DRUG DISCOVERY TODAY BIOSILICO

Global properties of biological networks

Martin G. Grigorov

This article discusses the most recent achievements in understanding the biological implications of the small-world and scale-free global topological properties of genetic, proteomic and metabolic networks. Most importantly, these networks are highly clustered and have small node-to-node distances. With their few very connected nodes, which are statistically unlikely to fail under random conditions, the proper functioning of these systems is maintained under external perturbations.

Reductionism has dominated chemistry and physics since their emergence as exact sciences during the early Renaissance. For a long time, biology occupied a marginal place within the family of natural sciences, as biological entities exhibit emergent properties that are not readily explainable by the properties of their constituent parts. In contrast to the success of physics in predicting the properties of regular arrays of electrons and atoms, interacting with well-defined forces, the understanding of complex biological systems was a challenge because interaction patterns are non-trivial and irregular. A new mathematical formalism was developed to represent and study the underlying complex networks, based on classical graph theory. The intent of this article is to review the fast progress of investigations in this field. The understanding of these principles will completely change our approach to the biodiscovery process by providing new tools to prioritize targets and leads. A perspective on how this might happen is given by an overview of all current fields of research such as the transcriptional, protein folding, protein expression, protein–protein, and metabolic networks that are known to date.

Modern graph theory originates in the city of Königsberg, located in the former Prussia. In the 1700s, the notable citizens of Königsberg amused themselves by wandering around the city's seven bridges to verify whether it was possible to walk a route that crossed each bridge exactly once, and return to the starting point. In 1736, the Swiss mathematician Leonard Euler represented this problem as a cyclic graph and demonstrated that it was not possible to visit each of its edges only once and return back [1]. Later on, the development of graph theory meandered between randomness [2,3] and order [4] – either purely statistical of fully analytical methods were used to model network representations of natural systems. The meaning attached to the words 'graph' and 'network' is usually the same, but lately the term 'network' is used to describe the physical or biological system itself, whereas 'graph' is used to depict the mathematical object representing the topology of the system. The most important property that characterizes the topology of huge networks is the connectivity distribution, estimated by counting the frequency $P(k)$ of nodes of degree k. Two other complementary quantities are the characteristic path length and the clustering coefficient. The former is defined as the length of the shortest path to connect one node to another, averaged over all pairs of nodes. The latter measures the average probability that two nodes with a mutual parent will be connected, or in other words, the average connectedness of local neighborhoods.

Experimental evidence revealed the limits of classical graph theory. An early example came with the

Martin G. Grigorov
Nestlé Research Center, BioAnalytical Science, CH-1000 Lausanne 26, Switzerland
e-mail: martin.grigorov@rdls.nestle.com

1359-6446/04/$ – see front matter PII: S1359-6446(04)03369-6

work of Milgram [5] who suggested in 1967 that the average number of social links needed to connect any two people in the United States was less than six, whereas classical random graph theory was predicting a uniform distribution of these. In 1998 Watts and Strogatz [6,7] designed a 'random rewiring' scheme to deliberately introduce increasing amounts of randomness into a lattice-like regular network. This model network was termed by Watts and Strogatz a 'small-world network'. It exhibited the short global separations probed by the experiment of Milgram, due to a high degree of node clustering. In an important contribution, Amaral and co-workers [8] classified the possible types of small-world networks into three different classes: first, scale-free networks, characterized by a vertex connectivity distribution that decays as a power law; second, broad-scale networks, characterized by a connectivity distribution that has a power law regime followed by a sharp cut-off; and third, single-scale networks, characterized by a connectivity distribution with a fast decaying tail.

In 1999, Barabási and Albert [9] showed that the degree of distribution of some real-world graphs was following a power law of the form $P(k) \sim k^{-\gamma}$, in contrast to the expected Poisson distribution $P(k) \sim \lambda^k/k!$ in a random graph*. To explain this finding, it was necessary to go beyond the modeling of topology and address the problems of the graph assembly and evolution. A first merit of Barabási and Albert [9] was that they questioned the fixed number of nodes in network modeling. Second, the authors noticed that in most of the models the probability that two nodes were connected was independent of the nodes' degree, by placing the new edges randomly. On the basis of these two observations, Barabási and Albert [9] proposed a new model in which a network was grown starting with a rather limited seed structure. In this structure, new nodes were added and connected to the already existing ones following the principle of preferential attachment. According to this 'rich-gets-richer' principle, the likelihood for attachment to a node was directly dependent on the node's degree. Few highly connected nodes emerge in such models, referred to as the 'hubs'. The numerical simulations confirmed that the structures evolved to scale-invariant topology described by a power law with an exponent of 3. There was certainly a temptation to claim universality for the exponent, but more recent studies concluded that this was not the case [10]. Most importantly, the scale-free architecture was found not only to ensure the network robustness against random failure of the nodes [11], but also to guarantee an efficient transport and flow processing by avoiding congestion. The dynamical robustness of these complex networks was found to be a direct consequence of their scale-free topology [12]. Therefore, it now seems that the scale-free topology resulted from specific aggregation driven by evolutionary selective processes [13]. In the following paragraphs we will see that these global topological properties are common to all major types of biological networks.

*where γ is referred to as the degree exponent, and represents, for a given network, the slope of the logarithmic graph of the frequency of nodes $P(k)$ of degree k versus the node degree k itself. In difference, the constant λ in the Poisson distribution of the connectivity in random networks is proportional to the characteristic scale of the random network, that is the average connectivity.

Genetic networks

Gene association networks

Snel and co-workers [14] devised a way to obtain protein networks by combining the pair-wise interactions as predicted by the conserved co-occurrence of their genes in operons. The resulting scale-free network contained one giant component, and a multitude of small, disjoint clusters. The giant component was a huge individual cluster that turned out to be a scale-free small-world network on its own. It consisted of locally highly connected subclusters that were related to each other by multifunctional essential linker proteins. All of the smaller subsets had a homogeneous functional composition. In a similar study, Yanai and DeLisi [15] used the methods of comparative genomics to generate the gene association in 43 known microbial genomes to find that the combined networks of functional links contained almost half of an organism's complete genetic complement.

Gene expression networks

Originally, gene expression networks were modeled as random Boolean switching networks, but it was realized that these networks possess a more complicated hierarchical structure that was proposed to lie at the origin of order in the cell [16]. The presence of a set of direct and indirect regulatory interactions between the genes that comprise a transcription regulatory network was noted by Featherstone and Brodie [17] and by Farkas and co-authors [18] as in the majority of single gene deletion *Saccharomyces cerevisiae* mutant strains, the expression of a variable number of other genes was altered. Van Noort, Snel and Huynenh [19] obtained similar results by investigating the gene co-expression network in *S. cerevisiae*, in which genes were linked when they were co-regulated. The topology of gene expression networks indicated the presence of regulatory hubs, while at the local level, network substructures such as motifs and modules were identified [20].

This characteristic structure of gene expression networks is the result of evolutionary self-organization via preferential node attachment. Different studies led to important conceptual developments in the general theory of natural networks. Jordan and co-workers [21] concluded that the genetic evolutionary rates were affected by the topology of the underlying network – the genes with numerous co-expressed partners were found to evolve more slowly than average. Gagen and Mattick [22] have shown that the models of scale-free networks allowing unconstrained growth over time were unable to explain

recent comparative genomics results on the expansion of prokaryote regulatory networks as a function of gene numbers. A new model proposed by the authors explained the observed accelerating, faster than linear (quadratic), growth of prokaryote genomes consisting of fewer than 10,000 genes encoded in DNA sequences of less than ~10 megabases.

The development of microarray technology to quantify the mRNAs that report for the levels of gene expression ushered in a major revolution in biology. The finding that gene expression events form a highly ordered structure provided the basis for the development of novel approaches to mine gene array data, but also to characterize whole cellular states. For example, del-Rio and co-workers [23] designed a methodology that focused on identifying the biological nexuses, representing the potentially important targets for modulation. The fluctuations in the measured mRNA levels of unperturbed cells under fixed conditions have often been evoked as an impediment to the extraction of information from expression profiles. Chen and co-workers [24] demonstrated that such expression fluctuations in *S. cerevisiae* strongly depended on gene function, and furthermore exhibited scale-free network structure. A study suggested a correspondence between a precise transcriptional profile and the related scale-free network [25], and proved the approach to facilitate the identification of disease genes. Another study proposed that network topologies derived from gene expression data might be used to characterize entire cellular states [26] by using the global pattern of gene co-expression events instead of the frequently used fold-change measure.

The presented scientific results suggest that gene networks have evolved to a scale-free and small-world architecture as the result of recurrent single duplication events [27], probably from a limited number of genes in an ancestral organism [28]. The uneven distribution of the connectivity made the living organisms resistant to the deleterious effects of mutations, while the short node-to-node distances guaranteed the efficient processing of regulatory signals. In this respect, microarray experiments would be useful for the fast identification of biological nexuses, which when ordered according to their importance for the network structure, would provide new priority lists for pharmaceutical modulation.

Protein networks

Networks in protein structural space

A polypeptidic chain consists of stable secondary structural elements, and folds instantaneously to the native three-dimensional structure. Vendruscolo and co-workers [29] have shown the existence of key residues acting as 'hubs' in the network of interactions that stabilized the transition state formed during protein folding. Later on, Greene and Higman [30] found that protein structural networks exhibited small-world, single-scale, and to some degree, scale-free properties when amino acids were represented by vertices, whereas edges were used to represent spatial proximity. Further, Atilgan and co-workers [31] complemented these results by showing the importance of the structural 'hubs' for a protein's function and stability. Furthermore, the authors discovered that the average shortest path lengths were highly correlated with residue fluctuations, providing a link between the spatial arrangement of the residues and protein dynamics. These results provide a completely new perspective in envisaging site-directed mutagenesis experiments, by providing priority lists of the candidate residues that are most likely to affect the stability of the target protein.

Because of their thermal motion, protein chains explore a complicated conformational landscape, where the biologically active native state is believed to represent the global energy minimum. The possibility of representing the molecular potential energy landscape as a network was first reported in a study of small molecular clusters [32]. In a similar way, Rao and Caflish [33] mapped the conformational space of a peptide on a network in which the energy minima were the nodes while the observed transitions among them during the simulation were the edges. The conformational space network displayed the typical small-world and scale-free properties that provided evidence for the existence of multiple folding pathways, interleaved due to the existence of 'hub' transition states that acted as linkers.

The variety of observed protein folds is the result of millions of years of divergent evolution. Interestingly, these different protein folds were found to be present to a different degree in genomes. Statistical studies established that these distributions are described by asymptotic power laws, typically associated with scale-free networks [34]. Based on a survey of the first 20 completely sequenced genomes, Qian and co-workers [35] identified the same trend independently of whether folds, protein families, or superfamilies were submitted to scrutiny. An interesting outcome of this work was the rough estimation of the fold composition of the initial genome, related to the common ancestor of modern organisms, which seemed to contain nearly 300 different folds [36]. This is in line with a previous computational estimation of the size of the minimal gene set [28]. The second-order statistics of genome co-occurrence of the different known types of structural domains was investigated by Wuchty [37] and Apic and co-workers [38,39]. Wuchty concluded that the resulting networks exhibited small-world and scale-free topologies with a high degree of local clustering accompanied by a few long-distance connections. Apic investigated the domain combinations in archaeal, eubacterial and eukaryotic proteomes. The authors concluded that recombination between common families, as compared with the invention of new families and recombination among these, has been a major contribution to protein evolution. Dokholyan and co-workers extended these studies to the structural level by investigating how folds were distributed within the

known protein structural universe [36]. The resulting network again possessed scale-free properties that were explained by a similar gene duplication-based divergent evolution from few precursor domains.

Networks of interacting proteins

Proteins are traditionally identified on the basis of their function and they are generally labeled as enzymes, receptors or structural units in cells. Proteins exert their function based on their three-dimensional structure, which they normally adopt instantaneously after synthesis in ribosomes. Currently, post-genomic science is abandoning the concept of single protein function by recognizing that proteins are the elements in complex protein–protein interaction networks. This new interpretation was formulated first by Jeong and co-workers [40] who investigated the direct physical interactions among several hundreds of *S. cerevisae* proteins. In the resulting network, the proteins were represented as vertices, where each pair was connected by an edge, provided that a physical interaction was identified within a large-scale yeast two-hybrid screen [41]. Wuchty obtained similar results, showing that a fairly small set of highly connected proteins and domains shaped the topology of the underlying network [42]. Similar topological properties were identified in the protein network of *Helicobacter pylori* [43] as well as in the networks of five additional species: *Escherichia coli, Caenorhabditis elegans, Homo sapiens, Mus musculus and Drosophila melanogaster* [44].

An important consequence of this inhomogeneous structure is the network's simultaneous tolerance to random errors, coupled with fragility against the removal of the most connected nodes. Jeong and co-workers found evidence for this behavior by modeling random mutations in the genome of *S. cerevisae* by removing arbitrarily selected yeast proteins. The authors identified a striking capacity of yeast to tolerate the deletion of a significant number of individual proteins from its proteome [45,46]. Later, Maslov and Sneppen [47] provided evidence of more subtle topological relationships in protein interaction networks. By quantifying the correlations between connectivities of interacting nodes and comparing them to a randomized network, the authors found that the links between the highly connected proteins were systematically suppressed, whereas those between highly connected and low-connected pairs of proteins were favored. It was suggested that this pattern decreases the likelihood of cross talk between the different functional modules of the cell and increases the overall robustness of the network by localizing the effects of deleterious perturbations.

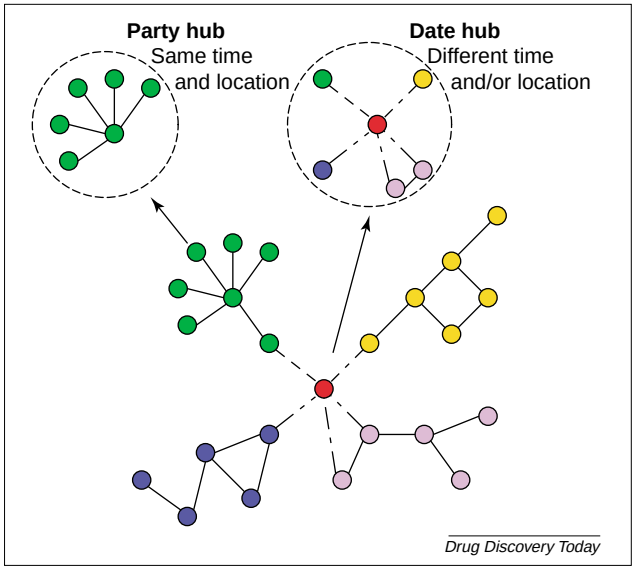

FIGURE 1

Global organisation principles of interactome dynamics, indicating when, where and how complexes form to drive functional changes. Interactions that occur synchronously in the same cellular comportments are depicted in plain lines, whereas the ones occurring at different times are drawn by broken lines.

An interesting aspect in rationalizing the information contained in protein interaction networks is related to the biological function of the most-connected proteins: the hubs. Han and co-workers [48] identified two types of hub: 'party' hubs, which interact within modules with most of their partners simultaneously; and 'date' hubs, which bind to different partners at different times or locations and therefore organize the proteome (cf. Figure 1). In a previous work, Park and co-workers [49] found that most protein families interacted with only one or two other families, whereas a few structural families, such as the immunoglobulins, were extremely versatile in their interactions and were connected to many other families. The interactions between protein domains in terms of the interactions between structural families of evolutionarily related domains, contained in protein structural databases were investigated [50]. The question raised by these studies was to understand what causes the hubs to be so highly interactive. Is the reason structural or evolutionary? Fernandez and co-workers [51] realized that although protein folding domains generally have conserved function across distant homologues, the wrapping of backbone hydrogen bonds is usually not conserved. The extent to which these hydrogen bonds are intramolecularly desolvated and thereby protected from water attack varies markedly. The authors postulated that insufficiently wrapped backbone hydrogen bonds in monomer domains should be adhesive, and therefore a determinant for interactivity, a result that was confirmed experimentally. Such an increased interactivity imposed an 'evolutionary brake' on the overall speed of sequence divergence as more and more residues became functionally indispensable.

Recently, Ramezanpour and co-workers [52] emphasized the important biological fact that proteins normally operate by forming complexes, rather than as isolated entities. Using the protein interaction network of *S. cerevisiae*, the authors found striking similarities and differences between the structural properties of the networks of proteins alone and protein complexes. Although the resulting network

of complexes was still small-world and scale-free, the expected correlations between the degrees and size of the neighboring complexes disappeared. The authors proposed a simple evolutionary model based on duplication and divergence of proteins that was in good agreement with the experimental data. Spirin and Mirny [53] who studied the multibody structure of protein interaction networks made a similar observation. These authors applied several computational algorithms to cluster the yeast interactome and discovered molecular modules that are densely connected with each other but sparsely connected with the rest of the network. Comparison with experimental data and functional annotation of genes revealed two types of modules. The first type are protein complexes that intervene in splicing and transcription, whereas the second class are composed of dynamic functional units such as signaling cascades or cell-cycle regulation machinery. These results provide strong support for the network modularity principle introduced by Hartwell and co-workers [54], suggesting that identified protein subsets constitute, in some sense, the 'building blocks' of molecular networks.

Evolution of protein interaction networks

An important aspect in studying the large-scale organization of protein networks is related to their evolutionary dynamics. The structural importance and the essential nature of the hub proteins suggest that they are likely to be conserved in evolution, implying a strong relationship between the number of interactions and their evolutionary distance to the related orthologs in other organisms. Wagner [55] first reported results on the evolution of the protein interaction network of *S. cerevisiae*. His results indicated that the persistence of redundant interaction partners in the network was the exception rather than the rule. Kunin [56] and co-workers investigated the dynamics of emergence of the protein interaction network of yeast by mapping the origins of proteins on an evolutionary tree. They demonstrated that the most-connected group of proteins date to the eukaryotic radiation, whereas the more ancient group of pre-eukaryotic proteins is less connected. These findings indicate that the evolution of function might be the reason for the differences in connectivity throughout evolutionary time. Wuchty [57] concluded that the relevance of the hubs for the network integrity is simultaneously reflected by a considerable probability of being evolutionarily conserved and essential. Similarly, Eisenberg and Levanon [58] demonstrated that the oldest proteins in a network are the best connected, and that the number of interactions a protein gains during its evolution is proportional to its connectivity. In addition, Wagner [59] found that the highly connected proteins show a greater rate of interaction turnover than proteins with few interactions.

Modularity of protein interaction networks

Network motifs have been suggested to be the elementary building blocks that carry out the key functions in networks [60]. Milo and co-workers [61] investigated the occurrence of these small elements in very complex biological networks and discovered that they appeared there much more frequently than in randomized networks. The same motifs are found in networks that perform information processing, even though they describe elements as different as biomolecules within a cell and synaptic connections between neurons in *C. elegans*. Motifs may thus define universal irreducible classes in biological networks, but their identification by non-overlapping clustering algorithms is limited by the fact that proteins are often participating in more than one complex, for example by being members of a 'date' or a 'party' hub.

Metabolic networks

In 2000, Barabási and co-workers [62] presented a systematic comparative analysis of the global properties of the metabolic networks from 43 organisms, representing all three domains of life. Despite the significant variations in their individual constituents and pathways, these metabolic networks had the same topological scaling properties. Furthermore, Wagner and Fell [63] and Raine and Norris [64] performed a graph theoretical analysis of the *E. coli* metabolic network that emphasized the centrality of the tricarboxylic acid cycle to metabolism. Both groups confirmed the conclusions of Barabási that metabolic networks had small-world and scale-free properties. By studying the metabolisms of 65 fully sequenced organisms, Ma and Zang [65] have shown that the underlying networks contained three major fully connected subsets: a substrate, a product and a third composite subnetwork. The largest fully connected subset of every metabolic network, the so-called giant component, was scale-free, contained less than one-third of the nodes, and its average path length approximated quite well the average path length of the whole network. Later, Schuster and co-workers [66] deconvoluted metabolic networks on the basis of their local connectivity. The authors performed a decomposition of the metabolism of *Mycoplasma pneumoniae* that provided evidence of the existence of 19 separate subnetworks.

Valuable results have been obtained in the general theory of scale-free networks based on metabolic model systems. Until now, most of the efforts aimed at rationalizing the robustness of natural networks to random failures considered only the failure of the vertices. However, the failure of single connections is arguably much more likely to occur than the extinction of entire nodes. Kaiser and Hilgetag [67] found that edge frequency in all of the shortest paths in a network yielded a particularly high correlation with vulnerability. In an interesting work, Csermely [68] demonstrated the existence of a scale-free distribution of the metabolic fluxes of *E. coli* and related the major metabolic reactions to the original minimal gene set of Mushegian and Koonin [28]. The author recognized the importance of the strong links, but suggested that the

weak ones might have a general role in the stabilization of the complex system.

The small-world architecture of metabolic networks was selected by evolution to minimize the transition times between metabolic states while the unevenly distributed number of enzymatic reactions per metabolite ensured the stability of the system to random mutations. Again, a restricted list of the 'hubs', supporting the whole metabolic network, provides the key nodes that control the behavior of the whole system. In this way, a new technique was designed to re-engineer the sensitivity in antibiotic-resistant *Pseudomonas aeruginosa* strains isolated in the clinic [69].

Discussion

When a new theory comes to life, it is common that it is wholeheartedly supported, but also strongly criticized, and this is happening with the new science of connectedness. Although it is now a commonly accepted fact that biological networks exhibit small-world and scale-free properties, several researchers are questioning the point on the basis of new experimental evidence. Probably the most important case was the recent claim of Arita [70] that the metabolic work of *E. coli* is not small. Indeed, the biochemical investigation of annotated carbon atomic traces in the metabolic reactions of *E. coli* revealed that the average path length of its metabolism is much longer than previously thought. Przulj and co-workers [71] analyzed protein–protein interaction networks of *S. cerevisiae* and *D. melanogaster* using a newly introduced measure of local network structure. The authors demonstrated that the currently accepted scale-free model of protein–protein interaction networks failed to fit the data in several respects and showed that a random geometric model was much more accurate.

An important argument in the controversy is related to the quality of the data used to derive global network properties. The completeness and the quality of the different datasets being used were often questioned. For example, an uncertainty exists about the relevance of protein interactions identified by yeast two-hybrid techniques because bait protein hybrids turned out to be more prolific binders than their prey counterparts. Among others, Goldberg and Roth [72] addressed the issue by showing that even if experimentally determined networks are susceptible to errors, one can still draw important inferences from them. Recently, Yook [73] and co-workers thoroughly addressed the issue of completeness of protein interaction networks. The authors compared four available datasets to uncover the network's generic large-scale properties and the impact of the proteins' function and cellular localization on the network topology. It was shown that each dataset supports a scale-free topology with hierarchical modularity, indicating that these features represented a robust and generic property of the protein interaction network. The uncovered systematic differences between the four protein interaction datasets reflects their relative coverage for different functional and localization classes.

The modular architecture of biological networks that has been unraveled sets new perspectives in envisaging the interaction and control of biological entities. Probably the most important outcome of all these discoveries is that we now have at hand new tools to prioritize biochemical experiments, targets and leads. The fact that protein structures themselves are organized around scale-free small-world principles enables the identification of key amino acids for site-directed mutagenesis experiments aimed at manipulating protein functionalities, such as enzymatic efficiency or thermostability. Similarly, the efficiency in re-engineering bacterial metabolisms to produce desired chemicals might be increased by focusing on the 'neural centers' controlling the networks of bioconversions.

The ultimate objective of modern food and drug industries is to develop the technologies necessary to guarantee a healthy life for biological entities (humans, pets), as well as to remedy any pathological deviation. The global properties of the underlying biological networks offer new ways to discover and validate therapeutic targets [74] as well as to design small molecular ligands to interact with them. The careful selection of several key proteins that are essential for the proper functioning of the targeted network allows the overcoming of the usual resistance of the network to the deleterious effects of single molecular agents [75]. Moreover, the nascent competences of chemogenomics are available to rationally design combinations of small molecules that are able to collectively shift network behavior [76]. The most recent development in this respect was the idea of targeting several key nodes in the disease-associated network by overlapping pharmacophores in single molecules [77]. In our quest of the ultimate drug, we have much to gain by learning from the organization and composition of highly optimized molecular cocktails that evolution offered us, such as milk [78,79].

Conclusions

The large-scale analysis of biological networks has begun to reveal the global organization of the cell. Now there is clear evidence that the cellular phenotypes observed at the macroscopic level depend on the collective characteristics of the underlying networks. The application of the new science of connectedness in life sciences has the potential to qualitatively change biological research. In the near future, the nature of the pharmacological targets will shift from single proteins, to functional protein complexes, to whole networks determining precise cellular states. In turn, the new therapeutics will no longer consist of single active molecules but will instead represent molecular cocktails with components that target the protein hubs in disease-associated molecular networks.

Acknowledgements

The author would like to thank Prof Dr Bruce German for the numerous discussions we had on the subject.

References

1 Euler, L. (1736) Solutio Problematis ad geometriam situs pertinensis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 8, 128–140
2 Erdos, P. and Rényi, R. (1959) On random graphs. *Pub. Mathem.* 6, 290–297
3 Solomonoff, R. and Rapoport, A. (1951) Connectivity of random nets. *Bull. Math. Biophys.* 13, 107–111
4 Cayley, A. (1857) On the theory of analytic forms called trees. *Philos. Mag.* 13, 19–30
5 Milgram, S. (1967) The small-world problem. *Psychology Today,* 60–67
6 Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature* 393, 440–442
7 Watts, D.J. (1999) *Small-worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press
8 Amaral, L.A. *et al.* (2000) Classes of small-world networks. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11149–11152
9 Barabási, L.A. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509–512
10 Goh, K-I. *et al.* (2002) Classification of scale-free networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12583–12588
11 Albert, R. *et al.* (2000) Error and attack tolerance of complex networks. *Nature* 406, 378–382
12 Toroczkai, Z. and Bassler, K.E. (2004) Network dynamics: jamming is limited in scale-free systems. *Nature* 428, 716
13 Aldana, M. and Cluzel, P. (2003) A natural class of robust networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8710–8714
14 Snel, B. *et al.* (2002) The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5890–5895
15 Yanai, I. and DeLisi, C. (2002) The society of genes: networks of functional links between genes from comparative genomics. *Genome Biol.* 3, research0064
16 Fox, J.J. and Hill, C.C. (2001) From topology to dynamics in biochemical networks. *Chaos* 11, 809–815
17 Featherstone, D.E. and Broadie, K. (2002) Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays* 24, 267–274
18 Farkas, I. J. *et al.* (2003) The topology of the transcription regulatory network in the yeast, *S. cerevisiae*. *Physica A* 318, 601–612
19 van Noort, V. *et al.* (2004) The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Rep.* 5, 280–284
20 Babu, M.M. *et al.* (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14, 283–291
21 Jordan, I.K. *et al.* (2004) Conservation and co-evolution in the scale-free human gene coexpression network. *Mol. Biol. Evol.* 21, 2058–2070
22 Gagen, M.J. and Mattick, J.S. (2004) Inherent size constraints on prokaryote gene networks due to "accelerating" growth. *Los Alamos National Laboratory, Preprint Archive, Quantitative Biology,* 1–15, arXiv:q-bio.MN/0312021
23 del-Rio, G. *et al.* (2001) Mining DNA microarray data using a novel approach based on graph theory. *FEBS Lett.* 509, 230–234
24 Chen, W.W. *et al.* (2004) An exact model of fluctuations in gene expression. *Los Alamos National Laboratory, Preprint Archive, Quantitative Biology,* 1–15, arXiv:q-bio.MN/0402021
25 Bortoluzzi, S. *et al.* (2003) Disease genes and intracellular protein networks. *Physiol. Genomics* 15, 223–227
26 Carter, S.L. *et al.* (2004) Gene expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20, 2242–2250
27 Amoutzias, G.D. *et al.* (2004) Convergent evolution of gene networks by single-gene duplications in higher eukaryotes. *EMBO Rep.* 5, 274–279
28 Mushegian, A.R. and Koonin, E.V. (2000) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273
29 Vendruscolo, M. *et al.* (2002) Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E* 65, 061910
30 Greene, L.H. and Higman, V.A. (2003) Uncovering network systems within protein structures. *J. Mol. Biol.* 334, 781–791
31 Atilgan, A.R. *et al.* (2004) Small-world communication of residues and significance for protein dynamics. *Biophys. J.* 86, 85–91
32 Doye, J.P.K. (2002) Network topology of a potential energy landscape: a static scale-free network. *Phys. Rev. Lett.* 88, 238701
33 Rao, F. and Caflisch, A. (2004) The protein folding network. (2004). *J. Mol. Biol.* 342, 299–306
34 Koonin, E.V. *et al.* (2002) The structure of the protein universe and genome evolution. *Nature* 420, 218–223
35 Qian, J. *et al.* (2001) Protein family and fold occurrence in genomes: power-law behaviour and evolutionary model. *J. Mol. Biol.* 313, 673–681
36 Dokholyan, N.V. *et al.* (2002) Expanding protein universe and its origin from the biological Big Bang. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14132–14136
37 Wuchty, S. (2001) Scale-free behavior in protein domain networks. *Mol. Biol. Evol.* 18, 1694–1702
38 Apic, G. *et al.* (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* 310, 311–325
39 Apic, G. *et al.* (2001) An insight into domain combinations. *Bioinformatics* 17, S83–89
40 Jeong, H. *et al.* (2001) Lethality and centrality in protein networks. *Nature* 411, 41
41 Uetz, P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in S. cerevisiae. *Nature* 403, 623–627
42 Wuchty, S. (2002) Interaction and domain networks of yeast. *Proteomics* 2, 1715–1723
43 Rain, J.-C. *et al.* (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature* 409, 211–215
44 Ng, K-L. and Huang, C-H. (2004) Global topological study of the protein-protein interaction networks. *Los Alamos National Laboratory, Preprint Archive, Quantitative Biology* 1–13, arXiv:q-bio.MN/0402027
45 Winzeler, E.A. *et al.* (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* 285, 901–906
46 Ross-Macdonald, P. *et al.* (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402, 413-418
47 Maslov, S. and Sneppen, K. (2002) Specificity and stability in topology of protein networks. *Science* 296, 910–913
48 Han, J.D. *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93
49 Park, J. *et al.* (2001) Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. *J. Mol. Biol.* 307, 929–938
50 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
51 Fernandez, A. *et al.* (2004) The nonconserved wrapping of conserved protein folds reveals a trend toward increasing connectivity in proteomic networks. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2823–2827
52 Ramezanpour, A. *et al.* (2003) Structural properties of S. cerevisiae protein complex network. *Los Alamos National Laboratory, Preprint Archive, Condensed Matter,* 1–14, arXiv:cond-mat/0304207
53 Spirin, V. and Mirny, L.A. (2003) Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12123–12128
54 Hartwell, L.H. *et al.* (1999) From molecular to modular cell biology. *Nature* 402, C47–C52
55 Wagner, A. (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* 18, 1283
56 Kunin, V. *et al.* (2004) Functional evolution of the yeast protein interaction network. *Mol. Biol. Evol.* 21, 1171–1176
57 Wuchty, S. (2004) Topology and evolution in the yeast protein interaction network. *Genome Res.* 14, 1310–1314
58 Eisenberg, E. and Levanon, E.Y. (2003) Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* 91, 138701/1-138701/4
59 Wagner, A. (2003) How the global structure of protein interaction networks evolves. *Proc. R. Soc. Lond. B. Biol. Sci.* 270, 457–466
60 Itzkovitz, S. *et al.* (2003) Subgraphs in random networks. *Phys. Rev. E* 68, 026127
61 Milo, R. *et al.* (2002) Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298, 824–827
62 Jeong, H. *et al.* (2000) The large-scale organization of metabolic networks. *Nature* 407, 651–654
63 Wagner, A. and Fell, D.A. (2001) The small-world inside large metabolic networks. *Proc. R. Soc. Lond. B. Biol. Sci.* 269, 1803–1810
64 Raine, D.J. and Norris, V. (2001) Network structure of metabolic pathways. *J. Biol. Phys. Chem.* 1, 89–94
65 Ma, H.W. and Zheng, A.P. (2003) The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 1423–1430
66 Schuster, S. *et al.* (2002) Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics* 18, 351–361
67 Kaiser, M. and Hilgetag, C.C. (2004) Edge vulnerability in neural and metabolic networks. *Biol. Cybern.* 90, 311–317
68 Csermely, P. (2004) Strong links are important, but weak links stabilize them. *Trends Biochem. Sci.* 29, 331–334

69 Gill, R. (2004) Evolution, characterization, and engineering of bacterial traits. In *Abstracts of Papers,* 227th ACS National Meeting, Anaheim, CA, USA, March 28–April 1, American Chemical Society.
70 Arita, M. (2004) The metabolic world of *E. coli* is not small. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1543–1547
71 Przulj, N. *et al.* (2004) Modeling interactome: scale-free or geometric? *Bioinformatics* 10.1093/bioinformatics/bth436
72 Goldberg, D.S. and Roth, F.P. (2003) Assessing experimentally derived interactions in a small-world. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4372–4376
73 Yook, S.H. *et al.* (2004) Functional and topological characterization of protein interaction networks. *Proteomics* 4, 928–942
74 Lappe, M. and Holm, L. (2004) Unraveling protein interaction networks with near-optimal efficiency. *Nat. Biotechnol.* 22, 98–103
75 Albert, R. *et al.* (2000) Error and attack tolerance of complex networks. *Nature* 406, 378–382
76 Sharom, J.R. *et al.* (2004) From large networks to small molecules. *Curr. Opin. Chem. Biol.* 8, 81–90
77 Morphy, R. *et al.* (2004) From magic bullets to designed multiple ligands. *Drug Discov. Today* 9, 641–651
78 Cavaletto, M. *et al.* (2004) The proteomic approach of human milk fat globule membrane. *Clin. Chim. Acta* 347, 41–48
79 German, B.J. *et al.* (2002) Bioactive components in milk. *Curr. Opin. Nutr. Metab. Care* 5, 653–658