

# CHAPTER 1

## SCALE-FREE NETWORKS IN BIOLOGY

**Eivind Almaas, Alexei Vázquez and Albert-László Barabási**

*Center for Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA*

*Almaas.1@nd.edu, avazque1@nd.edu, alb@nd.edu*

### **1. Introduction**

The last century brought with it unprecedented technological and scientific progress, rooted in the success of the reductionist approach. For many current scientific problems, however, it is not possible to predict the behavior of a system from an understanding of its (often identical) elementary constituents and their individual interactions. For these systems we need to develop new methods in order to gain insight into their properties and dynamics. During the last few years network approaches have shown great promise in this direction, offering new tools to analyze and understand a host of complex systems (1-7). A much studied example concerns communication systems like the internet and the world wide web, which are modeled as networks with nodes being the routers (8) or web pages (9) and the links are the physical wires or URL's, respectively. The network approach also lends itself to the analysis of societies, with people as nodes and the connections between the nodes representing friendships (10), collaborations (11,12), sexual contacts (13) or co-authorship of scientific papers (14,15) to name a few possibilities. It seems that the more we scrutinize the world surrounding us, the more we realize that we are hopelessly entangled in myriads of

interacting webs, and to describe them we need to understand the architecture of the various networks that nature and technology offers us.

Biological systems ranging from food webs in ecology to biochemical interactions in molecular biology can benefit greatly from being analyzed as networks. In particular within the cell the variety of interactions between genes, proteins and metabolites are well captured by network representations, especially with the availability of veritable mountains of interaction data from genomics approaches. In this *Chapter* we will discuss recent results and developments in the study and characterization of the structure and utilization of biological networks.

## 2. Characterizing Network Topology

There are by now many tools and measures available to study the structure of complex networks. In the following we will discuss three of the most fundamental quantities, the degree distribution, node clustering and hierarchy, and the issue of subgraphs and motifs. In addition, it is customary to investigate the betweenness-centrality (BC) of both nodes and links, and the network assortativity. The BC is related to the number of shortest paths going through either a node or a link, and hence a large BC value indicates that the node or link acts as a bridge by connecting different parts of the network (16). The assortativity describes the propensity of a node to be directly connected to other nodes with similar degree (17,18).

### 2.1. Degree Distribution

The representation of various complex systems as networks has revealed surprising similarities, many of which are intimately tied to power laws. The simplest network measure is the average number of nearest neighbors of a node, or the average degree  $\langle k \rangle$ . However, this is a rather crude property, and to gain further insight into the topological organization of real networks, we need to determine the variation in the nearest neighbors, given by the degree distribution  $P(k)$ . For a surprisingly large number of networks, this degree distribution is best characterized by the power law functional form (19) (Fig.1a);

$$P(k) \sim k^{-\gamma}. \quad (1)$$

Important examples include the metabolic network of 43 organisms (20), the protein interaction network of *S. cerevisiae* (21) *C. elegans* (22), *D. melanogaster* (23), and various food webs (24). If the degree distribution instead was single-peaked (e.g. Poisson or Gaussian) as in Fig. 1b, the majority of the nodes would be well described by the average degree and we can with reason talk about a “typical” node of the network. This is very different for networks with a power-law degree distribution; the majority of the nodes only have a few neighbors, while many nodes have hundreds and some even thousands of neighbors. Although average node degree values can be calculated for these networks since their size is finite, these values are not representative of a typical node. For this reason, these networks are often referred to as “scale-free”.

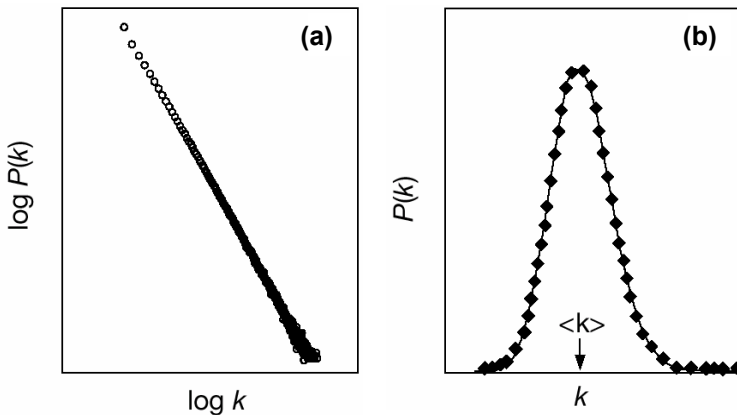


Figure 1. Characterizing degree distributions. For the power-law degree distribution (a), there exists no typical node, while for single peaked distributions (b), most nodes are well represented by the average (typical) node with degree  $\langle k \rangle$ .

## 2.2. Clustering Coefficient

A measure that gives insight into the local structure of a network is the so-called clustering of a node: the degree to which the neighborhood of a node resembles a complete subgraph (25).

For a node  $i$  with degree  $k_i$  the clustering is defined as

$$C_i = \frac{2n_i}{k_i(k_i - 1)}, \quad (2)$$

representing the ratio of the number of actual connections between the neighbors of node  $i$  to the number of possible connections. For a node which is part of a fully interlinked cluster  $C_i = 1$ , while  $C_i = 0$  for a node where none of its neighbors are interconnected. Accordingly, the overall clustering coefficient of a network with  $N$  nodes is given by  $\langle C \rangle = \sum C_i / N$ , quantifying a network's potential modularity. By studying the average clustering of nodes with a given degree  $k$ , information about the actual modular organization of a network can be extracted (26-29): For all metabolic networks available, the average clustering follows a power-law form as

$$C(k) \sim k^{-\alpha}, \quad (3)$$

suggesting the existence of a hierarchy of nodes with different degrees of modularity (as measured by the clustering coefficient) overlapping in an iterative manner (26). In summary, we have seen strong evidence that biological networks are both scale-free (20,21) and hierarchical (26).

### 2.3. Subgraphs and Motifs

A number of complex biological and non-biological networks were recently found to contain network motifs, representing elementary interaction patterns between small groups of nodes (subgraphs) that occur substantially more often than would be expected in a random network of similar size and connectivity (1,2). Theoretical and experimental evidence indicates that at least some of these recurring elementary interaction patterns carry significant information about the given network's function and overall organization (30-33). For example, transcriptional regulatory networks of cells (30,31,34,35 ; see Chapter 4), neural networks of *C. elegans* and some electronic circuits (31) are all information processing networks that contain a significant number of feed-forward loop motifs (see Chapter 2). However, in transcription-

regulatory networks these motifs do not exist in isolation but meld into motif clusters (36), while other networks are devoid of feed-forward loops altogether (31).

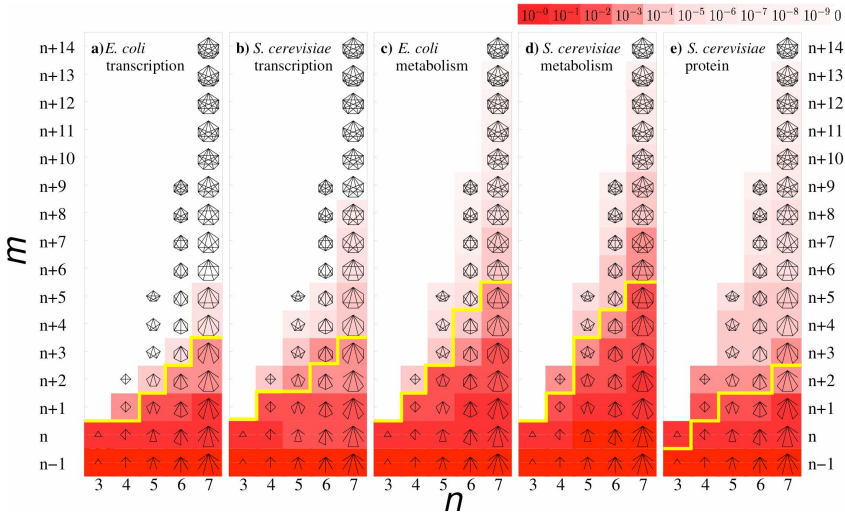


Figure 2. The phase diagrams organize the subgraphs based on the number of nodes ( $n$ , horizontal axis) and the number of links ( $m$ , vertical axis), each discrete point explicitly depicting the corresponding subgraph. The stepped yellow line corresponds to the predicted phase boundary separating the abundant Type I subgraphs (below the line) from the constant density Type II subgraphs (above the line). The background color is proportional to the relative subgraph count  $C_{nm} = N_{nm} / \sum_s N_{ns}$  of each  $n$ -node subgraph, the color code being shown in the upper right corner. Note that some  $(n, m)$  points in the phase diagram may correspond to several topologically distinguishable subgraphs. For simplicity, we depict only one representative topology in such cases. As the yellow phase boundary depends on the  $\gamma$  and  $\alpha$  exponents of the corresponding network, each phase diagram is slightly different. Yet, there is a visible similarity between the networks of the same kind: the phase diagrams of the two transcription or the two metabolic networks are almost indistinguishable.

The number  $N_{nm}$  of subgraphs with  $n$  nodes and  $m$  interactions expected for a network of  $N$  nodes can be estimated from the two key topological parameters of a network's large-scale structure: the degree exponent,  $\gamma$ , and the hierarchical exponent. In general we find that there are two subgraph classes: Type I subgraphs are those that satisfy  $(m-n+1)\alpha - (n-\gamma) < 0$ , their number being given by  $N_{nm}^I \sim N k_{\max}^{-[(m-n+1)\alpha - (n-\gamma)]}$ ,

where  $k_{\max}$  denotes the degree of the most connected node in the network. Type II subgraphs are those that satisfy  $(m-n+1)\alpha-(n-\gamma)>0$ , and their number is given by  $N_{nm}^{II} \sim N$ . As even for finite networks  $k_{\max} \gg 1$ , the typical number of Type I subgraphs is significantly larger than the number of Type II subgraphs ( $N_{nm}^I/N_{nm}^{II} \gg 1$ ). Moreover, for infinite systems ( $N \rightarrow \infty$ ) the relative number of Type II subgraphs is vanishingly small compared to Type I subgraphs, as  $N_{nm}^I/N_{nm}^{II} \rightarrow \infty$ .

This subdivision in Type I and II subgraphs is shown in Fig. 2 for five cellular networks: the metabolic networks of *E. coli* and *S. cerevisiae*, the regulatory networks of *E. coli* and *S. cerevisiae*, and the protein interaction network of *S. cerevisiae*; and different  $(n,m)$  subgraphs. The  $(m-n+1)\alpha-(n-\gamma)=0$  condition, predicted to separate the Type I and II subgraphs, appears as stepped yellow phase boundaries in the phase diagrams. For example, for the *E. coli* transcriptional regulatory network with  $\alpha=1$  and  $\gamma=2.1$  (Table 1) the phase boundary corresponds to a stepped-line with approximate overall slope  $1+1/\alpha=2.0$  and intercept  $-1-\gamma/\alpha=-3.1$  (Fig. 1a). The Type II subgraphs are those above this boundary, and should be either absent, or present only in very low numbers in the transcriptional regulatory network. In contrast, the Type I subgraphs below the boundary are predicted to be abundant. Comparing Figs. 2a-e we find that while the stepped phase boundaries for the different cellular networks differ due to the differences in the  $(\gamma,\alpha)$  exponents (Table 1), the observed densities in the real networks follow relatively closely the predicted phase boundaries. Occasional local deviations from the predictions can be attributed to the error bars of the  $(\gamma,\alpha)$  exponents (Table 1), which allow for some local uncertainties for the phase boundary. Figures 1a-e also indicate that, in agreement with the empirical findings (30-33), each cellular network is characterized by a distinct set of over-represented Type I subgraphs, raising the possibility of classifying networks based on their local structure (4). Yet, the phase diagrams demonstrate that knowledge of the two global topological parameters introduced in Sections 2.1 and 2.2 automatically uncovers the local structure of cellular networks, suggesting that a subgraph- or motif-based classification are equivalent with a classification based on the different  $(\gamma,\alpha)$  exponents characterizing these networks.

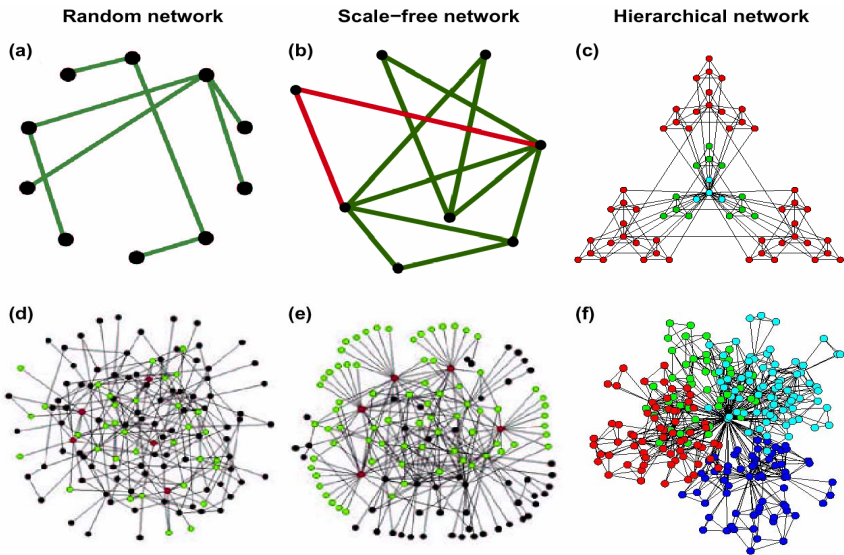


Figure 3. Graphical representation of three network models: **(a)** and **(d)** The ER (random) model, **(b)** and **(e)** the BA (scale-free) model and **(c)** and **(f)** the hierarchical model. The random network model is constructed by starting from  $N$  nodes before the possible node-pairs are connected with probability  $p$ . Panel **(a)** shows a particular realization of the ER model with 10 nodes and connection probability  $p=0.2$ . In Panel **(b)** we show the scale-free model at time  $t$  (green links) and at time  $(t+1)$  when we have added a new node (red links) using the preferential attachment probability (see Eq. (4)). Panel **(c)** demonstrates the iterative construction of a hierarchical network, starting from a fully connected cluster of four nodes (blue). This cluster is then copied three times (green) while connecting the peripheral nodes of the replicas to the central node of the starting cluster. By once more repeating this replication and connection process (red nodes), we end up with a 64-node scale-free hierarchical network. In Panel **(d)** we display a larger version of the random network, and it is evident that most nodes have approximately the same number of links. For the scale-free model, **(e)** the network is clearly inhomogeneous: while the majority of nodes has one or two links, a few nodes have a large number of links. We emphasize this by coloring the five nodes with the highest number of links red and their first neighbors green. While in the random network only 27% of the nodes are reached by the five most connected nodes, we reach more than 60% of the nodes in the scale-free network, demonstrating the key role played by the hubs. Note that the networks in **(d)** and **(e)** consist of the same number of nodes and links. Panel **(f)** demonstrates that the standard clustering algorithms are not that successful in uncovering the modular structure of a scale-free hierarchical network.

Table 1. The  $\gamma$  and  $\alpha$  exponents for five cellular networks, determined from a direct fit to the  $P(k)$  and  $C(k)$  functions.

	Transcription		Metabolic		Protein Interaction
	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>S. cerevisiae</i>
$\gamma$	$2.1 \pm 0.3$	$2.0 \pm 0.2$	$2.0 \pm 0.4$	$2.0 \pm 0.1$	$2.4 \pm 0.4$
$\alpha$	$1.0 \pm 0.2$	$1.0 \pm 0.2$	$0.8 \pm 0.3$	$0.7 \pm 0.3$	$1.3 \pm 0.5$

### 3. Network Models

As we have just seen, many biological networks are dominated by a scale-free distribution of nearest neighbors. Why is this power-law behavior so pervasive? To understand the cause of the scale-free degree distribution and the hierarchical network structure, we will in the following explain three models that serve as network paradigms. These models build on very different principles and, to varying degrees, are able to explain the observed network features.

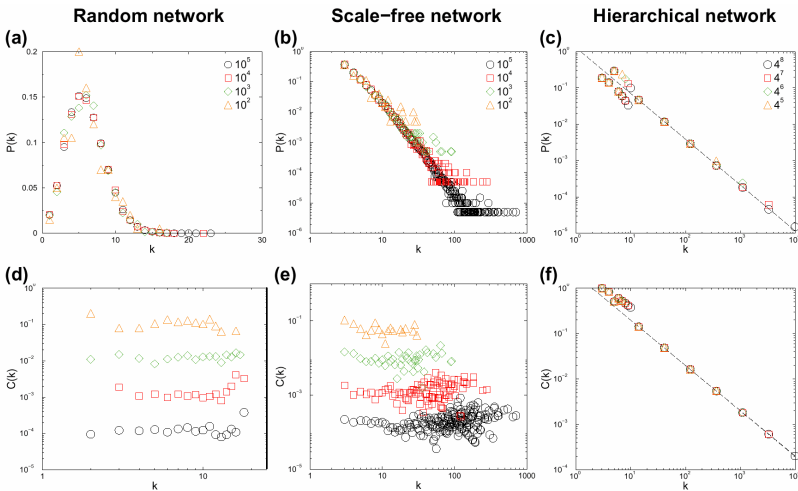


Figure 4. Properties of the three network models. **(a)** The ER model gives rise to a Poisson degree distribution  $P(k)$  (the probability that a randomly selected node has exactly  $k$  links) which is strongly peaked at the average degree  $\langle k \rangle$ . The degree distributions for the scale-free **(b)** and the hierarchical **(c)** network models do not have a peak, they instead decay according to  $P(k) \sim k^{-\gamma}$ . The average clustering coefficient for nodes with exactly  $k$  neighbors,  $C(k)$ , is independent of  $k$  for both the ER **(d)** and the scale-free **(e)** network model. **(f)** In contrast,  $C(k) \sim k^{-1}$  for the hierarchical network model.

### 3.1. Random Network Model

In discussing the origin of the observed power-law behavior, we need to first understand the properties of the simplest available network model. While graph theory initially focused on regular graphs, since the 1950's large networks with no apparent design principles were described as random graphs (37), proposed as the simplest and most straightforward realization of a complex network. According to this Erdos–Renyi (ER) model of random networks (38), we start with  $N$  nodes and connect every pair of nodes with probability  $p$ . This creates a graph with approximately  $pN(N-1)/2$  randomly distributed edges (Fig. 3a,d). The distribution of nearest neighbors follows a Poisson distribution (Fig. 4a), and consequently the average degree  $\langle k \rangle$  of the network describes the properties of a typical node. Furthermore, for this “democratic” network model, the clustering is independent of the node degree  $k$  (Fig. 4d). The ER model, although simple and appealing, does not capture the properties of neither the degree distribution nor the clustering coefficient observed in biological networks.

### 3.2. Scale-Free Network Model

In the network model of Barabási and Albert (BA), two key mechanisms, which both are absent from the classical random network model, are responsible for the emergence of a power-law degree distribution (19). First, networks grow through the addition of new nodes linking to nodes already present in the system. Second, there is a higher probability to link to a node with a large number of connections, a property called preferential attachment. These two principles are implemented as follows: starting from a small core graph consisting of  $m_0$  nodes, a new node with  $m$  links is added at each time step and connected to the already existing nodes (Fig. 3b,e). Each of the  $m$  new links are then preferentially attached to a node  $i$  (with  $k_i$  neighbors) which is chosen according to the probability

$$\Pi_i = k_i / \sum_j k_j . \quad (4)$$

The simultaneous combination of these two network growth rules gives rise to the observed power-law degree distribution (Fig. 4b). In Panel 3b, we illustrate the growth process of the scale-free model by displaying a network at time  $t$  (green links) and then at time  $(t+1)$ , when we have added a new node (red links) using the preferential attachment probability. Compared to random networks, the probability that a node is highly connected is statistically significant in scale-free networks. Consequently, many network properties are determined by a relatively small number of highly connected nodes, often called “hubs”. To make the effect of the hubs on the network structure visible, we have colored the five nodes with largest degrees red in Fig. 3d and 3e and their nearest neighbors green. While in the ER network only 27% of the nodes are reached by the five most connected ones, we reach more than 60% of the nodes in the scale-free network, demonstrating the key role played by the hubs. Another consequence of the hub’s dominance of the network topology is that scale-free networks are highly tolerant of random failures (perturbations) while being extremely sensitive to targeted attacks (39). Comparing the properties of the BA network model with those of the ER model, we note that the clustering of the BA network is larger, however  $C(k)$  is approximately constant (Fig. 4e), indicating the absence of a hierarchical structure.

### 3.3. Hierarchical Network Model

Many real networks are expected to be fundamentally modular, meaning that the network can be seamlessly partitioned into a collection of modules where each module performs an identifiable task, separable from the function(s) of other modules (40-43 ; see *Chapter 2*). Therefore, we must reconcile the scale-free property with potential modularity. In order to account for the modularity as reflected in the power-law behavior of  $C(k)$  and a simultaneous scale-free degree distribution, we have to assume that clusters combine in an iterative manner, generating a hierarchical network (26,29). Such a network emerges from a repeated duplication and integration process of clustered nodes (26), which in principle can be repeated indefinitely. This process is depicted in Panel 2c, where we start from a small cluster of four densely linked

nodes (blue). We next generate three replicas of this hypothetical initial module (green) and connect the three external nodes of the replicated clusters to the central node of the old cluster, thus obtaining a large 16-node module. Subsequently, we again generate three replicas of this 16-node module (red), and connect the 16 peripheral nodes to the central node of the old module, obtaining a new module of 64 nodes. This hierarchical network model seamlessly integrates a scale-free topology with an inherent modular structure by generating a network that has a power law degree distribution (Fig. 4c) with degree exponent  $\gamma = 1 + \ln 4 / \ln 3 \approx 2.26$  and a clustering coefficient  $C(k)$  which proves to be dependent on  $k^{-1}$  (Fig. 4f). However, note that modularity does not imply clear-cut sub-networks linked in well-defined ways (26,44). In fact, the boundaries of modules are often blurred considerably (see e.g. Fig. 3f).

### 3.4. Bose-Einstein Condensation and Networks

In most complex systems the nodes have differing abilities of attracting new links, which is independent of their number of nearest neighbors. For instance, some Web pages quickly acquire a large number of links through a mixture of good content and marketing, although they are just recently published on the World wide web. This competition for links can be incorporated into the scale-free model by adding a "fitness" parameter,  $\eta_i$ , to each node,  $i$ , describing its ability to compete for links at the expense of other nodes. For example, a Web page with good up-to-date content and a friendly interface would be expected to display a greater fitness than a low-quality page that is only updated occasionally. The probability  $\Pi_i$  that a new node connects to one with  $k_i$  links is then modified from Eq. (4) such that  $\Pi_i = \eta_i k_i / \sum_j \eta_j k_j$  (45).

The competition generated by the various fitness levels means that each node evolves differently in time compared with others. Indeed, the connectivity of each node is now given by  $k_i(t) \sim t^{\beta(\eta)}$ , where the exponent  $\beta(\eta)$  increases with  $\eta$ , and  $t$  is the time since the node was added to the network (45). Consequently, fit nodes (ones with large  $\eta$ ) can join the network at some later time and connect to many more links than less-fit nodes that have been around for longer.

Amazingly, such competitive-fitness models appear to have close ties with Bose-Einstein condensation, currently one of the most investigated problems in atomic physics. In a normal atomic gas, the atoms are distributed among many different energy levels. However in a Bose-Einstein condensate, all the atoms accumulate in the lowest energy state of the system and are described by the same quantum wave function. By replacing each node in the network with an energy level having energy  $\varepsilon_i = \exp(-\beta \eta_i)$ , the fitness model maps exactly onto a Bose gas (45). According to this mapping, the nodes map to energy levels while the links are represented by atoms in these levels. Additionally, the behavior of a Bose gas is uniquely determined by the distribution  $g(\varepsilon)$  from which the random energy levels (or fitnesses) are selected. One expects that the functional form of  $g(\varepsilon)$  depends on the system. For example, the attractiveness of a router to a network engineer comes from a rather different distribution than the fitness of a dot-com company competing for customers.

For a wide class of  $g(\varepsilon)$  distributions, a "fit-get-richer" phenomenon emerges (45). Although the fittest node acquires more links than its less-fit counterparts, there is no clear winner. On the other hand, certain  $g(\varepsilon)$  distributions can result in a Bose-Einstein condensation, where the fittest node does emerge as a clear winner. For these distributions, a condensate develops by acquiring a significant fraction of the links which is independent of the size of the system. In network language this corresponds to a "winner-takes-all" phenomenon. While the precise form of the fitness distribution for the Web or the Internet is not known yet, it is likely that  $g(\varepsilon)$  could be measured in the near future.

#### 4. Network Utilization

Despite their impressive successes, purely topologic approaches have important intrinsic limitations. For example, the activity of the various metabolic reactions or regulatory interactions differs widely, some being highly active under most growth conditions while others are switched on only for some rare environmental circumstances. Therefore, an ultimate description of cellular networks requires us to consider the intensity (i.e., strength), the direction (when applicable) and the temporal aspects of the

interactions. While we so far know little about the temporal aspects of the various cellular interactions, recent results have shed light on how the strength of the interactions is organized in metabolic and genetic-regulatory networks (46-48) and how the local network structure is correlated with these link strengths.

#### 4.1. Flux Utilization

In metabolic networks the flux of a given metabolic reaction, representing the amount of substrate being converted to a product within unit time, offers the best measure of interaction strength. Recent advances in metabolic flux-balance approaches (FBA, see also *Chapter 6*) (49-52) allow us to calculate the flux for each reaction, and they have significantly improved our ability to generate quantitative predictions on the relative importance of the various reactions, thus leading to experimentally testable hypotheses. The FBA approaches can be described as follows: Starting from a stoichiometric matrix model of an organism, e.g. one for *E. coli* contains 537 metabolites and 739 reactions (49-51), the steady state concentrations of all metabolites must satisfy

$$\frac{d}{dt}[A_i] = \sum_j S_{ij} v_j = 0 \quad (5)$$

where  $S_{ij}$  is the stoichiometric coefficient of metabolite  $A_i$  in reaction  $j$  and  $v_j$  is the flux of reaction  $j$ . We use the convention that if metabolite  $A_i$  is a substrate (product) in reaction  $j$ ,  $S_{ij} < 0$  ( $S_{ij} > 0$ ), and we constrain all fluxes to be positive by dividing each reversible reaction into two “forward” reactions with positive fluxes. Any vector of positive fluxes  $\{v_j\}$  which satisfies Eq. (5) corresponds to a state of the metabolic network, and hence, a potential state of operation of the cell.

Assuming that the cellular metabolism is in a steady state and optimized for the maximal growth rate (50,51), FBA allows us to calculate the flux for each reaction using linear optimization, providing a measure of each reaction’s relative activity (46). A striking feature of the resulting flux distribution from such modeling of both *H. pylori*, *E. coli*

and *S. cerevisiae* is its overall inhomogeneity: reactions with fluxes spanning several orders of magnitude coexist under the same conditions (Fig. 5a). This is captured by the flux distribution for *E. coli*, which follows a power law where the probability that a reaction has flux  $\nu$  is given by  $P(\nu) \sim (\nu + \nu_0)^{-\alpha}$ . This flux exponent is predicted to be  $\alpha = 1.5$  by FBA methods (46). In a recent experiment (53) the strength of the various fluxes of the *E. coli* central metabolism was measured, revealing (46) the power-law flux dependence  $P(\nu) \sim \nu^{-\alpha}$  with  $\alpha \cong 1$  (Fig. 5b). This power law behavior indicates that the vast majority of reactions have quite small fluxes, while coexisting with a few reactions with extremely large flux values.

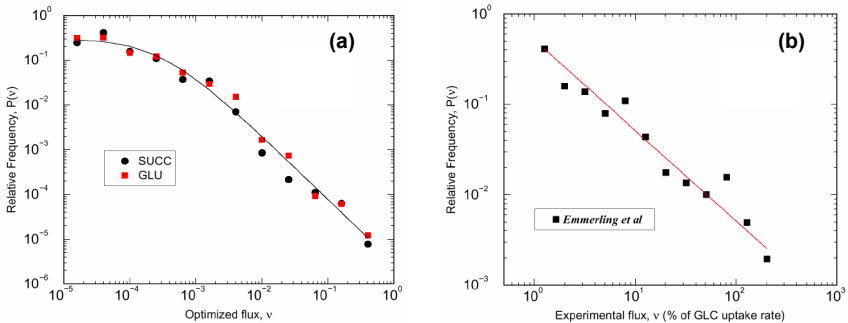


Figure 5. Flux distribution for the metabolism of *E. coli*. **(a)** Flux distribution for optimized biomass production on succinate (black) and glutamate (red) rich uptake substrates. The solid line corresponds to the power law fit  $P(\nu) \sim (\nu + \nu_0)^{-\alpha}$  with  $\nu_0 = 0.0003$  and  $\alpha = 1.5$ . **(b)** The distribution of experimentally determined fluxes (53) from the central metabolism of *E. coli* also displays power-law behavior with a best fit to  $P(\nu) \sim \nu^{-\alpha}$  with  $\alpha = 1$ .

The observed flux distribution is compatible with two quite different potential *local* flux structures (46). A homogeneous local organization would imply that all reactions producing (consuming) a given metabolite have comparable fluxes. On the other hand, a more delocalized “hot backbone” is expected if the local flux organization is heterogeneous, such that each metabolite has a dominant source (consuming) reaction. To distinguish between these two scenarios for each metabolite  $i$  produced (consumed) by  $k$  reactions, we define the measure (54,55)

$$Y(k, i) = \sum_{j=1}^k \left( \frac{\hat{v}_{ij}}{\sum_{l=1}^k \hat{v}_{il}} \right)^2, \quad (6)$$

where  $\hat{v}_{ij}$  is the mass carried by reaction  $j$  which produces (consumes) metabolite  $i$ . If all reactions producing (consuming) metabolite  $i$  have comparable  $\hat{v}_{ij}$  values,  $Y(k, i)$  scales as  $1/k$ . If, however, a single reaction's activity dominates Eq. (6), we expect  $Y(k, i) \sim 1$ , i.e.,  $Y(k, i)$  is independent of  $k$ . For the *E. coli* metabolism optimized for succinate and glucose uptake we find that both the *in* and *out* degrees follow the power law  $Y(k, i) \sim k^{-0.27}$ , representing an intermediate behavior between the two extreme cases (46). This suggests that the large-scale inhomogeneity observed in the overall flux distribution is increasingly valid at the level of the individual metabolites as well: for most metabolites, a single reaction carries the majority of the flux. Hence, the majority of the metabolic flux is carried along linear pathways – the metabolic high flux backbone (HFB) (46).

#### 4.2. Gene Interactions

One can also investigate the strength of the various genetic regulatory interactions provided by microarray datasets. Assigning each pair of genes a correlation coefficient which captures the degree to which they are co-expressed, one finds that the distribution of these pair-wise correlation coefficients follows a power law (47,48). That is, while the majority of gene pairs have only weak correlations, a few gene pairs display a significant correlation coefficient. These highly correlated pairs likely correspond to direct regulatory and protein interactions. This hypothesis is supported by the finding that the correlations are larger along the links of the protein interaction network and between proteins occurring in the same complex than for pairs of proteins that are not known to interact directly (56-59).

Taken together, these results indicate that the biochemical activity in both the metabolic and genetic networks is dominated by several 'hot links' that represent a few high activity interactions embedded into a web

of less active interactions. This attribute does not seem to be a unique feature of biological systems: hot links appear in a wide range of non-biological networks where the activity of the links follows a wide distribution (60,61). The origin of this seemingly universal property is, again, likely rooted in the network topology. Indeed, it seems that the metabolic fluxes and the weights of the links in some non-biological system (60,61) are uniquely determined by the scale-free nature of the network. A more general principle that could explain the correlation distribution data as well is currently lacking.

## 5. Conclusion

Power laws are abundant in nature, affecting both the construction and the utilization of real networks. The power-law degree distribution has become the trademark of scale-free networks and can be explained by invoking the principles of network growth and preferential attachment. However, many biological networks are inherently modular, a fact which at first seems to be at odds with the properties of scale-free networks. However, these two concepts can co-exist in hierarchical scale-free networks. In the utilization of complex networks, most links represent disparate connection strengths or transportation thresholds. For the metabolic network of *E. coli* we can implement a flux-balance approach and calculate the distribution of link weights (fluxes), which (reflecting the scale-free network topology) displays a robust power-law, independent of exocellular perturbations. Furthermore, this global inhomogeneity in the link strengths is also present at the local level, resulting in a connected “hot-spot” backbone of the metabolism. Similar features are also observed in the strength of various genetic regulatory interactions. Despite the significant advances witnessed the last few years, network biology is still in its infancy, with future advances most notably expected from the development of theoretical tools, development of new interactive databases and increased insights into the interplay between biological function and topology.

## References

1. Albert, R. and Barabási, A.L. (2002). Statistical mechanics of complex networks. *Rev Mod Phys.* 74, 47-97.
2. Strogatz, S.H. (2001). Exploring complex networks. *Nature.* 410, 268-76.
3. Dorogovtsev, S.N. and Mendes, J.F.F. (2003). *Evolution of networks : From biological nets to the Internet and WWW*, Oxford University Press, Oxford.
4. Bornholdt, S. and Schuster, H.G. (2003). *Handbook of graphs and networks: From the genome to the Internet*, Wiley-VCH, Berlin, Germany.
5. Newman, M.E.J., Barabási, A.L. and Watts, D. (Eds.) (2005). *The structure and growth of networks*, Princeton Univ Press, Princeton.
6. Ben-Naim, E., Frauenfelder, H. and Toroczkai, Z. (Eds.) (2004). *Complex networks*, Lect. Notes Phys., Springer Verlag, Berlin.
7. Pastor-Satorras, R. and Vespignani, A. (2004). *Evolution and structure of the Internet*, Cambridge Univ Press.
8. Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Comput Commu Rev.* 29, 251-62.
9. Albert, R., Jeong, H. and Barabási, A.L. (1999). Diameter of the World wide web. *Nature.* 401, 130-1.
10. Milgram, S. (1967). The small-world problem. *Psychology Today.* 2, 60-7.
11. Kochen, M. (1989). *The small-world*, Ablex, Norwood, N.J.
12. Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Application*, Cambridge University Press, Cambridge.
13. Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E. and Aberg, Y. (2001). The web of human sexual contacts. *Nature.* 411, 907-8.
14. Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proc Natl Acad Sci.* 98, 404-9.
15. Barabási, A.L., Jeong, H., Ravasz, R., Neda, Z., Vicsek, T. and Schubert, A. (2002). On the topology of the scientific collaboration networks. *Physica A.* 311, 590.
16. Goh, K.I., Kahng, B. and Kim, D. (2001). Universal behavior of load distribution in scale-free networks. *Phys Rev Lett.* 87, 278701.
17. Newman, M.E.J., (2002). Assortative mixing in networks. *Phys Rev Lett.* 89, 208701.
18. Pastor-Satorras, R., Vazquez, A. and Vespignani, A. (2001). Dynamical and correlation properties of the Internet. *Phys Rev Lett.* 87, 258701.
19. Barabási, A.L. and Albert, R. (1999). Emergence of scaling in random networks. *Science.* 286, 509-12.
20. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabási, A.L. (2000). The large-scale organization of metabolic networks. *Nature.* 407, 651-4.
21. Jeong, H., Mason, S.P., Barabási, A.L. and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. *Nature.* 411, 41-2.
22. Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., et al. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science.* 303, 540.

23. Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., Kuang, B., et al. A protein interaction map of *Drosophila melanogaster*. *Science*. 302, 1727.
24. Montoya, J.M. and Sole, R.V. (2002). Small-world patterns in food webs. *J Theor Biol*. 214, 405-12.
25. Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of small-world networks. *Nature*. 393, 440-2.
26. 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, 1551-5.
27. Ravasz, E. and Barabási, A.L. (2003). Hierarchical organization in complex networks. *Phys Rev E*. 67, 026112.
28. Dorogovtsev, S.N., Goltsev, A.V. and Mendes, J.F.F. (2002). Pseudofractal scale-free web. *Phys Rev E*. 65, 066122.
29. Vázquez, A., Pastor-Satorras, R. and Vespignani, A. (2002). Large-scale topological and dynamical properties of the Internet. *Phys Rev E*. 65, 066130.
30. Shen-Orr, S., Milo, R., Mangan, S. and Alon, U. (2002) *Nat Genet*. 31, 64-8.
31. Milo, R., Shen-Orr, S.S., Itzkovitz, S., Kashtan, N. and Alon, U. (2002). *Science*. 298, 824-27.
32. Mangan, S., Zaslaver, A. and Alon, U. (2003). *J Mol Biol*. 334, 197-204.
33. Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. and Alon, U. (2004). *Science*. 303, 1538-42.
34. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., et al. (2002). *Science*. 298, 799-804.
35. Hinman, V.F., Nguyen, A.T., Cameron, R.A. and Davidson, E.H. (2003). *Proc Natl Acad Sci. U.S.A.* 100, 13356-61.
36. Dobrin, R., Beg, Q.K., Barabási, A.L. and Oltvai, Z.N. (2004). *BMC Bioinformatics*. 5, 10.
37. Bollobas, B. (1985). *Random Graphs*. Academic Press, London.
38. Erdos, P. and Renyi, A. (1960). On the evolution of random graphs. *Publ Math Inst Hung Acad Sci*. 5, 17-61.
39. Albert, R., Jeong, H. and Barabási, A.L. (2000). Attack and error tolerance of complex networks. *Nature*. 406, 378-82.
40. Hartwell, L.H., Hopfield, J.J., Leibler, S. and Murray, A.W. (1999). From molecular to modular cell biology. *Nature*. 402, C47-52.
41. Rao, C.V. and Arkin, A.P. (2001). Control motifs for intracellular regulatory networks. *Annu Rev Biomed Eng*. 3, 391.
42. Hasty, J., McMillen, D., Isaacs, F. and Collins, J.J. (2001). Computational studies of gene regulatory networks: In numero molecular biology. *Nature Rev Genet*. 2, 268-79.
43. Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2001). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet*. 31, 64-8.
44. Holme, P., Huss, M. and Jeong, H. (2003). Subnetwork hierarchies of biochemical pathways. *Bioinformatics*. 19, 532-9.

45. Bianconi, G. and Barabási, A.L. (2001). Bose-Einstein condensation in complex networks. *Phys Rev Lett.* 86, 5632.
46. Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z.N. and Barabási, A.L. (2004). Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature.* 427, 839.
47. Kutznetsov, V.A., Knott, G.D. and Bonner, R.F. (2002). General statistics of stochastic processes of gene expression in eukaryotic cells. *Genetics.* 161, 1321-32.
48. Farkas, I.J., Jeong, H., Vicsek, T., Barabási, A.L. and Oltvai, Z.N. (2003). The topology of the transcription regulatory network in the yeast, *Saccharomyces cerevisiae*. *Physica A.* 318, 601-12.
49. Edwards, J.S. and Palsson, B.O. (2000). The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci.* 97, 5528-33.
50. Edwards, J.S., Ibarra, R.U. and Palsson, B.O. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol.* 19, 125-30.
51. 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, 186-9.
52. Segre, D., Vitkup, D. and Church, G.M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci.* 99, 15112-7.
53. Emmerling, M., Dauner, M., Ponti, A., Fiaux, J., Hochuli, M., Szyperski, T., Wuthrich, K., Bailey, J.E. and Sauer, U. (2002). Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *J Bacteriol.* 184, 152-64.
54. Barthelemy, M., Gondran, B. and Guichard, E. (2003). Spatial structure of the Internet traffic. *Physica A.* 319, 633-42.
55. Derrida, B. and Flyvbjerg, H. (1987). Statistical properties of randomly broken objects and of multivalley structures in disordered-systems. *J. Phys. A: Math Gen.* 20, 5273-88.
56. Dezsó, Z., Oltvai, Z.N. and Barabási, A.L. (2003). Bioinformatics analysis of experimentally determined protein complexes in the yeast, *Saccharomyces cerevisiae*. *Genome Res.* 13, 2450-4.
57. Grogoriev, A. (2001). A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 29, 3513-9.
58. Jansen, R., Greenbaum, D. and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Res.* 12, 37-46.
59. Ge, H., Liu, Z., Church, G.M. and Vidal, M. (2001). Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genet.* 29, 482-6.
60. Goh, K.-I., Kahng, B. and Kim, D. (2002). Fluctuation-driven dynamics of the internet topology. *Phys Rev Lett.* 88, 108701.
61. de Menezes, M.A. and Barabási, A.L. (2004). Fluctuations in network dynamics. *Phys Rev Lett.* 92, 028701.