

# 3 Global Properties of Networks

Ralf Steuer<sup>a</sup> and Gorka Zamora-López<sup>b</sup>

<sup>a</sup>Humboldt University Berlin, Institute for Theoretical Biology, Invalidenstr. 43, 10115 Berlin, Germany. <sup>b</sup>University of Potsdam, Institute for Physics, Nonlinear Dynamics Group, Am Neuen Palais 10, 14469 Potsdam, Germany.

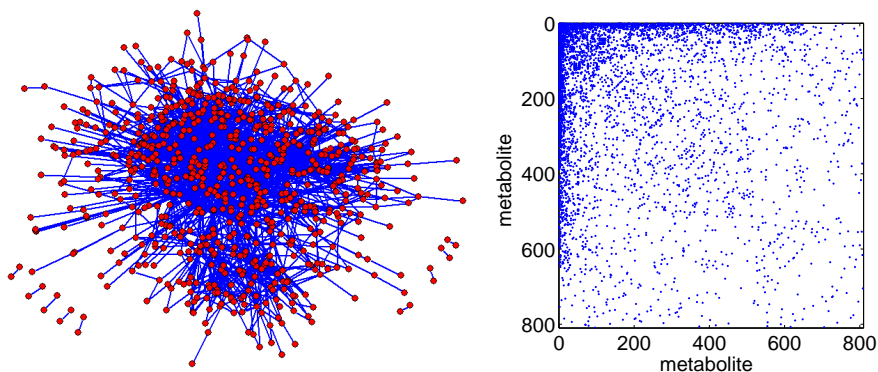
## 3.1 INTRODUCTION

Complex dynamical systems are often characterized by a large number of nonlinearly interacting elements, giving rise to emergent properties that transcend the principle of linear superposition. In particular within the biological sciences, one of the primary challenges is to investigate how the collective behavior of cells, tissues or organisms can be understood in terms of the properties of their molecular constituents.

To investigate this intricate connectivity of cellular systems, the analysis of complex networks has become an important part of molecular biology. A large number of biological phenomena and processes can be translated into the abstract concept of a complex network, making biological problems mathematically tractable. Prominent examples include the representation of transcriptional regulation as a network, where vertices represent genes or proteins and edges represent regulatory interactions, as well as cellular metabolism, where vertices represent metabolites and edges represent biochemical interconversions. However, beyond these rather straightforward examples, more abstract processes can sometimes also be translated into the language of complex networks. For example, different configurational states of a protein may be represented as vertices, with edges indicating transitions between them.

Once a biological process or phenomenon is represented by a network, the tools of complex network theory allow for a systematic characterization of its structural properties. The analysis of network topology then seeks to uncover the functional organization, the underlying design principles and unknown organizing principles of cellular systems. Indeed, as realized only rather recently, many empirically derived complex networks, ranging from technological and sociological to biological examples, share common topological features. The organizing principles of empirical networks often reflect crucial system properties, such as robustness, redundancy or other functional interdependencies between network elements. A quantitative analysis of the large-scale characteristics of complex networks thus contributes to a

ii



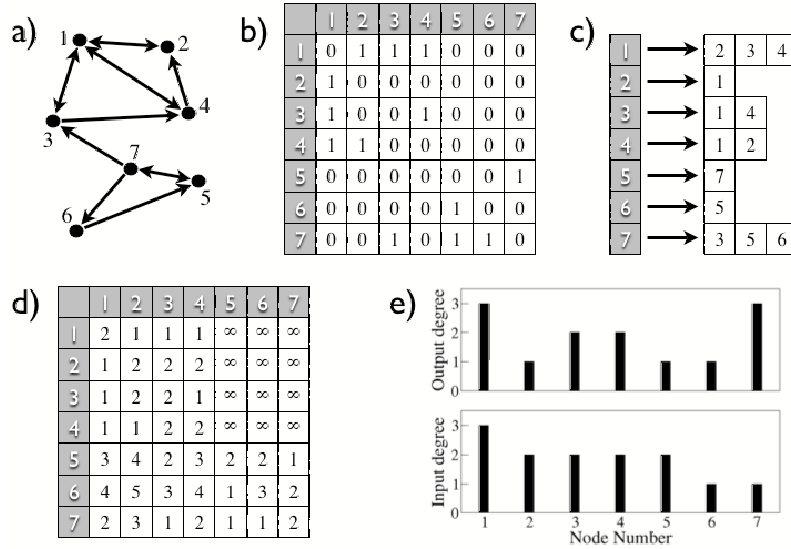
**Fig. 3.1** The substrate graph  $G_S$  of the *S. cerevisiae* metabolic network [23], consisting of  $N_V = 810$  vertices (metabolites) and  $N_E = 3419$  edges. Directional information is omitted. *Left*: A visualization of the substrate graph using the freely available software package Pajek [9]. *Right*: A visualization of the adjacency matrix, with vertices (metabolites) ranked according to their degree. Each dot indicates whether the corresponding vertices (metabolites) are connected by an edge. The figures are adapted from [61].

better understanding of the organization of cellular functions and has already made significant impact on our current view of molecular biology.

While not aiming at a comprehensive review, this chapter seeks to summarize and describe several basic measures and characteristics of network topology. The chapter is organized as follows: The main emphasis is placed on an overview of basic measures and indices that characterize the topology of networks, given within Section 3.2. In Section 3.3, several basic prototype models of complex networks are discussed. The subsequent Section 3.4 is devoted to a brief outline of global features of complex networks, such as hierarchies, modularity, attack tolerance and robustness. Finally, Section 3.5 provides notes on the statistical testing of network properties and describes several known pitfalls and possible misinterpretations in the statistical analysis of network properties. The working example throughout this chapter is a reconstructed version of the *S. cerevisiae* metabolic network [23], consisting of 810 metabolites and 843 reactions. The original bipartite graph was collapsed, such that two metabolites are connected if they participate in a common reaction. A graphical representation is shown in Fig. 3.1.

### 3.2 GLOBAL PROPERTIES OF COMPLEX NETWORKS

Following the nomenclature of Chapter 2, a network is formally represented by a graph  $G = (V, E)$ , consisting of a set  $V$  of  $N_V$  vertices and a set  $E$  of  $N_E$  edges. We distinguish between *undirected graphs*, whose vertices are connected by edges without any directional information and *directed graphs (digraphs)*, whose



**Fig. 3.2** Representations of complex networks. *a)* A directed network, consisting of  $N_V = 7$  vertices and  $N_E = 13$  directed edges. *b)* The adjacency matrix  $\mathbf{A}$  of the network. *c)* The set of adjacency lists, specifying to which other vertices each vertex connects. *d)* The distance matrix  $\mathbf{D}$  with elements  $d_{ij}$ . Note that the distances are not symmetric and may be infinite, indicating that not all vertices can be reached from all other vertices. *e)* The input degree  $k_i^{\text{in}}$  and output degree  $k_i^{\text{out}}$  of each vertex.

edges possess directional information. Additionally, in *weighted graphs*, each edge (directed or undirected) is associated with a scalar value, quantifying a possible interaction strength, a cost, or a flow on the respective edge.

In most cases, a network is represented by its *adjacency matrix*  $\mathbf{A}$ , with entries  $A_{ij} = 1$  indicating that there exists an edge between vertex  $n_i$  and  $n_j$ , and  $A_{ij} = 0$  otherwise. For undirected networks, the adjacency matrix is symmetric  $A_{ij} = A_{ji}$ . For weighted networks, the elements of the adjacency matrix are replaced by non-binary scalar values.

However, in particular for *sparse networks*, i.e. networks where the number of edges is much smaller than the number of possible edges  $N_E \ll N_V^2$ , the adjacency matrix becomes computationally inefficient in terms of memory allocation. Alternatively, the network can be specified by a set of *adjacency lists*, consisting of  $N_V$  lists that enumerate to which other vertices each vertex connects, see also Chapter 2. The adjacency matrix, as well as the adjacency lists, have their unique advantages and disadvantages in terms of computational efficiency. A schematic example of both representations is given in Fig. 3.2.

### 3.2.1 Distance, Average Pathlength and Diameter

In a network consisting of  $N_V$  vertices, the *distance*  $d_{ij}$  between any two vertices  $n_i$  and  $n_j$  is given by the length of the *shortest path* between the vertices, i.e., the minimal number of edges that need to be transversed to travel from vertex  $n_i$  to  $n_j$ . The shortest path between two vertices does not have to be unique, often there exist several alternative paths with identical pathlength. For directed networks, the distance between two vertices  $n_i$  to  $n_j$  is usually not symmetric  $d_{ij} \neq d_{ji}$ . Likewise, for directed, as well as *disconnected networks*, i.e., networks consisting of two or more isolated components, there might not always be a path that connects vertex  $n_i$  to  $n_j$ . In such a case, the distance between the respective vertices is infinite  $d_{ij} = \infty$ . See Fig. 3.2 for examples.

The *diameter*  $d_m = \max(d_{ij})$  of a network is defined as the maximal distance of any pair of vertices. The *average* or *characteristic* pathlength  $d = \langle d_{ij} \rangle$  of a network is defined as the average distance between all pairs of vertices. In the case of infinite distances, the average inverse pathlength  $d_{\text{eff}} = \langle 1/d_{ij} \rangle$ , also referred to as *efficiency*, can be used to specify the average pathlength within the network. In this case, a fully connected network  $d_{ij} = 1 \forall i, j$  has an efficiency  $d_{\text{eff}} = 1$ , whereas large distances and disconnected components (using the limit  $1/d_{ij} = 0$  for  $d_{ij} = \infty$ ) reduce the efficiency of the network.

The situation is slightly less straightforward if weighted networks are considered. Then, we are faced with the possibility to take additional information into account. For example, within a network of train connections, the shortest pathlength (distance) between two stations can be defined according to physical distances, or, taking travel time into account, by the total time needed to travel from one station to another. Furthermore, the fastest connection must not always be the cheapest, thus we might wish to define the distance between two stations according to the amount of money needed to travel from one station to another. In either case, the term distance between vertices can be generalized to accommodate additional scalar information, given by a *weight factor* that is associated with each edge.

Computationally, the estimation of the distance between two vertices is not trivial. Within the extensive literature on the shortest paths problem, the most common choices are the *Dijkstra* and the *Floyd-Warshall* algorithm [6]. The Dijkstra algorithm returns the lowest cost path between a source vertex  $n_i$  and all other vertices in the network in  $\mathcal{O}(N_V^2)$  time. For efficiency reasons, the algorithm return just *one* shortest path, enumerating all shortest paths between two vertices is computationally more tricky and expensive. To calculate the all-to-all distances, the Floyd-Warshall algorithm is the method of choice. The algorithm returns the distance matrix in  $\mathcal{O}(N_V^3)$ . Both algorithms straightforwardly allow to incorporate weighted edges. Negative weights may induce cycles that reduce the cost of a path each time the cycle is traversed. In this case, the definition of the 'lowest cost' path has to be modified. Note that distances, pathlength and diameter also depend on network size and density (number of vertices and links) and are therefore no genuine classifiers that straightforwardly allow to compare different networks.

### 3.2.2 Six Degrees of Separation: Concepts of a Small World

One of the striking properties of almost all empirical networks is that, despite their huge size of sometimes several millions of vertices, the average pathlength is usually surprisingly small. For example within cellular metabolism, represented by a network of metabolites (vertices) linked by biochemical reactions (edges), the average pathlength between two metabolites is only approximately  $d \approx 3$ , independent of the specific organism [22, 33, 71]. A recent study of the World-Wide Web (WWW), represented by a network of web documents (vertices) that are connected by directed hyperlinks (URLs), estimated that the average pathlength between any two vertices is only  $d \approx 16$  [1], extrapolated for a network of 200 million documents.

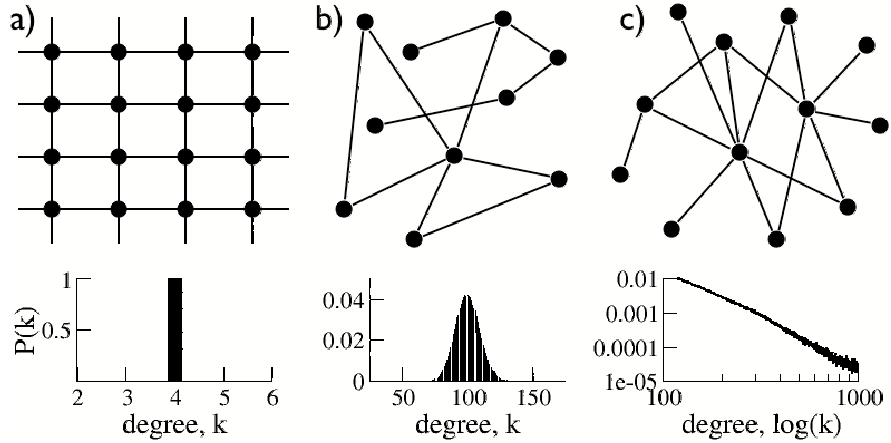
The term *small world* network itself originated in the social sciences, reflecting the assertion that within networks of social acquaintances (or friendships) all people (vertices) on the planet are separated from each other by just a small number of intermediate friends or acquaintances ('*six degrees of separation*', although the specific value six must not be taken too literally).

However, strictly speaking, the term small-world is not a genuine network property, i.e. there is no measure or statistical test that allows to check whether a given specific empirical network belongs to the class of small world networks. As stated above, the average distance between vertices also depends on the size of the network: The more vertices a network has, the more distant the vertices tend to be. The small-world property is thus mainly understood to apply to network models whose average pathlength  $d$  increases slower or equal than the logarithm of the network size  $d \sim \log N_V$  for  $N_V \rightarrow \infty$ . A further distinction includes *ultra-small* networks [13], whose average pathlength scales as  $d \sim \log \log N_V$ .

### 3.2.3 The Degree Distribution

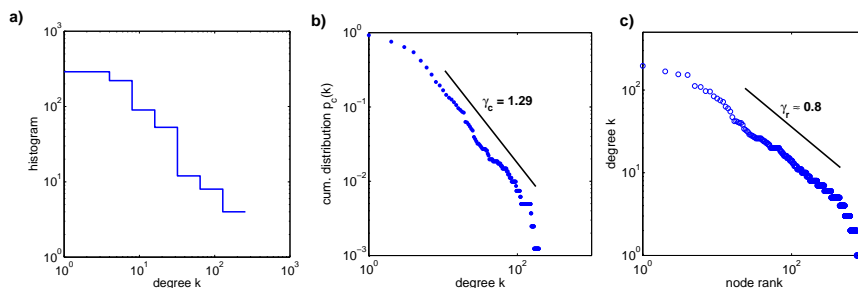
One of the most basic properties of a vertex  $n_i$  is its *degree*  $k_i$ , defined as the number of edges adjacent to the vertex. In a network without *self-loops* (edges that connect a vertex to itself) and *multiple links* (two vertices are connected by more than one edge) the degree equals the number of neighbors of the vertex. In the case of directed networks, we distinguish between the *input degree*  $k_i^{\text{in}}$  and the *output degree*  $k_i^{\text{out}}$ . Taking all vertices of a network into account, we can ask for the probability  $p(k)$  that the degree of a randomly chosen vertex equals  $k$ . The *degree distribution*  $p(k)$  has become one of the most prominent characteristics of network topology. One of the key discoveries that triggered the renewed interest in complex network theory was that the distribution  $p(k)$  of many empirical networks approximately follows a power law  $p(k) \sim k^{-\gamma}$ , where  $\gamma$  denotes the *degree exponent*. In contrast to the until then prevailing picture, where vertices are connected randomly and each vertex has approximately the same number of links, many empirical networks are strongly inhomogeneous: While the vast majority of vertices only possess a small number of links, a small number of vertices ('*hubs*') are highly connected. Examples of prototypical degree distributions are depicted in Fig. 3.3.

Though being one of the most basic characteristics of network architecture, a statisti-



**Fig. 3.3** Degree distributions of complex networks. *a)* A lattice-like network. Each vertex has the same degree  $k$  (for periodic boundary conditions or large networks, such that vertices at the border can be neglected). *b)* An Erdős-Rényi random network. The degree distribution is homogeneous, the degrees of the vertices are centered around the average value. *c)* A scale-free network. The degree distribution is highly inhomogeneous and follows a power law of the form  $p(k) \sim k^{-\gamma}$ , where  $\gamma$  denotes the *degree exponent*. While most vertices only have a low number of connections, a smaller number of vertices is highly connected.

cally stringent numerical estimation of the degree distribution is far from trivial [25]. In the simplest case,  $p(k)$  can be straightforwardly estimated from an (usually binned) histogram of degrees. However, for many real networks with strongly inhomogeneous degree distributions, the simple histogram approach provides insufficient statistics at high degree vertices and is a notorious source of misinterpretations [25]. More reliable in terms of numerical estimation is the *cumulative degree distribution*  $p_c(k)$ , defined as the probability that a randomly chosen vertex has a degree larger than  $k$ . The cumulative degree distribution  $p_c(k)$  is a monotonously decreasing function of  $k$  and its estimation requires no binning. For a power-law distribution  $p(k) \sim k^{-\gamma}$ , the cumulative degree distribution is of the form  $p_c(k) \sim k^{-(\gamma-1)}$ . An exponential distribution  $p(k) \sim \exp(-k)$  corresponds to an invariant cumulative distribution  $p_c(k) \sim \exp(-k)$ . Computationally even more straightforward is to rank the vertices according to their degree and plot the degree versus the rank of each vertex. Examples of different representations of the degree distribution are shown in Fig 3.4. It should be noted that all empirical networks necessarily show deviations from an strict mathematical degree distribution. In particular for power-law distributions, the size (number of vertices) of the network puts constraints on the estimation of the degree exponent. Highly connected vertices are rare, and their probability is thus difficult to estimate for small networks. Likewise, the number of vertices with small degree is restricted by network size. Consequently, the formula  $p \sim k^{-\gamma}$  often only applies to an intermediate region of the empirical degree distribution and has to be adjusted with an exponential cut-off at high degrees. More importantly, as shown



**Fig. 3.4** Different representations of the degree distribution of the metabolite substrate network described in Fig 3.1. *a*) A binned histogram. Shown is the number of vertices with a degree  $k$ , using a logarithmic binning. *b*) The cumulative degree distribution  $p_c(k)$ , i.e. the probability that a vertex has a degree larger or equal  $k$ . Note that the cumulative distribution does not require binning, but is obtained from the (normalized) number of vertices with degree larger or equal  $k$ . *c*) The rank plot of metabolites, ranked according to their degree  $k$ . A power-law of the form  $p \sim k^{-\gamma_r}$  in the rank plot corresponds to a degree exponent  $\gamma = 1 + \frac{1}{\gamma_r} \approx 2.3$  in the original degree distribution  $p(k)$  and  $\gamma_c \approx 1.3$  in the cumulative distribution. The straight lines are not fitted and only serve as a guide to the eye.

in the recent literature, the reported degree exponent of many empirical networks correlates with network size and thus might not reflect the actual exponent of the underlying networks [18, 17]. Furthermore, for small degree exponents the variance of the degree distribution is infinite, thus any empirical sample of vertex degree is no 'typical' observation.

However, for many biological problem it is often more important to note that the degree distribution is highly inhomogeneous and long-tailed, as opposed to the question whether the degree distribution fits a power-law in a strict statistical sense. For weighted networks, the concept of degree can also be extended to account for the weights of the edges by defining the *strength* of a vertex as the sum of the absolute values of the weights.

### 3.2.4 Assortative Mixing and Degree Correlations

Despite its importance in the topological characterization of complex networks, the degree distribution itself does provide only little information about the internal structure and organization of the network. More interesting is thus to look for correlations between the degrees of adjacent vertices. A network is called *disassortative* if vertices with high degree connect preferentially to vertices with low degree. Vice versa, a network is called *assortative* if vertices with high degree preferentially also connect to other vertices with high degree. As pointed out in the recent literature [49], social networks tend to be assortative, i.e. persons (vertices) with many friends (connections) tend to be also connected to other persons with many friends, while most technological and biological networks are disassortative.

Formally, the degree correlation can be obtained from the joint probability distribution  $p(k_i, k_j)$  that two connected vertices  $n_i$  and  $n_j$  have degree  $k_i$  and  $k_j$  respectively. For uncorrelated degrees the joint probability is given by the product of the marginal degree distributions  $p(k_i, k_j) = p(k_i)p(k_j)$ . A measure for the deviation from statistical independence is given by the *mutual information* [64, 66].

Unfortunately, a direct numerical estimation of  $p(k_i, k_j)$  is computationally demanding and often not feasible due to the limited size of the (empirical) network (but see also [66] for the numerical estimation of probability distributions and a discussion of finite size effects). More straightforward is thus to consider the Pearson correlation coefficient between the degree of two adjacent vertices. The correlation coefficient or *assortativity coefficient*  $r$  lies in the range  $-1 \leq r \leq 1$ , with  $r < 0$  corresponding to a disassortative network and  $r > 0$  to an assortative network. Note that the assortativity coefficient  $r$ , similar to the usual Pearson correlation, has its limits for strongly inhomogeneous degree distributions and fails to correctly quantify nonlinear degree correlations, i.e. networks that are assortative for low degree vertices and disassortative for high degree vertices.

Another popular, and closely related, measure to evaluate degree correlations is the *average neighbor degree* [53]. For each vertex  $n_i$  the average degree  $k_{i,nn} = \frac{1}{k_i} \sum_{j=1}^{N_V} A_{ij} k_j$  of its neighbors is calculated. Subsequently, these values are averaged for all vertices having the same degree  $k$ , resulting in the average neighbor degree  $k_{nn}(k)$ . See Fig. 3.5 for examples of vertex degree correlations.

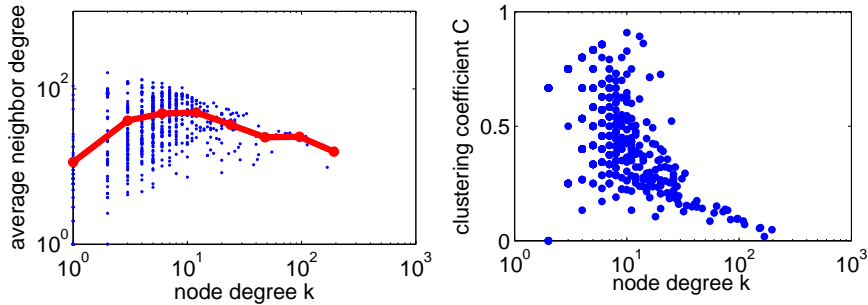
To evaluate the degree correlations for weighted and directed networks requires slight modifications in the respective definitions. In the case of directed networks, two distinct correlation indices are most interesting: (i) Do the in-degrees  $k_i^{\text{in}}$  of vertices correlate with their neighbors out-degrees  $k_i^{\text{out}}$ , and (ii) do the out-degrees  $k_i^{\text{out}}$  of vertices correlate with their neighbors in-degrees  $k_i^{\text{in}}$ ? In the case of weighted networks, the degrees can again be replaced by their weighted counterparts.

### 3.2.5 The Clustering Coefficient

Another basic measure that accounts for the internal structure of a network is the *clustering coefficient*  $C$ . The clustering coefficient relates to the local cohesiveness of a network and measures the probability that two vertices with a common neighbor are connected. In the case of undirected networks, given a vertex  $n_i$  with  $k_i$  neighbors, there exist  $E_{\text{max}} = k_i(k_i - 1)/2$  possible edges between the neighbors. The clustering coefficient  $C_i$  of the vertex  $n_i$  is then given as the ratio of the actual number of edges  $E_i$  between the neighbors to the maximal number  $E_{\text{max}}$ ,

$$C_i = \frac{2E_i}{k_i(k_i - 1)}. \quad (3.1)$$

See Fig. 3.6 for a schematic example. Note that, strictly speaking, the clustering coefficient  $C_i$  is not a property of the vertex  $n_i$  itself, but rather a property of its neighbors. The global or mean clustering coefficient  $C = \langle C_i \rangle$  of the network is the average cluster coefficient of all vertices.



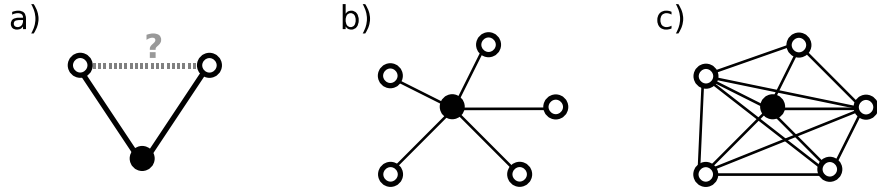
**Fig. 3.5** Vertex degree correlation in the substrate graph. *Left:* The average neighbor degree  $k_{i,nn}$  of each vertex  $n_i$ , plotted versus the degree  $k_i$ . The solid line gives the (binned) average over all vertices with the same degree  $k$ . For large degrees a weak negative correlation is observed. *Right:* The clustering coefficient  $C_i$  of each vertex versus the degree  $k_i$ . Highly connected vertices exhibit a low clustering coefficient, i.e., highly connected vertices preferentially connect to vertices that are not mutually connected, indicating a hierarchical structure.

Many empirical networks exhibit a rather high clustering coefficient, indicating a local cohesiveness and a tendency of vertices to form clusters or groups. Indeed, for example in social networks, it seems intuitive that two persons (vertices) who have a common friend are much more likely to be also friends, as compared to two randomly chosen persons. Interestingly, this also directly relates to the notion of degree correlations and dynamics on networks. As persons that share a common friend are likely to become acquainted themselves, they will acquire new friends over time. In particular, a highly connected person will induce new connections among his friends (neighboring vertices). In this sense, within social networks, a situation with disassortative degree correlations and low clustering coefficients is dynamically unstable and must be expected to evolve gradually towards more clustering and thus assortative degree correlations.

However, despite its conceptual simplicity, the interpretation and statistical testing of the clustering coefficient holds some pitfalls, which are discussed in more detail in the Section 3.5. Furthermore, the clustering coefficient depends on the number of edges within the network. To claim a nontrivial local clustering within the network, an estimated value of  $C$  thus has to be compared to an appropriate null model to validate whether the value is indeed statistically significant, i.e., whether the respective network indeed exhibits a higher degree of clustering than a corresponding random network. Difficulties also arise for specific types of graphs, such as bipartite graphs, that exhibit a nontrivial clustering coefficient inherent to the bipartite structure [1, 52], see Section 3.5 for a detailed discussion.

An alternative, but equivalent, definition of  $C$  can be given with respect to the number of triads (triples of vertices where each vertex is connected to both others) within a network. Note that the number of edges between the neighbors of a vertex is equal to the number of triads that vertex is part of. The global clustering coefficient is then

x



**Fig. 3.6** The clustering coefficient relates to the local cohesiveness of a network. *a)* The clustering coefficient is defined as the probability that two vertices with a common neighbor are connected. *b)* A highly connected vertex with a low clustering coefficient, indicating a (at least locally) hierarchical structure. *c)* A vertex with high clustering coefficient  $C_{\text{vertex}} = 0.8$

defined as the proportion of triads in a network with respect to the total number of connected triples (triples where at least one vertex is connected to both others).

$$C = \frac{3 \times \text{number of triads}}{\text{number of connected triples}} \quad (3.2)$$

The factor 3 accounts for the fact that each triad contributes to 3 connected triples [1]. A characterization of the clustering coefficient with respect to the number of triads holds some advantages with respect to numerical estimation and can be generalized to other structures, such as the number of squares [31].

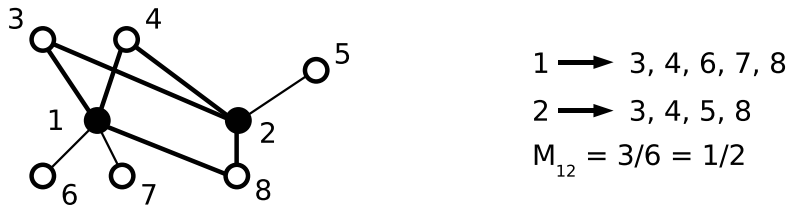
Of particular interest is also the correlation of the clustering coefficient  $C_i$  with other properties of a vertex  $n_i$ . For example, as described by Newman [49], many empirical networks exhibit a negative correlation between the degrees  $k_i$  and the clustering coefficients  $C_i$ , indicating a modular structure of the network. See Fig. 3.5 for an example.

### 3.2.6 The Matching Index

Within many empirical networks, two vertices that are functionally similar do not necessarily have to be connected. For example, within a network of protein interactions, two proteins that are involved in the regulation of similar processes and should be considered as closely related, must not necessarily bind to each other. Correspondingly, the normalized *matching index*  $M_{ij}$  quantifies the 'similarity' between two vertices based on the number of common neighbors shared by two vertices  $n_i$  and  $n_j$ .

$$M_{ij} = \frac{\sum \text{common neighbors}}{\sum \text{total number of neighbors}} = \frac{\sum_{k,l}^N A_{ik} A_{jl}}{k_i + k_j - \sum_{k,l}^N A_{ik} A_{jl}} \quad (3.3)$$

Note that for the measure to be properly normalized, the denominator only counts the number of distinct neighbors, i.e. neighbors that are shared by both vertices are only counted once. One of the virtues of the matching index is that it can



**Fig. 3.7** Vertices that are functionally related do not necessarily have to be connected. The matching index counts number of common neighbors shared by two vertices, normalized by the total number of distinct neighbors. The right panel shows the adjacency list of the vertices  $n_1$  and  $n_2$ , along with the corresponding matching index  $M_{12}$ .

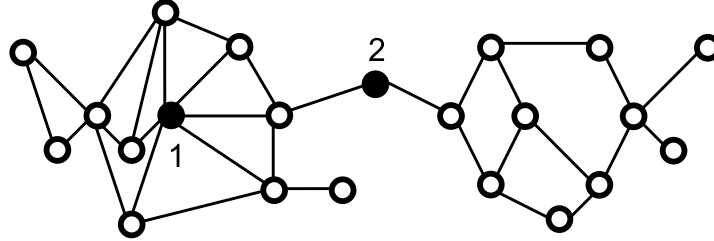
be straightforwardly applied to networks consisting of different types of vertices, such as bipartite graphs. For example, two transcription factors may regulate the expression of similar genes, without necessarily regulating (or binding to) each other. A schematic illustration of the matching index is given in Fig 3.7.

The matching index can be generalized beyond the immediate neighbors of a vertex or extended to multiple vertices [40]. Furthermore, at the most general level, two vertices can be regarded (or defined) as 'similar' if their distance to all other vertices within the network is approximately the same, irrespective of whether they are directly connected or not [75]. An advantage of this definition lies in the fact that the actual pair-wise similarity of two vertices must not be specified. The definition only draws upon the notion that two entities (vertices) must be considered 'similar', if they perceive the rest of the world (here the distance to all other vertices within the network) in a similar way.

### 3.2.7 Network Centralities

Closely related to distance measures, network centrality indices seek to characterize each vertex or edge with respect to their position within the network. Centrality measures will be discussed in more detail in Chapter 4 of this book, here we will only briefly outline some basic features.

Intuitively, a basic measure of the importance of a vertex  $n_i$  is its degree  $k_i$  (degree centrality). And indeed, several studies on biological network report a significant relationship between vertex degree and functional importance of vertices [2]. For example, within protein interaction networks, the removal of highly connected proteins is more likely to have lethal effects than removal of proteins with only a small number of links [32]. However, the degree is clearly not the only determinant of the functional importance of a vertex. Often more relevant, is the contextual location of the vertex within the network. For example, we can ask from which vertex a signal should be sent to reach all other vertices in minimal time. Or, vice versa, which vertices can be reached fastest from any other vertex within the network? In this respect, the *closeness centrality* specifies which vertices have the shortest paths to all



**Fig. 3.8** The degree of a vertex does not necessarily reflect importance with respect to function of a network. While vertex  $n_1$  has a high degree, its removal does not necessarily affect communication within the network. However, removal of vertices with low degree may have significant effects on communication or mass flow within the network, as seen for vertex  $n_2$ .

others, measured for example by the (inverse of the) average distance from a vertex to all other vertices. For detailed definitions see Chapter 4 in this book.

Probably the most well-known centrality measure is the *betweenness centrality* ( $BC$ ). The betweenness centrality can be defined with respect to vertices and edges, and measures how often a vertex or edge is present in the set of all shortest paths. As can be seen in Fig. 3.8, low degree vertices can be crucial to establish communication or mass flow within a network. Thus, with respect to robustness properties of a network, a selective attack on vertices with high  $BC$  was often found to be more relevant than a removal of vertices with high degree. Computationally, the estimation of the betweenness centrality is rather demanding and described in Chapter 4 of this book.

### 3.2.8 Eigenvalues and Spectral Properties of Networks

An important property of network topology are the spectral properties of the adjacency matrix  $\mathbf{A}$ . Though as yet only hardly used in biological research, the spectra of random graphs are among the oldest characteristics of network topology with a plethora of applications in many branches of physics [1].

For an undirected graph, the symmetric adjacency matrix  $\mathbf{A}$  has  $N_V$  real eigenvalues  $\lambda_i$ . The spectral density  $\rho(\lambda)$ ,

$$\rho(\lambda) = \frac{1}{N_V} \sum_{i=1}^{N_V} \delta(\lambda - \lambda_i), \quad (3.4)$$

approaches a continuous function for increasing network size  $N_V \rightarrow \infty$ . An extensive amount of work about the mathematical properties of the spectral density is available, including the famous Wigner semicircle law [21, 1].

Of more relevance to the biological sciences, the eigenvalues of network matrices are becoming increasingly important with respect to two different fields of research: First, in networks of coupled oscillators, i.e., in networks where each vertex corre-

sponds to an oscillator coupled to other oscillators via an adjacency matrix, the global dynamics of the system are determined by the structure of the adjacency matrix. In particular, the stability of the synchronized state, i.e., the state of the network where almost all vertices oscillate synchronously, can be related to the eigenvalues of the *Laplacian* matrix of the network [54], defined in close analogy to the adjacency matrix. Recent studies also take into account the effect of weighted edges [74].

Second, along similar lines, the eigenvalues of network matrices determine the stability and local dynamics of networks composed of interacting elements. For example, the vertices of a metabolic network denote metabolites, whose concentrations change according to the adjacent edges (metabolic reactions). Formally this system is represented by a differential equation for all metabolite concentrations. However, at least locally, this (usually unknown) system of differential equations can be approximated by a weighted interaction matrix, denoted as the Jacobian  $\mathbf{J}$  of the system. The Jacobian matrix already governs essential aspects of the dynamics and predicts specific dynamic behavior even if detailed knowledge about the underlying reactions and interactions is not available [63, 65, 68].

### 3.3 MODELS OF COMPLEX NETWORKS

The various network indices discussed until now characterize and quantify the topological structure of a given network. However, to understand and elucidate whether an estimated value indeed corresponds to nontrivial structure within the network requires to consider basic prototype models of complex networks. We emphasize that none of the models described below aims to mimic the detailed features of any real network. Rather they represent minimal models, each invented to exhibit distinct generic features of complex networks. The purpose of prototype models is twofold: First, they provide null models to understand whether an observed feature is a generic feature of certain network classes or whether it deviates from what could be expected for a simplistic model. Second, prototype models often provide insight on how certain features of complex networks arise from the construction rules of the prototype models, allowing to probe to what extent (for example evolutionary) mechanisms can account for the observed features of empirical networks. Again, more detailed mathematical treatises on random network models are given elsewhere [1, 17, 50], here we only outline the basic ideas.

#### 3.3.1 The Erdős-Rényi Model

Probably the most basic model of a random network is given by the Erdős-Rényi (ER) network [20]. The ER network consists of  $N_V$  vertices, connected by  $N_E$  (undirected) edges which are chosen randomly from the set of  $N_V(N_V - 1)/2$  possible edges (excluding multiple connections and links from a vertex to itself). The probability  $p$  that two randomly chosen vertices are connected is thus  $p = \frac{2N_E}{N_V(N_V-1)}$ . Alternatively, the ER model can be defined as a set of  $N_V$  vertices, with each pair of vertices connected with an equal probability  $p \leq 1$ . The number of edges  $N_E$  is then

a random variable, with the expectation value  $\langle N_E \rangle = pN_V(N_V - 1)/2$  [1].

The ER model has been the primary subject of random graph theory, resulting in extensive knowledge about its mathematical properties and typical features. Here we only summarize some basic properties.

The degree distribution of the ER model is given by a *binomial distribution* that becomes approximately *Poissonian* in the limit of large networks ( $N_V \rightarrow \infty$ ). The probability of a vertex to have degree  $k$  is

$$p(k) \simeq e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad (3.5)$$

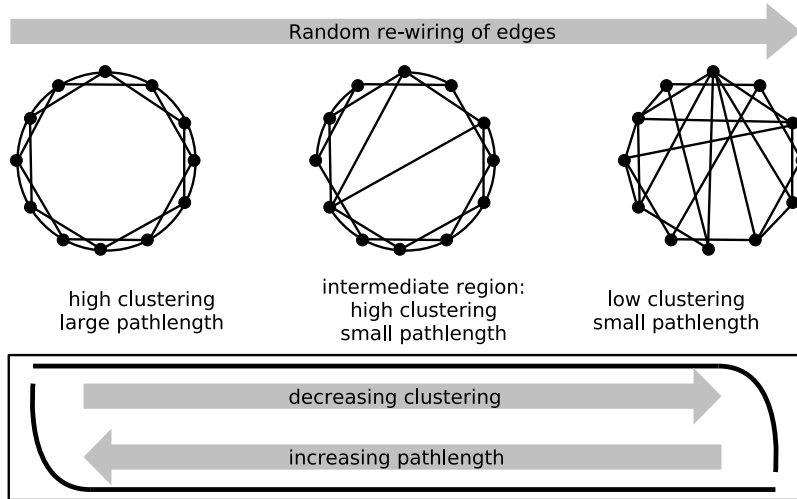
with  $\langle k \rangle = pN_V$  denoting the *average degree*. A typical realization of the ER model is rather homogeneous, most vertices have a similar degree, distributed approximately symmetrically around the average degree  $\langle k \rangle$ , as shown in Figure 3.3b. Most analytical work on the ER model has concentrated on questions related to percolation theory, i.e., the connectedness of the network and the emergence of paths that enable a traversal of the whole network. For small  $p$  the network is disconnected and consists of a large number of isolated components [50]. At  $p \approx 1/N_V$  (thus for average degree  $\langle k \rangle \approx 1$ ) a phase transition occurs, giving rise to a *giant-component* that encompassed most of the vertices of the network. For  $p \geq \log(N_V)/N_V$  all vertices are connected for almost all realizations of the random network. The ER model exhibits the *small-world* property. Above the percolation threshold, the average pathlength is very small and scales as the logarithm of the number of vertices  $l \sim \log N_V$  (with  $\langle k \rangle$  kept constant for increasing number of vertices). By construction, the clustering coefficient of the ER network  $C = p = \langle k \rangle/N_V$ , i.e. the probability that two vertices with a common neighbor are connected equals the probability that any pair of randomly chosen vertices are connected. The ER model does not show any local cohesiveness. Likewise, the degree of connected vertices is uncorrelated, the ER model does not display degree correlations.

The Erdős-Rényi model remains one of the most important prototype models in graph theory. However, the main limitations for a direct comparison of network properties with empirical networks are its homogeneous degree distribution, the absence of local structure and the lack of degree correlations. A close variant of the ER model, the *configuration model*, will be discussed in Section 3.5.

### 3.3.2 The Watts-Strogatz Model

While the Erdős-Rényi model correctly reproduces the *small-world* property, it fails to account for the local clustering that characterizes many empirical networks. In particular for social networks, i.e. networks of mutual friendships or acquaintances, most studies indicate a clustering coefficient that is orders of magnitude higher than the value obtained for a corresponding ER network.

In one of the seminal papers of complex network theory, Watts and Strogatz proposed a model for coexistence of local structure on the one hand, and a small average pathlength on the other hand [72]. The starting point of the model is the limiting case of

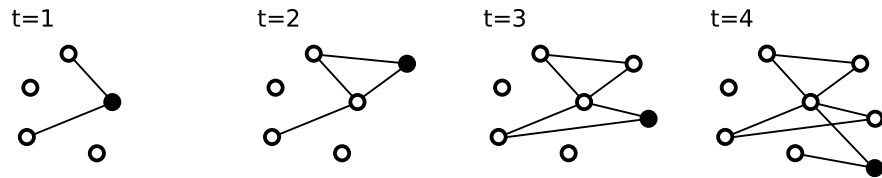


**Fig. 3.9** The Watts-Strogatz model: Starting point is a regular network, constructed such that each vertex is connected to its two nearest neighbors, resulting in a maximal clustering coefficient  $C = 1$ . With probability  $p_{rew}$  links are randomly rewired. In the limit  $p_{rew} \rightarrow 1$  the ER model is recovered.

a regular lattice-like network: Each vertex (arranged on an one-dimensional ring in the original model) is connected to its  $n/2$  nearest neighbors. In social terms, this would resemble a strictly local medieval-like world, where each person only knows people in his or her immediate vicinity, such as neighbors and people in nearby villages. Consequently, the model exhibits strong local cohesiveness (a high clustering coefficient), but the spread of information is slow, i.e. the average pathlength scales linearly with system size.

Extending the regular lattice-like network, shortcuts between distant vertices are introduced, i.e. with a probability  $p_{rew}$  a link is rewired, such that one end is detached from its original vertex and connected to a randomly chosen vertex. In social terms, this would correspond to a merchant or traveler, who is also acquainted to a small number of more distant people within the country. As the probability  $p_{rew}$  increases and more links are rewired, the model approaches a random network of the ER type. In the limit  $p_{rew} \rightarrow 1$  the ER model is recovered. The network thus again exhibits no local structure (small clustering coefficient) and the average pathlength scales as the logarithm of network size.

One of the intriguing result of the WS model is that already a very small number of shortcuts ( $p_{rew} \ll 1$ ) is sufficient to rapidly decrease the average pathlength [28]. On the other hand, for small  $p_{rew}$ , the local clustering remains almost unaffected and the clustering coefficient only decreased significantly for  $p_{rew} \rightarrow 1$ . Thus for an intermediate region of  $p_{rew}$ , the WS model exhibits a coexistence of high local clustering and short average pathlength (small-world property), as also observed in many empirical networks. A schematic representation of the WS model is given in



**Fig. 3.10** The Barabási-Albert model [7]: Starting with an initial small network, consisting of  $N_0$  unconnected vertices, a new vertex is introduced at each timestep and connected with  $m < N_0$  edges (here shown with  $m = 2$ ).

Fig. 3.9.

The main significance of the WS model results from the fact that it emphasizes a difference between local and global properties of networks. The clustering coefficient, a local property, is determined by the immediate neighborhood of a vertex and is almost unaffected by the introduction of additional 'shortcuts' within the network. On the other hand, the average pathlength, a global property, rapidly decreases upon the introduction of just a few shortcuts. This has, for example, profound implications on the spread of infectious diseases across continents. A change in average pathlength to distant vertices is not detectable at the local level, i.e., your social neighborhood might remain almost unaltered, while the 'distance' (in network terms) to infected persons can rapidly decrease with only a small number of transcontinental travellers. However, apart from the coexistence of high local clustering and short average pathlength, the WS models captures almost no other feature found in empirical networks. Its importance as a null model for biological networks thus remains limited.

### 3.3.3 The Barabási-Albert Model

Among the most important limitations of the models discussed above is that neither captures or accounts for the inhomogeneous degree distribution found in many empirical networks. To this end, Barabási and Albert [7] proposed a simple network model that gives rise to a scale-free degree distribution and still provides the conceptual basis for most current network models described in the literature.

Closely related to (and actually a simplification of) an earlier model by Price [4, 16, 45, 50], the BA model is based on two essential ingredients: *i*) Growth: In contrast to the models discussed above, the BA model does not assume that the number of vertices within the network is fixed. Mimicking the dynamics of many real networks, vertices are continuously added and the network grows as a function of time. *ii*) Preferential attachment: New edges are not introduced randomly, but the probability that a vertex receives a new edge depends on its present degree  $k_i$ , again reflecting dynamic properties of real networks.

The growth process is organized as follows: Starting with an initial small network, consisting of  $N_0$  unconnected vertices, a new vertex is introduced at each timestep. The new vertex is connected to with  $m \leq N_0$  edges to the already present vertices. The probability  $\rho(n_i)$  that an already present vertex  $n_i$  receives a new edge is

proportional to its degree  $k_i$ :

$$\rho(n_i) = \frac{k_i}{\sum_j k_j} . \quad (3.6)$$

After  $t$  timesteps, the network consists of  $N_V(t) = N_0 + t$  vertices, connected by  $N_E(t) = mt$  edges. Due to the preferential attachment mechanism, older vertices tend to have accumulated more links, and thus have an even higher probability to receive yet more links (a 'rich-get-richer' dynamics). Likewise, new vertices only have a small number of links, and thus a low probability of receiving additional links. A schematic illustration of the growth process is given in Fig. 3.10.

In the long time limit  $t \gg 1$ , the BA model exhibits a scale-free degree distribution  $p(k) \sim k^{-\gamma_{\text{BA}}}$  with a degree exponent  $\gamma_{\text{BA}} = 3$  that is invariant with time. The degree exponent is independent of the free parameters  $m$  and  $N_0$ . The BA model captures the small world property, BA networks are found to have shorter average pathlength than ER and WS models of the same size and density. The degrees are uncorrelated and analytical estimates of the clustering coefficient are available [36]. One of the merits of the BA model is that it provides a possible mechanism to explain the observed scale-free distribution of many empirical networks. Indeed, the time evolution of many empirical networks is governed by preferential attachment-like processes. For example, within a social network, people (vertices) with already many friends (edges) are more likely to acquire new friends, as compared to people with few edges. Likewise, already famous actors will obtain more offers to act in a new movie than young unknown actors. Scientific papers that are already frequently cited are more likely to be read and cited again than less frequently cited papers.

Importantly, the preferential attachment rule also provides several testable predictions for complex networks. For example, if metabolic networks are reported as scale-free, then, according to this growth rule, highly connected metabolites should have an early evolutionary origin. Indeed, as emphasized by Wagner and Fell [71], many of the highly connected metabolites, mainly intermediates of the TCA cycle and glycolysis, as well as some ubiquitous co-factors, are among the evolutionary oldest.

However, explanations in terms of evolutionary mechanisms also hold some pitfalls which are unfortunately rarely – if ever – discussed in the literature. Most importantly, not only the formation of a network itself, but often also the acquisition of data about the network is governed by similar mechanisms. For example, minor movies with famous actors are more likely to be included in the respective databases than local movies starring only unknown actors. Likewise, putative biochemical regulations or reactions adjacent to the TCA cycle are more likely to be investigated, and thus reported in publications, than putative regulations within the outskirts of metabolism. In this sense, an observed feature of an empirical network might also always reflect properties of the data acquisition process, rather than genuine properties of the network itself.

### 3.3.4 Extensions of the BA Model

The BA model constitutes the conceptual basis for a large variety of extensions and modifications and has triggered an exceptional amount of further work in the complex network models. Most extensions can roughly be subdivided into two (though often overlapping) categories: *i)* Modifications that aim to generate networks with specific tunable features, such as different degree exponents, tunable cluster coefficients or degree correlations. *ii)* Modifications that aim to mimic the evolutionary growth processes of specific networks in more detail, such as aging in social networks or capacity restrictions in transportation networks.

For example, the exponential cutoff at high degrees observed in many real networks can be accounted for by aging of vertices, i.e., vertices that have been present for a given time  $T$  stop acquiring new edges or are removed from the network [3] – as could be expected in social networks. Similar, an airport within a transportation network will not acquire new connections beyond a certain capacity, again resulting in an exponential cutoff at high degrees. Other processes to modify the properties of the network include re-wiring of edges according to defined rules. For example, within a social network people that have a common friend are more likely to become acquainted themselves, resulting in an increased local clustering of the network [15]. Further extensions and modifications include memory effects and high clustering [36], degree correlations [56, 73], tunable degree exponents [38], information accessibility [48], among many more. An overview of early modifications and extensions of the original BA model can also be found in Table III of [1].

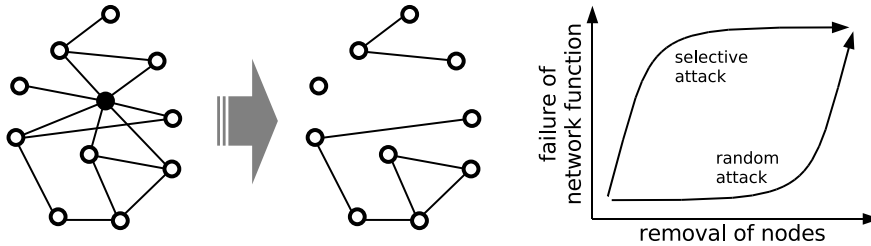
## 3.4 ADDITIONAL PROPERTIES OF COMPLEX NETWORKS

Within the first section, most emphasis was placed on quantitative measures that describe the properties of individual vertices and edges. However, complex networks are also characterized by emergent features that transcend the properties of individual vertices and relate to the organization of the network as a whole. In the following the basic emergent global properties of complex networks, such as robustness or modularity, are outlined.

### 3.4.1 Structural Robustness and Attack Tolerance

Most biological systems share a common feature: robustness [8, 35, 60, 69]. Constituting one of the fundamental organizing principles of biology, cellular networks must be able to maintain their function in the face of constant perturbations and fluctuations that affect the internal or external parameters of the system. In the context of complex network analysis, robustness is mainly understood as the persistence of topological network properties, such as average pathlength or connectedness, upon removal of vertices or links [1, 2, 10, 33].

Indeed, most empirical networks show a surprising tolerance against removal of vertices. Focusing on topological aspects of robustness only, a number of studies



**Fig. 3.11** The 'robust, yet fragile' nature of scale-free networks. Properties of scale-free networks are highly robust against random removal of vertices, but vulnerable against selective intentional removal of vertices.

revealed significant differences between distinct network topologies upon removal of vertices or edges [2, 29]. In general, we have to distinguish between random and intentional attacks on network topology. While for ER networks, due to the homogeneity of vertex properties, the response to random and intentional attacks is roughly similar, the situation for scale-free networks is markedly different. Most properties of scale-free networks were found to be exceptionally robust against random removal of vertices. However, at the same time, scale-free networks are vulnerable with respect to intentional attacks. This difference is due to the heterogeneous degree distribution. Low degree vertices are far more frequent than high degree vertices, but only play a minor role in overall network topology. While random attacks will most likely affect low degree vertices, a selective attack on high degree vertices has far more dramatic consequence on global network indices [2, 1, 29]. A schematic illustration is given in Fig. 3.11.

In general, the difference between random and intentional attacks is at the core of most current research on network robustness. The 'robust, yet fragile' nature of complex systems refers to the fact that many complex systems are robust against random attacks, while they remain fragile against selective attacks. In particular highly optimized systems are extremely robust against anticipated attacks, while optimization concomitantly leads to vulnerability against unanticipated perturbations, related to the principle of 'highly optimized tolerance' (HOT) [12].

It should be noted though, that the restriction to topological aspects of robustness only allows for a rather restricted view on network robustness. Dynamic aspects of functional robustness thus receive increasing interest recently [35, 47, 60, 63, 65].

### 3.4.2 Modularity, Community Structures and Hierarchies

Related to the idea of functional robustness is the notion of modules and community structures within complex networks. In general, it is assumed that many complex networks are built up from (interacting and possibly overlapping) *modules* or *communities*. The detection of such community structures has attracted substantial interest recently and defines an important aspect of complex network analysis [57, 75, 76].

Unfortunately, there is no generally accepted definition of what constitutes a module or community structure within a complex network. As a working definition, modules or communities are often understood as subsets of vertices that are densely connected among each other, but are only sparsely connected to other vertices outside the community. In this respect, probably the most influential definition of network modularity comes from the work of Newman and Girvan, who define a score function for network modularity based on this principle [24, 51].

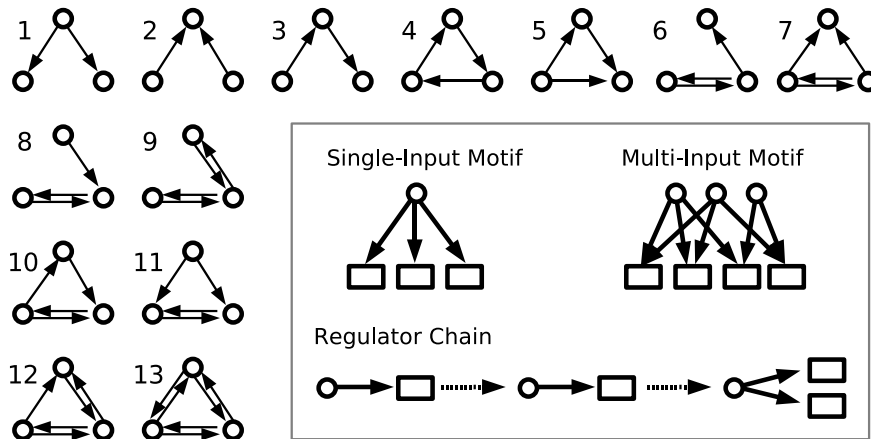
Computationally, the detection of community structures is closely related to the clustering of elements, i.e., the assignment of 'entities' into distinct categories based on a suitably defined notion of similarity. In the context of complex network analysis, similarity of vertices can be defined in different manners, for example with respect to the shortest path between two vertices, the total number of paths between vertices (weighted according to pathlength), among several other possibilities. Likewise, and again resembling the situation in clustering algorithms, modules can be detected using agglomerative or divisive methods, i.e. by grouping similar vertices into larger units or by iteratively breaking down a complex network into smaller and smaller subsets. Although there is a plethora of proposed novel algorithms specifically designed to detect modules within complex networks, it should be noted that a recent study [26, 27] demonstrates that standard ways of clustering data, e.g. based on k-means or hierarchical clustering, often outperform newer methods.

The relevance of modularity in complex networks can be seen from two different perspectives: First, modularity represent a inherent design principle of many complex networks. Indeed, in particular for biological networks shaped by evolution, it seems plausible that modular networks have advantages in terms of their ability to adjust to new evolutionary constraints by incorporating new functionality within the network [34, 26]. Second, a decomposition into modules is often helpful to understand the functional organization of complex networks. For example, many technological networks, such as integrated electronic circuits, can not be rationalized on the basis of individual resistors or transistors. Only by the construction of functional modules, such as shift registers, counters, gates, inverters, each consisting of many lower-level elements, an effective manipulation and design can be achieved.

As already implicit in the example of an integrated electronic circuit, modules are not restricted to one specific level of representation. Rather, modules may overlap and consist of smaller modules, giving rise to a hierarchical organization of modularity within complex networks [14, 57, 70, 58]

### 3.4.3 Subgraphs and Motifs in Networks

Closely related to the detection of community structures or functional modules is the notion of *motifs* as building blocks of complex networks [31, 39, 43, 44, 59]. Providing a bridge between local vertex-related properties and global functional properties of networks, the basic idea of network motifs is that large complex networks are essentially composed of small interlinked subgraphs. Similar to the notion of sequence motifs, we can thus look for re-occurring patterns within the network, i.e., small sets of vertices that exhibit an identical local topological structure. Indeed, as



**Fig. 3.12** All 13 possible three-vertex motifs, according to [44, 59]. *Inset:* The concept of motifs can be generalized to account for specific repeated subgraphs within a complex network. Shown are possible motifs of the yeast transcription regulatory network, with circles denoting regulators and rectangles denoting gene promoters, according to [39].

shown in previous studies [44], the transcriptional interaction network of *Escherichia coli* is essentially composed of repeated appearances of three highly overrepresented network motifs. Furthermore, similar studies reveal that the distribution of motifs is characteristic for certain classes of networks, i.e., networks with similar overall functionality (such as communication networks or food-webs) also exhibit a highly similar motif distribution [43]. Figure 3.12 shows all possible directed subgraphs composed of three vertices. Unfortunately, the systematic enumeration of network motifs is computationally demanding. A more detailed description of network motifs, also accounting for computational aspects, is given in later chapters of this book.

In general, the concept of network motifs is not restricted to subgraphs with a fixed number of vertices. Rather, it allows to look for any re-occurring patterns, including more complicated topological structures, such as multi-input motifs, regulator chains, or dense overlapping regulons (DORs) [59, 39]. The most intriguing aspect of network motifs, however, results from the asserted connection of local topological structure to dynamic function of a network motif. While such a correspondence between topology and function, i.e., that each motif has also a specific 'hard-wired' function, was claimed in several early studies on network motifs [44, 59], later results indicate that this is not the case. Though there is increasing interest in the analysis of dynamic aspects of complex networks, the precise relationship between structure and function of complex networks is still largely unclear [37, 30, 55, 68, 63]. Additional aspects in the interpretation of motif distributions, including possible pitfalls in their statistical estimation, are discussed in the next section.

### 3.5 STATISTICAL TESTING OF NETWORK PROPERTIES

The most crucial, and probably most widely underestimated aspect of complex network analysis is the statistical testing of network properties. As yet, all network indices were described as numerical quantities that can usually be straightforwardly estimated from any given network topology. However, in most applications, network indices are also associated with a biological meaning or interpretation – we seek to uncover those features of the network that are characteristic of the underlying system. Thus, for example in the case of a metabolic network, the question is not whether the clustering coefficient takes a specific numerical value, but rather whether this value distinguishes the network from other networks of similar size, i.e. whether the metabolic network can be regarded as 'highly clustered'. Only in the latter case, i.e. if the clustering coefficient indeed deviates from what could be expected for networks of similar size, it represents a distinctive feature of the respective network. But then, how should such a deviation be detected or quantified? What values of clustering coefficient should be considered 'usual' or 'typical' for a network of given size?

One answer to these questions, in addition to the prototype models described in the last section, lies in the formulation of appropriately randomized null-models of complex networks. We create an ensemble of surrogate networks, usually of identical size and density, and compare the values of network indices obtained from the empirical network to those obtained from the ensemble of randomized surrogate networks.

More general, statistical testing always means to set up a null hypothesis, i.e., a process or mechanism that is assumed to account for an observed feature, and a subsequent test whether the observed feature is actually consistent with the null-hypothesis. In the context of complex network analysis, a possible null-hypothesis is, for example, that an observed clustering coefficient is consistent with values arising from Erdős-Rényi networks of the similar size. Given a certain probability threshold, the null-hypothesis is rejected if the probability to actually find the observed clustering coefficient for Erdős-Rényi networks is below the defined threshold. In this case, the deviation of the clustering coefficient with respect to the null-hypothesis is *significant*.

However, apart from some straightforward cases, the statistical testing of network properties holds several potential pitfalls and possible sources of misinterpretations. In the following, we briefly outline some of the most widely used null-models for complex network analysis and point out possible ambiguities in their interpretation.

#### 3.5.1 Generating Networks and Nullmodels

The most basic null-model is a network of identical size (number of vertices and edges) but lacking any other internal structure. Conceptually equivalent to an Erdős-Rényi network of the same size, such an ensemble can be constructed by randomly rewiring links within the network - as already done in the construction of the Watts-Strogatz model.

Usually more appropriate, however, is to preserve the degree distribution of the empirical network. An ensemble of randomized surrogate networks with preserved



**Fig. 3.13** Generating random networks with preserved degree distribution: At each iteration, two edges ( $a \rightarrow b$ ) and ( $c \rightarrow d$ ) are selected at random and rewired as ( $a \rightarrow d$ ) and ( $c \rightarrow b$ ), provided that the respective edges do not already exist. Note that for undirected networks, there are two possible ways to rewire the links.

degree distribution is obtained by iteratively swapping randomly selected edges, as schematically depicted in Fig. 3.13: For a directed network, at each iteration two edges ( $a \rightarrow b$ ) and ( $c \rightarrow d$ ) are selected at random and rewired as ( $a \rightarrow d$ ) and ( $c \rightarrow b$ ), provided that the respective edges do not already exist. Repeating this sufficient times, i.e. such that most edges have a statistical chance to be selected, the resulting network has a preserved degree distribution, but lacks any other internal structure of the initial empirical network. The approach can be generalized to account for other features of complex networks. For example, the analysis of network motifs [44, 59] makes use of a similar approach to generate networks with a preserved three-vertex motif distribution, swapping two edges if and only if the resulting motif distribution is preserved. Closely related to network randomization is the *configuration model* of complex networks: To construct a network with a specified degree distribution, each vertex is assigned a number  $k_i$  of adjacent edges, such that the total number of assigned edges is even. Subsequently, pairs of the, as yet unconnected, 'stubs' are randomly chosen and connected [52, 59].

In any case, a network index  $Q$  of interest is subsequently compared against the values found for the ensemble of surrogate networks, and can be evaluated according to a significance score

$$S = \left| \frac{Q^{\text{network}} - \langle Q^{\text{surrogate}} \rangle}{\sigma_{\text{surrogate}}} \right|, \quad (3.7)$$

where  $\langle Q^{\text{surrogate}} \rangle$  denotes the average found within the ensemble of surrogate networks and  $\sigma_{\text{surrogate}}$  the standard deviation. Unfortunately, the construction of randomized networks that preserve features other than the degree or motif distribution is far from straightforward.

### 3.5.2 The Conceptualization of Cellular Networks

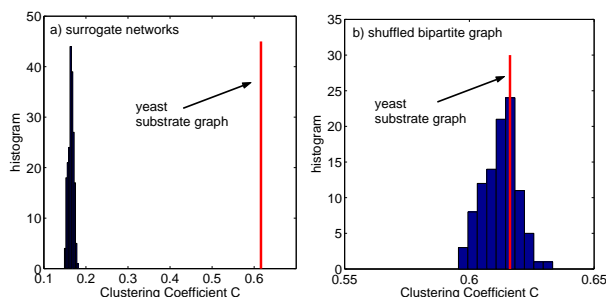
The most difficult problem of complex network analysis is often the choice of an appropriate null-model or null-hypothesis. By definition, all statistical tests rely on the definition of a null-model, and any notion of 'significance' is only defined with respect to this null-model. Thus, if the null-model is erroneous or trivial, so will be any result obtained from an evaluation of the significance of network properties

against this null-model. In particular, the choice of the null-model is additionally complicated by the fact that networks are usually abstract representations of more complex biological processes. While this 'reduction in complexity' is often necessary to make biological questions mathematically tractable, it also holds the temptation to neglect properties of the underlying system – resulting in an erroneous or misleading interpretation of network properties.

An illustrative example of such a case was discussed in the context of a recent study of motif distributions in complex networks [43]. Therein, the significance profile of small subgraphs (motifs) within a neuronal network was evaluated and compared to simple degree-preserving randomized networks. The study concluded that the neural information processing networks exhibits a highly characteristic significance profile for its motif distribution, suggesting evolutionary mechanisms that result in key circuit elements to perform specific tasks. However, as pointed out later [5], a neural network, i.e., a network of neurons connected by synapses, is not just a network of vertices connected by edges. Rather, neurons have a spatial position and a tendency to form local clusters, hence neighboring neurons have greater chance of forming connections than distant neurons. As the spatial properties are not reflected in the null hypothesis, the statistical test misclassifies a completely random but spatially clustered network as one that is nonrandom and exhibits significant network motifs. Indeed, a simple toy model that preserves the spatial position of neurons and connects neurons preferentially to nearby neurons is able to reproduce an almost identical significant profile for the motif distribution, without the need to invoke any evolutionary mechanisms to select for specific functional tasks [5].

The example serves to illustrate two important aspects of statistical testing of network properties: *i*) The observed motif distribution within the neural information processing network was indeed highly significant, as compared to degree-preserving random networks of the same size – there was no computational error involved. The significance profile of motifs is thus indeed a true characteristic of the system. *ii*) However, the significance profile of motifs tests against a trivial null-hypothesis: the assumption of a completely random network. Strictly speaking, rejecting this null-hypothesis only trivially proves that neural networks are not random. However, only little can be learned about the possible biological function of network motifs from the sole fact that their distribution is not random. As exemplified here for the neural network, the significant deviation from random networks is most likely a simple and straightforward consequence of the (neglected) spatial structure of the system.

Thus, as a general rule, network properties that are found to be significant with respect to simple randomized networks must not necessarily be important with respect to function. Even though the statistical estimation of significance might be technically correct, significance here only implies deviations from randomness – which is often a trivial consequence of the underlying process or system itself. Two important classes of complex networks where the construction of the networks itself implies deviations from randomness, and thus implies highly significant network properties, are discussed below.



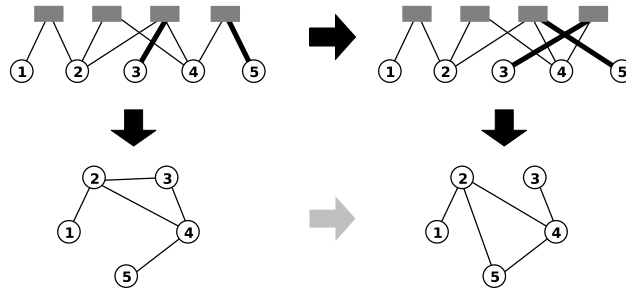
**Fig. 3.14** The clustering coefficient of the *S. cerevisiae* substrate graph, depicted in Figure 3.1. *Left plot:* The clustering coefficient is compared against randomized networks with preserved degree distribution. The yeast substrate graph shows a highly significant clustering, as compared to randomized networks generated according to the scheme shown in Figure 3.13. *Right plot:* The clustering coefficient is compared against randomized networks that preserve the original bipartite structure. In this case, no significant deviation between the clustering coefficient of the yeast substrate graph and its shuffled counterparts are detected.

### 3.5.3 Bipartite Graphs

A large number of complex networks in biology and other fields are derived from bipartite graphs. A prominent example is the metabolic network, consisting of two distinct types of vertices: substrates and reactions. Likewise, one of the benchmark examples of complex network analysis, the movie-actor network, originates from a bipartite graph: Based on a public movie database ([www.imdb.org](http://www.imdb.org)), two actors are connected by an edge if they have performed in a movie together. The movie-actor network often serves as a prototypical (and computationally much better accessible) example of a social acquaintance network. In particular, the coexistence of high local clustering and small average pathlength, typical for social networks, has been demonstrated using the movie-actor network as a paradigmatic example [72]. Comparing the observed clustering coefficient of the movie-actor network to Erdős-Rényi networks of the same size, indeed reveals a highly significant clustering that is orders of magnitudes higher than for corresponding random networks.

However, by construction, the movie-actor network is composed of fully-connected subnetworks only: All actors performing in one movie are mutually connected and form a fully-connected subgraph. Since an actor participates in several movies, the cliques for each movie are then joined together and organized into the final movie-actor network. Thus, by construction, the underlying bipartite structure clearly distinguishes the network from actual social acquaintance networks and already implies a 'significant' local structure.

A similar reasoning holds for metabolic networks: Using the same scheme as previous studies [71], the substrate-graph of a metabolic network is constructed by connecting two metabolites if they share a common reaction. The resulting network, analyzed in a number of recent studies [22, 33, 71, 41], shows a significantly higher clustering coefficient than could be expected for random networks with preserved



**Fig. 3.15** To preserve the underlying bipartite structure of a network, a modified randomization scheme is proposed. Using the original bipartite graph, two edges are randomly selected and rewired. Note that this leads to nontrivial modifications of the substrate graph projection.

degree distribution. Repeating the analysis for the *S. cerevisiae* substrate graph, already used in the previous sections and depicted in Figure 3.1, confirms these results: The observed clustering coefficient of the metabolic network significantly deviates from the values obtained for the ensemble of random networks with preserved degree distributions, as shown in Figure 3.14 (left plot). However, does this result really imply a significant clustering of the network beyond the fact that it was derived from a bipartite graph? To test this assertion, a scheme that generates random instances of the original network, but always retains the underlying bipartite structure, was proposed [61]. The randomization scheme is shown in Figure 3.15. Note that the randomization of the bipartite graph leads to nontrivial modifications when only the projection onto the substrate graph is considered. Again comparing the resulting clustering coefficients of the randomized networks to the clustering coefficient obtained from the yeast metabolic network, reveals no significant deviation. Thus we have to conclude that the metabolic network of *S. cerevisiae* considered here shows no significant clustering beyond the trivial clustering imposed by the construction rule of the network - namely that it was constructed as a projection from a bipartite graph.

### 3.5.4 Correlation Networks

A similar reasoning also holds for *correlation networks*, representing an increasingly popular way to investigate and visualize large-scale empirical data [11, 19, 42, 46, 62, 67]. A correlation network is constructed by connecting two vertices, such as metabolites or proteins, if their respective concentration is correlated over a suitable set of experiments. That is, the correlation matrix  $\mathbf{C}$  is converted into a binary (adjacency) matrix  $\mathbf{A}$ , with elements  $A_{ij} = 1$  whenever the correlation between two elements  $n_i$  and  $n_j$  exceeds a given threshold, usually defined with respect to a given significance level.

However, networks that are constructed from correlation matrices possess, by con-

struction, a nontrivial structure and are highly clustered. In particular, correlations are transitive, i.e. if an entity  $A$  correlates with two entities  $B$  and  $C$ , then  $B$  and  $C$  are also (likely to be) correlated themselves. Formally, this transitivity can be expressed as an inequality that holds between any triplet of correlation coefficients [61]. To test for the (nontrivial) structure of correlation networks thus implies to generate networks that preserve all features inherent to correlation matrices, but are random otherwise. Specifically, randomized networks with preserved degree distributions, generated according the scheme shown in Figure 3.13, violate the inherent structure of correlation networks. If recent studies [19] thus report an 'unsuspected assortative feature' for biological correlation networks, this is most probably a result of an inappropriate null-hypothesis, rather than a distinctive feature of the process in question. However, no satisfying solution to the problem of constructing appropriate null-models for correlation networks is known. The interpretation of correlation networks is discussed in more detail in [62].

These examples serve to illustrate that a proper interpretation of network properties, even when tested for significance against randomized networks, is sometimes far from straightforward. Unfortunately, to what extent the underlying structure of the original network trivially determines its topological features, is often not as obvious as in the case of bipartite graphs or correlation networks.

### 3.6 SUMMARY

The analysis of complex networks will continue to play a major role in many scientific disciplines. In particular the analysis of large systems, consisting of many interacting elements, often necessitates an abstract representation of the system and its interactions in terms of a complex network. In respect thereof, the theory of complex networks, emanating from the mathematical field of graph theory, has attracted renewed interest over the last decade. While many concepts of complex network theory have their roots in graph theory, novel problems have also triggered the development of a variety of novel concepts, shedding light on the topological organization and design principles of complex networks. In this chapter, we have provided a brief overview over some typical properties and concepts used in the analysis of large complex networks. Though certainly no comprehensive review, we have focused mainly on the best known indices to characterize the topological structure of complex networks, such as distance and average pathlength, the degree distribution, the clustering coefficient, assortative mixing and network centralities. Several of the indices and concepts described in this chapter will play an important role in the remainder of this book and are discussed in more detail in later chapters. However, complex network analysis does not stop at the description and numerical estimation of individual indices characterizing the topological structure. An important aspect of complex network theory is also the development of paradigmatic prototype models that elucidate and explain how topological features arise from the construction rules and design principles of complex network. Of further interest are also global features of networks that result from the local properties of vertices and

often relate to the function of a network as a whole. In this respect, we have briefly discussed concepts of network functionality and organization, such as robustness, attack tolerance, modularity and hierarchies within complex networks. One of the most important aspects of complex network analysis, however, is the statistical testing of network properties. Any interpretation of the functional relevance of topological structures is only valid as far as it has been properly tested and validated against appropriate null-models. Not always receiving the attention it deserves, a sound validation of network properties is far from trivial. As demonstrated in this chapter, an inappropriate choice of null-models will lead to 'significant' results, even when the network property in question is actually a simple and straightforward consequence of the way the network is constructed. In this sense, the great simplification achieved by the representation of a complex system as a network, concomitantly gives rise to a large number of potential pitfalls and misinterpretations in the analysis. One should always keep in mind that complex biological processes, such as protein binding or cellular metabolism, are not actually networks. Rather, they are physical or biological systems that can be represented as complex networks – making biological problems mathematically tractable, but only at the cost of also omitting some aspects of their entirety.

### 3.7 EXERCISES

1. Name several typical characteristics of vertices and describe how they can be estimated. In particular, give examples for vertex properties that can be estimated using only information about the vertex and its neighbors and properties that relate to the topology of the network as a whole.
2. Give examples of typical degree distributions. How is the degree exponent defined and how can it be estimated from a given empirical network.
3. Why does a negative correlation between vertex degree and clustering coefficient indicate a hierarchical structure?
4. Are functionally related vertices always connected? If not, what are alternative approaches to define 'relatedness' of vertices within a complex network?
5. What are 'small-world' networks? Is the Erdős-Rényi model a 'small-world' network?
6. What is the average clustering coefficient of Erdős-Rényi random networks?
7. Outline how the Barabási-Albert network model is constructed. What are the main differences to Erdős-Rényi and Watts-Strogatz network models?
8. Outline how an ensemble of random networks can be constructed that preserves the degree distribution of a given empirical network.

9. Outline a scheme to randomize a network, such that the three-vertex motif distribution is preserved.
10. What does the term 'significant' mean in the context of statistical validation of network properties?



## References

1. R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.
2. R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378 – 382, 2000.
3. L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci. (U. S. A.)*, 2000.
4. M. Arita. Scale-freeness and biological networks. *J. Biochem.*, 138(1):1–4, 2005.
5. Y. Artzy-Randrup, S. J. Fleishman, N. Ben-Tal, and L. Stone. Comment on "network motifs: Simple building blocks of complex networks" and "superfamilies of evolved and designed networks". *Science*, 305(5687):1107c, 2004. doi: 10.1126/science.1099334.
6. J. Bang-Jensen and G. Gutin. *Digraphs: Theory, Algorithms and Applications*. Springer Monographs in Mathematics. Springer-Verlag, London, 2000.
7. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, October 1999.
8. A.-L. Barabási and Z. N. Oltvai. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004.
9. V. Batagelj and A. Mrvar. Pajek - program for large network analysis. *Connections*, 21(2):47–57, 1998.
10. D. S. Callaway, M. E. J. Newman, S. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85(25):5468–5471, 2000.
11. D. Camacho, A. de la Fuente, and P. Mendes. The origin of correlations in metabolomics data. *Metabolomics*, 1:53–63, 2005.
12. J. M. Carlson and J. Doyle. Highly optimized tolerance: Robustness and design in complex systems. *Phys. Rev. Lett.*, 84(11):2529–2532, 2000.

13. R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Phys. Rev. Lett.* 90, 058701 (2003) [4 pages], 90:058701, 2003.
14. Luciano da F. Costa. The hierarchical backbone of complex networks. *Phys. Rev. Lett.*, 93:098702, 2004.
15. Jörn Davidsen, Holger Ebel, and Stefan Bornholdt. Emergence of a small world from local interactions: Modeling acquaintance networks. *Phys. Rev. Lett.*, 88(12):128701, 2002.
16. Derek J. de Solla Price. Network of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science*, 149:510–515, 1965.
17. S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51(4):1079 – 1187, 2002.
18. S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E*, 63(6):062101, 2001.
19. V. M. Eguiluz, D. R. Chialvo, G. Cecchi, M. Baliki, and A. V. Apkarian. Scale-free brain functional networks. *Phys. Rev. Lett*, 94:018102, 2005. arXiv:cond-mat/0309092v2.
20. P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.
21. Illés J. Farkas, Imre Derényi, Albert-László Barabási, and Tamás Vicsek. Spectra of 'real-world' graphs: Beyond the semicircle law. *Phys. Rev. E*, 64(2):026704, 2001.
22. D. A. Fell and A. Wagner. The small world of metabolism. *Nature Biotechnology*, 18:1121–1122, 2000.
23. J. Förster, I. Famili, P. Fu, B. O. Palsson, and J. Nielsen. Genome-scale reconstruction of the *saccharomyces cerevisiae* metabolic network. *Genome Research*, 13:244–253, 2003.
24. M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Nat. Acad. Sci. USA*, 99(12):7821–7826, 2002.
25. M. L. Goldstein, S. A. Morris, and G. G. Yen. Problems with fitting to the power-law distribution. *The European Physical Journal B*, 41(2):255–258, 2004. DOI:10.1140/epjb/e2004-00316-5.
26. M. Gustafsson. *Large-scale topology, stability and biology of gene networks*, volume 1256 of *Linköping Studies in Science and Technology Thesis*. Linköpings universitet, Norrköpping, 2006.

27. M. Gustafsson, M. Hörnquist, and A. Lombardi. Comparison and validation of community structures in complex networks. *Physica A*, 367:559–576, 2006.
28. H. Herzel. How to quantify 'small-world networks'? *Fractals*, 6(4):301–303, 1998.
29. P. Holme, B. J. Kim, C N. Yoon, and S K. Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65:056109, 2002.
30. P. J. Ingram, M. P. H. Stumpf, and J. Stark. Network motifs: structure does not determine function. *BMC Genomics*, 7:108, 2006.
31. S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, and U. Alon. Subgraphs in random networks. *Phys. Rev. E*, 68(2):026127, 2003.
32. H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature (Brief Communications)*, 411:41–42, 2001. doi:10.1038/3507513.
33. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic network. *Nature*, 407:651–654, 2000.
34. N. Kashtan and U. Alon. Spontaneous evolution of modularity and network motifs. *Proc. Natl. Acad. Sci. USA*, 102(39):13773–13778, 2005.
35. H. Kitano. Biological robustness. *Nat. Rev. Genetics*, 5:826–837, 2004.
36. K. Klemm and V. M. Eguíluz. Highly clustered scale-free networks. *Phys. Rev. E*, 65:036123, 2002.
37. E. Klipp, W. Liebermeister, and C. Wierling. Inferring dynamic properties of biochemical reaction networks from structural knowledge. *Genome Informatics Series*, 15(1):125–137, 2004.
38. P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85(21):4629–4632, 2000.
39. T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298:799–804, 2002.
40. A. Li and S. Horvath. Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics 2007* 23(2):222-231; doi:10.1093/bioinformatics/btl581, 23(2):222–231, 2007.
41. H. Ma and An-Ping Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 19(2):270–277, 2003.

42. P. Mendes, D. Camacho, and A. de la Fuente. Modelling and simulations for metabolomics data analysis. *Biochem. Soc. Trans.*, 33(6):1427–1429, 2005.
43. R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004. doi: 10.1126/science.1089167.
44. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002.
45. M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2004.
46. K. Morgenthal, W. Weckwerth, and R. Steuer. Metabolomic networks in plants: Transitions from pattern recognition to biological interpretation. *BioSystems*, 83(2-3):108–117, 2006.
47. M. Morohashi, A. E. Winn, M. T. Borisuk, H. Bolouri, J. Doyle, and H. Kitano. Robustness as a measure of plausibility in models of biochemical networks. *J. Theor. Biol.*, 216:19–30, 2002.
48. S. Mossa, M. Barthélémy, H. E. Stanley, and L. A. N. Amaral. Truncation of power law behavior in scale-free network models due to information filtering. *Phys. Rev. Lett.*, 88:138701[4 pages], 2002.
49. M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, 2002.
50. M. E. J. Newman. The structure and function of complex networks. *SIAM REVIEW*, 45(2):167–256, 2003.
51. M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.
52. M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118 [17 pages], 2001.
53. Romualdo Pastor-Satorras, Alexei Vázquez, and Alessandro Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.*, 87(25):258701, 2001.
54. L.M. Pecora and T.L. Carroll. Master stability function for synchronized coupled systems. *Phys. Rev. Lett.*, 80(10):2109–2112, 1998.
55. R. J. Prill, P. A. Iglesias, and A. Levchenko. Dynamic properties of network motifs contribute to biological network organization. *PLoS Biol.*, 3(11):1881–1892, 2005.

56. A. Ramezani, V. Karimipour, and A. Mashaghi. Generating correlated networks from uncorrelated ones. *Phys. Rev. E*, 67:046107, 2003.
57. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.
58. Erzsébet Ravasz and Albert-László Barabási. Hierarchical organization in complex networks. *Phys. Rev. E*, 67(2):026112, 2003.
59. S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *escherichia coli*. *Nature Genetics*, 31:64–68, 2002.
60. J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle III, and J. Doyle. Robustness of cellular function. *Cell*, 118:675–685, 2004.
61. R. Steuer. *Nonlinear Dynamics and Molecular Biology: From Gene Expression to Metabolic Networks*. PhD thesis, University Potsdam, 2005.
62. R. Steuer. On the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics*, 7(2):151–158, 2006.
63. R. Steuer. Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry*, (submitted, in press), 2007.
64. R. Steuer, C. O. Daub, J. Selbig, and J. Kurths. Measuring distances between variables by mutual information. In D. Baier and K.-D. Wernecke, editors, *Innovations in Classification, Data Science, and Information Systems*, volume Proc. 27th Annual GfKI Conference, University of Cottbus, March 12 - 14, 2003, pages 81–90. Springer-Verlag, Heidelberg-Berlin, 2004.
65. R. Steuer, T. Gross, J. Selbig, and B. Blasius. Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. USA*, 103(32):11868–11873, 2006.
66. R. Steuer, J. Kurths, C. O. Daub, J. Weise, and J. Selbig. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, 18(Suppl. 2):S231–S240, 2002.
67. R. Steuer, J. Kurths, O. Fiehn, and W. Weckwerth. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*, 19(8):1019–1026, 2003.
68. R. Steuer, A. Nunes-Nesi, A. R. Fernie, T. Gross, B. Blasius, and J. Selbig. From structure to dynamics of metabolic pathways: Application to the plant mitochondrial TCA cycle. *Bioinformatics*, doi: 10.1093/bioinformatics/btm065(advance access), 2007.
69. L. J. Sweetlove and A. R. Fernie. Regulation of metabolic networks: Understanding metabolic complexity in the systems biology era. *New Phytologist*, 168:9–24, 2005.

xxxvi REFERENCES

70. A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen. Hierarchy measures in complex networks. *Phys. Rev. Lett.*, 92(17):178702, 2004.
71. A. Wagner and D. A. Fell. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B.*, 268:1803–1810, 2001.
72. D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
73. R. Xulvi-Brunet and I. M. Sokolov. Reshuffling scale-free networks: from random to assortative. *Phys. Rev. E*, 70:066102, 2004.
74. C. Zhou, A. E. Motter, and J. Kurths. Universality in the synchronization of weighted random networks. *Phys. Rev. Lett.*, 96:034101, 2006.
75. H. Zhou. Network landscape from a Brownian particle's perspective. *Phys. Rev. E*, 67:041908, 2003.
76. E. Ziv, M. Middendorf, and C. H. Wiggins. Information-theoretic approach to network modularity. *Phys. Rev. E*, 71:04611, 2005.