Chapter 7
Analysis of *E. coli* Network

Hawoong Jeong

Contents

Abstract Diverse complex systems such as cells, Internet and society can be mapped into networks by simplifying each constituent as a node and their interaction as a link. Traditionally it has been considered that these networks are random, but recent series of studies show that they are far from being random and have common inhomogeneous topology through generic self-organizing process. In this chapter, we briefly introduce the network analysis methods which were re-developed in statistical physics community recently. First, we introduce basic complex network models such as Erdős-Rényi model, small-world model, scale-free model which were developed to describe complex systems. And then, we applied these methods to biological system, such as metabolic network and protein-protein interaction network of *E. coli*. We measure the global and local characteristics of the network structure. Finally we briefly review recent works on biological networks especially on dynamic aspect.

H. Jeong (✉)
Department of Physics, Institute for the BioCentury, KAIST, Daejeon, 305-701, Korea
e-mail: hjeong@kaist.edu

S.Y. Lee (ed.), *Systems Biology and Biotechnology of Escherichia coli*,
DOI 10.1007/978-1-4020-9394-4_7,

7.1 Introduction: Complex Bio-Networks

During the latter half of the 20th century, biology has been dominated by reductionist approaches that have provided a wealth of knowledge about individual cellular components and their functions. Typically, these approaches have entailed careful examination of a limited number of individual components in a biological system, hypothesis building based on the empirical observations, and further experiments to test these hypotheses. Reflecting the value of following this approach, biomedical researchers from a range of disciplines have deliberately restricted their analyses to well-defined systems with relatively few components, implicitly attempting to reduce biological phenomena to the behavior of individual molecules.

Despite the enormous success of the reductionist approach, a discrete biological function can only rarely be attributed to an individual molecule. Indeed, most biological functions arise from complex interactions among its various components (individual proteins, nucleic acids, small molecules, etc.). The need for more comprehensive approaches that address the full complexity of a biological system has now surfaced, largely with the emergence of genomics, in which the entire DNA sequences for a number of organisms now allows the definition of their gene portfolios. Extrapolation between genomes has accelerated the definition of what amounts to a "parts catalog" of cellular components in a large number of organisms. Also, large-scale efforts for studying the effects of systematic gene disruptions and for measuring expression levels of all genes under different conditions by microarray and proteomics approaches for entire genomes are well underway.

In turn, these advances have created an unprecedented opportunity towards developing a comprehensive understanding of biological systems, in part through the identification of the fundamental logic and derivative constraints that limit cell behavior. While the datasets available to us are far from being complete, they do offer a critical mass and coherency for such analyses, and for the subsequent capacity for model development and prediction through simulation of the ensuing model. Therefore, it has been studied to identify such underlying constraints and to model in quantitative terms the structure and functional (including regulatory) properties of the complex biological networks that maintain proper functioning various organisms. This analysis is aided by the coincidence of two recent scientific developments: the emergence of databases containing integrated data on the topology of various networks of biological significance, and the recent advances in understanding and quantifying the topology of complex (non-biological) networks which we are going to review in the next sections.

This chapter has been organized as follows. In Section 7.2, we introduce several basic network models which were developed to describe the ubiquitous complex networks found in real world. In Section 7.3, we analyzed metabolic network and protein-protein interaction network of *E. coli* in details. Section 7.4 includes recent advance in *network-biology* especially about the dynamic aspect of bio-network analysis. Most of this chapter was taken from recent papers written by the author (Eom et al. 2006, Jeong 2003, Kim et al. 2007).

7.2 Simple Models of Complex Networks

Modeling complex networks has a long history, and has been particularly active as a branch of combinatorial graph theory. However, the study of random networks in association with the real-world networks such as information systems, economic systems, and biological systems has begun recently. In this section, we briefly review a few important theoretical network models, and discuss recent empirical results on the network topology, which indicate the need for new approaches in understanding network development and describing their topology.

7.2.1 Erdős-Rényi Random Network Model

The most investigated random network model has been introduced by two Hungarian mathematicians, Erdős and Rényi (ER) (Bollobas 1985, Erdős and Rényi 1960) (see Fig. 7.1a), who were the first to study the statistical aspect of random

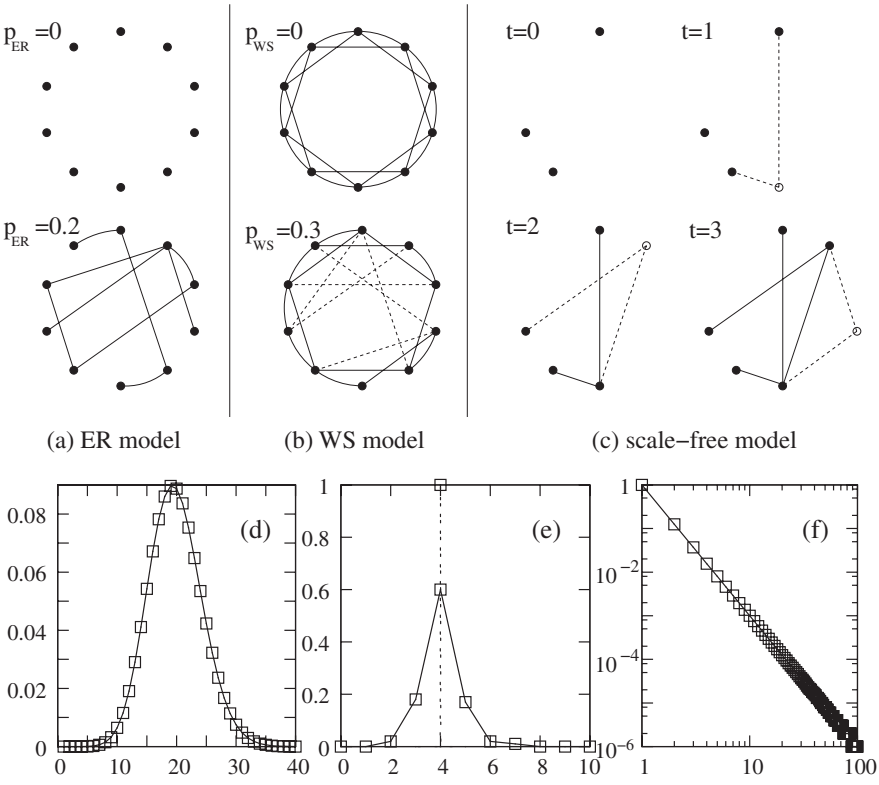

Fig. 7.1 Examples of model networks (**a**) Erdős-Rényi network, (**b**) Watts-Strogatz Small-world network, (**c**) Barabasi-Albert scale-free network. Typical degree distribution of (**d**) ER (**e**) SW (**f**) SF networks

graphs using the probabilistic method. The popularity of the ER model lies in its simplicity: It assumes that all vertices are equivalent, and any pair of vertices is connected with the same probability p_{ER}. ER discovered that many properties of random graphs, such as the emergence of trees or cycles, appear quite suddenly at a threshold value $p_{ER}(N)$. Within the physical literature, the ER model is known as infinite-dimensional percolation, belonging to the universality class of the mean field percolation (Stauffer and Aharony 1992). To compare the ER model with other network models, we need to focus on the connectivity distribution. As ER have shown, the probability that a vertex has k edges follows a Poisson distribution $P(k) = e^{-\lambda}k^{\lambda}/k!$, where the expectation value of degree $\langle k \rangle = \lambda$ is $(N-1)p_{ER}$, therefore ER network exhibits random and homogeneous structure (See Fig. 7.1d). However, it was found that degree distribution of most real world networks is far from being random which leads us to develop new network model.

7.2.2 Small-World Network Model

In 1998 Watts and Strogatz (WS) reported that many systems display both a high degree of local clustering reminiscent of finite-dimensional lattices (for example, a square lattice), and small-world phenomena characterizing random networks. Local clustering describes the tendency of groups of nodes to be all connected to each other, while small-world phenomena describes the property that any two nodes in the system can be connected by relatively short paths. To account for the transition from the local order to the small world behavior, they introduced the small-world network model (see Fig. 7.1b) (Watts and Strogatz 1998), which has been investigated rather intensely lately (Barthelemy and Amaral 1999a, Suki et al. 1998). In this model, starting from a regular lattice, each link between nodes is rewired with probability p_{WS}, such that long range link can be formed to ensure small-world characteristics. The connectivity distribution of the WS model depends on the parameter p_{WS}: for $p_{WS} = 0$, $P(k)$ is narrowly peaked at the average connectivity of the regular lattice, while for finite p_{WS}, $P(k)$ gets broader, converging to the Poisson connectivity distribution of the ER random graph (See Fig. 7.1e), which again turns out to be not appropriate to describe the inhomogeneous topology of the real world networks.

7.2.3 Barabasi-Albert Scale-Free Network Model

All existing network models we have considered so far fail to incorporate two generic aspects of real networks. First, they assume that networks have a fixed number of nodes. In contrast, most networks form and grow by the continuous addition of new nodes, that link to the nodes already present in the system. For example, the Internet expands by the attachment of new communication devices and routers to the system, and the World-Wide Web (WWW) grows by the addition of new web pages and domains. Second, the models assume that the probability that two

vertices are connected is random and uniform. In contrast, most real networks exhibit preferential connectivity. For example, new Internet domains are preferentially linked to major highly connected routers (nodes) to obtain broader bandwidth, or a newly created webpage will more likely link to well known, popular webpages with already high connectivity. Consequently, the probability with which a new node is connected to the existing nodes is not uniform, but there is a higher probability to be connected to a node that already has a large number of links (Fig. 7.1c). Barabasi et al. demonstrated that these two ingredients are sufficient to explain the inhomogeneous power-law distribution observed in real networks (Barabasi and Albert 1999). The network generated by this model evolves into a scale-invariant state, the probability that a node has k edges following $P(k) \sim k^{-3}$, i.e., a power-law with an exponent $\gamma = 3$ (See Fig. 7.1f). Furthermore, the Barabasi's group showed that excluding any of the two ingredients will eliminate the power-law connectivity (Albert and Barabasi 2000) and they developed a continuum theory (Barabasi et al. 1999) that allowed them to calculate the exponent γ, and predict the dynamics of the scale-free network. And they have also shown that the power-law distribution is robust against various local actions on the network structure, such as establishing links between existing nodes, or rerouting existing links from one node to another (Albert and Barabasi 2000). While these events can modify the scaling exponent γ, they do not eliminate the inhomogeneous nature of the network connectivity. The user's main goal is to maximize the benefits of the online environment, which can be best achieved by connecting to nodes where the best service is available, a flocking attitude that eventually leads to a few highly connected nodes and power laws. Consequently, complex communication networks inevitably evolve to develop scale-free network connectivity, and thus display topological inhomogeneities. (Albert et al. 1999b, Huberman and Adamic 1999)

7.3 Topology of Biological Networks

It is increasingly appreciated that the robustness of various cellular processes is rooted in the dynamic interactions among its many constituents (Barkai and Leibler 1997, Bhalla and Iyengar 1999, Yi et al. 2000), such as proteins, DNA, RNA, and small molecules. The existence of complex interactions among various components of a cell or simple microorganisms has long been appreciated, but in the absence of large-scale databases and a sufficiently developed theoretical framework, no meaningful analysis of these interactions was deemed possible. However, recent large-scale sequencing projects coupled with systematic two-hybrid analyses have provided complete sequence information for a number of genomes, and also allowed the development of protein interaction-(Rain et al. 2001a, Uetz et al. 2000) and integrated pathway-genome databases (Kanehisa and Goto 2000, Karp et al. 1999, Overbeek et al. 2000) that provide organism-specific connectivity maps of metabolic- and, to a lesser extent, various other cellular networks. Yet, due to the large number and the diversity of the constituents and reactions forming such networks, these maps are extremely complex, offering only limited insight into the

organizational principles of these systems. Our ability to address in quantitative terms the structure of these cellular networks, however, has benefited from recent advances in understanding the generic properties of complex networks (Albert et al. 2000, Watts and Strogatz 1998), which will be described in this section.

7.3.1 Network Analysis Methods

Until recently, complex networks have been modeled using the classical random network theory (Bollobas 1985, Erdős and Rényi 1960) which assumes that each pair of nodes (i.e., constituents) in the network is connected randomly with probability p. This process leads to a statistically homogeneous network, in which most nodes have approximately the same number of links, $\langle k \rangle$ (Fig. 7.1a). On the other hand, recent empirical studies on the structure of the World-Wide Web (Albert et al. 1999a), Internet (Faloutsos et al. 1999), and social networks (Barabasi and Albert 1999) have demonstrated that these systems are described by scale-free networks (Barabasi and Albert 1999) (Fig. 7.1c), for which degree distribution *P(k)* follows a power-law, i.e. $P(k) \sim k^{-\gamma}$. Unlike exponential networks, scale-free networks are extremely heterogeneous, their topology being dominated by a few highly connected nodes (hubs) which link the rest of the less connected nodes to the system (Fig. 7.1c). This degree distribution *P(k)* is a good measure for analyzing connectivity of the complex network and also has been applied to several biological networks as well.

Another basic measure for network analysis is a clustering coefficient. The clustering coefficient C_i of node i is the ratio of the total number y of the links connecting its nearest neighbors to the total number of all possible links between all these nearest neighbors, $C_i = {}^{2y}/_{k_i(k_i-1)}$ where k_i is the degree of node i. The clustering coefficient of a network, C, is the average of this value over all the nodes. Most real networks have much larger value of clustering coefficient than model networks such as ER or BA network due to, e.g., the community or modular structure (Dorogovtsev and Mendes 2002). Finally, the assortativity r, which measures the correlation between degrees of node linked to each other, is defined as the Pearson correlation coefficient of degrees between pairs of nodes (Newman 2002). Positive values of r stand for the positive degree-degree correlation which means that nodes with large degrees tend to be connected to one another. Most social networks have this positive degree correlation $r > 0$ (assortative mixing), like the co-authorship network of arxiv.org network (Newman 2001). On the other hand, most biological and technological networks show negative degree correlation $r < 0$ (disassortative mixing), including protein interaction network (PIN) and Internet AS network. If there is no degree correlation among nodes (neutral), as in the case of BA model, the value of r is in the vicinity of 0. There is another convenient way to check the degree correlation, which is measuring the quantity $\langle k_{nn} \rangle = \Sigma_{k'} k' p(k'/k)$, i.e. the average degree of nearest neighbors of nodes with degree k (Pastor-Satorras et al. 2001). Assortative mixing is represented by a positive slope of the $\langle k_{nn}(k) \rangle$ graph, while the others by horizontal (neutral) or a negative slope (disassortative).

While these quantities are measures for global properties of the network, local properties of the network have been analyzed via *motif* analysis. Subgraph patterns and network motifs have been applied recently to understand the local structure of complex networks (Milo et al. 2004, 2002, Vazquez et al. 2004). Subgraph patterns consist of more than three nodes and the links connecting only these nodes, which represent the minimum subnetworks of complex networks. Examples of triad subgraph patterns are shown in Fig. 7.4a. Network motifs are the subgraph patterns that occur in a complex network at numbers that are significantly higher than those in a random network (Milo et al. 2002). These are believed to represent the simplest building blocks of complex networks and the topologically characteristic interaction patterns within complex networks. Recently, it was also shown that certain motifs have been enhanced through the evolution of a network, which supports the functional importance of the motifs (Vazquez et al. 2004). For example, in transcription networks, a biochemical network responsible for regulating the expression of genes in cells, the network motifs are thought to be circuit elements that perform key information processing functions (Mangan and Alon 2003, Milo et al. 2002, Shen-Orr et al. 2002). The feed-forward loop, one motif of transcription networks, can act as a circuit that reduces noise and responds only to a persistent signal.

The following algorithm is used to obtain the network motifs (Milo et al. 2002). We scanned for all possible three-node subgraphs in the network and recorded the number of occurrences of each subgraph. To identify a statistically significant subgraph pattern, we compared the network to an ensemble of suitably randomized networks. Each node in the randomized networks contained the same number of incoming and outgoing links as the corresponding node in the original network. In addition, the randomized networks that were used to estimate the significance of n-node subgraphs were generated to preserve the same number of appearances of all $(n-1)$ node subgraphs as in the original network. For each subgraph i, the statistical significance of the subgraph is described by the Z score $Z_i = (N_i^{real} - \langle N_i^{rand} \rangle)/std(N_i^{rand})$. N_i^{rand} is the number of appearances of the subgraph i in the network, and $\langle N_i^{rand} \rangle$ and $std(N_i^{rand})$ are the average and standard deviation of its appearances in the ensemble of randomized networks, respectively. The subgraph pattern exhibiting a high Z score is the statistically significant pattern. In this analysis, the network motifs were selected when those subgraph patterns have a Z score greater than 2.

With this well-developed theoretical framework in hand and with the availability of detailed databases, we are now in position to initiate the analyses of complex bio-networks. Some of the first questions we asked included the following: What is the topological structure of metabolic and other cellular networks in global and local perspective? (See Fig. 7.2) What are the biologically and topologically relevant quantities that characterize them? Are there generic and common structural characteristics that apply to all cells, including both prokaryotes and eukaryotes? How are the specificity and the differential properties of various organisms reflected in the structure of these networks? In the following section we will summarize our results obtained on the large-scale structure of biochemical reaction pathways and protein interaction networks, especially for the case of *E. coli*, main topic of this book.

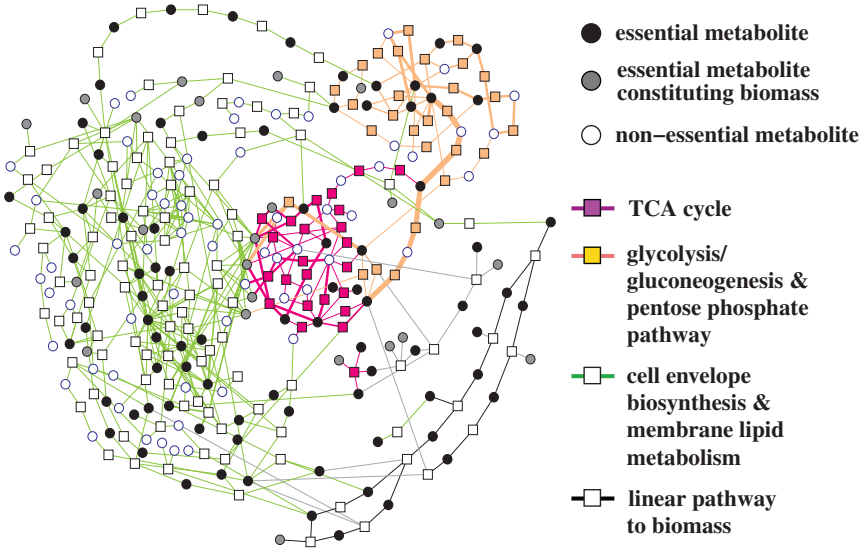

Fig. 7.2 Metabolic network including the central pathways and the membrane formation pathways. Circles denote essential and non-essential metabolites distinguished by the colors (*black*: essential metabolite; *gray*: essential metabolite constituting biomass; and *white*: non-essential metabolite). Cofactors are not drawn here because the number of the associated reactions is too large for visual examination. Each box represents the metabolic reaction for different functional classes specified by different colors and line styles

7.3.2 Metabolic Network of E. coli

To address the large-scale structural organization of metabolic networks, we have examined the topologic properties of the core metabolic network of 43 different organisms based on data deposited in the WIT (now ERGO) database (Jeong et al. 2000, Overbeek et al. 2000). In the metabolic network, nodes are substrates which are connected to each other through the actual metabolic reactions (Fig. 7.4B). As illustrated in Fig. 7.3a, results convincingly indicate that in *E. coli* the probability that a given substrate participates in k reactions follows a power-law distribution, i.e., the *E. coli* metabolic network belong to the class of scale-free networks. Furthermore, it is found that scale-free networks describe the metabolic networks in all organisms in all three domains of life, including 6 Archaea, 32 Bacteria, and 5 Eukaryotes, indicating the generic nature of this structural organization. Also, essentially identical results were obtained when we examined the topologic properties of the information transfer pathways of the 43 different organisms based on 'Information transfer' portions of data deposited in the WIT/ERGO database (Overbeek et al. 2000). Another general feature of many complex networks is their small-world character (Strogatz 2001, Watts and Strogatz 1998), i.e., any two nodes in the system can be connected by relatively short paths along existing links. In metabolic networks these paths correspond to the biochemical pathway connecting two substrates. The degree of interconnectivity of a metabolic network can be

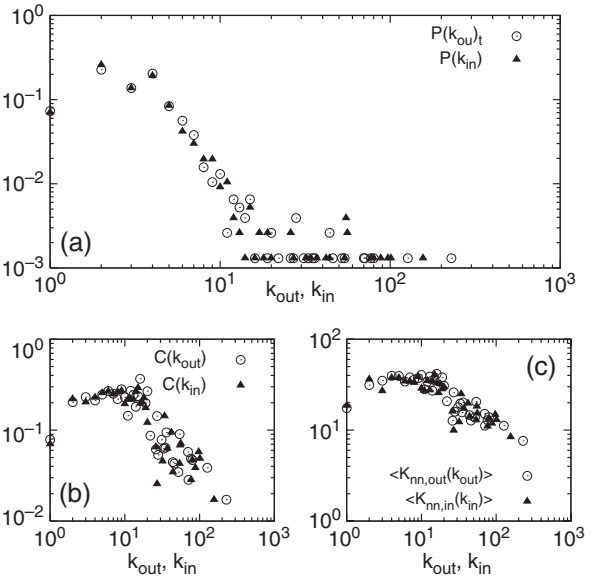

Fig. 7.3 Topological properties of *E. coli* metabolic network. **(a)** Degree distribution P(k), showing inhomogeneous structure for both in and out degrees, **(b)** Clustering coefficient C(k), showing typical decreasing behavior as a function of degree k like many other biological networks, **(c)** Assortativity, average degree of neighbor node $\langle K_{nn}(k)\rangle$, showing dissortative mixing again like many other biological networks

characterized by the network diameter, defined as the shortest biochemical pathway averaged over all pairs of substrates. For all non-biological networks they examined to date the average connectivity of a node is fixed, which implies that the diameter of the network increases logarithmically with the addition of new nodes (Barabasi and Albert 1999, Barthelemy and Amaral 1999b, Watts and Strogatz 1998). In contrast, we find that the diameter of the metabolic network is the same for all 43 organisms, irrespective of the number of substrates found in the given species (Jeong et al. 2000). This is surprising and unprecedented, and is possible only if with increasing organism complexity individual substrates are increasingly connected in order to maintain a relatively constant metabolic network diameter. Within the special characteristics of living systems this attribute may increase an organism's fitness to efficiently respond to external changes or internal errors. For example, the transition time between two metabolic steady states is apparently largely governed by time constants involved in changing the enzyme concentrations (Cascante et al. 1995), an attribute which could be best achieved when only a few alternative biochemical reactions need to be activated. In Fig. 7.3b, clustering coefficient of *E. coli* metabolic network shows $C(k) \sim k^{-\alpha}$ which represents the hierarchical and modular structures embedded in the biological networks (Ravasz et al. 2002). L ike other biological network, metabolic network of *E. coli* shows dissortative mixing (Fig. 7.3c), such that substrates with larger degrees (hubs) tend to interact with substrates with smaller degrees.

We also examined the triad subgraph patterns of metabolic networks of 43 organisms and identified their network motifs including *E. coli*. In this analysis, the direction of each link implies direction from an input substrate (educt) to an output substrate (product) (Fig. 7.4b) (Eom et al. 2006). We found that all metabolic networks have their own network motifs. To provide a more quantitative analysis,

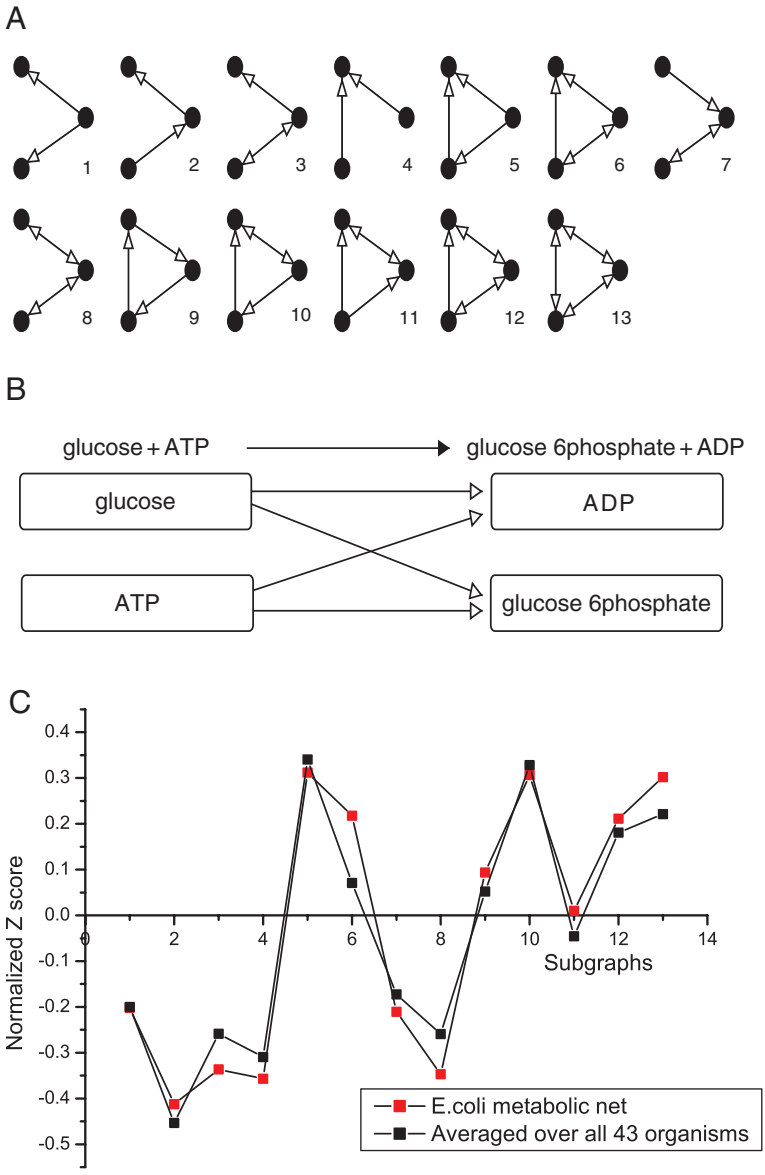

Fig. 7.4 Local properties of *E. coli* metabolic network. (**A**) Motif profile, all possible 13 types of three node connected subgraphs. (**B**) Graphical reorientation of a chemical reaction. (**C**) The triad significance profiles (TSPs) of metabolic networks. TSPs for *E. coli* and other organisms found in WIT database were plotted

we investigated the local structure of metabolic networks of each organism in detail and identified the significance profile (SP) of each metabolic network (Milo et al. 2004). The SP is the vector of Z scores normalized to a unit length, of which the i-th component is given by $SP_i = Z_i/(\Sigma_j Z_j^2)^{1/2}$. The SP of a given network represents the relative significance of the subgraphs in that network. It is important to compare networks of different sizes because network motifs in large networks tend to have higher Z scores than network motifs in small networks (Milo et al. 2004). The triad significance profile (TSP) for each metabolic network is presented in Fig. 7.4c. The TSPs of these networks are found to be almost insensitive to a removal of 20% of edges or to an addition of 20% new edges randomly, representing that the results are robust to possible missing or false-positive data errors. All metabolic networks showed similar TSPs and three network motifs of triads 5, 10, and 13 were found frequently. These motifs, especially 5 and 10, are well-known feed-forward loop and its variation of function is a prevalence of short detours in metabolic network (Gleiss et al. 2001, Heinrich and Schuster 1996). In contrast, triads 2, 4, and 8 were anti-motifs that were significantly underrepresented. The correlation coefficient between the TSPs of metabolic networks in 43 organisms was about 0.78 showing that metabolic networks have the same topological structure in both large-scale organization (inhomogeneous power-law degree distribution) and in local organization (sharing common topological substructures).

7.3.3 Protein Interaction Network of E. coli

Next example of biological network is protein interaction network (PIN). Proteins are traditionally defined by their individual actions as catalysts, signaling molecules, or building blocks of cells and microorganisms. However, recent integrative approaches view their role as an element in a network of protein–protein interactions with a 'contextual' or 'cellular' function within functional modules (Eisenberg et al. 2000, Hartwell et al. 1999). To uncover this role, it is important to assess the position of a protein within the protein–protein interaction network. We first have assessed the topologic characteristics of system-wide protein–protein interaction network found in the yeast, *S. cerevisiae*, and the bacterium, *H. pylori*, obtained mostly by systematic two-hybrid analyses (Ito et al. 2001, Rain et al. 2001b, Xenarios et al. 2000). Due to its size, a complete map of the yeast and *H. pylori* networks, while informative, in themselves offers little insight into their large-scale characteristics (See Fig. 7.5). Like other bio-networks, the probability that a given yeast protein interacts with k other yeast proteins follows a power-law (Jeong et al. 2001) with an exponential cutoff (Barthelemy and Amaral 1999a). This exponential cutoff is due to the physical limitation of the binding sites in the protein structure. A similar result was obtained for *H. pylori* as well. This indicates that the network of protein interactions in both a bacterium and an eukaryotic cell forms a highly inhomogeneous scale-free network. An important known consequence of the inhomogeneous structure is the network's simultaneous tolerance against random errors coupled with fragility

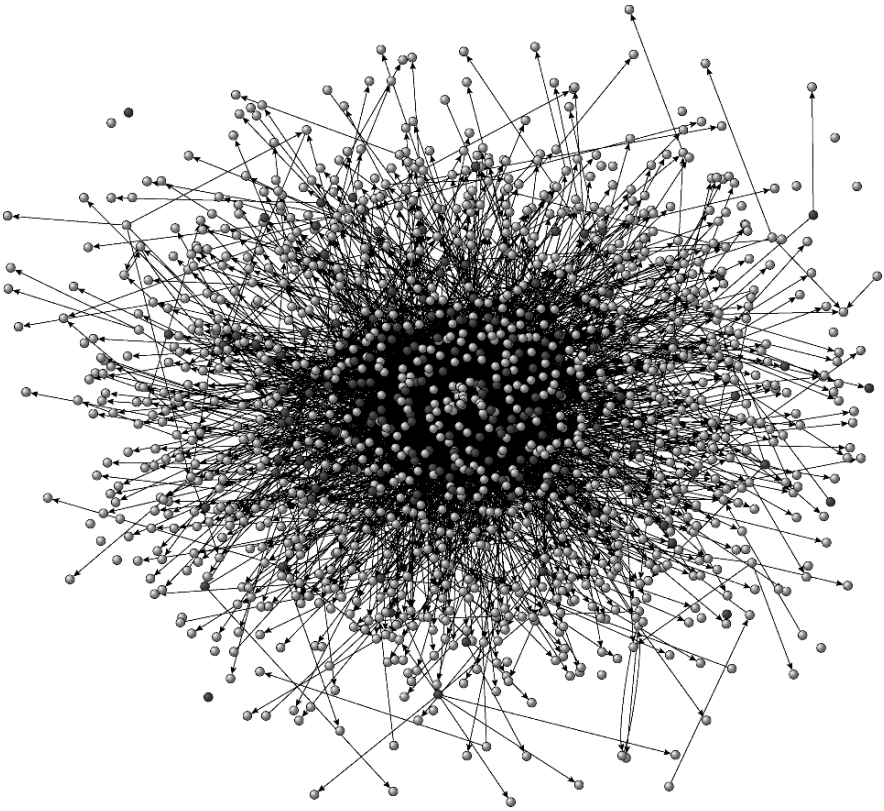

Fig. 7.5 Protein-Protein interaction network, essential (*gray*) and non-essential (*white*) proteins were connected through physical bindings

against the removal of the most connected nodes (Barabasi and Albert 1999). Yet, if there is indeed a biologically relevant functional link between topology and error tolerance, on average less connected proteins should prove less essential than highly connected ones. We calculated this correlation and showed that the likelihood that removal of a protein will prove to be lethal clearly correlates with the number of interactions the protein has. For example, while proteins with five or less links constitute 93% of the total number of proteins they find that only 21% of them are essential. In contrast, only 0.7% of the yeast proteins with known phenotypic profile have more than 15 links but single deletion of 62% of these proves lethal. This implies that highly connected proteins with a central role in the network's architecture are three times more likely to prove essential than proteins with low number of links to other proteins (Jeong et al. 2001).

We also analyzed PPI network of *E. coli* using protein complex data by G. Butland et al. (Butland et al. 2005). We found that again degree distribution of *E. coli* PPI network shows inhomogeneous scale-free degree distribution (Fig. 7.6a) and proteins with larger degrees are more essential than proteins with

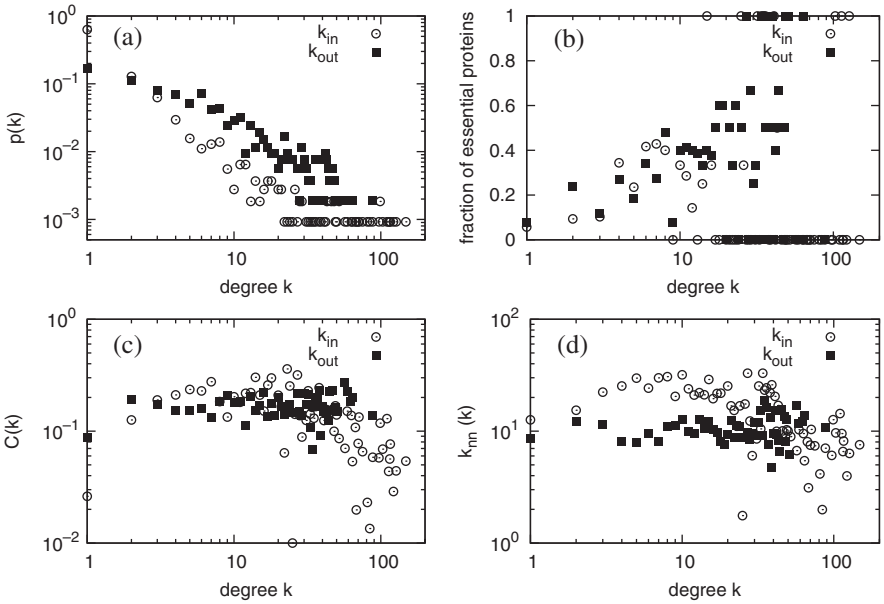

Fig. 7.6 Topological properties of protein-protein interaction network. (**a**) Degree distribution, (**b**) degree vs essentiality showing that as degree increases, lethality of the protein also increases. (**c**) Clustering coefficient C(k), (**d**) Assortativity, average degree of neighbor nodes

smaller degrees (Fig. 7.6b). Interestingly in Fig. 7.6c, it is observed that clustering coefficient *C(k)* shows relatively neutral behavior implying *E. coli* protein interaction network doesn't have hierarchical characteristics. Also the assortativity of *E. coli* protein interaction network seems to be neutral for outgoing link while it is dissortative for incoming link like many other biological networks (see Fig. 7.6d). Since PPI network by Han et al. is directed, we applied motif analysis algorithm to find relevant subgraph pattern hidden in *E. coli* protein interaction network. As seen in Fig. 7.7, quite different from the metabolic network, motif 11 is found more frequently and motif 10 is suppressed. However, motifs 5, 6, 12, 13 are shared with the metabolic network of *E. coli*.

7.4 Beyond Static Graph Analysis

So far, we have only considered spatial (geometrical) inhomogeneity of the complex networks, however it is also important to deal with temporal heterogeneity of the complex network. Links between nodes in the network can vary over time, for example, not every reaction in the metabolic network is active all the time. And the activity of each link in the metabolic network or regulatory network can be different in time such that some of them are highly active under most conditions while others are activated for certain specific conditions. Therefore, to fully understand the

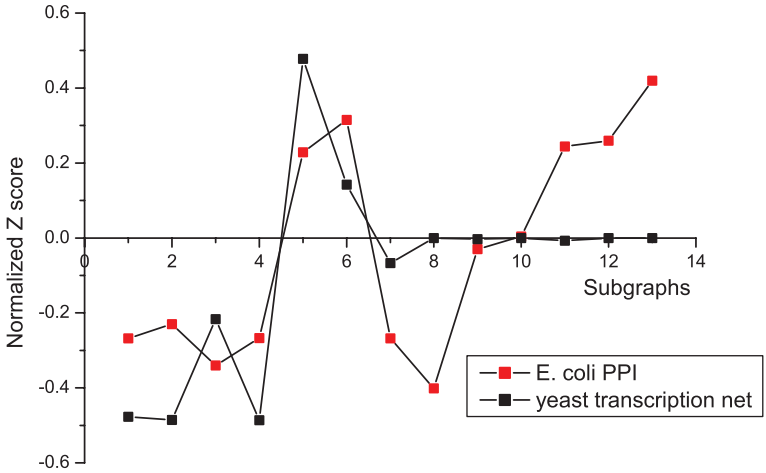

Fig. 7.7 Local properties of protein-protein interaction network of *E. coli*. TSPs for *E. coli* and yeast transcription network were plotted together

biological networks we have to consider the weight and direction and the temporal change of the network components. In this respect, we will introduce recent studies on dynamic aspect of metabolic networks using flux balance analysis (FBA) and protein interaction networks of *E. coli* in this section.

7.4.1 Understanding the Robustness of Metabolic Network

As complex biological systems are very robust to genetic and/or environmental changes on all levels of organization, their inherent robustness has been of great interest in biology as well as in engineering theory (Wagner 2005). The biological function of *E. coli* metabolism can be sustained against single-gene or even multiple-gene mutation possibly by utilizing the redundant pathways (Papp et al. 2004, Reed and Palsson 2004). While the investigations on the topological and functional/phenotypic properties of metabolic networks have been increasingly populated as shown in previous sections, (Almaas et al. 2004, Covert et al. 2004, Guimera and Nunes Amaral 2005, Papp et al. 2004) they still provide a limited understanding of the metabolic robustness despite its biological significance. In this section, we focus on the interplay between such robustness and the underlying metabolism, and how the robustness can be accomplished at the level of the metabolites which are the fundamental entities (Raymond and Segre 2006, Schmidt et al. 2003) integrated/dissipated by the metabolic processes. To this end, we constructed the computational models at a system level, and simulated them with a constraints-based flux analysis (Price et al. 2004).

To explore the robustness of *E. coli* metabolism from the metabolite perspective, we should identify the metabolites which are substantial in cellular functions. In this regard, all intracellular metabolites are classified into two categories, essential

and non-essential metabolites according to the phenotypic effects on cell survival when the consumption rate of the given metabolite is suppressed to zero. The resultant list of essential metabolites is identified under many different environments which are specified by combinations of several C, P, N, and S sources, and aerobic/anaerobic conditions (Kim et al. 2007). By disrupting multiple genes around essential/non-essential metabolites *in vivo*, we could validate the predicted effects of the metabolite essentiality on cell survival. For example, the associated genes of an essential metabolite, tetrahydrofolate, were selected for the multiple-gene disruption. Each single and double gene deletion mutant (ΔpurN, ΔlpdA, ΔglyA, and ΔpurN ΔlpdA) could still survive albeit with some growth rate changes, but simultaneous deletions of the triple genes (ΔpurN ΔlpdA ΔglyA) did not allow the cell to grow at all, reflecting that the combinatory suppression of the tetrahydrofolate pool is indeed fatal to the cell. On the contrary, 1-deoxy-D-xylulose 5-phosphate had been identified as a non-essential metabolite *in silico*, and experimental removals of all the reactions producing the metabolite by constructing Δdxs ΔxylB caused the only slight change and even increase of growth rate compared with wild type. Throughout these experiments, the measured growth rates of the gene deletion mutants relative to that of the wild type were found to be consistent with the *in silico* predictions. These results indicate that deletion strains for essential metabolites can suffer from the deleterious impact on cellular functions, while those for non-essential metabolites show the negligible influence on the actual growth. We also investigated the inherent network property of essential metabolites to elucidate the correlation between the structural property and functional behavior from the metabolite perspective. We found that essential metabolites are likely to be connected with more reactions than non-essential ones. Furthermore, the metabolic networks of 227 organisms with fully sequenced genomes disclose that the metabolites essential for various growth conditions are commonly distributed across the organisms, showing the high degree of phylogenetic conservation.

To better understand the robustness of the cellular metabolism from the metabolite perspective, it is necessary to quantify the usage of all relevant fluxes to a single metabolite. In this sense, we introduce the flux-sum (Φ) of the metabolite, which is defined as the summation of all incoming or outgoing fluxes for given metabolite i as follows:

$$\Phi_i = \sum_{j \in P_i} S_{ij} v_j = -\sum_{j \in C_i} S_{ij} v_j = \frac{1}{2} \sum_j \left| S_{ij} v_j \right|$$

where S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, and j is the flux of reaction j. P_i denotes the set of reactions producing metabolite i, C_i the set of reactions consuming metabolite i. Under the stationary assumption, Φ_i is the mass flow contributed by all fluxes producing (consuming) metabolite i. Based on this measure pertaining to the behavioral characteristic of metabolites, we can analyze the robustness of *E. coli* metabolism to maintain the cellular functions against the genetic mutations. The sensitivity to genetic perturbation for a given metabolite can be quantified by evaluating the relative fluctuation of Φ_i in response to each deletion

of non-lethal reactions: $\sqrt{\langle \Phi_i^2 \rangle - \langle \Phi_i \rangle^2} / \langle \Phi_i \rangle$ where $< \ldots >$ denotes the average over the reaction deletions. It turns out that the essential metabolites are more likely to have small relative fluctuations. This indicates that flux-sums of essential metabolites are relatively insensitive to genetic perturbation compared with those of non-essential ones. Indeed, 94.3% of total metabolites found in the fluctuation range of less than 0.0875 are essential, while there are only non-essential metabolites in the twenty highest ranks in relative fluctuations. Thus, essential metabolites are resistant to the internal variation compared with non-essential ones by maintaining the basal mass flow of the corresponding metabolite, thereby leading to the robustness of the cellular metabolism.

To clarify such resistance of essential metabolites against the internal disturbance, the severe perturbation was conducted by deleting the most contributing reaction to the flux-sum for a given essential metabolite. Remarkably, for many essential metabolites, the resultant flux loss is mostly recovered by the fluxes of other remaining reactions, thereby leading to very small change of the flux-sum, in spite of removing the dominant reaction5 with the largest flux value. For instance, the flux-sum of an essential metabolite, carbamoyl phosphate, is reproducible by other fluxes even when the largest flux from carbamate kinase is eliminated; other reaction, carbamoyl-phosphate synthase can compensate such flux loss fully, thus resulting in the recovery of 98.9% of the basal flux-sum. For many essential metabolites, the flux-sum is only changed much less than the reduced flux corresponding to the deleted reaction. Accordingly, even though the reaction with relatively high flux is eliminated, the flux-sum can be compensated by other fluxes around the essential metabolite, recovering such flux loss. Moreover, using the stoichio-similarity, we could develop the method to predict the most probable reaction which would recover the flux-sum after disruption. Hence, we believe that cellular robustness can be elucidated by such functional property of metabolic network manifesting the resilience of essential metabolites against the disturbed flux configuration.

Essential metabolites play a pivotal role in the cell survival, steadily maintaining the mass flow to produce or consume the metabolites against any internal disturbance within the cell. In other sense, this metabolite perspective on the robustness of *E. coli* provides us the cellular-level fragility: the failure of maintaining the flux-sum of a single essential metabolite can suppress the whole cellular growth drastically. Especially, for most essential metabolites (85%), reducing the flux-sum by half below the basal level intentionally leads to the growth rate down to half or even less, while only 28.9% of active non-essential metabolites have the same effect on the cell growth for such flux-sum perturbation.

The functional robustness of metabolic networks reflects the resistance towards internal defects and environmental fluctuations as an end product of a long evolutionary process. Such fault-tolerance or robustness may be a key to cell survival against environmental or genetic change. In this regard, a metabolite-based perspective could provide us a new guideline to interpret the cellular robustness. Essential metabolites substantial to the cell survival are capable of rerouting metabolic fluxes while sustaining their usage level. This capability of the essential metabolites leads

to the quite dramatic tolerance to a wide range of internal disturbances. From a therapeutic point of view, disrupting (knock-out) the multiple non-lethal genes around an essential metabolite can lead to fatal cell damage; even attenuating (knock-down) the relevant genes may cause the same effect. Thus, synthetic lethal mutations (Tucker and Fields 2003, Wong et al. 2004) can be systematically identified in conjunction with experimental screening techniques available (Ooi et al. 2003, Tong et al. 2001), thereby facilitating the discovery of drug targets for the genetic therapy.

7.4.2 Beyond the Static Graph Analysis: Spatio-Temporal Dynamics

The topological data and approach discussed in previous sections represent a partial snapshot of the metabolism. Indeed, the topology of the metabolic network provides only the genome-encoded potential metabolic activity of an organism. The actual function of its metabolic network, however, is realized through the genetic regulatory network that functionally activates and inactivates various enzymes or groups of enzymes that catalyze biochemical reactions embedded in the metabolic network topology. Thus, for an in-depth characterization of metabolism we need to develop a better understanding of the regulatory network and its dynamics, as well. An important limitation of any modeling effort is the lack of availability of enzyme kinetic data, making impossible the full dynamic characterization of these pathways. However, already available microarray data does give us important qualitative information on the correlation between the enzymatic activities of different pathways. In this sense, there are several studies to analyze the available microarray data to infer information about correlations between the various components of the *E. coli* metabolism. These studies will offer valuable information on the dynamical features of its metabolism that has never been included in previous modeling efforts. One of simple but interesting works on temporal aspect of complex network was found in protein interaction network. For the case of protein interaction network, it was verified that considering dynamic aspect is crucial to understand the lethality of the node properly. That is, although it was shown that highly connected proteins (hubs) are more essential (lethal) than less connected proteins, recent study shows that all hub proteins are not equivalent. Han et al. showed (Han et al. 2004) that there are two different categories for the hub proteins, first one is ‘party’ hubs which interact with their partner proteins simultaneously, the other is ‘date’ hubs which in contrast, interact with different proteins at different locations and times using a filtered yeast interactome (FYI), compiled from different sets of yeast mRNA expression data to find the difference. (See Fig. 7.8) They found that date hub is more important than party hub such that when party hubs are removed from the system, general connectivity of the network remains still unaffected while the removal of date hubs breaks network into pieces so that proteins cannot interact with each other. Therefore, it is very important to consider spatial and temporal information when we analyze the

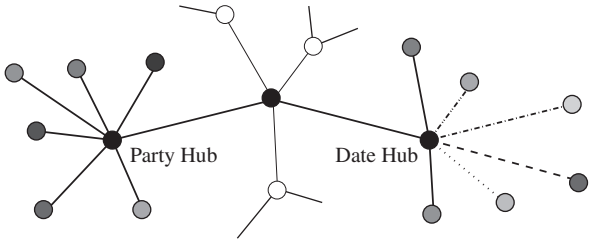

Fig. 7.8 Two different types of the hub. Party hub interacts with many proteins at the same time and location while date hub interacts with many proteins at different time and locations

bio-network. In this sense, spatio-temporal dynamic analysis should be applied to biological system along with static graph analysis.

Despite of the significant advance in network science during last decade, we are still far from understanding the biological system even for simple organism like *E. coli*. However, network biology which is still in its infancy, will give us an insight to find a way to understand the biological system along with large scale data sets generated and integrated into the database extensively.

Acknowledgments I'd like to thank many colleagues for their assistance in preparing this chapter, including Dr. P.-J. Kim, Y.-H. Eom, S. H. Lee for their help on data analyses and new illustrations, and Dr. R. Albert, Dr. B. Tomber, Dr. S. P. Mason, Dr. A.-L. Barabasi, Dr. Z. N. Oltvai, Dr. S. Lee, Dr. D.-Y. Lee, T. Y. Kim, K. H. Lee, Dr. S. Y. Lee, Dr. S. Park for their kind supports.

References

Albert R, Barabasi AL (2000) Topology of evolving networks: local events and universality. Phys Rev Lett 85(24):5234–7

Albert R, Jeong H, Barabasi A-L (1999a) The diameter of the World Wide Web. Nature 401:130–1

Albert R, Jeong H, Barabasi A-L (1999b) Emergence of Scaling in Random Networks Nature 400:130

Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406(6794):378–82

Almaas E, Kovacs B, Vicsek T et al. (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. Nature 427(6977):839–43

Barabasi A-L, Albert R (1999) Emergence of Scaling in Random Networks Science 286:509

Barabasi A-L, Albert R, Jeong H (1999) Emergence of Scaling in Random Networks Physica A 272:173

Barkai N, Leibler S (1997) Robustness in simple biochemical networks. Nature 387(6636):913–7

Barthelemy M, Amaral LAN (1999a) Small-World Networks: Evidence for a Crossover Picture. Phys Rev Lett 82:3180–2

Barthelemy M, Amaral LAN (1999b) Small-World Networks: Evidence for a Crossover Picture. Phys Rev Lett 82:3180–3183

Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. Science 283(5400):381–7

Bollobas B (1985) Random Graphs. Academic Press, London

Butland G, Peregrin-Alvarez JM, Li J et al. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433(7025):531–7

Cascante M, Melendez-Hevia E, Kholodenko B et al. (1995) Control analysis of transit time for free and enzyme-bound metabolites: physiological and evolutionary significance of metabolic response times. Biochem J 308(Pt 3):895–9

Covert MW, Knight EM, Reed JL et al. (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429(6987):92–6
Dorogovtsev SN, Mendes JFF (2002) Evolution of networks. Adv Phys 51:1079–1187
Eisenberg D, Marcotte EM, Xenarios I et al. (2000) Protein function in the post-genomic era. Nature 405(6788):823–6
Eom YH, Lee S, Jeong H (2006) Exploring local structural organization of metabolic networks using subgraph patterns. J Theor Biol 241(4):823–9
Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5A:17–61
Faloutsos M, Faloutsos P, Faloutsos C (1999) On power-law relationships of the internet topology. ACM SIGCOMM'99, Boston, MA
Gleiss PM, Stadler PF, Wagner A et al. (2001) Relevant Cycles in Chemical ReactionNetworks. Adv Complex Syst 1:1–000
Guimera R, Nunes Amaral LA (2005) Functional cartography of complex metabolic networks. Nature 433(7028):895–900
Han JD, Bertin N, Hao T et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430(6995):88–93
Hartwell LH, Hopfield JJ, Leibler S et al. (1999) From molecular to modular cell biology. Nature 402(6761 Suppl):C47–52
Heinrich R, Schuster S (1996) The Regulation of Cellular Systems. Chapman & Hall, New York
Huberman BA, Adamic LA (1999) Growth dynamics of the World-Wide Web. Nature 400:131
Ito T, Chiba T, Ozawa R et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 98(8):4569–74
Jeong H (2003) Complex scale-free networks Physica A 321:226–37
Jeong H, Mason SP, Barabasi AL et al. (2001) Lethality and centrality in protein networks. Nature 411(6833):41–2
Jeong H, Tombor B, Albert R et al. (2000) The large-scale organization of metabolic networks. Nature 407(6804):651–4
Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30
Karp PD, Krummenacker M, Paley S et al. (1999) Integrated pathway-genome databases and their role in drug discovery. Trends Biotechnol 17(7):275–81
Kim PJ, Lee DY, Kim TY et al. (2007) Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism. Proc Natl Acad Sci USA 104(34):13638–42
Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci USA 100(21):11980–5
Milo R, Itzkovitz S, Kashtan N et al. (2004) Superfamilies of evolved and designed networks. Science 303(5663):1538–42
Milo R, Shen-Orr S, Itzkovitz S et al. (2002) Network motifs: simple building blocks of complex networks. Science 298(5594):824–7
Newman MEJ (2001) Scientific collaboration networks. Phys Rev E64:016131
Newman MEJ (2002) Assortative Mixing in Networks. Phys Rev Lett 89:208701
Ooi SL, Shoemaker DD, Boeke JD (2003) DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nat Genet 35(3):277–86
Overbeek R, Larsen N, Pusch GD et al. (2000) WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Res 28(1):123–5
Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429(6992):661–4
Pastor-Satorras R, Vázquez A, Vespignani A (2001) Dynamical and Correlation Properties of the Internet. Phys Rev Lett 87:258701
Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2(11):886–97
Rain J-C, Selig L, De Reuse H et al. (2001a) The protein-protein interaction map of *Helicobacter pylori*. Nature 409:211

Rain JC, Selig L, De Reuse H et al. (2001b) The protein-protein interaction map of *Helicobacter pylori*. Nature 409(6817):211–5

Ravasz E, Somera AL, Mongru DA et al. (2002) Hierarchical organization of modularity in metabolic networks. Science 297(5586):1551–5

Raymond J, Segre D (2006) The effect of oxygen on biochemical networks and the evolution of complex life. Science 311(5768):1764–7

Reed JL, Palsson BO (2004) Genome-scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res 14(9):1797–805

Schmidt S, Sunyaev S, Bork P et al. (2003) Metabolites: a helping hand for pathway evolution? Trends Biochem Sci 28(6):336–41

Shen-Orr SS, Milo R, Mangan S et al. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 31(1):64–8

Stauffer D, Aharony A (1992) Percolation Theory. Taylor & Francis, London

Strogatz SH (2001) Exploring complex networks. Nature 410(6825):268–76

Suki B, Alencar AM, Sujeer MK et al. (1998) Life-support system benefits from noise. Nature 393(6681):127–8

Tong AH, Evangelista M, Parsons AB et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294(5550):2364–8

Tucker CL, Fields S (2003) Lethal combinations. Nat Genet 35(3):204–5

Uetz P, Giot L, Cagney G et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403(6770):623–7

Vazquez A, Dobrin R, Sergi D et al. (2004) The topological relationship between the large-scale attributes and local interaction patterns of complex networks. Proc Natl Acad Sci 101:17940

Wagner A (2005) Robustness and Evolvability in Living Systems. Princeton University Press, Princeton

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–2

Wong SL, Zhang LV, Tong AH et al. (2004) Combining biological networks to predict genetic interactions. Proc Natl Acad Sci USA 101(44):15682–7

Xenarios I, Rice DW, Salwinski L et al. (2000) DIP: the database of interacting proteins. Nucleic Acids Res 28(1):289–91

Yi TM, Huang Y, Simon MI et al. (2000) Robust Perfect Adaptation in Bacterial Chemotaxis through Integral Feedback Control. Proc Natl Acad Sci USA 97:4649–53