and the *out-degree*, denoting the number of outgoing links. For instance, if gene x regulates n other genes, it will have an out-degree of $k_{\text{out}} = n$, whereas if it is being regulated by m other genes, its in-degree will be $k_{\text{in}} = m$. Accordingly, the degree distribution in such networks is split into the incoming distribution $P(k_{\text{in}})$ and the outgoing distribution $P(k_{\text{out}})$. As an example consider node number four in Fig. 1, which in the undirected version of the graph (a) has a total degree of $k = 5$, while in the directed version (b) it is characterized by $k_{\text{in}} = 2$ and $k_{\text{out}} = 3$.

Network Paths and the Small World Phenomena

A crucial feature of any biological network is its ability to maintain a flow of information, mass or energy, between all of its nodes. From the graph theoretical perspective this requirement translates into the existence of a *network path* connecting all (or most) of the nodes in the network. By network path we refer to a route leading from one node to another by passing solely over existing links (Fig. 1). Such a group of interconnected nodes constitutes a connected component, and if indeed a large fraction of the nodes in the graph can be reached from one another via these network paths, the graph is said to have a *giant connected component*. This seemingly remote feature, appears rather frequently, and does not require much high level organization for it to be observed. In fact, for an Erdős-Rényi graph, a giant component will emerge as long as $\langle k \rangle \geq 1$. Moreover, in the case where $1 \ll \langle k \rangle \ll \ln N$, this giant component is likely to encompass almost all of the nodes in the network [6,27].

Networks impose a unique metric, by which the distance between nodes can be measured. Consider the network path between a pair of nodes. Its length is defined as the number of links it crosses. The length of the shortest path between a pair of nodes, x and y, is the network distance, l_{xy}. In Fig. 1(a) we display the shortest path between the nodes one and eight. The network distance between these two nodes is $l_{1,8} = 4$. If no such path exists, then the pair of nodes cannot communicate with one another and their distance is set to be infinite. The distance between a pair of nodes parameterizes the potential to propagate information from one to the other, such that close nodes are more likely to affect one another than far away nodes. By averaging over all pairs of nodes in the network, one obtains the *average path length* of the network, $\langle l \rangle$, which offers a measure of the network's overall connectivity. If the network includes infinite paths, namely has isolated components, the averaging is commonly carried out over the nodes belonging to the giant connected component. For an Erdős-Rényi network (as well as almost all other random networks) we find that the average path length is strikingly small, compared to the network size. To understand this we focus on the surrounding of a typical node in the network. It has $\langle k \rangle$ nearest neighbors, all at a distance of $l = 1$ from it; each of these neighbors has on average $\langle k \rangle$ neighbors of its own, so that

there are roughly $\langle k\rangle^2$ nodes at a distance of $l = 2$. Following this logic we find that the number of nodes within a given distance from any typical node inflates exponentially, leading to a logarithmic dependence of the average path length on the network size, N [27]

$$\langle l\rangle \approx \frac{\ln N}{\ln\langle k\rangle}. \quad (1)$$

This simplified argument overlooks the possibility that some of the edges might be redundant, namely that some of the links emerging from nodes at a distance l might link to other nodes which are at the same distance. Still, for large N it captures the behavior of the network at the vicinity of a node, where the probability of linking to an already *used* node, creating a loop, is very small.

In a directed network, network paths are commonly defined to propagate only in the direction of the edges. This way the paths reflect the flow of information between the nodes. The network distance between a pair of nodes is no longer symmetrical, as displayed in Fig. 1(b). The distance from node one to node eight in the directed network is $l_{1,8} = 6$ (as opposed to $l_{1,8} = 4$ in the undirected version). However no path exists in the opposite direction, rendering $l_{8,1}$ to be infinite.

Clustering Coefficient

Certain networks show a tendency to form clusters of interconnected nodes [28]. In such networks if the nodes y and z are both connected to some other node, x, it is likely that there is also a direct link between y and z themselves. To quantify this we denote by c_x, the number of links connecting x's nearest neighbors to one another. This number can take values ranging from zero, in case that no such links exist, to $k_x(k_x - 1)/2$ in case that all of the pairs among x's k_x nearest neighbors are connected. The clustering coefficient is thus defined as $C_x = 2c_x/k_x(k_x - 1)$, taking values which are between zero and unity [29-30]. For instance, the clustering of node four in Fig. 1 is $C_4 = \frac{2\times2}{5\times4} = 0.2$. Averaging over all the nodes in the network, one obtains the average clustering coefficient $\langle C\rangle$.

For an Erdős-Rényi graph the probability of all pairs of nodes to be linked is uniform, regardless of whether they share a mutual neighbor or not. The average clustering coefficient is thus $\langle C \rangle_{\mathrm{ER}} = p$, which, for a sparse graph, is typically very small.

4 Successes and Failures of the Erdős-Rényi Model

Biological Small Worlds

The Erdős-Rényi network model is greatly oversimplified, and thus overlooks many important features observed in real biological networks. Nevertheless, this model does prove successful in predicting the overall connectivity observed is practically all analyzed biological network. These networks all feature a giant connected component, such that almost all pairs of nodes are connected by finite network paths. Moreover, the average path length is found to be consistent with Eq. (1), so that the interacting nodes in the network are typically just a few steps away from one another, meaning that cellular networks, like many other networks in nature, feature the small world effect. For instance, in metabolism, it was found that most pairs of metabolites can be linked by paths, averaging approximately three edges by length. These extremely short average path lengths are not unique to any specific species. They were found in as many as forty three different species, ranging from the evolutionary reduced metabolic network of parasitic bacterium to the highly developed networks of large multicellular organisms [31]. Similar, albeit less dramatic, results apply for protein and genetic interaction networks, were the average path length ranges from about four to eight [32-33].

Deviations from the Erdős-Rényi Model

In most cellular networks a tendency to form cliques is observed, where the neighbors of one node tend to be themselves connected. Thus the average clustering coefficient, $\langle C \rangle$, of most cellular networks is significantly larger than that of an equivalent Erdős-Rényi network. By an equivalent network, we refer to an Erdős-Rényi network with the same size and average degree. For instance,

protein-protein interaction networks feature a clustering coefficient which is typically about an order of magnitude higher than that observed in their randomly rewired equivalents [32]. Similar findings characterize metabolic networks as well [34-35].

The emergence of high clustering provides the first hint towards the recognition that the Erdős-Rényi model cannot account for the topological properties of realistic networks. However, the most significant indication in that direction comes from the degree distributions observed in actual networks. In contrast to the Poisson degree distribution, which is the fingerprint of the Erdős-Rényi graph, cellular networks consistently follow a power-law degree distribution [2,5,31,36-45], predicting that the probability for a randomly chosen node to have exactly k links is given by

$$P(k) \sim k^{-\gamma}, \qquad (2)$$

where γ takes values, which are typically between two and three. This finding has profound implications on the architecture of biological networks, as well as on their evolution and their functionality. We discuss these implications in the next section.

5 Scale-Free Nature of Cellular Networks

As the structure of cellular networks was elucidated, it became evident that their topology does not obey the typical narrow distribution observed for many other quantities in nature. Instead of the commonly found Poisson, Gaussian and exponential distributions, cellular networks feature a power-law degree distribution. The first evidence for this came from metabolic networks, where we take the nodes to represent the metabolites, and the directed links to represent the enzyme-catalyzed chemical reactions between them. The analysis of metabolic networks from as many as forty three different organisms revealed that they are all characterized by a power-law degree distribution [31]. Similar findings followed from the study of protein-protein interaction networks [32], and

transcriptional regulatory networks [38,44].

The Scale-Free Property

In contrast with the Poisson (and other narrow) distributions, the power-law distribution is not concentrated around its mean. Networks characterized by such a degree distribution are thus highly non-uniform - most of the nodes have only a few links, whereas a few nodes have a disproportionately large number of links. These highly connected nodes, often called *hubs*, are the glue that binds the majority of low degree nodes together. The presence of these hubs, which is strictly banned in narrow degree distributions, is observed in practically all the analyzed cellular networks, ranging from the ultra-reactive pyruvate and coenzyme A in metabolic networks, to the insulin receptor in protein-protein interaction networks [2,5]. This can be seen in Fig. 2(a), where we display the protein-protein interaction network of *S. cerevisiae*. In this representation the node size is proportional to its degree, so that the clearly visible variability in the node sizes illustrates the heterogeneity in their degrees. While most proteins participate in one, two or three interactions, a few hubs participate in well above ten, and the degree of some even exceeds a hundred.

Location of Fig. 2

These highly heterogeneous topologies differ essentially from the classical Erdős-Rényi networks, in that they do not have a *typical node*. In an Erdős-Rényi network, the degrees of most nodes are in the vicinity of the average degree. The scarcity of nodes with any given degree can be estimated by comparing with the average degree of the network. In that sense the average degree provides a characteristic scale by which the rest of the nodes should be measured. In contrast, a power-law degree distribution, of the form of Eq. (2), allows for the coexistence of nodes with an extremely broad range of degrees, freeing the network of any typical scale. The cellular networks are thus *scale-free* (SF) networks [36]. Graphically, the power-law degree distribution forms a straight line when plotted on logarithmic axes, with the slope providing the scaling exponent, γ [Fig. 3]. This yields an intuitive illustration for the concept of the SF topology. It shows, graphically, that one

cannot assign a typical scaling, since the graph duplicates itself regardless of the scaling used in the horizontal axis (denoting the degrees).

Of particular significance in the characterization of the degree distribution is the value of the scaling exponent, γ. The broadness of the distribution becomes larger as the value of γ becomes smaller. This means that smaller γ values characterize more degree heterogeneous networks. More specifically, Eq. (2) features three different regimes as the value of γ is changed. To observe this we consider the value of the nth moment, $\langle k^n \rangle$, of the distribution, given by $\int k^n P(k) dk$. Note that for an infinite network, where k ranges from zero to infinity, this integral diverges if $n - \gamma \geq -1$. For $\gamma \leq 2$ the divergence is observed already at the level of the first moment. In such cases, the distribution is so broad, that the average is undefined. In practice, when the network is finite, this will take form in a topology where almost all the nodes have degrees significantly lower than the average, and a small minority of nodes will have such a high degree, that they connect directly to a significant fraction of the nodes in the network, namely their degree is of order N. For $2 < \gamma \leq 3$ (typical of most cellular networks) the distribution has a finite average, but the second moment diverges. This means that the variance, $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, becomes undefined, capturing the high variability of the distribution. These mathematical pathologies are removed once $\gamma \geq 3$, as when the scaling exponent is above this threshold, for many practical purposes, the scale-free nature of the distribution is no longer relevant. These three regimes are also expressed in the average path length of the network, as we discuss in more detail below.

Location of Fig. 3

Network Integrity and the Role of Hubs

The SF topology allows for a disproportionate number of highly connected nodes. These nodes play a crucial role in the structural integration of the network. To understand this we consider the majority of nodes in the network, which have only a few links. They are all likely to be connected to the hubs by very short paths (of one or two edges). In addition, any selected pair of hubs is also

likely to be very close to one another, due to the large number of links that they have. The result is that in SF networks the path between nodes becomes even shorter than in Erdős-Rényi networks, the hubs playing the role of network *shortcuts* [46-48]. In fact, in a SF network, where the scaling exponent is $2 < \gamma < 3$ the average path length satisfies $\langle l \rangle \sim \ln \ln N$, adding an additional logarithmic correction to the average path length characteristic of Erdős-Rényi networks [Eq. (1)]. For $\gamma = 3$ it is found that $\langle l \rangle \sim \ln N / \ln \ln N$, and for $\gamma > 3$ the result is like that of the Erdős-Rényi network. Cellular networks, for which the scaling exponent is usually between two and three, are thus *ultra-small worlds* [49].

The analysis above, regarding the importance of the hubs as the structural backbone of SF networks, has some surprising implications on the robustness of cellular networks to random perturbations. Our intuition leads us to view complex systems as highly intricate structures, which depend strongly on the proper functionality of all of their components. When a significant fraction of their nodes fail, these systems are expected to become dysfunctional. In contrast, biological networks prove to be astoundingly resilient against component failure [50-54]. From the topological perspective this can be attributed to their SF topology, and its hub-based backbone. Scale-free networks have been shown to maintain their structural integrity even under the deletion of as many as 80% of their nodes. The remaining 20% will still form a connected component [55-57]. This is while in an Erdős-Rényi network the removal of nodes beyond a certain fraction, inevitably results in the network disintegrating into small isolated components [27]. The source of this topological resilience of SF networks is rooted in their inherent non-uniformity. The vast majority of nodes in SF networks has merely one or two links, playing a marginal role in maintaining the integrity of the network. Most random failures will occur on these unimportant nodes, and thus will not significantly disrupt the network's functionality. The relative scarceness of the hubs, and, on the other hand, their central role in maintaining the network's structural integrity, ensures that random failures will rarely break down the network.

The robustness of cellular networks, which relies strongly on the hub nodes, is, however, a double-edged sword. While allowing the networks to withstand a large number of random failures, it makes

it extremely vulnerable to intentional interventions. The removal of just a small number of key hubs will cause the SF network to break down into isolated dysfunctional clusters [56-57]. Supporting evidence for this comes from the small number of lethal genes found in many organisms, and, on the other hand, by the relatively large number of hubs found among these genes [39,41,58-65].

The Origins of the Scale-Free Topology

The SF topology is found to be a universal feature of many real networks, both in the context of biology and in social and technological systems [5]. This ubiquitous topological feature not only characterizes the architecture of a given network, but also serves as an indicator for its formation process. This idea is captured by the Barabási-Albert model, which attributes the emergence of a SF topology to the presence of two fundamental formation processes: network growth and preferential attachment [36]. By growth we refer to the fact that networks are not static. They evolve in time by constantly adding new nodes and new links. By preferential attachment we state that nodes are more likely to link to already highly connected nodes. For a more accurate definition, consider an evolving network, where at each time-step, a single new node is introduced, drawing m new links to any of the existing nodes. According to the preferential attachment mechanism the new node will choose to connect to the existing node, x, with a probability proportional to x's current degree, namely

$$P(x) = \frac{k_x}{\sum_i k_i}, \tag{3}$$

where the sum in the denominator is over all nodes in the current state of the network. These two processes, growth and preferential attachment, give rise to the observed power-law degree distributions. It can be shown that any one of these processes alone is insufficient and does not yield the desired SF topology. Network growth is required, as otherwise the network reaches saturation, and the degree distribution becomes nearly Gaussian. The preferential attachment mechanism is

needed to support the formation of hubs [36]. By this mechanism, if a node has many links, it is more likely to acquire new links, creating a state where the *rich get richer*. The result is that the more connected nodes gain new links at a higher rate, and eventually emerge as hubs. Eliminating the preferential attachment mechanism leads to an exponential distribution, much less broad than a power-law.

Preferential Attachment in Biological Networks

The realization of the Barabási-Albert model in the formation of cellular networks is rooted in the process of gene duplication [66-71]. This process is clearly responsible for network growth, as duplicated genes produce duplicate proteins, and thus introduce new nodes into the network. The more delicate point is that gene duplication also adheres to the rules of preferential attachment. To understand this, consider an interaction network which grows via node duplication. At each time step, a random node is chosen, say x, and an identical node, $\tilde{x}$, is created. This newly created *duplicate* node will have exactly the same interactions as the original node. This means that each of x's nearest neighbors will receive a new edge. Therefore the distribution of new links in the network is biased towards the more connected nodes. Indeed, a node with many nearest neighbors is more likely to have one of its neighbors chosen for duplication. In fact, for a given node with degree k, the probability for a randomly chosen node to be linked to it is directly proportional to k. Thus its probability to gain a link in the growth process is also proportional to k, consistently with Eq. (3).

One of the predictions of the Barabási-Albert model is that nodes can become well connected by virtue of being older. A node that was introduced early in the history of the network will have more time to accumulate links, and, by the rich get richer mechanism, enhance its chances of becoming a hub [36]. In metabolic networks, we find that the hubs do, indeed, tend to be older. Some examples are coenzyme A, NAD and GTP, remnants of the RNA world, which are among the most connected substrates of the metabolic network [34]. Similar findings rise from the analysis of protein-protein interaction networks, where, on average, the evolutionary ancient proteins are characterized by higher degrees [72-73]. This offers direct empirical evidence to the preferential attachment

hypothesis.

6 Hierarchy and Modularity

The ability of complex systems to properly function and carry out vital tasks requires the cooperation of many independent components. In many artificial networks this is commonly achieved by relying on a hierarchical design. The network is layered, and nodes at one level orchestrate the behavior of their subordinates, belonging to a level below. In that sense we tend to picture network hierarchy as related to a tree-like topology. However, the idea of having distinct hierarchical layers of nodes stands in sharp contrast with the scale-free nature of the cellular networks. The presence of hubs, which directly connect to a large fraction of the nodes in the network, will inevitably break down the layered topology. We thus have to adopt a different notion of hierarchy to account for the functional design of biological networks.

The conceptual idea is that the functionality of these elaborate networks can be broken into distinct, relatively isolated tasks [35,74-78]. From a structural point of view, this will be expressed in networks that are composed of highly interconnected sub-graphs, or modules. The hierarchical disposition of a given node can be characterized by the number of such sub-graphs to which it belongs. This way, a node which is placed low in the hierarchy will participate in just one functional task, and thus belong to just one module. Higher in the hierarchy we find nodes that bridge between two or three different modules. Eventually, at the highest level of the hierarchy will reside the hubs, which do not belong to any specific sub-graph, but rather connect many sub-graphs, that will otherwise be isolated. The quantifiable fingerprint of such a hierarchical design can be found in $C(k)$, which describes the dependence of the clustering coefficient on the degree [35,45,79-80]. Low degree nodes will tend to belong to a specific module, and thus feature a high clustering coefficient – indeed, almost all their neighbors will themselves be part of the same module. The hubs, on the other hand, will be connected to many nodes from different modules, and accordingly will tend to have a low clustering coefficient.

The analysis of cellular networks shows clear evidence of hierarchical topology. The dependence of the clustering coefficient on the degree features a power-law scaling, $C(k) \sim k^{-\beta}$. This has been observed for metabolic networks [35], protein-protein interaction networks [32] and regulatory networks, with β taking values, typically between one and two.

Party vs. Date Hubs

We have already acknowledged the crucial role that the hubs play in the integration of the network. In the above discussion we further emphasized their importance when the network has modular structure, as the mediators between separate modules. In this context, an interesting distinction between two types of hubs has been proposed [81]. The first type, named *party* hubs, corresponds to our usual perception of hubs – nodes that interact with many other nodes simultaneously. The second type, *date* hubs, bind to their partners at different times or at different cellular locations. While the party hubs tend to interact within a module, it is the date hubs which typically connect between separate modules. So that it is mainly the latter that serve as the integrators of the network. In the analysis of the yeast protein-protein interaction network these two types of hubs were, indeed, identified [81]. When the date hubs were systematically removed, the network split into small disconnected modules. In contrast, the removal of party hubs, while diluting the modules themselves, harmed the overall integrity of the network to a much lesser extent.

7 Degree Correlations

It is commonly observed in networks that similar nodes tend to connect to one another. This feature, termed assortative mixing, can be related to any characteristic of a node, and in particular to the node's degree. For instance, in social networks individuals with many friends tend to link to others who too have a high degree. However, as shown in Fig. 2(b), in the featured protein-protein interaction network the opposite is true: the network is disassortative, which means that the hubs tend to avoid each other, leading to a network where highly connected nodes are surrounded by low degree nodes [82]. This disassortativity is observed in most biological networks, including the

metabolic and regulatory networks, and is, in fact, a property shared by technological networks, such as the power grid or the Internet [83-84].

To classify a network as assortative or disassortative we first need to define our expectations of a *neutral* network. What we are aiming at, is to characterize the expected correlations between the degrees of nearest neighbors in the absence of any assortative bias. To do this we consider the random selection of an edge in the network, and calculate the probability that at one of its ends resides a node with degree of k and at the other a node with a degree of k'. Let us first calculate the probability for the first node, namely we are seeking the probability to find a node with k links at the end of a randomly selected edge. This is essentially different than the direct selection of a random node, since it gives an advantage to nodes with a higher degree. The reason is simply because such nodes have a larger number of edges to which they are attached. For instance, to reach a node with a single edge through this procedure, one must pinpoint the one edge leading to it. On the other hand there are k potential edges through which a k degree node can be reached, making this outcome k times more likely. Thus the desired probability is proportional to the abundance of k degree nodes, as well as to the degree itself, namely it is $q_k = kP(k)/\langle k \rangle$, where the denominator is used as a normalization constant. In a neutral network, the degree distribution of the nodes that lay at the other end of the selected edge is independent of q_k. Thus, in the absence of degree correlations, the probability that a randomly selected edge links between two nodes with a degree of k and k' is simply $Q_{kk'}^{\mathrm{Neu}} = q_k q_{k'}$. To evaluate the assortativity of the network we compare the observed probability $Q_{kk'}$ to $Q_{kk'}^{\mathrm{Neu}}$. For an assortative network, the observed probability will show a positive bias along the diagonal, where the value of k is close to that of k'. Disassortativity will be expressed as a negative bias along the diagonal, and a tendency to have more links where $k \neq k'$.

Another, more compact description, of the degree correlations can be viewed by observing the average degree of a node's nearest neighbors. We denote this average by K_{nn}. We then average over all nodes with a given degree, k, to obtain $K_{\mathrm{nn}}(k)$, namely the average degree of the neighbors surrounding a typical node with k links. In a neutral network, K_{nn} should not depend on k, but if degree correlations are present, they will be expressed in a monotonic increase or decrease in

$K_{\mathrm{nn}}(k)$. In the protein-protein interaction network displayed in Fig. 2, this dependency is clearly visible – the average degree of the hub's nearest neighbors are between one and two, and yet the low degree nodes are almost all connected to the hubs, so that for them K_{nn} is much greater. Indeed the analysis shows that for this network $K_{\mathrm{nn}}(k) \sim k^{-\alpha}$, where $\alpha \approx 0.24$ [85].

One can obtain an even more compact parameterization for a network's assortativity, by referring to the Pearson-correlation coefficient measured between the degrees of pairs of connected nodes. This can be explicitly done by extracting the correlation coefficient for k and k' from the distribution given by $Q_{kk'}$. The result is [83]

$$r = \frac{\langle kk' \rangle - \langle k \rangle \langle k' \rangle}{\sigma^2}, \tag{4}$$

where σ^2 is the variance obtained from the distribution q_k. The parameter r takes values between 1, for a perfectly assortative network, and -1, when the network is perfectly disassortative. For the yeast protein-protein interaction network, shown in Fig. 2(a), it measures $r = -0.156$, confirming that the network is, indeed, disassortative.

It remains unclear what is the mechanism responsible for the disassortative nature of biological networks. It cannot be accounted for by the Barabási-Albert mechanism, which does not yield any degree correlations [83]. From a functional point of view, it highlights the modular structure of biological networks, possibly strengthening even further the central role of the hubs. It was also shown that disassortativity harms the resilience of the network, and makes it more vulnerable to the intentional removal of hubs, since in such networks, the majority of low degree nodes are connected solely to the hubs. On the other hand, disassortativity has a positive contribution to the integrity of the network when it is not under attack, as typically a disassortative network will feature a larger giant connected component than an assortative or a neutral one [83]. This once again emphasizes the resilience of cellular networks against random failure, compared to their vulnerability against

selected node removal.

8 Human Disease Network

The applications of graph theory to systems biology can go beyond the mapping of the concrete network systems found within the cell. Graphs could also be used as a means of organizing biological information in a way that could potentially spark new insights. An innovative example is provided by the network approach to the study of human diseases [86]. In this approach two networks are constructed. The first network is the human disease network. In this network the nodes represent genetic disorders, and the edges link disorders which are associated with mutations in the same gene. The second network is the disease gene network. Here the nodes represent genes, and the edges link genes which are associated with the same disorder. Both networks are found to be highly clustered, showing that diseases, as well as disease genes, tend to divide into modules, or families, of related disorders. In the disease gene network genes that contribute to the same disorder tend to be correlated in many other ways as well. They have an increased tendency to be expressed together in specific tissues, they typically display high coexpression levels, and, in many cases, they share common cellular and functional characteristics, as annotated in the Gene Ontology [86]. This network is also in close relation with the protein-protein interaction network, as disease related genes are very likely to have their products interact together through physical binding. A surprising feature that this analysis reveals is the distinction between lethal genes and disease genes. In many cases, the products of lethal genes are highly connected nodes in the protein interaction network [41]. This emphasizes the importance of the hubs for the proper function of the network. In contrast, disease genes tend to avoid the hubs, and the vast majority of them are nonessential. It has been suggested that this is driven by natural selection, enabling the proliferation of mutations only if they harmed non vital genes [86]. Confirming evidence comes from the fact that somatic mutations, which do not harm the organism's reproduction, are indeed more frequently related to hub genes. In a broader perspective, this network approach to human diseases offers a tool for the understanding of general patters in genetic disorders, and could potentially reveal connections which are not apparent in the study of individual disorders [87-91].

9 The Building Blocks of Cellular Networks

In the previous section we discussed the macroscopic aspects of the hierarchical topology. We have shown that the hierarchy in cellular networks is closely intertwined with their modular structure. Indeed, from a functional point of view, biology is full of examples of modularity, where a distinct group of proteins, genes or metabolites is responsible for the execution of some basic biological operations. Topologically, as discussed earlier, this is expressed in the emergence of various sub-graphs composed of highly interlinked groups of nodes. The high clustering typical of cellular networks provides the quantitative evidence for this modular network structure. In this section we focus on the typical recurring structures of these sub-graphs, and their meaning. In a sense, we are lowering the altitude of the bird's eye perspective by which we viewed the networks until now, going from the macroscopic analysis, to a more focused look at the building blocks of our complex systems.

Sub-graphs and Motifs

In order to conduct a fruitful analysis of network modules, we need first to develop a scheme by which we can identify what are meaningful modules. For instance, consider a tetrahedral sub-graph, which is a fully connected set of four nodes as shown in Fig. 4(a). We can evaluate the abundance of this sub-graph in our network, but this will not be sufficient in order to tag it as a significant functional module. The randomness in the network topology makes it probable that such a module is due to appear in the network by chance. We thus consider a certain module to be a significant *motif* if is overrepresented in that network, that is more abundant that expected by chance alone [92-93]. The idea is that if the network has the tendency to over represent a certain module, there must be an evolutionary or functional need for it. Since natural selection discriminates on the basis of functional criteria, it will be these motifs that are likely to be capable of carrying important biological functions.

Randomized Networks

As we stated above, for a certain module to qualify as a motif, it must be more abundant then expected by chance. However, we have not accurately defined what we mean by this criterion. It might, at first glance, seem intuitive to use the Erdős-Rényi networks as the grounds for comparison. However a more careful look shows that this is not sufficient. The reason is that the expected frequency of a given sub-graph is dictated by the degree distribution of the graph as a whole. Consider, for instance, the tetrahedral module discussed above. This module consists of four nodes and six links. In order for such a module to emerge, first we need to have a node with a degree of at least three. Then there must be at least three additional links among this node's nearest neighbors. The likelihood of the first condition is dictated by $P(k)$, and the likelihood of the second is determined by $C(k)$. In the broader sense what this means is that the macroscopic features of the network, given by $P(k)$ and $C(k)$, are in close relations with the detailed structure of its modules. In the context of the current discussion, it states that the abundance of a given sub-graph is not independent of $P(k)$. Thus, in order to deem a certain module as overrepresented in a particular network, we must compare its abundance to that of a randomized network with the same degree distribution [94]. Such a randomized network can be constructed by randomly rewiring all the links in the original network, preserving each node's degree, and hence $P(k)$, but deleting fine structure, such as the recurrence of motifs.

Auto-Regulation and the Feed-Forward Loop

We now briefly discuss two noted examples of highly recurring motifs found in transcriptional regulatory networks. The first motif is the negative auto-regulator, which is one of the simplest and most abundant network motifs found in *E. coli* [95-96]. It includes a single transcription factor, which represses its own transcription. Graphically, this motif, shown in Fig. 4(b), is simply a single node loop. It was shown to have two important functions. The first function is response acceleration. Compared to alternative regulating processes, such as protein degradation, the process of auto-regulation allows for a faster response to signals. This was shown both theoretically and

experimentally by employing synthetic gene circuits in *E. coli* [97]. The second advantage is that the motif increases the stability of the gene product concentration against stochastic noise. It thus reduces the variations in protein levels between different cells [98-99].

Another motif frequently encountered in regulatory networks, is the feed-forward loop [100]. This motif consists of three nodes, x, y and z, where x is directly linked to both y and z, and in addition y is also directly linked to z [see Fig. 4(c)-(d)]. The direct links can symbolize the activation or the inhibition of the target gene, or any combination thereof. Thus eight different versions of this motif can be constructed, each with a different biological function [101]. To demonstrate the functional importance of this motif, we focus below on two different versions of the motif. The first is a coherent feed-forward loop, observed in the arabinose utilization and in the flagella systems of *E. coli* [102-103], and the second is an incoherent feed-forward loop, which appears in the galactose system of *E. coli* [104].

Location of Fig. 4

In the coherent feed-forward loop, all the directed links represent the process of activation. Thus the gene x activates both genes y and z, and yet gene y itself activates z once again. This might seem redundant, but can be shown to have important functional implications. Consider the case where the target gene, z, can only be activated if it receives a signal from both x and y. Using a computational analogy, we say that it serves as an *AND* gate, as it yields a *positive output* only when both of its *inputs* are positive. The motif will feature a time lag from when x is activated to when z responds. This is because z will be activated only after a sufficient concentration of y products have been produced. The result is that short sporadic expressions of the x gene will die off before z is ever activated. This motif, thus, functions as a filter, ignoring stochastic short-term perturbations, and responding only to persistent ones. The complementary feature rises when the target gene serves as an *OR* gate. In this case, z is activated by either x or y. Here the delayed response will appear if x suddenly seizes to be expressed, in which case z will still remain active for some time, as long as a sufficient abundance of y's product persists. Thus the stability of z's expression in ensured against

sudden short term drops is the production of x. This type of behavior is observed in the flagella system in *E. coli*, where a persistent activation of the flagella is maintained, even under transient loss of the input signal [105].

A surprising, but nevertheless prevalent, version of the feed-forward motif is the incoherent feed-forward loop. Here while x activates both y and z, the link between y and z is inhibitory. This seemingly contradictory wiring leads to an interesting functional feature. Consider a sudden activation of the gene x, due to, say, an external signal. As a result both y and z will be activated too. For a short time after x's activation, the expression levels of z will be constantly rising, due to its activation by x. However, after a sufficient amount of y products have been produced, the expression of z will be suppressed, due to its inhibition by y. This version of the motif, thus, translates a persistent signal induced by x into a spike of activation of the target gene, z.

10 Going Beyond Topology

Despite its success, the purely topological approach possesses inherent limitations in the race for understanding cellular networks. In focusing on topology alone, we have neglected the fact that *not all edges are created equal*. In practice the dynamical functionality of a complex network is probably affected not just by the binary pattern of who is connected to whom, but also by the nature of this connection and by its strength. Indeed, in a realistic biological network, several reactions are more dominant than others – a feature that is overlooked by topology based analyses. To obtain a more effective description we assign different *weights* to the edges, based on the intensity of the interaction [106-107]. This gives rise to weighted networks, where the link between a pair of nodes i, j, is no longer represented by the discrete state of present versus absent, but by a continuous number, w_{ij}, evaluating its importance.

Assigning the Weights

In metabolic networks, the most natural measure for the weight of a given reaction is its flux,

namely the rate by which a substrate is being converted to its product. The flux-balance approach has proven to be very successful in retrieving these fluxes [108-109]. In this approach one writes a set of equations for the metabolic fluxes, based on the assumption that all the metabolic reactions are balanced, that is to say that the concentrations of the reactants are at a steady state. This amounts to a set of linear algebraic equations for the fluxes. Typically these equations are underdetermined, as they include more variables than equations. To further narrow the solution space one imposes biological and chemical constraints. These constraints may emerge from experimental data, for instance, if some of the fluxes can be directly measured. Other constraints may be thermodynamic in nature, for instance, if a certain reaction is known to be irreversible because the product has a much lower free energy than the substrates. Finally, after characterizing the diminished solution space to which the fluxes are constrained, the specific solution is chosen to be the one that optimizes the predefined biological function (*e.g.* maximal growth rate). For more detailed information regarding flux-balance methods see Chapter 15.

In transcriptional regulatory networks one can rely on microarray datasets to express the strength of a connection between a pair of genes. The coexpression of a pair of genes can be evaluated by measuring the correlations in their expression patterns. Alternatively, one can search for local similarities in the perturbed transcriptome profile of the genes, and use that to infer the network connections and their weights [110-111].

Characterizing the Weighted Topology

Metabolic flux balance analysis has been applied to the metabolic network of *E. coli*, and the complete weighted network has been obtained. Similarly to the topological findings, the weighted reactions were found to display strikingly high variability [112]. The reaction weights, based on the calculated fluxes, range over several orders of magnitude. The weight distribution, just like the degree distribution, follows a power-law, $P(w) \sim w^{-\alpha}$, which, for *E. coli*, has a scaling exponent of $\alpha \approx 1.5$. While the specific fluxes for the different reactions depend on the environmental conditions, the aggregate behavior, captured by the weight distribution, remains unchanged under

various environmental conditions. The fact that $\alpha < 2$ emphasizes the broadness of the observed weight distribution, as, mathematically, the mean value obtained for such a distribution diverges. In practice, one finds that almost all fluxes are below the average, and a few fluxes measure orders of magnitude above it, dominating the dynamics of the system. This provides an interesting illustration for the biochemical activity of metabolism. It suggests that under any given conditions metabolism is dominated by a small set of highly active reactions, embedded in a background of mostly dim chemical activity. A similar pattern of a highly uneven load distribution occurs in the regulatory network of *S. cerevisiae*. Also there most genes have weak correlations, while a few pairs show quite significant correlation coefficients [113].

Topology Correlated Weights

The connectedness of a node, characterized topologically, is captured by its degree. The weighted network analog is the node's *strength*. The strength is defined as the sum of all weights assigned to the node's set of links, namely $s_i = \sum_j w_{ij}$, where the absence of a link between a pair of nodes is denoted by setting the corresponding weight to zero. To characterize the relationship between the network topology and the link weights, we measure the dependence of the strength on the node degrees, namely $s(k)$. This function typically takes the form of $s(k) \sim k^{\beta}$, where a linear dependence, namely $\beta = 1$, reflects the absence of such degree-strength correlations, meaning that the weights are evenly distributed among the edges, so that a node acquires more strength simply because it has more links. However, in real systems it is commonly observed that $\beta > 1$. This implies that nodes with a higher degree tend to also have links with higher weight. This feature was explicitly observed in the *E. coli* metabolic network [114], where it was found that highly weighted links favored highly connected nodes. For each pair of connected nodes the weight, w_{ij}, is measured via flux balance analysis, finding that it features a power-law dependence on the degrees of the two linked nodes, *i.e.* $\langle w_{ij} \rangle \sim \left(k_i k_j\right)^{\theta}$, with $\theta \approx 0.5$.

Location of Fig. 5

Controllability

The functional state of a cellular network at any given time can be characterized by the concentrations of the reacting molecules – be them proteins, metabolites or any other bio-molecules. This defines a vast state space that the network could explore. However, there are cases, where some parts of this vast state space are restricted due to the dynamics of the network. For instance, consider the simple case of just two interacting metabolites, X and Y, in which X produces Y, namely $X \rightarrow Y$. The state of this system is described by a point in the two dimensional space given by $\vec{N} = (N_X, N_Y)$, where $N_{X(Y)}$ is the concentration of X (Y). However, by assigning a certain value for N_X, the result for N_Y becomes predetermined via Y's production by X. The system is, thus, confined to a small sub-space of the complete two dimensional state space. If one wished to steer the system to any desired state, the node X must be explicitly *driven* by an external input. We say [115-116] that this system can be *controlled* (that is manipulated into any desired state), with a single *driver node*, X. In general if we wish to control a complex network, we first need to identify the set of driver nodes that, if driven by different signals, can offer full control over the network. We are particularly interested in identifying the minimum number of driver nodes, whose control is sufficient to fully control the system's dynamics [117]. An illustration for simple three node networks is shown in Fig. 5.

Applying the above concepts to transcriptional regulatory networks reveals that they are typically difficult to control. This is expressed by their relatively large number of driver node. As an example, for a typical transcriptional regulatory network, approximately 80% of the nodes are driver nodes, indicating that in order to steer these networks, the majority of the nodes have to be explicitly controlled [117]. A higher extent of controllability is observed for metabolic networks, where the fraction of driver nodes is typically around one third. From a topological point of view, dense and homogeneous networks are relatively easy to control, while sparse and inhomogeneous ones are hard to control. This implies that the degree distribution plays an important role in determining the controllability of a network. More specifically, scale-free networks, which are highly non-regular, will feature a large number of driver nodes, and thus be difficult to control.

Interestingly, controllability is not governed by the hubs, as the drives nodes tend to avoid the high degree nodes.

The results presented above might seem to defy out intuitive perception of biological networks as systems which are expected to be firmly controlled. However, when examined once again, they might offer some deeper insight on the nature of control in biological systems. The fact that there are many driver nodes, and that they are typically the less central nodes in the network, is an expression of the highly constrained nature of these networks. It shows that the cellular networks, are not *free* to explore the entire state space, but are rather confined to a restricted area of this space. Thus the only way in which these networks can be driven into a predefined final state is by explicitly driving almost each one of their nodes by an external signal. Especially the low degree nodes, whose state is otherwise governed by the hubs. In a sense, one can interpret these results as the strategy of cellular networks to circumvent external control, and maintain their function, even if a large number of nodes is being influenced.

Differential Networks

A given network topology may give rise to a variety of dynamical behaviors under different dynamical rules or environmental conditions. External stimuli may trigger the activity of different parts of the network, further impacting its dynamical functionality. However, a much broader range of dynamical behaviors could be achieved under the rule of a non-static topology, where the structure of the network itself can react to environmental changes and external stimuli. Recently, it has been found that cellular networks indeed take advantage of this source of dynamical diversity, altering the architecture of the network itself under different conditions or biological states, such as tissue type, disease state or the surrounding environment [118].

11 From Structure to Dynamics

The path taken throughout this chapter outlines, in some sense, the approach of network biology

towards its future challenges. The chapter begins by describing the network representation of cellular systems, examining their topological properties. We then follow with a discussion regarding motifs, weighted networks and controllability, which addresses the dynamics and function of these networks. In this spirit, we end this chapter, with what is probably the most pressing challenge of this area of research – the *bridging between structure and dynamics*. We are currently at a stage where the topological aspects of cellular networks have been thoroughly elucidated, and their evolutionary origins fairly understood. However, we still lack a complete theory, which could interpret the topological findings into a set of dynamical predictions, from which the actual functionality of the networks could be inferred [5]. Below we stress, in a very broad fashion, the strategic path that could meet this challenge [119-120].

The most fundamental question we must address is whether the gap between structure and dynamical behavior could at all be bridged. We need to take into account, that the topology is one actor in a highly detailed cast of network characteristics. In the most detailed description of these cellular systems, all the interactions can take on different reaction processes, and different strengths. By reaction processes we refer to the types of the interactions, e.g. chemical, regulatory, etc., and by strengths we mean that similar processes may occur at different rates. So while structurally, we denote all the various interactions by network links, one should ask: is the process of genetic regulation really comparable to that of physical binding? Is it guaranteed that two structurally identical networks will express similar behavior even if they differ in some other details? Or perhaps, these details, which are overlooked by the graph theoretic analysis, are not important.

A more constructive approach towards the above questions, is asking how far *can* one actually progress with structure alone. It seems pretty clear that a complete time dependent dynamics of the system would require the incorporation of all of the details mentioned above, and is thus beyond the scope of network biology. On the other hand, what *could* be achievable based on a structural analysis is a macroscopic understanding of the network dynamics. More specifically, network science is not expected to be successful in predicting the behavior of a specific set of nodes. It could, however, provide answers to general questions regarding the network as a whole. Questions

such as: who are the most effective nodes in this network? Does this network support long range interactions, or is the impact of nodes contained locally? Are the concentrations of the nodes in this network stable or governed by fluctuations? Will a small perturbation cause a macroscopic failure? At least some of these characteristics, and others like them, might be determined by the topology of the network, regardless of the other details of the interactions. And if that is the case, we should be able to address these important questions with the tools of network analysis.

In a broader perspective, applying graph theory to the study of complex systems is aimed at bringing about an intuitive, visual and mathematical, toolkit for their understanding. In that sense, the challenge of this approach is to devise a set of intuitive dynamical interpretations to the already defined set of topological features. The idea is to assign a functional meaning to characterizations such as a *broad degree distribution*, *high clustering*, *small world*ness, etc.. Along this path, the future research is to challenge some of the common wisdoms regarding these attributions between structure and dynamics. For instance, the intuitive notion that in a small world topology, all the nodes are impacted by one another, since they are just a few reactions away; or the common perception that the hubs are the most influential nodes in the network. Once these statements, and others like them, are examined, they will bring forth a *new intuition* on the meaning of different structural attributes. Then, by analyzing the structure of a network, researchers will be able to make general assessments regarding its expected dynamics.

The rapidly improving experimental techniques in biology will hopefully enable to test the dynamical predictions derived from network analysis. However, even where the existing experimental procedures are insufficient, help might arrive from unexpected sources. Perhaps the greatest success of the network approach thus far is in revealing the universal nature of the topology of networks, cellular and others, providing a set of tools and criteria by which to classify and characterize the structure of these diverse systems [5]. A similar degree of universality in the dynamics of networks, if found, will provide us with a parallel set of unifying principles, allowing us to describe, using a common platform, various dynamical processes, and make meaningful predictions on the behavior of networks from diverse fields. These universal dynamical aspects

could then be inferred from one system to the other. Metaphorically, this expands the boundaries of the classical biology laboratories far beyond their traditional *walls*. As data is currently collected in vast amounts from biological, social and technological systems, the abilities that network science opens to learn from one system about the other provide a crucial source of empirical strength. A strength, that may one day help make *complex* systems, slightly more *simple*.

Bibliography

1. Alm E., and Arkin A. (2003). Biological networks. *Current Opinion in Structural Biology* **13**,193-202.
2. Barabási A.L., and Oltvai Z.N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews – Genetics* **5**, 101-113.
3. Bray D. (2003) Molecular networks: the top-down view. *Science* **301**, 1864-1865.
4. Vidal M, Cusick M.E., and Barabási A.L. (2011). Interactome networks and human disease. *Cell* **144**, 986-998.
5. Albert R., and Barabási A.L., (2002). The statistical mechanics of complex networks. *Review of Modern Physics* **74**, 47-97.
6. Newman M. (2003). The structure and function of complex networks. *SIAM Reviews* **45**, 167-256.
7. Schadt E.E., Lamb J., Yang X., Zhu J., Edwards S., Guhathakurta D., Sieberts S.K., Monks S., Reitman M., Zhang C., Lum P.Y., Leonardson A., Thieringer R., Metzger J.M., Yang L., Castle J., Zhu H., Kash S.F., Drake T.A., Sachs A., Lusis A.J. (2005). An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics* **37,** 710-717.
8. Chuang H.Y., Lee E., Liu Y.T., Lee D., Ideker T. (2007). Network-based classification of breast cancer metastasis. *Molecular Systems Biology* **3**, 140-150.
9. François K. (2007). "Biological Network. Complex Systems and Interdisciplinary Science – Volume 3", 1st ed. World Scientific Publishing Company, Hackensak, New-Jersey.
10. Ihmels J., Friedlander G., Bergmann S., Sarig O., Ziv Y., and Barkai N. (2002). Revealing modular organization in the yeast transcriptional network. *Nature Genetics* **31**,370-377.
11. Stelzl U., Worm U., Lalowski M., Haenig C., Brembeck F.H., Goehler H., Stroedicke M., Zenkner M., Schoenherr A., Koeppen S., Timm J., Mintzlaff S., Abraham C., Bock N., Kietzmann S., Goedde A., Toksöz E., Droege A., Krobitsch S., Korn B., Birchmeier W., Lehrach H., and Wanker E.E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957-968.
12. Uetz P., Giot L., Cagney G., Mansfield T.A., Judson R.S., Knight J.R., Lockshon D.,

Narayan V., Srinivasan M., Pochart P., Qureshi-Emili A., Li Y., Godwin B., Conover D., Kalbfleisch T., Vijayadamodar G., Yang M., Johnston M., Fields S., and Rothberg J.M. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. (2000) *Nature* **403**, 623-630.

13. Rain J.C., Selig L., De-Reuse H., Battaglia V., Reverdy C., Simon S., Lenzen G., Petel F., Wojcik J., Schächter V., Chemama Y., Labigne A., Legrain P. (2001). The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211-216.
14. Giot L., Bader J.S., Brouwer C., Chaudhuri A., Kuang B., Li Y., Hao Y.L., Ooi E., Godwin B., Vitols E., Vijayadamodar G., Pochart P., Machineni HHH., Welsh M., Kong Y., Zerhusen B., Malcolm R., Varrone Z., Collis A., Minto M., Burgess S., McDaniel L., Stimpson E., Spriggs F., Williams J., Neurath K., Ioime N., Agee M., Voss E., Furtak K., Renzulli R., Aanensen N., Carrolla S., Bickelhaupt E., Lazovatsky S., DaSilva E., Zhong J., Stanyon C.A., Finley Jr R.L., White K.P., Braverman M., Jarvie T., Gold S., Leach M., Knight J., Shimkets R.A., McKenna M.P., Chant J., Rothberg J.M. (2003). A Protein Interaction Map of *Drosophila melanogaster*. *Science* **302**, 1727-1736.
15. 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.
16. Costanzo M.C., Crawford M.E., Hirschman J.E., Kranz J.E., Olsen P., Robertson L.S., Skrzypek M.S., Braun B.R., Hopkins K.L., Kondu P., Lengieza C., Lew-Smith J.E., Tillberg M., Garrels J.I. (2001). YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res*. **29**, 75-84.
17. Li S., Armstrong C.M., Bertin N., Ge H., Milstein S., Boxem M., Vidalain P.O., Han J.D., Chesneau A., Hao T., Goldberg D.S., Li N., Martinez M., Rual J.F., Lamesch P., Xu L., Tewari M., Wong S.L., Zhang L.V., Berriz G.F., Jacotot L., Vaglio P., Reboul J., Hirozane-Kishikawa T., Li Q., Gabel H.W., Elewa A., Baumgartner B., Rose D.J., Yu H., Bosak S., Sequerra R., Fraser A., Mango S.E., Saxton W.M., Strome S., Van-Den-Heuvel S., Piano F., Vandenhaute J., Sardet C., Gerstein M., Doucette-Stamm L., Gunsalus K.C., Harper J.W., Cusick M.E., Roth F.P., Hill D.E., Vidal M. (2004). A map of the interactome network of

the metazoan *Caenorhabditis elegans*. *Science* **303**, 540-543.

18. Costanzo M., Baryshnikova A., Bellay J., Kim Y., Spear E.D., Sevier C.S., Ding H., Koh J.L., Toufighi K., Mostafavi S., Prinz J., St. Onge R.P., VanderSluis B., Makhnevych T., Vizeacoumar F.J., Alizadeh S., Bahr S., Brost R.L., Chen Y., Cokol M., Deshpande R., Li Z., Lin Z.Y., Liang W., Marback M., Paw J., San Luis B.J., Shuteriqi E., Tong A.H., van Dyk N., Wallace I.M., Whitney J.A., Weirauch M.T., Zhong G., Zhu H., Houry W.A., Brudno M., Ragibizadeh S., Papp B., Pál C., Roth F.P., Giaever G., Nislow C., Troyanskaya O.G., Bussey H., Bader G.D., Gingras A.C., Morris Q.D., Kim P.M., Kaiser C.A., Myers C.L., Andrews B.J., Boone C. (2010). The genetic landscape of a cell. *Science* **327**, 425-456.
19. De-Las Rivas J., Fontanillo C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. (2010). *PLoS Computational Biology* **6**, e100807.
20. Stumpf M.P., Thorne T., de-Silva E., Stewart R., An H.J., Lappe M., Wiuf C. (2008). Estimating the size of the human interactome. Proc. Natl. Acad. Sci. 105, 6959-6964.
21. Tong A.H.Y., Lesage G., Bader G.D., Ding H., Xu H., Xin X., Young J., Berriz G.F., Brost R. L., Chang M., Chen Y., Cheng X., Chua G., Friesen H., Goldberg D.S., Haynes J., Humphries C., He G., Hussein S., Ke L., Krogan N., Li Z., Levinson J.N., Lu H., Ménard P., Munyana C., Parsons A.B., Ryan O., Tonikian R., Roberts T., Sdicu A.M., Shapiro J., Sheikh B., Suter B., Wong S.L., Zhang L.V., Zhu H., Burd C.G., Munro S., Sander C., Rine J., Greenblatt J., Peter M., Bretscher A., Bell G., Roth F.P., Brown G.W., Andrews B., Bussey H., Boone C. (1999). Global mapping of the yeast genetic interaction network. *Science* **286**, 509-521.
22. Salgado H., Gama-Castro S., Peralta-Gil M., Díaz-Peredo E., Sánchez-Solano F., Santos-Zavaleta A., Martínez-Flores I., Jiménez-Jacinto V., Bonavides-Martínez C., Segura-Salazar J., Martínez-Antonio A., and Collado-Vides J. (2006). RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Research*, **34**, D394-D397.
23. Farkas I., Jeong H., Vicsek T., Barabási A.L., Oltvai Z.N. (2003). The topology of the transcription regulatory network in the yeast, *Sacchromyces cervisiae*. *Physica A* **318**, 601-612.

24. Erdős P., and Rényi A. (1959). On Random Graphs I. *Publicationes Mathematicae* **6**, 290–297.

25. Erdős P., and Rényi A. (1960). The Evolution of Random Graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**, 17–61.

26. Gilbert E.N. (1959). Random Graphs. *Annals of Mathematical Statistics* **30**, 1141–1144.

27. Newman M.E.J. (2010). "Networks – an Introduction", 1st ed. Oxford University Press, New York.

28. Holland P.W., and Leinhardt S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies* **2**, 107–124.

29. Luce R.D., and Perry A.D. (1949). A method of matrix analysis of group structure. *Psychometrika* **14**, 95–116.

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

31. Jeong H., Tombor B., Albert R., Oltvai Z.N., and Barabási A.L. (2000). The large-scale organization of metabolic networks. *Nature* **407**, 651-655.

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

33. Xu K., Bezakova I., Bunimovich L., and Yi S.V. (2011). Path lengths in protein-protein interaction networks and biological complexity. *Proteomics* **11**, 1857-1867.

34. Wagner A., and Fell D.A. (2001). The small world inside large metabolic networks. *Proc. Biological Science* **268**,1803-1810.

35. 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**,1551-1556.

36. Barabási A.L., and Albert R. (1999). Emergence of scaling in random networks. *Science* **286**, 509-521.

37. Mason O., and Verwoerd M. (2007). Graph theory and networks in Biology. *IET Systems Biology* **1**, 89-119.

38. Guelzim N., Bottani S., Bourgine P., and Képès F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics* **31**, 60-63.

39. Wagner A. (2001). The yeast protein interaction network evolves rapidly and contains few

redundant duplicate genes. *Molecular Biology and Evolution* **18**,1283-1292.

40. Arita M. (2005). Scale-Freeness and Biological Networks. *J. Biochemistry* **138**, 1-4.
41. Jeong H., Mason S.P., Barabási A.L., and Oltvai Z.N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41-43.
42. Wuchty, S. (2001) Scale-free behavior in protein domain networks. *Molecular Biology and Evolution* **18**, 1964-1702.
43. Apic G., Gough J. and Teichmann S.A. (2001). An insight into domain combinations. *Bioinformatics* **17**, S83-S89.
44. Featherstone D.E., and Broadie K. (2002). Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays* **24**,267-274.
45. Barabási A.L., Dezső Z., Ravasz E., Yook S.H., and Oltvai, Z. (2003). Scale-free and hierarchical structures in complex networks, *AIP Conference Proceedings* **661**, 1-16.
46. Chung F., and Lu L. (2002). The average distance in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99**, 15879-15882.
47. Bollobás B., and Riordan O. (2004). The diameter of a scale-free random graph. *Combintorica* **24**, 5-34.
48. Chen F., Chen Z., Wang X., and Yuan Z. (2008). The average path length of scale-free networks. *Communications in Nonlinear Science and Numerical Simulation* **13**, 1405-1410.
49. Cohen R., and Havlin S. (2003). Scale-free networks are ultra-small. *Physical Review Letters* **90**, 058701:1-058701:4.
50. Winzler E.A., Shoemaker D.D., Astromoff A., Liang H., Anderson K., Andre B., Bangham R., Benito R., Boeke J.D., Bussey H., Chu A.M., Connelly C., Davis K., Dietrich F., Dow S.W., El-Bakkoury M., Foury F., Friend S.H., Gentalen E., Giaever G., Hegemann J.H., Jones T., Laub M., Liao H., Liebundguth N., Lockhart D.J., Lucau-Danila A., Lussier M., M'Rabet N., Menard P., Mittmann M., Pai C., Rebischung C., Revuelta J.L., Riles L., Roberts C.J., Ross-MacDonald P., Scherens B., Snyder M., Sookhai-Mahadeo S., Storms R.K., Véronneau S., Voet M., Volckaert G., Ward T.R., Wysocki R., Yen G.S., Yu K., Zimmermann K., Philippsen P., Johnston M., Davis R.W. (1999). Functional characterization of the *Saccharomyces cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901-906.

51. Giaever G., Chu A.M., Ni L., Connelly C., Riles L., Véronneau S., Dow S., Lucau-Danila A., Anderson K., André B., Arkin A.P., Astromoff A., El-Bakkoury M., Bangham R., Benito R., Brachat S., Campanaro S., Curtiss M., Davis K., Deutschbauer A., Entian K.D., Flaherty P., Foury F., Garfinkel D.J., Gerstein M., Gotte D., Güldener U., Hegemann J.H., Hempel S., Herman Z., Jaramillo D.F., Kelly D.E., Kelly S.L., Kötter P., LaBonte D., Lamb D.C., Lan N., Liang H., Liao H., Liu L., Luo C., Lussier M., Mao R., Menard P., Ooi S.L., Revuelta J.L., Roberts C.J., Rose M., Ross-MacDonald P., Scherens B., Schimmack G., Shafer B., Shoemaker D.D., Sookhai-Mahadeo S., Storms R.K., Strathern J.N., Valle G., Voet M., Volckaert G., Wang C.Y., Ward T.R., Wilhelmy J., Winzeler E.A., Yang Y., Yen G., Youngman E., Yu K., Bussey H., Boeke J.D., Snyder M., Philippsen P., Davis R.W., Johnston M. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391.
52. Gerdes S.Y., Scholle M.D., Campbell J.W., Balázsi G., Ravasz E., Daugherty M.D., Somera A.L., Kyrpides N.C., Anderson I., Gelfand M.S., Bhattacharya A., Kapatral V., D'Souza M., Baev M.V., Grechkin Y., Mseeh F., Fonstein M.Y., Overbeek R., Barabási A.L., Oltvai Z.N., Osterman A.L. (2003). Experimental determination and system-level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriology* **185**, 5673-5684.
53. Yu B.J., Sung B.H., Koob M.D., Lee C.H., Lee J.H., Lee W.H., Kim M.S., and Kim S.C. (2002). Minimization of the Escherichia coli genome using a Tn5-targeted Cre/*loxP* excitation system. *Nature Biotechnology* **20**, 1018-1023.
54. Kolysnychenko V., Plunkett G., Herring C.D., Fehér T., Pósfai J., Blattner F.R., and Pósfai G. (2002). Engineering a reduced Escherichia coli genome. *Genome Research* **12**, 640-647.
55. Albert R., Jeong H., and Barabási A.L. (2000). Error and attack tolerance of complex networks. *Nature* **406**, 378-382.
56. He X., and Zhang J. (2006). Why do hubs tend to be essential in protein networks? *PLoS Genetics* **2**, 826-834.
57. Havlin S., and Cohen R. (2010). "Complex Networks – Structure, Robustness and Function", 1st ed. Cambridge University Press, New York.
58. Yu H., Greenbaum D., Xin Lu H., Zhu X., and Gerstein M. (2004). Genomic analysis of essentiality within protein networks. *Trends in Genetics* **20**, 227-231.

59. Hahn M.W., and Kern A.D. (2005). Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular Biology and Evolution* **22**, 803-809.

60. Kamath R.S., Fraser A.G., Dong Y., Poulin G., Durbin R., Gotta M., Kanapin A., Le Bot N., Moreno S., Sohrmann M., Welchman D.P., Zipperlen P., and Ahringer J. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231-238.

61. Prachumwat A., and Li W.H. (2006). Protein function, connectivity, and duplicability in yeast. *Molecular Biology and Evolution* **23**, 30-39.

62. Yamada T., and Bork P. (2009). Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nature Reviews Molecular Cell Biology* **10**, 791-803.

63. Liang H., and Li W.H. (2007). Gene essentiality, gene duplicability and protein connectivity in human and mouse. *Trends in Genetics* **23**, 375-383.

64. Wuchty S. (2004). Evolution and topology in the yeast protein interaction network. *Genome Research* **14**, 1310-1314.

65. Fraser H.B. (2005). Modularity and evolutionary constraint on proteins. *Nature Genetics* **37**, 351-353.

66. Rzhetsky A., and Gomez S.M. (2001) Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome. *Bioinformatics* **17**, 988–996.

67. Qian J., Luscombe N.M., and Gerstein M. (2001). Protein family and fold occurrence in genomes: power-law behavior and evolutionary model. *J. Molecular Biology* **313**, 673–681.

68. Bhan A., Galas D.J., and Dewey T.G. (2002). A duplication growth model of gene expression networks. *Bioinformatics* **18**, 1486–1493.

69. Pastor-Satorras R., Smith E., and Sole R. (2003). Evolving protein interaction networks through gene duplication. *J. Theoretical Biology* **222**, 199–210.

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

71. 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.

72. Wagner, A. (2003). How the global structure of protein interaction networks evolves. *Proc.*

Royal Society of London B **270**, 457–466.

73. Eisenberg, E., and Levanon, E.Y. (2003). Preferential attachment in the protein network evolution. *Physical Review Letters* **91**, 138701.
74. Hartwell L.H., Hopfield J.J., Leibler S., and Murray A.W. (1999). From molecular to modular cell biology. *Nature* **402**, C47–C52.
75. Wall M.E., Hlavacek W.S., and Savageau M.A. (2004). Design of gene circuits: lessons from bacteria. *Nature Reviews Genetics* **5**, 34–42.
76. Alon U. (2003). Biological networks: the tinkerer as an engineer. *Science* **301**, 1866–1867.
77. Alberts B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294.
78. Ravasz E., and Barabási A.L. (2003). Hierarchical organization in complex networks. *Physical Review E Statistical Nonlinear Soft Matter Physics* **67**, 026112.
79. Deisboeck T.S., Yasha-Kresh J., and Kepler T.B. (Eds.). (2005). "Complex Systems in Biomedicine", 1st ed. Kluwer Academic Publishing, New York, New York.
80. Dorogovtsev S.N., Goltsev A.V., and Mendes J.F.F. (2002). Pseudofractal scale-free web. *Physical Review E Statistical Nonlinear Soft Matter Physics* **65**, 066122.
81. 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., and Vidal M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88-93.
82. Maslov S., and Sneppen K. (2002). Specificity and stability in topology of protein networks. *Science* **296**, 910–913.
83. Newman M.E.J. (2002). Assortative mixing in networks. *Physical Review Letters* **89**, 208701.
84. Vázquez A., Pastor-Satorras R., and Vespignani A. (2002). Large-scale topological and dynamical properties of the Internet. *Physical Review E* **65**, 066130-066142.
85. Colliza V., Flammini A., Maritan A., and Vespignani A. (2005). Characterization and modeling of protein-protein interaction networks. *Physica A* **352**, 1-27.
86. Goh K., Cusick M.E., Valle D., Childs B., Vidal M., and Barabási A.L. (2007). The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685-8690.
87. Oti M., Snel B., Huynen M.A., and Brunner, H.G. (2006). Predicting disease genes using

protein-protein interactions. *J. Medical Genetics* **43**, 691–698.

88. Lage K., Karlberg E.O., Størling Z.M., Olason P.I., Pedersen A.G., Rigina O., Hinsby A., Tümer Z., Pociot F., Tommerup N., Moreau Y., Brunak S. (2007). A human phenome interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology* **25**, 309-316.
89. Franke L., Bakel H., Fokkens L., de-Jong E.D., Egmont-Petersen M., and Wijmenga C. (2006). Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *The American Journal of Human Genetics* **78**, 1011-1025.
90. Sharma A., Chavali S., Tabassum R., Tandon N., and Bharadwaj D. (2010). Gene prioritization in Type 2 Diabetes using domain interactions and network analysis. *BioMed Central Genomics* **11**, 84-94.
91. Köhler S., Bauer S., Horn D., and Robinson P.N. (2008). Walking the interactome for prioritization of candidate disease genes. *American Journal of Human Genetics* **82**, 949-958.
92. Shen-Orr S.S., Milo R., Mangan S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* **31**, 64–68.
93. Milo R., Shen-Orr S.S., Itzkovitz S., Kashtan N., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827.
94. Vázquez A., Dobrin R., Sergi D., Eckmann, J.P., Oltvai Z.N., and Barabási A.L. (2004). The topological relationship between the large-scale attributes and local interactions patterns of complex networks. *Proc. Natl. Acad. Sci. USA* **101**, 17940-17945.
95. Savageau M.A. (1974). Comparison of classical and autogenous systems of regulation in inducible operons. Nature 252, 546-549.
96. Thieffry D., Huerta A.M., Pérez-Rueda E., and Collado-Vides J. (1998). From Specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. *BioEssays* **20**, 433-440.
97. Rosenfeld N., Elowitz M.B., Alon U. (2002). Negative autoregulation speeds the response times of transcription networks. *J. Molecular Biology* **323**, 785-793.

98. Becskei A., and Serrano L. (2000). Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593.

99. Dublanche Y., Michalodimitrakis K., Ku¨mmerer N., Foglierini M., and Serrano L. (2006). Noise in transcription negative feedback loops: simulation and experimental analysis. *Molecular Systems Biology* **2**, 41.

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

101. Alon U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics* **8**, 450-461.

102. Kalir S., Mangan S., and Alon U. (2005). A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Molecular Systems Biology* **1**, 2005.0006.

103. Mangan S., Zaslaver A., and Alon U. (2003). The coherent feed-forward loop serves as a sign-sensitive delay element in transcription networks. *J. Molecular Biology* **334**, 197–204.

104. Mangan S., Itzkovitz S., Zaslaver A., and Alon U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J. Molecular Biology* **356**, 1073–1081.

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

106. Savageau M. (1976). Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology (Addison-Wesley, Reading, 1976).

107. Fell D.A. (1997). Understanding the Control of Metabolism (Portland, London, 1997).

108. Edwards J.S., Ibarra, R.U., and Palsson, B.O. (2001) In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology* **19**, 125–130.

109. Ibarra R.U., Edwards J.S., and Palsson B.O. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**, 186–189.

110. Zhang B., and Horvath S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applied Genetics and Molecular Biology* **4**, 17.

111. de-la Fuente A., Brazhnik P., and Mendes P. (2002). Linking the genes: inferring quantitative gene networks from microarray data. *Trends in Genetics* **18**, 395–398.

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

113. Balaji S., Babu M.M., Iyer L.M., Luscombe N.M., and Aravind L. (2006). Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Molecular Biology* **360**, 213-227.

114. Macdonald P.J., Almaas E., and Barabási A.L. (2005). Minimum spanning trees of weighted scale-free networks. *Europhysics Letters* **72**, 308-314.

115. Kalman R.E. (1963). Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics* **1**, 152‑192.

116. Slotine, J.-J., and Li W. (1991). Applied Nonlinear Control (Prentice-Hall, 1991).

117. Liu Y., Slotine J.-J., and Barabási A.L. (2010). Controllability of complex networks. *Nature* **473**, 167-173.

118. Ideker T., and Krogan N.J. (2012). Differential network biology. *Molecular Systems Biology* **8**, 565.

119. Barzel B., and Biham O. (2009). Quantifying the connectivity of a network: the network correlation function method. *Physical Review E* **80**, 046104-046115.

120. Barzel B. and Biham O. (2011). Binomial moment equations for stochastic reaction systems. *Physical Review Letters* **106**, 150602.

Figure Captions

Figure 1:

(a) An undirected network which includes ten nodes. The network path between nodes number one and eight is emphasized (red). (b) In the directed version of this network, the path must advance in accordance with the direction of the edges.

Figure 2:

(a) The yeast protein-protein interaction network. The size of the nodes is proportional to their degree. The heterogeneity in the node sizes serves as a visual expression of the scale-free nature of the degree distribution. (b) A small portion of the network reveals that it is disassortative, namely that hubs are typically surrounded by low degree nodes.

Figure 3:

The degree distribution of the Erdös-Rényi network (A1) follows a Poisson (A2), as opposed to the cellular networks (B1) and (C1) for which it follows a power-law (B2) – (C2). While the Poisson distribution is concentrated around the mean, the degrees observed in the cellular networks range over several orders of magnitude. This non-uniformity of the cellular networks is expressed in the coexistence of dense and sparse patches visible in the network graph. For the metabolic network, where the edges are directed, both the in-degree and the out-degree distributions are plotted. In all graphs the dots represent the data, while the solid lines are fits to Poisson and to power-law distributions accordingly.

Figure 4:

Network motifs: (a) The hypothetical tetrahedral motif; (b) the auto-regulator; (c) the coherent feed-forward loop and (d) the incoherent feed-forward loop.

Figure 5:

(a) The state of a three node network is given by the concentration assigned to each of the nodes. This defines a point in the three dimensional state-space. Controlling the network means

steering is from any initial state to any desired final state. (b) In this network two of the three nodes must be explicitly controlled in order to manipulate the network. (c) Here it is sufficient to control just one node.