CHAPTER 3

Correlations in Complex Networks

M. Ángeles Serrano[1], Marián Boguñá[2], Romualdo Pastor-Satorras[3] and Alessandro Vespignani[1]

[1] *School of Informatics and Biocomplexity Institute, Indiana University, Eigenmann Hall, 1900 East Tenth Street, Bloomington, IN 47406, USA*

[2] *Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

[3] *Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*

3.1. *Introduction*

Most real networks exhibit the presence of non trivial correlations in their connectivity pattern. Indeed, empirical measurements bring evidence to the fact that, in some instances, high or low degree vertices of the network tend to preferentially connect to other vertices with similar degree. In this situation, correlations are named *assortative* and are typically observed in social networks.[41] On the other hand, connections in many technological and biological networks[42] attach vertices of very different degree with stronger likelihood. Correlations are in this case referred to as *disassortative*. The overall origin of the appearance of these correlations is not yet completely understood, neither is the reason for the distinction in real systems between assortative and disassortative behavior. Correlations, however, drastically impact the topological properties of networks, encoding the blueprint of structural organization and are customarily used as a method to classify real nets. Moreover, correlations do not only have a topological relevance but may impact a variety of related problems such as percolation phenomena, resilience and robustness, spreading processes, or communication efficiency, to name just a few. For these reasons, several strategies have been proposed to model correlated networks. The most general practical algorithms allow the construction of networks matching any desired correlation pattern. Other just generate correlations of a fixed signature.

In the following sections, we will focus on the characterization and modeling of correlations in undirected unweighted complex networks. In particular, we will devote our attention to the statistical characterization of these features in large scale networks. In the section, we review some important and useful general analytical results concerning the topological characterization of random networks. In the third section we recall a number of specific metrics. In particular, two vertices correlations will be characterized by the average degree of nearest neighbors as a function of the vertex degree, $\bar{k}_{nn}(k)$, and correlations among three vertices will be described by several clustering measures, in particular the average clustering coefficient for vertices of a given degree $\bar{c}(k)$. Real networks are discussed in the fourth section, where we present some well-known and representative examples of correlated structures, such as the science collaboration network of physicists submitting papers to a preprint database, the Pretty-Good-Privacy web of trust between users of digital communications, the world-wide air transportation network (all of them assortative), and the Internet at the Autonomous System level, the protein interaction network of the yeast *S. Cerevisiae*, and the world trade web of commercial exchanges between countries in the world (all of them disassortative). Finally, recent developments in the modeling of correlated networks will be discussed in the fifth section. We distinguish between disassortative correlations, derived as an implicit consequence from the formulation of some classical models, and assortative correlations which should be specifically introduced in theoretical constructions. Several more general and rigorous frameworks able to reconstruct a wide range of correlation patterns are also presented. Finally, we conclude by providing an outlook on current and future developments.

3.2. *Detailed balance condition*

Although several possibilities could be considered, the conditional probabilities $P(k', k'', \ldots, k^{(n)} \,|\, k)$ that a vertex of degree k is simultaneously connected to a number n of other vertices with corresponding degrees k',k'',. . . ,$k^{(n)}$ might be the simplest theoretical functions that encode degree correlation information from a local perspective. A network is said to be uncorrelated when the conditionality on k does not apply and, therefore, the only relevant function is just the degree distribution $P(k)$. Otherwise, $P(k)$ cannot be considered in isolation and degree correlations must be taken into account through the conditional probability functions up to the pertinent order. In particular, two vertices and three vertices degree correlations are respectively encoded by the conditional probabilities $P(k' \mid k)$ and $P(k', k'' \mid k)$.

Due to the fact that edges join pairs of nodes, the key functions at the lowest level are $P(k)$ and $P(k' \mid k)$. Theoretically, they can have any form with only two constraints. First, they must be normalized, i.e.

$$\sum_k P(k) = \sum_{k'} P(k' \mid k) = 1. \tag{33}$$

Second, all edges must point from one vertex to another, so that no edges with dangling ends exist in the network. Thus, the total number of edges pointing from vertices of degree k to vertices of degree k' must be equal to the number of edges that point from vertices of degree k' to vertices of degree k. In other words, these functions must obey a degree detailed balance condition:[43]

$$kP(k'|k)P(k) = k'P(k|k')P(k'), \tag{34}$$

stating the closure of the network through the physical conservation of edges among vertices. To prove this condition, we will follow an intuitive derivation.[44]

Let N_k be the number of vertices of degree k, so that $\sum_k N_k = N$, where N is the total number of nodes in the network. In the thermodynamic limit, we can calculate the degree distribution as a frequency distribution[d], that is,

$$P(k) \equiv \lim_{N\to\infty} \frac{N_k}{N}. \tag{35}$$

Additionally, to complete the topological characterization of the network, we need also to specify how the different degree classes are connected to each other. To this end, let us consider the symmetric matrix $E_{kk'}$ accounting for the total number of edges between vertices of degree k and vertices of degree k' for $k \neq k'$. The diagonal values E_{kk} are equal to two times the number of connections between vertices in the same degree class, $k = k'$. This matrix meets the following identities:

$$\sum_{k'} E_{kk'} = kN_k, \tag{36}$$

$$\sum_{k,k'} E_{kk'} = \langle k \rangle N = 2E, \tag{37}$$

[d]For the sake of simplicity, in what follows we will obviate the limit.

where $\langle k \rangle$ is the average degree and E is the total number of edges in the network. The first identity simply states that the total number of edges emanating from vertices of degree k is k times the number of vertices in this degree class. The second identity just states that the sum of the degrees of all nodes in the network is equal to two times the number of edges.

The first identity allows to write the conditional probability as

$$P(k' \mid k) = \frac{E_{k'k}}{kN_k}. \tag{38}$$

On the other hand, from the second identity we can define the joint degree distribution as

$$P(k, k') = \frac{E_{kk'}}{\langle k \rangle N}, \tag{39}$$

where the symmetric function $(2 - \delta_{k,k'})P(k, k')$ is equal to the probability that a randomly chosen edge connects two vertices of degrees k and k'. The conditional probability can be easily related to the joint degree distribution, namely

$$P(k' \mid k) = \frac{\langle k \rangle P(k, k')}{kP(k)}. \tag{40}$$

The symmetry of $P(k, k')$ leads directly from the previous equation to the detailed balance condition:

$$kP(k' \mid k)P(k) = k'P(k \mid k')P(k') = \langle k \rangle P(k, k'). \tag{41}$$

The pre-factors k and k' in this equation account for the multiplicative nature of networks as random processes and the whole relation stands as the closure condition for networks with no detached edge ends and with no isolated vertices. On the technical side, the detailed balance condition constraints the possible form of the conditional probability $P(k' \mid k)$ once $P(k)$ is given, and vice versa.

Making use of this important relation and the normalization condition, $P(k)$ can also be written as a function of the joint degree distribution[e]:

$$P(k) = \frac{\langle k \rangle}{k} \sum_{k'} P(k, k'). \tag{42}$$

Among all the networks one can consider, Markovian networks are particularly important.[43] This class of network is completely defined by its

[e]Notice that this relation excludes vertices of degree 0, which are never considered in real complex networks.

degree distribution $P(k)$ and the first conditional probability $P(k'|k)$. In other words, such networks belong to a statistical ensemble which is maximally random under the constraint of having a given degree distribution and a given first conditional probability. In this case, the joint distribution $P(k,k')$ conveys all the relevant topological information since both $P(k)$ and $P(k'|k)$ can be derived from it. In turn, all higher-order correlations can also be expressed as a function of these fundamental functions. In particular, the three vertices conditional probability can be written as $P(k',k''|k) = P(k'|k)P(k''|k)$ and the same applies to higher order correlation functions.

The meaning of the term Markovian network that we use in this chapter is borrowed from the theory of Stochastic Processes. In this field, a stochastic process $X(t)$ is called Markovian if the probability to find the process at the position $X(t) = x$ at time t only depends on its position at the previous time $t' < t$. Then, the process is completely characterized by the probability density function $p(x,t)$ of being at x at time t and the transition probability density $p(x,t|x',t')$ of being at x at time t, provided that the process was at x' at time t'. If we identify $P(k)$ with $p(x,t)$ and $P(k'|k)$ with $p(x,t|x',t')$, we can define Markovian networks in a similar manner. One can force even more the analogy and find another connection between Markovian networks and Markovian stochastic processes. Suppose, for instance, a particle that randomly diffuses through the network, uniformly choosing at each time step one of its neighbors to continue its walk. If the underlying network is Markovian, the stochastic process constructed from the sequence of degrees of the visited vertices follows a Markovian jump process with a transition probability given by $P(k|k')$ and a steady state distribution given by $kP(k)/\langle k\rangle$. Notice that the meaning of Markovian network should not be confused with the notion of *Markov graph.*[45]

3.3. *Empirical measurement of correlations*

At the level of two vertices degree correlations, the most straightforward measure consists in a direct inspection of the two-dimensional histograms of the joint degree distribution $P(k',k)$[46,47] or the conditional probability $P(k' \mid k)$. However, such histograms in finite size systems are highly affected by statistical fluctuations and are thus not good candidates to evaluate empirical correlations. In order to characterize degree correlations, it is then more convenient to adopt other standards, which nevertheless will eventually depend on these functions. A most useful approach consists in defining a one-parameter function encoding the signature of correlations. In the case of two vertices correlations, such function is defined as the average nearest neighbors degree (ANND) of nodes with degree k, $\bar{k}_{nn}(k)$.[48] It considers the mean degree of the neighbors of a vertex as a function of its degree k. When this function increases with k, the network is named assortative, with vertices associating preferentially to other vertices of similar degree. When $\bar{k}_{nn}(k)$ instead decreases, the network is named disassortative, with high-degree vertices attaching preferentially to other low-degree ones. Hence, this is a representation which gives a clear interpretation of pair correlations and at the same time can provide further information about hierarchical organization in networks. Finally, the scalar Pearson correlation coefficient of the degrees of vertices at the ends of edges is used to summarize the level of correlation with a single number[f].[41,42]

Despite the increasing attention in the literature about the measurement of $P(k', k)$, the first correlation observable appearing in the literature is the network transitivity or clustering coefficient,[6,7] a scalar which quantifies the probability that two vertices with a common neighbor are also connected to each other. This concept has its roots in sociology and, in the language of social networks, it measures the likelihood that the friend of your friend is also your friend. Therefore, it is in fact a measure of three vertices correlations although, curiously, it is among the first studied structural properties of networks, together with the small-world effect or the degree distribution. This definition and other alternatives[49,50] have been broadly used to quantify in a statistical sense the deviation of real networks, strongly clustered, from the behavior of classical random graphs.

Since clustering measures triangles in a network, it seems also natural to pose the question of how to measure higher order loops (closed paths). This issue is particularly important in order to asses if a network can be assumed to be Markovian, since, in this case, the loop structure must be very well described by the two vertices correlations. A number of authors have paid attention to loops of length four and above. However, there are technical difficulties when one tries to separate the independent contributions of the different motifs.[51–55] This is the main reason why triangles –and not higher order loops– have been chosen as a measure of correlations.

In this section we will concentrate on the broadly accepted and used statistical correlation observables in the analysis of large scale networks,

[f]The Pearson coefficient is computed as the correlation coefficient of the joint distribution $P(k, k')$.

the average nearest neighbors degree and the degree dependent clustering coefficient, focusing on their theoretical grounds and significance.

3.3.1. *Two vertices correlations: ANND*

The average nearest neighbors degree, $\bar{k}_{nn}(k)$, of vertices of degree k is defined as a smoothed conditional probability:[48]

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k' \,|\, k), \tag{43}$$

so that the statistical fluctuations that usually disturb the evaluation of $P(k' \,|\, k)$ are damped.

Real networks usually tend to display one of two different patterns: either $\bar{k}_{nn}(k)$ is a monotonous increasing function of k or, on the contrary, it is a monotonous decreasing function of k. At the level of correlation properties, this segregation allows the classification of networks based on their ANND behavior:[41]

- Assortative networks exhibit $\bar{k}_{nn}(k)$ functions increasing with k, which denotes that vertices are preferentially connected to other vertices with similar degree. Examples of assortative behavior are typically found in many social structures.
- Disassortative networks exhibit $\bar{k}_{nn}(k)$ functions decreasing with k, which implies that vertices are preferentially connected to other vertices with very different degree. Examples of disassortative behavior are typically found in several technological networks, as well as in communication and biological networks.

This measure provides a sharp evidence for the presence or absence of correlations since, in the case of uncorrelated networks, it is easy to demonstrate that this quantity should not depend on k. In fact, the uncorrelated ANND value is found to coincide with the heuristic parameter $\kappa = \langle k^2 \rangle / \langle k \rangle$, independently introduced to characterize the level of heterogeneity of networks.[12] For homogeneous networks $\kappa \sim \langle k \rangle$, whereas for scale-free (SF) networks with unbounded degree fluctuations it diverges in the thermodynamic limit. As a consequence, it comes to be a key parameter characterizing the properties of networks and the processes running on top of them.

Here, we deduce $\bar{k}_{nn}^{unc}(k)$ from the detailed balance and the normalization conditions. Summing Eq. 41 over k and recalling that $P^{unc}(k'|k)$ does not depend on this variable, we obtain that

$$P^{unc}(k'|k) = \frac{k'P(k')}{\langle k \rangle}, \tag{44}$$

from where we have

$$\bar{k}_{nn}^{unc}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{45}$$

Therefore, a function $\bar{k}_{nn}(k)$ showing any explicit dependence on k signals the presence of degree correlations in the system.

As in the case of uncorrelated networks, it is also possible to derive some general exact results concerning the behavior of $\bar{k}_{nn}(k)$ in the case of SF networks with a degree distribution of the form $P(k) \sim k^{-\gamma}$ for $k \in [1, k_c]$. The cut-off value k_c is a consequence of the finiteness of the network and diverges in the thermodynamic limit.[56] The specific dependency of k_c on N depends, in general, on the details of the model. Let once again exploit the detailed balance and the normalization conditions. By multiplying by a k factor both terms of Eq. 41 and summing over k' and k up to k_c, we obtain

$$\langle k^2 \rangle = \sum_{k'} k'P(k') \sum_{k}^{k_c} kP(k \,|\, k') = \sum_{k'} k'P(k')\bar{k}_{nn}(k', k_c), \tag{46}$$

where we have made explicit the dependence on k_c. In scale-free networks with exponent $2 < \gamma < 3$ the second moment of the degree distribution diverges as $\langle k^2 \rangle \sim k_c^{3-\gamma}$, and therefore

$$\sum_{k'} k'P(k')\bar{k}_{nn}(k', k_c) \sim \frac{A}{3-\gamma} k_c^{3-\gamma}, \tag{47}$$

where A is a constant pre-factor depending on the details of $P(k)$. As a consequence, the left hand side of this equation must bear divergences[g].

In the case of disassortative correlations, the divergence should just be contained in the k_c dependence of $\bar{k}_{nn}(k', k_c)$, since $\bar{k}_{nn}(k', k_c)$ is decreasing in k' and furthermore $k'P(k')$ is an integrable function.

When correlations are assortative, however, there may be singularities associated with the sum over k' depending on the rate of growth of the increasing $\bar{k}_{nn}(k', k_c)$. Nevertheless, it can be demonstrated that, even for strong growth rates, the divergence associated to the explicit k_c dependence is predominant.[56]

---

[g]For $\gamma = 3$ the arguments are still valid although more involved.

Therefore, one can conclude, just from the detailed balance and the normalization conditions, that in SF networks with $2 < \gamma \leq 3$ the function $\bar{k}_{nn}(k', k_c)$ must diverge when $k_c \to \infty$ in a nonzero measure set, regardless of the character and level of the correlations present in the network. This fact is, for instance, fundamental in determining the properties of epidemic spreading processes in correlated scale-free networks.[56]

From a practical point of view, when studying real SF networks, one can always take advantage of the fact that the divergence of the function ANND is independent of the underlying correlation structure, so that $\bar{k}_{nn}(k)$ can be always normalized by the uncorrelated value $\bar{k}_{nn}(k)_{unc} = \frac{\langle k^2 \rangle}{\langle k \rangle}$. This finite size correction makes comparable the ANND functions of different real networks.

As we have mentioned at the beginning of this section, it is also possible to obtain information on the nature of two vertices correlations by examining a single scalar quantity, the Pearson correlation coefficient of the degrees of the vertices at the end of edges.[41] The Pearson coefficient r can be defined as follows:

$$r = \frac{\langle kk' \rangle_e - \langle k \rangle_e^2}{\langle k^2 \rangle_e - \langle k \rangle_e^2}, \tag{48}$$

where $\langle kk' \rangle_e$ is the average of the product the degrees at the end points of all edges and $\langle k^n \rangle_e$ is the average of the n-th power of the degree at the end of any edge[h]. These averages can be expressed in terms of the joint degree distribution as

$$\langle kk' \rangle_e = \sum_{kk'} kk' P(k, k'), \tag{49}$$

$$\langle k^n \rangle_e = \sum_{kk'} \frac{k^n + k'^n}{2} P(k, k'). \tag{50}$$

Using the detailed balance condition Eq. 41, we obtain the following relation between the Pearson coefficient and the ANND function

$$r = \frac{\langle k \rangle \sum_k k^2 \bar{k}_{nn}(k) P(k) - \langle k^2 \rangle^2}{\langle k \rangle \langle k^3 \rangle - \langle k^2 \rangle^2} \tag{51}$$

[h]In the original definition,[41] r was defined in terms of the averages of the excess degree, that is, discounting the connection from the considered edge. It is easy to see, however, that both definitions yield the same result.

For uncorrelated networks, with $\bar{k}_{nn}^{unc}(k) = \langle k^2 \rangle / \langle k \rangle$, we obtain $r = 0$, while $r < 0$ ($r > 0$) is interpreted as a signature of dissasortative (assortative) two vertices correlations. While the Pearson coefficient can be useful to give a single value measure of the character of correlations, its efficiency suffers from some drawbacks as compared with the ANND function. On the one hand, it misses the possible hierarchical structure of correlations that is explicitly evident in the k dependence of the ANND. On the other hand, for SF networks it strongly depends on the size of the network. To see this, consider a dissasortative SF network with $2 < \gamma < 3$ and degree cut-off k_c. In this case, we have $\langle k^n \rangle \sim k_c^{n+1-\gamma}$. Since the network is dissasortative, $r < 0$ and we have $\langle k \rangle \sum_k k^2 \bar{k}_{nn}(k) P(k) < \langle k^2 \rangle^2$, so at leading order

$$|r| \sim \frac{\langle k^2 \rangle^2}{\langle k^3 \rangle} \sim k_c^{2-\gamma}, \tag{52}$$

which tends to zero in the thermodynamic limit for $2 < \gamma \leq 3$. This indicates that one has to be very cautious when drawing conclusions about the nature of correlations in SF networks based only on the information provided by the Pearson coefficient.

To finish this section, we discuss another consideration that must be taken into account, and which refers to the distinction between the purely uncorrelated case and the maximally random case achievable when respecting the degree distribution. It turns out that completely uncorrelated networks are not always feasible due to architectural constraints. Given a certain degree distribution $P(k)$, finite size effects could condition in some cases the closure of the network to either the presence of multiple and self-connections or disassortative two vertices correlations.[84–86] Bounded degree distributions, in which $\langle k^2 \rangle$ is finite, present maximum degree values k_c below or around a structural cut-off k_s, so that physical networks can indeed be constructed as uncorrelated. However, when dealing with unbounded degree distributions and diverging fluctuations in the infinite network size limit (for instance, scale-free degree distributions with $2 < \gamma < 3$ as observed in many real systems), $k_c > k_s$ and then structural correlations are important and cannot be avoided. In that case, one can just consider the maximally random network with a given degree distribution $P(k)$. For bounded degree distributions with actual cut-offs below the structural one, the maximally random network will indeed correspond to the uncorrelated case. However, for unbounded degree distributions with divergent second moment and actual cut-off well above the structural one, the closure of the maximally random network forces the conservation of structural correlations. Whereas correlation measures provide information about the overall presence of correlations in the network, the comparison with the maximally random case discounts the structural effects, so that physical correlations can be detected.

3.3.2. *Three vertices correlations: Clustering*

Correlations among three vertices can be measured by means of the conditional probability $P(k', k'' \,|\, k)$ that a vertex of degree k is simultaneously connected to two vertices with degrees k' and k''. Only in the case of Markovian networks, this function can be expressed in terms of two vertices correlations through the relation $P(k', k'' \,|\, k) = P(k' \,|\, k)P(k'' \,|\, k)$.

As previously indicated, the conditional probabilities $P(k', k'' \,|\, k)$ or $P(k'' \,|\, k)$ are difficult to estimate directly from real data, so other assessments have been proposed. All of them are based in the concept of clustering, which refers to the tendency to form triangles (loops of length 3) in the neighborhood of any given vertex.

The clustering in a network quantifies the likelihood that vertex j is connected to vertex l, if vertices j and l are simultaneously connected to vertex i. Watts and Strogatz originally proposed a scalar local measure for clustering, which is known as the clustering coefficient.[6] It is computed for every vertex i as the ratio of the number of edges e_i existing between the k_i neighbors of i and the maximum possible value, i.e.:

$$c_i = \frac{2e_i}{k_i(k_i - 1)}. \tag{53}$$

The clustering coefficient of the whole network C is then defined as the average of all individual c_i's, $C = \sum_i c_i/N$. Watts and Strogatz also pointed out that real networks display a level of clustering typically much larger than the value for a classical random network of the same size, $C_{rand} = \langle k \rangle/N$.

The clustering coefficient has been redefined in a number of ways, for instance as a function of triples in the network (triples are defined as subgraphs which contain exactly three nodes) and reversing the order of average and division in Eq. 53:[49,57]

$$C_\Delta = \frac{3 \times \text{number fully connected triples}}{\text{number triples}}. \tag{54}$$

This definition corresponds to the concept of the fraction of transitive triples introduced in sociology long time ago.[7]

Although overall scalar measures are helpful as a first indication of clustering, it is always more informative to work with quantities which explicitly depend on the degree. As in the case of two vertices correlations, an uniparametric function $\bar{c}(k)$[50] can also be computed. In practice, the degree-dependent local clustering $\bar{c}(k)$ is calculated as the clustering coefficient averaged for each degree class k. Formally, it is defined as the probability that two vertices, neighbors of a vertex of degree k, are linked to each other. Hence, it can be written as a function of the three vertices correlations:

$$\bar{c}(k) = \sum_{k',k''} P(k',k'' \mid k) r_{k'k''}(k), \tag{55}$$

where $r_{k'k''}(k)$ is the probability that the vertices of degree k' and degree k'' are connected given that they both are neighbors of the same vertex of degree k. The corresponding scalar measure is the mean clustering coefficient

$$\bar{c} = \sum_k P(k)\bar{c}(k), \tag{56}$$

which is related to the clustering coefficient by[i]:

$$C = \frac{\bar{c}}{1 - P(0) - P(1)}. \tag{57}$$

For Markovian networks, $\bar{c}(k)$ can be expressed as a function of the two vertices degree correlations, giving the asymptotic expression:[58,59]

$$\bar{c}(k) = \frac{\langle k \rangle^3}{N k^2 P^2(k)} \sum_{k',k''>1} \frac{(k''-1)(k'-1)P(k'',k')P(k'',k)P(k',k)}{k'k''P(k')P(k'')}. \tag{58}$$

In the case of uncorrelated networks, $\bar{c}(k)$ is independent of k. Furthermore, all the measures collapse and reduce to C.[58,60,61]

$$\bar{c}(k) = C = C_\Delta = \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}. \tag{59}$$

Therefore, a functional dependence of the local clustering on the degree can be attributed to the presence of a complex structure in the three vertex correlation pattern. Indeed, it has been observed that $\bar{c}(k)$ exhibits a power-law behavior $\bar{c}(k) \sim k^{-\alpha}$ for several real scale-free networks. Hence, $\bar{c}(k)$ has been proposed as a measure of hierarchical organization and modularity in complex networks.[62]

---

[i]Notice that we have implicitly assumed that $\bar{c}(0) = \bar{c}(1) = 0$ whereas in the definition of C we only consider an average over the set of vertices with degree $k > 1$. This fact explains the difference between both measures.

3.4. *Networks in the real world*

Degree correlations are ubiquitous in real networks, denoting the presence of structural organization and hierarchy. Usually, empirical networks show a highly clustered architecture and two vertices correlations are present as well, which demonstrates that nodes in networks do not mix randomly. What is more, among a number of theoretical possibilities, pair correlations commonly display one out of only two well-defined mixing patterns. As discussed, this observation has led to the segregation of most real networks into two universality classes, assortative and disassortative, depending on whether their ANND function is an increasing or a decreasing function of k, respectively.

Empirical networks are often classified in several loose general categories as well, and within a given class most networks are found to display the same type of correlations.[42] Indeed, among other specific features, many social networks are assortative,[41,63,64] such as, for instance, company director networks,[65] co-authorship and collaboration networks,[66–68] or the network of email address books.[69] On the contrary, most biological networks (protein-protein interaction network in the yeast cell,[70] metabolic networks in bacteria,[71] food webs[72]) or technological networks (the Internet at the Autonomous System level,[12] the network of hyperlinks between pages in the World Wide Web,[73] *etc.*) appear to be disassortative. In some cases, it is difficult, if not impossible, to classify real networks into single categories, especially for systems related to human action or when functionality is also taken into account. Since different sets of classification criteria can be defined, and although in some cases classifications are unquestionable, here we prefer to treat specific examples instead of whole categories in order to avoid any potential conflict. Next, we will examine the details concerning correlations of a number of well-known and representative real networks.

Protein interaction networks

Biological structures are among the most complex systems that can be represented as networks, and simplified models turn out to be very useful to understand how they organize and evolve. Cells themselves are very intricate systems comprising millions of molecules acting in a coherent manner as open systems exchanging matter, energy and information with the environment. Therefore, the cellular network, albeit one, is commonly reduced to three different sub-webs: the metabolome, or the ensemble of all metabolites and the reactions that they enter, the genome, or the set of all genes

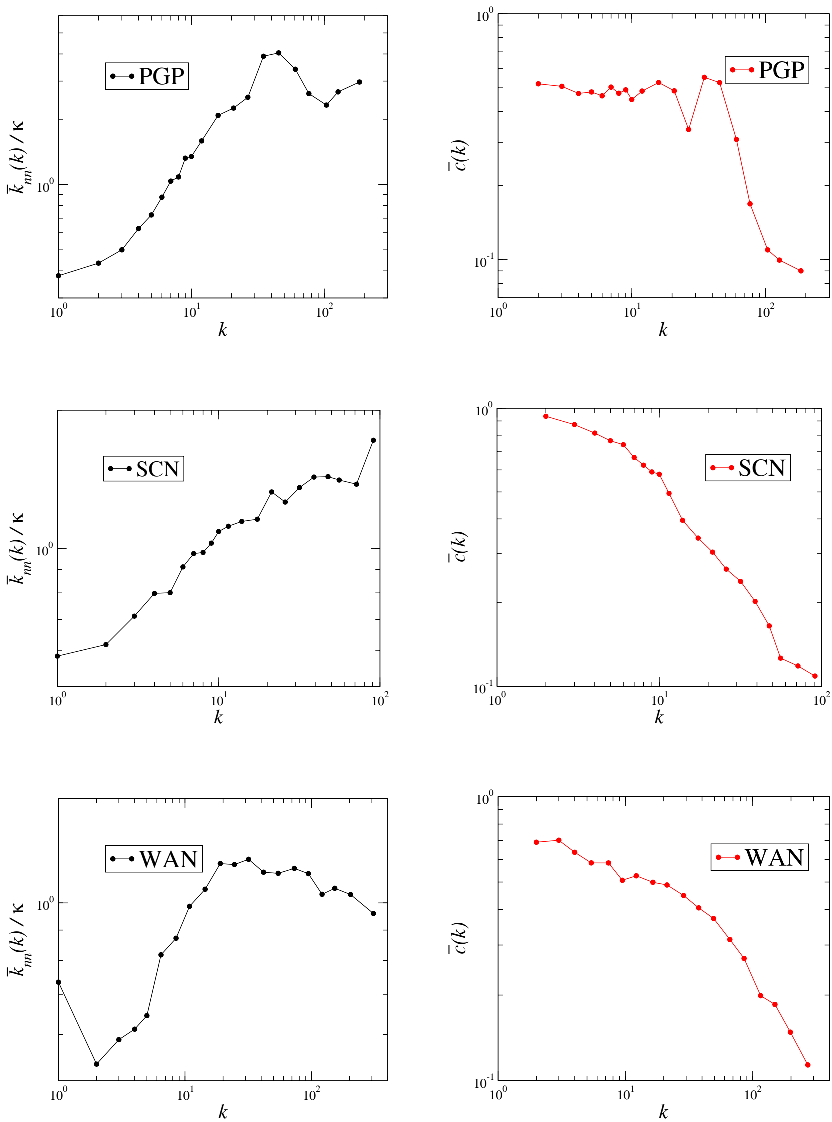

Fig. 11. Assortative real networks. The average nearest-neighbor degree is shown in the column on the left scaled by the heterogeneity parameter κ. The column on the right exhibits the clustering coefficient as a function of the vertex degree. PGP is the Pretty-Good-Privacy web of trust between users of digital communications, SCN stands for the scientific collaboration network of researchers co-authoring academic papers in the cond-mat e-Print archive, and WAN is the world-wide airport transportation network. For the data publicly available visit the site http://www.cosin.org.

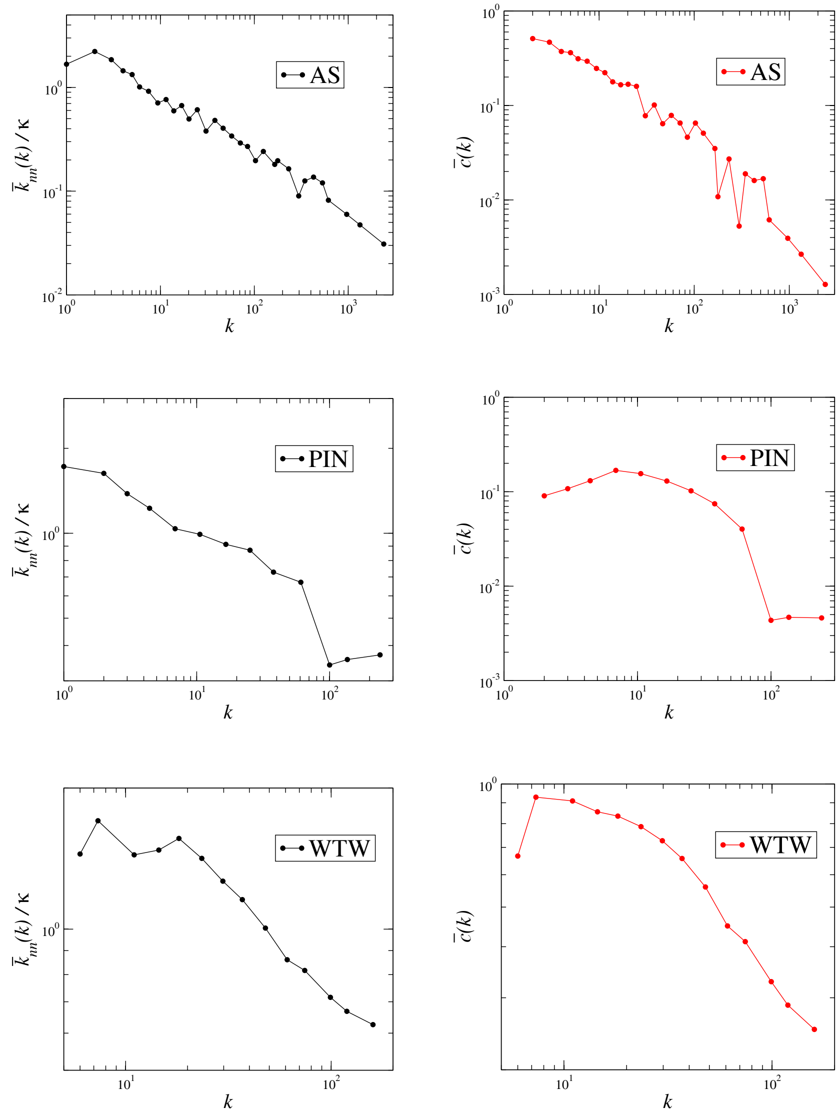

Fig. 12. Disassortative real networks. The average nearest-neighbor degree is shown in the column on the left scaled by the heterogeneity parameter κ. The column on the right exhibits the clustering coefficient as a function of the vertex degree. AS stands for the Internet map at the Autonomous System level, PIN is the protein interaction network of the yeast *S. Cerevisiae*, and WTW is the world trade web of commercial exchanges between countries in the world. For the data publicly available visit the site http://www.cosin.org.

in a cell which can interact by affecting each other's level of expression, and the proteome, or group of proteins and their interactions by physical contact. Metabolic webs and protein interactions networks (PINs) are better known for the most simple cells, such as the yeast *S. Cerevisiae* or the bacteria *E. Coli*. Although different data sets can provide varying results, there is enough evidence to ensure that these networks exhibit a nontrivial topological structure with a statistical abundance of hubs and presence of correlations.

Here, we inspect in more detail the PIN of the yeast *S. Cerevisiae* constructed from data, obtained with different experimental techniques, at the Database of Interacting Proteins (http://dip.doe-mbi.ucla.edu).[74] The network has 4713 proteins and 14846 interactions for data collected until April 2003. The degree distribution is heavy tailed, with a power-law of exponent $\gamma \simeq 2.5$ and an exponential cut-off. A signature of hierarchy is the disassortative behavior of its ANND function, as shown in Fig. 12. For most values of k the decay is power-law like with an exponent ~ 0.24. On the other hand, the degree-dependent clustering coefficient does not show a clear functional form. However, the value of the clustering coefficient C for the whole network is 0.09, five times larger than the corresponding value for a comparable random graph. This suggests the presence of structural organization.

Scientific collaboration network

The organization of social communities has been an extensively studied topic in social sciences. Recently, however, it has been possible to take advantage of the progress made in Information Technology to access and manage extensive and reliable data sets in various kind of social structures, for instance clubs, organizations, or collaborative teams. Among others, professional communities have been analyzed from large databases as complex collaboration networks. Examples are the already classic collaboration network of film actors,[68] the company directors network[65] and the network of co-authorship among academics.[66,67] These are in fact bipartite networks,[7] although the one mode projection is usually used so that members are tied through common participation in one or more films, boards of directors, or academic papers.

As an illustration, here we consider the scientific collaboration network (SCN) reconstructed from the submitted papers to the condensed matter physics section of the e-Print Archive (http://xxx.lanl.gov/archive/cond-mat) between 1995 and 1998. The network has 15179 scientists with an average number of collaborators $\langle k \rangle = 5.67$. The analysis of correlations confirms the commonly accepted expectations for social networks. The presence of assortative pair correlations is denoted by the increasing trend of the function $\bar{k}_{nn}(k)$ in Fig. 11, which indicates that researchers with a relatively large number of collaborators tend to be connected among them. The mean clustering coefficient is very high with a value of $\bar{c} = 0.64$. Furthermore, the degree-dependent local clustering follows a clear decay with increasing k, indicating the existence of some hierarchy[50] or modularity.[62]

3.4.1. *Pretty-good-privacy web of trust*

The web of trust[75] between users of the Pretty-Good-Privacy (PGP) encryption algorithm[76,77] is one of the largest reported non-bipartite graphs one can build from a social network emerging in the technological world. On the technical side, the PGP software encrypts files or email messages which may only be opened by the intended recipients. Moreover, it allows to protect also identities. The sender of a digital communication signs the outgoing document so that the recipients know for certain who the author is. The cryptographic system uses two keys associated to each user, a public key known to everyone and a private or secret key known only to the recipient of the message. The public and private keys are related in such a way that only the public key can be used to encrypt messages and only the corresponding private key can be used to decrypt them, being computationally infeasible to deduce one key from the other. When A wants to send a secure message to B, it must use B's public key to encrypt the message. B then uses its private key to decrypt it.

Provided that pairs of keys can be generated by everyone, users should verify that a given key belongs really to the person stated in the key. This requires authentication of the public key, which implies a signing procedure where a person signs the public key of another, meaning that she trusts the other person is who she claims to be. This procedure generates a web of peers that have signed public keys of another based on trust, the so-called web of trust of PGP.

The undirected web of trust (with an edge between peers who have mutually signed their keys) as it was on July 2001 (http://www.dtype.org) comprises 57243 public keys and its average degree is $\langle k \rangle = 2.16$. It shows a scale-free degree distribution with an exponent $\gamma = 2.6$ for small degrees $k < 40$, and a crossover towards another power law with a higher exponent,

~ 4, for large values of the degree. This indicates that, in contrast to many technological networks or social collaboration networks, the PGP is not a scale-free network but has a bounded degree distribution.

However, as for many other social networks, it shows assortative mixing and a large clustering coefficient $C = 0.4$.[75] In Fig. 11 we analyze the correlations of the PGP network as measured by the function $\bar{k}_{nn}(k)$. The growing trend confirms the assortative character of the connections between users. Remarkably, the function $\bar{k}_{nn}(k)$ has an approximately linear behavior, at least for not very large values of k. In Fig. 11, we also plot the clustering coefficient as a function of the degree, $\bar{c}(k)$. Despite the short range of values of k shown in the plot (due to the limited size of the network and the bounded nature of the degree distribution), we can observe that $\bar{c}(k)$ is a nearly independent function of the degree for most k values. This absence of structure is surprisingly in contrast to many other real networks in which $\bar{c}(k)$ has been shown to be a decreasing function of the degree.[62]

The internet at the autonomous system level

The Internet has become a paradigm in complex networks science. Its own organization as a networked system of physical connections between computers makes the graph abstraction a natural representation. However, its intricate ever-evolving structure forces to opt for coarse-grained descriptions so that it is usually examined at the level of routers (special devices that transfer the packets of information across the Internet's different networks) or at the level of Autonomous Systems (ASs) (which are defined as independently administered domains which autonomously determine internal communication and routing policies for Internet communications[12]). Several projects, CAIDA (http://www.caida.org/) and DIMES (http://www.netdimes.org/) among others, have been gathering and analyzing data on the Internet at different levels. In particular, measurements on the Internet structure and topology allow to recreate maps that display connectivity information.

Here, we review one of these Internet maps, which reconstructs the Internet topology at the AS level, from data collected by the *Oregon route-views* project (http://www.routeviews.org/). The map is dated May 2001 and comprises 11174 nodes with an average degree $\langle k \rangle = 4.2$. Statistical measures on this map provide evidence of the large-scale heterogeneity of the Internet, characterized by the small-world property and a scale-free degree distribution with exponent $\gamma \simeq 2.1$. It also clearly reveals its hierarchical structure. More precisely, degree-degree correlations are strongly disassortative and exhibit a heavy tail that can be fitted by a power-law decay with a characteristic exponent close to 0.55, as shown in Fig. 12. The clustering coefficient $\bar{c}(k)$ for nodes of degree k is also displayed. The power-law behavior is not so sharp in this case, but nevertheless the curve also shows a very clear heavy tail. The scalar clustering coefficient is $C = 0.3$.[48,50,78] All these features rule out the possibility of a purely random graph structure or a regular architecture.

The world airport network

The World Airport Network (WAN)[79] is a representative example of a large transportation infrastructure which can be examined under the perspective of complex networks theory. At the level of functionality, the WAN is also a communication network bringing passengers from one side of the world to another.

The database of the International Air Transportation Association (http://www.iata.org) compiles information about direct flights between world airports, and the number of available seats in each flight. For the year 2002, a network with 3880 nodes and 18810 edges can be reconstructed from the data. The topology of the network exhibits the small-world property and a scale-free degree distribution of exponent $\gamma \simeq 2$, which presents an exponential cut-off induced by physical restrictions in the number of connections that a single airport can handle.

Regarding correlations, the topological $\bar{k}_{nn}(k)$ in Fig. 11 surprisingly shows assortative behavior for small degrees and a plateau for higher degrees, which denotes the absence of noticeable topological correlations for large k's. This picture changes notably if the weighted character of the network is taken into account. Then, the ANND function appears to be assortative in the whole k spectrum. With reference to clustering, low degree vertices present a much higher interconnected neighborhood than hubs, as can be seen in Fig. 11 showing a decaying $\bar{c}(k)$. That means that large airports act as bridges on the international and intercontinental scale. The weighted version follows the same trend with a much more limited decay.

The world trade web

The network of trade relationships between different countries in the world can be classified as an economic system where the activity is governed by optimization criteria and competition and cooperation forces. Publicly available import, export, and gross domestic product databases (http://www.intracen.org/menus/countries.htm, http://www.tswoam.co.uk/world) provide the information to analyze the international trade system as a complex network. Nodes in the world trade web (WTW)[80] represent countries and edges appear between them whenever a commercial channel exists. Despite its relatively small size ($N = 179$ and $\langle k \rangle \simeq 18$ in the undirected version) this socioeconomic structure displays the typical properties of complex networks, namely, the small-world property and scale-free degree distribution with $\gamma \simeq 2.6$ for high degrees.

Correlations also match clear patterns and reflect a discerning hierarchical organization, where countries that belong to influential areas connect to other influential areas through hubs. As can be observed in Fig. 12, the function $\bar{k}_{nn}(k)$ clearly depend on k, with a power law decay of exponent $\simeq 0.5$. This result means that the WTW is a disassortative network where highly connected countries tends to connect to poorly connected ones. There exists a high positive correlation between the number of trade channels of a country and its wealth (measured by the per capita Gross Domestic Product) so that, as expected, highly connected nodes correspond to rich countries and poorly connected nodes to poor ones. The socio-economic implication of disassortativity is then that poor countries do not trade to each other but they do that only with rich countries. Hierarchy is also reflected by the high level of local cohesiveness. Fig. 12 shows the clustering coefficient of the undirected WTW as a function of the vertex degree. As is distinctly seen, this function has a strong dependence on k, with a power law behavior of exponent $\simeq 0.7$. The clustering coefficient averaged over the whole network is $C = 0.65$, greater by a factor 2.7 than the value corresponding to a random network of the same size. Surprisingly, these results point to a high similarity between the WTW and other completely different types of networks, for instance the Internet.

3.5. *Modeling correlations*

All the empirical evidences reported in the previous section about the hierarchical architecture of real networks should be included in models aiming to help us to understand how these complex systems self-organize and evolve. Models that neglect correlations will inevitably fail to trustworthy recreate actual systems. In this section, we will review how disassortative correlations arise in the classical configuration model and in scale-free growing networks as a by-product. Then, we will go over recent efforts in the construction of models attending to correlations. Some of them are intended to reproduce specific correlation behaviors and others, more ambitious, are devoted to set up a general framework to study the origin of correlations in random networks.

3.5.1. *Disassortative correlations*

Models reproducing disassortative correlations can be divided into two main classes referring to static and dynamic algorithms. In the first category, the classical configuration model[81–83] provides correlations for scale-free degree distributions although, *a priori*, it was supposed to generate uncorrelated networks. In the second group, growing scale-free networks display disassortative correlations between the degrees of neighboring vertices, which spontaneously appear as a consequence of the asymmetry in the history of nodes introduced at different times.

3.5.2. *The configuration model*

The configuration model (CM) is a classical algorithm to construct random networks with a specific degree distribution $P(k)$ settled *a priori*. This is a static model where the total number of nodes in the network N remains constant. For each one of these nodes, a random number k_i is drawn from the probability distribution $P(k)$ and is assigned to it in the form of stubs or ends of edges emerging from that vertex. Several constraints apply. The first one states that no vertex can have a degree larger than $N - 1$. The second is that the sum $\sum_i k_i$ must be even and is imposed by the closure condition. The network is constructed by connecting pairs of these edge ends chosen uniformly at random. The result of this assembly is a random network with degrees distributed according to $P(k)$, by definition.

Given the random nature of the assignment of stubs, it was expected that the ensuing network was uncorrelated, and it is in fact the case if the degree distribution is bounded or multiple connections and tadpoles (self-connections) are allowed. On the other hand, the CM indeed generates disassortative correlations when fluctuations diverge in the infinite-network-size limit, for instance, when the expected degree distribution is scale-free with exponent $2 < \gamma \leq 3$, and no more than one edge is allowed between any two vertices.[47,84]

If the degree distribution has a finite second moment $\langle k^2 \rangle$, the fraction of multiple edges and tadpoles resulting from the construction process vanishes in the thermodynamic limit and, as a consequence, they can be neglected. For scale-free degree distributions with exponent $2 < \gamma \leq 3$, the weight of these multiple edges with respect to the overall number of edges is small but cannot be ignored since they are not evenly distributed among all the degree classes. In the thermodynamic limit, a finite fraction of multiple edges and tadpoles will remain among high degree vertices.[85] There are theoretical and technical reasons to try to avoid multiple edges in some instances, but imposing the restriction on the algorithm that multiple edges are prohibited originates the presence of disassortative correlations.[47,84]

The origin of this phenomenon can be traced back to be a cut-off effect,[85] with the maximum degree ruling the presence or absence of correlations in a random network with no multiple or self-connections. These facts have been taken into account in the construction of a procedure, the uncorrelated configuration model,[86] to generate uncorrelated scale-free networks with no multiple and self-connections.

3.5.3. *Growing models*

Real networks, as everything else in the world we experience, are far from being static. Their evolution is relevant, specially when the time scale of the occurrence of structural changes in the network is of the same magnitude of the characteristic time associated to processes taking place on top of them. Therefore, dynamic models are more appropriate to describe reality and they can further contribute to the understanding of the mechanisms that shape the topological properties of complex networks.

These dynamical models are typically devised as growing networks models, where nodes and edges are gradually added to the network and connected following specific attachment rules. This kind of theoretical construction has succeed in explaining the scale-free structure observed in real nets applying mechanism such as the preferential attachment rule.[87]

A number of authors have worked out analytic studies on this sort of networks. All of them are centered on solving the basic dynamical equations governing the network evolution and take the network size $N(t)$ as the natural time scale. Aside the degree distribution and other first order properties, degree correlations have also been examined.[58,88,89] Before going into further details, let us first briefly revise the standard procedure which assembles this sort of networks:

- At each time step, a new node with m edges is added to the network.
- Ends of the new edges are distributed among old vertices. Each

vertex i has a probability $\Pi(k_i)$ of getting new edges, where k_i is its degree.

In the original Barabási-Albert model,[87] the probability $\Pi(k_i)$ is proportional to k_i, and the system evolves into a steady power-law degree distribution with the form $P(k) \sim k^{-3}$. Many variations have been introduced. In particular, the preferential attachment probability has been generalized and allowed to grow more slowly or faster than linearly with the degree. Only in the linear case, the ensuing degree distribution is power-law, but its exponent can be modulated by introducing an additional constant factor in the attachment probability, *i.e.*, $\Pi(k_i) = k_i + A$. Then, a scale-free degree distribution of the form $P(k) \sim k^{3+A/m}$ is obtained,[90] which for the range of values $-m < A < \infty$ yields degree exponents $2 < \gamma < \infty$ Other ingredients can be incorporated in order to account for a power-law degree distribution of exponent $2 < \gamma < 3$, such as edge disappearance[91] or wiring processes.[92] Summarizing, the class of growing scale-free networks models is described by power-law degree distributions of the form $P(k) \sim k^{\gamma}$, with an average degree at time t given by $\langle k(t,t') \rangle \sim (t/t')^{\beta}$, for a node introduced at time t'. The exponents γ and β are related through $\gamma = 1 + 1/\beta$.

As the network grows, it can be proved that correlations between the degrees of neighboring vertices spontaneously appear. The first theoretical derivation of this result[93] was obtained by calculating the number of nodes of degree k attached to an ancestor node of degree k'. In the framework of the rate equation approach, this joint distribution does not factorize so that correlations exit. This characterization of degree correlations is indeed measuring $P(k,k')$. With respect to measures of the average nearest neighbors degree function, it is found that, in the large k limit,

$$\bar{k}_{nn}(k) \sim N^{(3-\gamma)/(\gamma-1)} k^{-(3-\gamma)} \tag{60}$$

for $\gamma < 3$. That is, two vertices correlations are disassortative and characterized by a power-law decay.[58,88] On the other hand, it can also be proved that for $\gamma = 3$, the ANND function converges to a constant value independent of k and proportional to $\ln N$, and therefore, the Barabási-Albert model lacks appreciable correlations.

3.5.4. *Assortativity generators*

Unlike disassortative correlations, which are inherent to the very construction of some general models, assortative correlations must be specifically forced. The special character of this type of mixing is also patent in the implications for issues such as percolation or network resilience. Extensive numerical simulations show that assortative networks percolate more easily than disassortative ones. Concerning resilience, simulations also prove that assortative networks display robustness through redundancy against targeting hubs, since high degree vertices tend to be clustered together in groups of high cohesiveness. On the contrary, such attacks are much more effective in disassortative networks, where hub connections are broadly distributed.

The basic model generating assortative networks[41,42] proposes a specific Monte Carlo sampling scheme equivalent to the Metropolis-Hasting method.[94] The degree distribution can be computed from the distribution of excess degrees $q(k_e)$[j], which on its turn must be calculated from a given edge distribution $e(k_e, k'_e)$ representing the fraction of edges in the network between nodes with excess degree k_e and nodes of excess degree k'_e:

$$q(k) = \sum_{k'} e(k', k) \tag{61}$$

$$P(k) = \frac{q(k-1)/k}{\sum_{k'} q(k'-1)/k'}, \tag{62}$$

where nodes of degree zero are not considered. Once the degree distribution is known, the classical configuration model[81–83] can be applied to assemble the network. The algorithm generating the assortative mixing works then in two repeated steps:

- Two edges are selected at random, named $(1, 2)$ and $(3, 4)$ after the vertices they connect. The excess degrees q_1, q_2, q_3, q_4 of those vertices are calculated.
- The two edges $(1, 2)$ and $(3, 4)$ are replaced by the new ones $(1, 3)$ and $(2, 4)$ with probability 1 if $e(q_1, q_3)e(q_2, q_4) \geq e(q_1, q_2)e(q_3, q_4)$. Otherwise, the swap is performed with probability $p = [e(q_1, q_3)e(q_2, q_4)]/[e(q_1, q_2)e(q_3, q_4)]$.

Finally, the correlation structure of the resulting network will depend on the choice of $e(k', k)$. A uniparametric assortative family can be obtained from

$$e(k', k) = q(k)q(k') + r\sigma_q^2 m(k, k'), \tag{63}$$

where σ_q is the standard deviation of the distribution $q(k)$, $m(k, k')$ is any symmetric matrix that has all rows and columns sums zero and is normalized, and the parameter r is the assortative coefficient.

[j]The excess degree is defined as $k_e = k - 1$.

3.5.5. *Modeling clustered networks*

When it was realized that correlations were unavoidable in an accurate characterization of real networks, most modeling efforts merely focused on the reproduction of two point correlations typified by the average nearest neighbors degree. This finds a justification in the fact that many models are assumed to observe the Markovian property, not only because analytic analysis simplifies but also because several real networks, such as the Internet at the Autonomous System level, indeed share this attribute.[55] These systems are those whose topology is completely defined by the degree distribution $P(k)$ and the first conditional probability $P(k \mid k')$, so that all higher-order correlations can be expressed as a function of these two. Some examples of these types of models will be discussed in the following subsection, and the analytic expression for the degree-dependent clustering coefficient will be provided there along with the ANND function. Nonetheless, all these Markovian models fail to maintain clustering in the thermodynamic limit. An independent modeling of clustering is thus required.

The simplest more general approach follows the philosophy of the configuration model, which gives maximally random networks with a given degree distribution $P(k)$. Instead of fixing $P(k)$, one could fix the function $P(k, k')$ so to construct a network with an expected two vertices degree correlations and otherwise maximally random. It can be demonstrated that the clustering of these networks again vanishes in the thermodynamic limit without exception. However, scale-free networks with divergent second moment deserve special attention once more. The decay of their clustering with the increase of the network size is so slow that relatively large networks with an appreciable high cohesiveness can be obtained.

Growing linear preferential attachment models also yield vanishing $\bar{c}(k)$ in the thermodynamic limit, from which new variations are needed in order to recreate the empirically observed values. As an illustrative example of the prescriptions that have been used to generate clustering in scale-free growing networks, one of the proposed models[95] reproduces a large clustering coefficient by adding nodes which connect to the two extremities of a randomly chosen network edge, thus forming a triangle. The resulting network has the power-law degree distribution of the Barabási-Albert model $P(k) \sim k^{-3}$, with $\langle k \rangle = 4$, and since each new vertex induces the creation of at least one triangle, the model generate networks with finite clustering coefficient. A generalization on this model[88] which allows to tune the average degree to $\langle k \rangle = 2m$, with m an even integer, considers new nodes connected to the ends of $m/2$ randomly selected edges. Two vertices and three vertices correlations can be calculated analytically through a rate equation formalism. The average nearest neighbors degree is again equal to the one obtained for the Barabási-Albert model, which indicates a lack of two vertices correlations. On the other hand, the clustering spectrum is here finite in the infinite size limit and scales as k^{-1},

$$\bar{c}(k) = \frac{2k - m}{k(k-1)}, \tag{64}$$

and the overall clustering coefficient for large m is

$$C(m) \simeq 2m^2 - 3m - 4/3 + 2m^2(2-m)\ln\frac{m}{m-1}. \tag{65}$$

Bipartite representations[7] constitute a special case since they provide high levels of clustering by construction. In bipartite networks, two types of nodes are present, such as for instance actors and films in the collaboration network of cinematographic productions. Links associate nodes in one category with nodes in the second, in the previous example, actors with films. The one-mode projection only preserves one of the two kinds of nodes connected among them whenever they were linked to the same second type node in the original bipartite composition, say only actors are preserved in the one-mode projection and linked among them whenever they play in the same movie. It is clear that this construction will produce fully connected subsets of actors appearing in the same films, so that the number of triangles in the network, and so the clustering, will be very high. On the other hand, nothing can be said about the dependence of the clustering with the degree and each pattern must be evaluated separately. Indeed, most social networks are represented as the one-mode projection of originally bipartite graphs. Then, the high levels of clustering measured in those networks are strongly affected by the network construction.

3.5.6. *Random graphs with attributes*

Aside partial models, several works attempt to establish a general framework for understanding and modeling correlations. Most of them are based on breaking the similarity of nodes by the introduction of a new stochastic characterization where vertices may come in different types. All these models generate ensembles of random networks which are able to reproduce a wide range of asymptotic topological properties, including different classes of correlation behaviors.

3.5.7. *Hidden color models*

The idea of inhomogeneity in the characterization of vertices is at the heart of the transition from regular latices to random graphs, where vertices have no longer a predefined degree but a stochastic one described by a probabilistic distribution. A further sophistication leads to the so called inhomogeneous random graphs models, where vertices may come in different types and edges appear between the different classes with different probabilities. In this context, the first unifying theoretical doctrine is the hidden color formalism for the generation of colored degree-based sparse random networks.[96–98] Notice these graphs should not be confused with the *colored random graph*.[45]

Graphs with hidden colors are constructed on the basis of the classical configuration model and hence are also a static class of models. The key idea is to define a color space $l = \{1, \ldots, l_a, \ldots, L\}$ and to assign one of these colors to each vertex's edge end or stub. Then, the coloring of a vertex i is given by $k_{li} = (k_{1i}, \ldots, k_{Li})$, where the number of stubs k_{ai} of a given color a is got from the colored degree distribution p_l defining the relative frequencies of vertices with different colored degrees. Finally, the color preference matrix $T_{L\times L}$ controls the relative abundance of edges between color pairs. The resulting ensemble of stub-colored graphs is well-defined if the coloring is considered unobservable. Hence, the coloring can be seen as a set of hidden variables introduced with the purpose of inducing a nontrivial correlation structure in the resulting graphs.

This general framework allows the analytical calculation, in the thermodynamic limit, of global and local properties for a large class of models, which are seen to contrast to those of standard degree-driven random graph (DRG) models. Edge correlations are studied through the generating function formalism by counting the expected number of triangles or three-cycles, $n_\triangle$, wedges or three-chains, $n_\wedge$, edges or two-chains, $n_|$ and m-chains. While the result for $\langle n_\wedge\rangle$ and $\langle n_|\rangle$ is identical to that obtained for plain degree-driven models, $\langle n_\wedge\rangle = \mathrm{N}E/2$ and $\langle n_|\rangle = N\langle l\rangle/2$ (where $\langle l\rangle = \sum_{a=1}^{L}\sum_l p_l l_a$), the non-colored number of triangles and k-chains are found to be different. For instance, in the case of triangles:

$$\langle n_\triangle\rangle_{HC} = \frac{(\mathrm{T}E)^3}{6} \tag{66}$$

$$\langle n_\triangle\rangle_{DRG} = \frac{\mathrm{E}^3}{6\langle l\rangle^3}, \tag{67}$$

where E is the matrix of second order combinatorial moments, $E \equiv E_{ab} =$

$\partial_a \partial_b \hat{p}_l(x = 1)$, with $\hat{p}_l$ the Laplace transform of p_l. Thus, for the degree-driven random graphs one has $\langle n_\triangle \rangle_{DRG} = \langle \wedge \rangle^3 / [6 \langle n_| \rangle^3]$, a relation which is absent in the hidden colors scenario.

The clustering coefficient C can also be computed from the count of the expected number of triangles and three-chains:

$$C_{HC} = \frac{(\mathrm{T}E)^3}{\mathrm{N}E} \tag{68}$$

$$C_{DGR} = \frac{(E)^2}{N\langle l \rangle^3}. \tag{69}$$

Although C_{HC} indeed scales as $\mathcal{O}(N^{-1})$, the finite quantity NC_{HC} has a nontrivial dependence on the color preference matrix T, an example of the increased correlation possibilities of hidden color models over DRG models.

3.5.8. *Fitness or hidden variables models*

A powerful and systematic subclass of the family of models described above is introduced as a class of correlated random networks with fitness or hidden variables.[58] Again, a hidden variable h, which can be defined in a discrete or a continuous space, plays the role of a tag assigned to the vertices, and completely determines the topological properties of the network through their probability distribution and the probability to connect pairs of vertices.

The procedure, which generates correlated undirected random networks without loops or multiple edges, is as follows:

- Each vertex i is assigned a variable h_i, independently drawn from a probability distribution $\rho(h)$.
- An undirected edge is created between a pair of vertices i and j following a connection probability $r(h_i, h_j)$, where $r(h, h') \geq 0$ is a symmetric function of h and h'.

The resulting networks are Markovian at the hidden variable level, which makes possible the calculation of analytical expressions for the most important structural properties, such as the degree distribution, the ANND function for two vertices correlations, and the clustering coefficient for three vertices correlations. The clue is in the conditional probability (the propagator) $g(k|h)$ that a vertex with initial hidden variable h ends up connected to other k vertices, which enables to write expressions in the degree-space as a function of distributions in the hidden variables space. For instance,

$$P(k) = \sum_h g(k|h)\rho(h) \tag{70}$$

$$\langle k\rangle = \sum_k kP(k) = \sum_h \bar{k}(h)\rho(h), \tag{71}$$

where $\bar{k}(h) = \sum_k kg(k|h)$ is the average degree of nodes with hidden variable h. The generating function formalism can be applied to find an explicit expression for the propagator:

$$\ln \hat{g}(z|h) = N\sum_{h'} \rho(h')\ln[1-(1-z)r(h,h')]. \tag{72}$$

Even without solving this equation, one can find that:

$$\bar{k}(h) = N\sum_{h'} \rho(h')r(h,h') \tag{73}$$

$$\langle k\rangle = N\sum_{h,h'} \rho(h)r(h,h')\rho(h'), \tag{74}$$

and these results are valid for sparse and non-sparse networks.

Pair degree correlations can be calculated as

$$\bar{k}_{nn}(k) = 1 + \frac{1}{P(k)}\sum_h g(k|h)\rho(h)\bar{k}_{nn}(h) \tag{75}$$

$$\bar{k}_{nn}(h) = \sum_{h'} \bar{k}(h')p(h'|h), \tag{76}$$

where $\bar{k}_{nn}(h)$ is the ANND of a vertex of hidden variable h.

The degree dependent clustering is

$$\bar{c}(k) = \frac{1}{P(k)}\sum_h \rho(h)g(k|h)c_h,\ \ k = 2,3,\dots \tag{77}$$

$$c_h = \sum_{h',h''} p(h'|h)r(h',h'')p(h''|h), \tag{78}$$

with c_h the clustering coefficient of a vertex h.

Furthermore, this analysis provides a new algorithm for the construction of random networks with a correlation structure determined *a priori*:

- Assign to each vertex i an integer random variable $\tilde{k}_i$, $i = 1,\dots,N$, drawn from the theoretical probability distribution $P_t(k)$.
- For each pair of vertices i and j, draw an indirect edge with probability $r(\tilde{k}_i,\tilde{k}_j) = \langle k\rangle P_t(\tilde{k}_i,\tilde{k}_j)/NP_t(\tilde{k}_i)P_t(\tilde{k}_j)$.

In the large-k limit, the degree structure of the ensuing network will be distributed according to the probability $P_t(k)$, with correlations given by $P_t(k,k')$.

Despite its static character, another of the main achievements of this general approach concerns its application to the mapping of growing networks into a particular kind of hidden variables model, where the hidden variable associated to each vertex corresponds to its injection time. All known results for growing models can be recovered from the hidden variables formalism.

3.5.9. *Fitness and preferential attachment models*

The original fitness model[99] appeared as an attempt to loosen the preferential attachment rule in the Barabási-Albert model so that degree distributions with exponents different from 3 could be obtained.

The fitness associated to each vertex is defined as a stochastic parameter η_i picked out from a probability distribution $\rho(\eta)$. The fitness embodies properties, different from the degree, that may also influence the probability Π of node i of gaining new edges, which is computed as

$$\Pi(k_i, \eta_i) = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \tag{79}$$

Even if the distribution $\rho(\eta)$ is the simplest one, that is uniform in the interval $[0, 1]$, the model generates a network displaying a non-trivial degree distribution, and for some more complex alternatives the model also reproduces structural correlations.

Inspired by the idea of fitness, a new mechanism leading also to scale-free networks is obtained if the preferential attachment rule in terms of the degree is eliminated and only the fitness remain.[19] Since the fitness is a non-evolving quantity, the network can then be built as static. Although previous in time, this intrinsic fitness model is a particular example of the general class of models with hidden variables, where the fitness is distributed exponentially and nodes are joined whenever the sum of the fitness of the endpoints is larger than a given constant threshold ζ, so that $r(h, h')$ is the Heaviside step function $\Theta(x)$:

$$\rho(h) = e^{-h}, \; h \; \epsilon \; [0, \infty] \tag{80}$$

$$r(h, h') = \Theta(h + h' - \zeta). \tag{81}$$

Within the hidden variables formalism, analytical expressions can be computed for the main properties of the model.[58] Two point correlations are disassortative:

$$\bar{k}_{nn}(k) = 1 + \frac{N^2 e^{-\zeta}}{k} \left[1 + \zeta + \ln\left(\frac{k}{N}\right) \right] \Theta_k(N e^{-\zeta}, N). \tag{82}$$

The clustering coefficient is

$$\bar{c}(k) = \Theta_k(Ne^{-\zeta}, Ne^{-\zeta/2}) \tag{83}$$

$$+ \frac{N^2 e^{-\zeta}}{k^2}\left[1 + \zeta + 2\ln\left(\frac{k}{N}\right)\right]\Theta_k(Ne^{-\zeta/2}, N), \tag{84}$$

which reflects the fact that the clustering is equal to its maximum value 1 for all vertices with $h < \zeta/2$. On the other hand, for $Ne^{-\zeta/2} \leq k \leq N$ it decreases as k^{-2} but modulated by a logarithmic correction term.

3.6. *Outlook*

In general, as we have shown in the previous sections, uncorrelated random graphs do not match real networks, which indeed in most cases show non-trivial topological correlations encoding the properties of the underlying hierarchical architecture or community structure. While uncorrelated random networks are greatly valuable to provide null hypotheses for network structures, correlated models can provide a more faithful image of reality. Moreover, any deep understanding of the ordering principles governing the formation and evolution of networks must take into account correlations, clustering and other topological attributes. Despite the intense research activity witnessed by the various results reported in this chapter, several directions have yet to be fully explored. The characterization and modeling of correlations in directed networks is surely at an early stage due to various technical complications both in the mathematical tools and the data gathering. The effect of correlations on networks physical properties has been analyzed only in a handful of systems. Finally, the origin and meaning of correlations spur also the question of which phenomena and dynamical aspects rule the development of these features. The physics of the dynamical processes occurring on networks (traffic flows, communication transmission etc.) has as well a role in determining specific correlation patterns. It is therefore important to start bridging the topological properties of networks with the dynamics acting on them, finding the interplay of these various elements and their interaction rules. In this book, a chapter is devoted to recent studies on weighted networks and the interaction among topological features and weighted quantities representing the interactions or traffic carried by the edges.

References

1. D. B. West, *Introduction to Graph Theory* (2^{nd} *Edition*), Prentice Hall (2001).
2. B. Bollobás, *Graph Theory, An Introductory Course*, Springer-Verlag, New York, Heidelberg, Berlin (1979).
3. R. Diestel, *Graph Theory*, Springer-Verlag, New York, Heidelberg, Berlin (1997-2000).
4. G. Caldarelli, *Scale-free Networks*, Oxford University Press, Oxford (2007).
5. S. Milgram, *Psychology Today*, **2**, 56-63 (1967).
6. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
7. S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge (1994).
8. R. Albert and A.-L. Barabási, *Review of Modern Physics* **74**, 47-97 (2002).
9. S. N. Dorogovtsev and J. F. F. Mendes, *Advances in Physics* **51**, 1079-1187 (2002).
10. M. E. J. Newman, *SIAM Review* **45**, 167-256 (2003).
11. S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford (2003).
12. R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*, Cambridge University Press, Cambridge (2004).
13. P. Erdős and A. Rényi, *Publicationes Mathematicae* **6**, 290-297 (1959).
14. P. Erdös and A. Renyi, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17-61 (1960).
15. A.-L. Barabási and R. Albert, *Science* **286**, 509-511 (1999).
16. E. A. Bender and E. R. Canfield, *Journal of Combinatorial Theory A* **24**, 296-307 (1978).
17. M. Molloy and B. Reed, *Random Struct. Algorithms* **6**, 161 (1995).
18. K.-I. Goh, B. Khang and D. Kim, *Physical Review Letters* **87**, 278701 (2001).
19. G. Caldarelli, A. Capocci, P. De Los Rios and M.A. Muñoz, *Physical Review Letters* **89**, 258702 (2002).
20. M. Boguña and R. Pastor-Satorras, *Physical Review E* **68**, 036112 (2003).
21. V. D. P. Servedio G. Caldarelli and P. Buttà, *Physical Review E* **70**, 056126 (2004).
22. A. Fronczak, P. Fronczak and J. A. Holyst, *Physical Review E* **70**, 056110 (2004).

23. M. Boguña, R. Pastor-Satorras, A. Díaz-Guilera and A. Arenas, *Physical Review E* **70**, 056122 (2004).
24. I. Ispolatov, A. Yuryev, I. Mazo and S. Maslov, *Nucleic Acid Research* **33**, 3629-3635 (2005).
25. H. Tangmunarunkit, J. Doyle, R. Govindan, S. Jamin, S. Shenker and W. Willinger, *Computer Communication Review* **31**, 7-10 (2001).
26. G. Zipf, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, MA (1949).
27. R. Albert, H. Jeong and A.-L. Barabási, *Nature* **401**, 130-131 (1999).
28. P. L. Krapivski, S. Redner and F. Leyvraz, *Physical Review Letters* **85**, 4629 (2000).
29. J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan and A. S. Tomkins, *Lecture Notes in Computer Science* **1627**, 1-18 (1999).
30. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. S. Tomkins and E. Upfal, Stochastic models for the Web graph, in *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (FOCS)*, 57-65 (2000).
31. S. Ohno, *Evolution by Gene Duplication*, Springer, Berlin (1970).
32. A. Wagner, *Molecular Biological Evolution* **18**, 1283-1292 (2001).
33. A. Vázquez, *Europhysics Letters* **54**, 430 (2001).
34. A. Vázquez, A. Flammini, A. Maritan and A. Vespignani, *ComPlexUs* **1**, 38-44 (2003).
35. R. V. Solé, R. Pastor-Satorras, E. D. Smith and T. Kepler, *Advances in Complex Systems* **5**, 43 (2002).
36. G. Bianconi and A.-L. Barabási, *Europhysics Letters* **54**, 436-442 (2001).
37. G. Bianconi and A.-L. Barabási, *Physical Review Letters* **86**, 5632-5635 (2001).
38. A. Capocci, G. Caldarelli and P. De Los Rios, *Physical Review E* **68**, 047101 (2003).
39. A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, *Proc. Natl. Acad. Sci. (USA)* **101**, 3747-3752 (2004).
40. A. Barrat, M. Barthélemy and A. Vespignani, *Physical Review Letters* **92**, 228701 (2004).
41. M. E. J. Newman, *Physical Review Letters* **89**, 208701 (2002).
42. M. E. J. Newman, *Physical Review E* **67**, 026126 (2003).
43. M. Boguñá and R. Pastor-Satorras, *Physical Review E* **66**, 047104 (2002).
44. R. Pastor-Satorras, M. Rubi and A. Díaz-Guilera (editors), *Proceedings of the Conference "Statistical Mechanics of Complex Networks"*, Springer (2003).
45. O. Frank and D. Strauss, *J. Amer. Stat. Assoc.* **81**, 832 (1986).
46. S. Maslov and K. Sneppen, *Science* **296**, 910 (2002).
47. S. Maslov, K. Sneppen and A. Zaliznyak, *Phys. A* **333**, 529 (2004).
48. R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Physical Review Letters* **87**, 258701 (2001).
49. A. Barrat and M. Weigt, *Eur. Physics J. B* **13**, 547 (2000).

50. A. Vázquez, R. Pastor-Satorras and A. Vespignani, *Physical Review E* **65**, 066130 (2002).
51. P. M. Gleiss, P. F. Stadler, A. Wagner and D. A. Fell, *Adv. Comp. Sys.* **4**, 207 (2001).
52. A. Fronczak, J. A. Holyst, M. Jedynak and J. Sienkiewicz, *Physica A* **316**, 688 (2002).
53. G. Caldarelli, R. Pastor-Satorras and A. Vespignani, *Eur. Phys. J. B* **38**, 183 (2004).
54. G. Bianconi and A. Capocci, *Physical Review Letters* **90**, 078701 (2003).
55. G. Bianconi, G. Caldarelli and A. Capocci, *Physical Review E* **71**, 066116 (2005).
56. M. Boguñá, R. Pastor-Satorras and A. Vespignani, *Physical Review Letters* **90**, 028701 (2003).
57. M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Physical Review E* **64**, 026118 (2001).
58. M. Boguñá and R. Pastor-Satorras, *Physical Review E* **68**, 036112 (2003).
59. S. N. Dorogovtsev, *Physical Review E* **69**, 027104 (2004).
60. M. E. J. Newman, in *Handbook of Graphs and Networks: From the Genome to the Internet*, Ed. S. Bornholdt and H. G. Schuster, Wiley-VCH, Berlin (2003), p. 35.
61. Z. Burda, J. Jurkiewicz and A. Krzywicki, *Physical Review E* **70**, 026106 (2004).
62. E. Ravasz and A.-L. Barabási, *Physical Review E* **67**, 026112 (2003).
63. A. Vázquez, M. Boguñá, Y. Moreno, R. Pastor-Satorras and A. Vespignani, *Physical Review E* **6**, 046111 (2003).
64. A. Capocci, G. Caldarelli and P. De Los Rios, *Physical Review E* **68**, 047101 (2003).
65. G. F. Davis, M. Yoo and W. E. Baker, *Strategic Organization* **1**, 301 (2003).
66. M. E. J. Newman, *Physical Review E* **64**, 016131 (2001).
67. M. E. J. Newman, *Physical Review E* **64**, 016132 (2001).
68. L. A. N. Amaral, A. Scala, M. Barthélemy and H. E. Stanley, *Proc. Nat. Acad. Sci.* **97**, 11149 (2000).
69. M. E. J. Newman, S. Forrest and J. Balthrop, *Physical Review E* **66**, 035101 (2002).
70. H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai, *Nature* **411**, 41 (2001).
71. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, *Nature* **407**, 651 (2000).
72. N. D. Martinez, *Ecol. Monogr.* **61**, 367 (1991).
73. R. Albert, H. Jeong and A.-L. Barabási, *Nature* **401**, 130 (1999).
74. V. Colizza, A. Flammini, A. Maritan and A. Vespignani, *Physica A* **352**, 1 (2005).
75. M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera and A. Arenas, *Physical Review E* **70**, 056122 (2004).
76. S. Garfinkel, *PGP: Pretty Good Privacy*, O'Reilly and Associates, Cambridge MA (1994).
77. W. Stallings, *BYTE* **20**, 161 (1995).

78. M. Faloutsos, P. Faloutsos and C. Faloutsos, *Comput. Commun. Review* **29**, 251 (1999).
79. A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, *Proc. Nat. Acad. Sci.* **101**, 3747 (2004).
80. M. A. Serrano and M. Boguñá, *Physical Review E* **68**, 015101(R) (2003).
81. A. Bekessy, P. Bekessy and J. Komlos, *Stud. Sci. Math. Hungar.* **7**, 343 (1972).
82. E. A. Bender and E. R. Canfield, *J. Comb. Theory A* **24**, 296 (1978).
83. M. Molloy and B. Reed, *Rand. Str. and Algth.* **6**, 161 (1995).
84. J. Park and M. E. J. Newman, *Physical Review E* **68**, 026122 (2003).
85. M. Boguñá, R. Pastor-Satorras and A. Vespignani, *Eur. Physics J. B* **38**, 205 (2004).
86. M. Catanzaro, M. Boguñá and R. Pastor-Satorras, *Physical Review E* **71**, 027103 (2005).
87. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
88. A. Barrat and R. Pastor-Satorras, *Physical Review E* **71**, 036127 (2005).
89. G. Szabó, M. Alava and J. Kertész, *Physical Review E* **67**, 056102 (2003).
90. S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, *Physical Review Letters* **85**, 4633 (2000).
91. S. N. Dorogovtsev and J. F. F. Mendes, *Europhysics Letters* **52**, 33 (2000).
92. R. Albert and A.-L. Barabási, *Physical Review Letters* **85**, 5234 (2000).
93. P. L. Krapivsky and S. Redner, *Physical Review E* **63**, 066123 (2001).
94. D. Strauss, *SIAM Review* **28**, 513 (1986).
95. S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, *Physical Review E* **63**, 062101 (2001).
96. B. Söderberg, *Physical Review E* **66**, 066121 (2002).
97. B. Söderberg, *Physical Review E* **68**, 015102(R) (2003).
98. B. Söderberg, *Physical Review E* **68**, 026107 (2003).
99. G. Bianconi and A.-L. Barabási, *Europhysics Letters* **54**, 436 (2001).
100. S. L. Pimm, *Food Webs*, The University of Chicago Press, 2nd edition, Chicago (2002).
101. A. E. Krause, K. A. Frank, D. M. Mason, Z. R. E. Ulanowic and W. W. Taylor, *Nature* **426**, 282 (2003).
102. M. Granovetter, *American Journal of Sociology* **78**, 1360 (1973).
103. R. Guimerà, S. Mossa, A. Turtschi and L. A. N. Amaral, *Proceedings National Academy Science (USA)* **102**, 7794 (2005).
104. R. Guimerà and L. A. N. Amaral, *European Physical Journal B* **38**, 381 (2004).
105. W. Li and X. Cai, *Physical Review E* **69**, 046106 (2004).
106. C. Li and G. Chen, cond-mat/0311333.
107. D. Garlaschelli, S. Battiston, M. Castri, V. D. P. Servedio and G. Caldarelli, *Physica A* **350**, 491 (2004).
108. E. Almaas, B. Kovács, T. Viscek, Z. N. Oltvai and A.-L. Barabási, *Nature* **427**, 839 (2004).
109. S. H. Yook, H. Jeong, A.-L. Barabási and Y. Tu, *Physical Review Letters* **86**, 5835 (2001).

110. J.-P. Onnela, J. Saramäki, J. Kertész and K. Kaski, *Physical Review E* **71**, 065103 (2005).
111. B. Derrida and H. Flyvbjerg, *Journal Physics A* **20**, 5273 (1987).
112. M. Barthélemy, B. Gondran and E. Guichard, *Physica A* **319**, 633 (2003).
113. S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai and N. Wang, *Nature* **429**, 180 (2004).
114. L. Hufnagel, D. Brockmann and T. Geisel, *Proceedings National Academy of Science (USA)* **101**, 15124 (2004).
115. G. Chowell, J. M. Hyman, S. Eubank and C. Castillo-Chavez, *Physical Review E* **68**, 066102 (2003).
116. S. Zhou and R. J. Mondragon, The missing links in the BGP-based AS connectivity maps, *PAM2003 — The Passive and Active Measurement Workshop* (http://www.pam2003.org), San Diego, USA, April 2003.
117. A. de Montis, M. Barthélemy, A. Chessa and A. Vespignani, *Structure of Inter-cities traffic: A weighted network analysis*, submitted to *Env. Plan. J. B* (2005).
118. A. Barrat, M. Barthélemy and A. Vespignani, *Journal of Statistical Mechanics*, P05003 (2005).
119. A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert and T. Vicsek, *Physica A* **311**, 590 (2002).
120. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, *Nature* **407**, 651 (2000).
121. A. Wagner and D. A. Fell, *Proceedings Royal Society London B* **268**, 1803 (2001).
122. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A.-L. Barabási, *Science* **297**, 1551 (2002).
123. D. Zheng, S. Trimper, B. Zheng and P. M. Hui, *Physical Review E* **67**, 040102 (2003).
124. W. Jeżewski, *Physica A* **337**, 336 (2004).
125. E. Almaas, P. L. Krapivsky and S. Redner, *Physical Review E* **71**, 036124 (2005).
126. K. Park, Y.-C. Lai and N. Ye, *Physical Review E* **70**, 026109 (2004).
127. T. Antal and P. L. Krapivsky, *Physical Review E* **71**, 026103 (2005).
128. S. Wang and C. Zhang, *Physical Review E* **70**, 066127 (2004).
129. A. Barrat, M. Barthélemy and A. Vespignani, *Physical Review E* **70**, 066149 (2004).
130. A. Barrat, M. Barthélemy and A. Vespignani, *Lecture Notes in Computer Science* **3243**, 56 (2004).
131. R. V. R. Pandya, cond-mat/0406644.
132. S. H. Yook, H. Jeong and A.-L. Barabási, *Proceedings National Academy of Science (USA)* **99**, 13382 (2002).
133. M. Barthélemy, *Europhysics Letters* **63**, 915 (2003).
134. S. N. Dorogovtsev and J. F. F. Mendes, cond-mat/0408343.
135. S. N. Dorogovtsev, J. F. F. Mendes and A. N. N. Samukhin, *Physical Review E* **63**, 062101 (2001).
136. G. Bianconi, *Europhysics Letters* **71**, 1029-1035 (2005).

137. W.-X. Wang, B.-H. Wang, B. Hu, G. Yan and Q. Ou, *Physical Review Letters* **94**, 188702 (2005).
138. M. Li, J. Wu, D. Wang, T. Zhou, Z. Di and Y. Fan, cond-mat/0501665.
139. S. H. Strogatz, *Nature* **410** 268-276 (2001).
140. S. Bornholdt and H. G. Schuster (editors), *Handbook of Graphs and Networks - From the Genome to the Internet*, Wiley-VCH, Berlin (2002).
141. M. E. J. Newman, *European Physics Journal B* **38**, 321-330 (2004).
142. P. Holme, M. Huss and H. Jeong, *Bioinformatics* **19**, 532-538 (2003).
143. M. Boss, H. Elsinger, M. Summer and S. Thurner, cond-mat/0309582 (2003).
144. M. Girvan and M. E. J. Newman, *Proceedings of the National Academy of Sciences (USA)* **99**, 7821-7826 (2002).
145. G. W. Flake, S. Lawrence, C. L. Giles and F. M. Coetzee, *IEEE Computer* **35**, 66-71 (2002).
146. J.-P. Eckmann and E. Moses, *Proceedings of the National Academy of Sciences (USA)* **99**, 5825-5829 (2002).
147. H. Zhou and R. Lipowsky, Preprint (2005).
148. M. R. Garey and D. S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W. H. Freeman, New York (1979).
149. B. W. Kernighan and S. Lin, *The Bell System Tech. J* **49**, 291-307 (1970).
150. M. Fiedler, *Czechoslovak Mathematical Journal* **23**, 298-305 (1973).
151. S. Boettcher and A. G. Percus, *Physical Review E* **64** (2001).
152. R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt and A. Arenas, *Physical Review E* **68**, 065103 (2003).
153. P. Gleiser and L. Danon, *Advances in Complex Systems* **6**, 565-573 (2003).
154. A. Arenas, L. Danon, A. Díaz-Guilera, P. M. Gleiser and R. Guimerà, *European Physical Journal B* **38**, 373-380 (2004).
155. F. Radicchi, C. Castellano, F. Cecconi, V. Loreto and D. Parisi, *Proceedings of the National Academy of Sciences (USA)* **101**, 2658-2663 (2004).
156. C. Bron and J. Kerbosch, *Communications of the ACM*, 575-577 (1973).
157. R. Guimerà, M. Sales and L. N. A. Amaral, *Physical Review E* **70**, 025101 (2004).
158. U. Brandes, *Journal of Mathematical Sociology* **25**, 163-177 (2001).
159. M. E. J. Newman and M. Girvan, *Physical Review E* **69**, 026113 (2004).
160. S. Fortunato, V. Latora and M. Marchiori, *Physical Review E* **70**, 056104 (2004).
161. V. Latora and M. Marchiori, cond-mat/0402050 (2004).
162. J. Scott, *Social Network Analysis, A Handbook*, SAGE Publications (2000).
163. A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Upper Saddle River, NJ (1988).
164. J. Hopcroft, O. Khan, B. Kulis and B. Selman, *Proceedings of the National Academy of Sciences (USA)* **101**, 5249-5253 (2004).
165. J. P. Bagrow and E. M. Bollt, *Physical Review E* **72**, 046108 (2005).
166. M. E. J. Newman, *Physical Review E* **69**, 066133 (2004).
167. A. Clauset, M. E. J. Newman and C. Moore, *Physical Review E* **70**, 06111 (2004).

168. C. P. Massen and J. P. K. Doye, cond-mat/0412469 (2004).
169. J. Duch and A. Arenas, cond-mat/0501368 (2005).
170. S. Boettcher and A. G. Percus, *Physical Review Letters* **86**, 5211-5214 (2001).
171. B. Bollobás, *Modern Graph Theory*, Springer, New York (1998).
172. A. Pothen, *SIAM Journal on Matrix Analysis and Applications* **11**, 430-452 (1990).
173. A. Pothen, Graph partitioning algorithms with applications to scientific computing, in *Parallel Numerical Algorithms*, Kluwer Academic Press (1996).
174. L. Donetti and M. A. Muñoz, *Journal of Statistical Mechanics: Theory and Experiment*, P10012 (2004).
175. G. H. Golub and C. F. V. Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore (1996).
176. A. Capocci, V. D. P. Servedio, G. Caldarelli and F. Colaiori, in *Algorithms and Models for the Web-Graph: Third International Workshop, WAW 2004, Rome, Italy, October 16, 2004, Proceeedings*, Lecture Notes in Computer Science **3243**, 181-188 (2004).
177. M. Steyvers and J. B. Tenenbaum, cond-mat/0110012 (2001).
178. F. Wu and B. Huberman, *Eurpean Physics Journal B* **38**, 331-338 (2004).
179. J.-P. Eckmann, E. Moses and D. Sergi, *Proceedings of the National Academy of Sciences (USA)* **101**, 14333-14337 (2004).
180. H. Zhou, *Physical Review E* **67**, 041908 (2003).
181. H. Zhou, *Physical Review E* **67**, 061901 (2003).
182. H. Zhou and R. Lipowsky, *Lecture Notes in Computer Sciences* (in press) (2004).
183. M. Latapy and P. Pons, cond-mat/0412568 (2004).
184. J. H. Ward, *Journal of the American Statistical Association* **53**, 263-244 (1963).
185. J. Reichardt and S. Bornholdt, *Physical Review Letters* **93**, 218701 (2004).
186. M. Blatt, S. Wiseman and E. Domany, *Physical Review Letters* **76**, 3251-3254 (1996).
187. W. Tutte, *Proceedings of the London Mathematical Society (Third Series)* **13**, 743-768 (1963).
188. D. Knuth, *Communication of the ACM* **6**, 555-563 (1963).
189. International Symposium on Graph Drawing, `http://graphdrawing.org`
190. G. Di Battista, P. Eades, R. Tamassia and I. G. Tollis, *Graph Drawing. Algorithms for the Visualization of Graphs*, Prentice Hall (1999).
191. J. Michael and P. Mutzel (eds.), *Graph Drawing Software*, Mathematics and Visualization, Springer-Verlag (2003).
192. T. Kamada, *Visualizing Abstract Objects and Relations*, World Scientific (1989).
193. M. Kaufmann and D. Wagner (eds.), *Drawing Graphs: Methods and Models*, Lecture Notes in Computer Science **2025**, Springer-Verlag (2001).
194. Sugiyama, Kozo, *Graph Drawing and Applications for Software and Knowledge Engineers*, World Scientific (2002).

195. P. Eades, *Congressus Numerantium* **42**, 149-160 (1984).
196. U. Brandes, in *Drawing Graphs: Methods and Models*, Springer Lecture Notes in Computer Science, **2025**, M. Kaufmann and D. Wagner (eds.), Springer-Verlag (2001), pp. 71-86.
197. K. M. Hall, *Management Science* **17**, 219-229 (1970).
198. U. Brandes, P. Kenis and D. Wagner, *IEEE Transactions on Visualization and Computer Graphics* **9**, 241-253 (2003).
199. U. Brandes and S. Cornelsen, *Journal of Graph Algorithms and Applications* **7**, 181-201 (2003).
200. U. Brandes and D. Wagner, in *Special Issue on Graph Drawing Software*, Springer Series in Mathematics and Visualization, M. Jünger and P. Mutzel (eds.), Springer-Verlag (2003), pp. 321-340, `http://www.visone.de/`
201. R. Brockenauer and S. Cornelsen, in *Drawing Graphs: Methods and Models*, Lecture Notes in Computer Science, **2025**, M. Kaufmann and D. Wagner (eds.), Springer (2001), pp. 71-86.
202. V. Batagelj and M. Zaveršnik, *Generalized cores*, Preprint 799, Universtiy of Ljubljana (2002).
203. S. B. Seidman, *Social Networks* **5**, 269-287 (1983).
204. M. Baur, U. Brandes, M. Gaertler and D. Wagner, in *Proceedings of the 12th International Symposium on Graph Drawing (GD'04)*, Lecture Notes in Computer Science, **3383**, Springer (2005), pp. 43-48.
205. I. Alvarez-Hamelin, M. Gaertler, R. Görke and D. Wagner, Halfmoon – A new Paradigm for Complex Network Visualization TR 2005-29, Informatics, University Karlsruhe (2004).
206. S. Brin and L. Page, *Computer Networks and ISDN Systems* **30**, 107-117 (1998).
207. J. Kleinberg, *Journal of the ACM* **46**, 604-632 (1997).
208. S. Dill, R. Kumar, K. McCurley, S. Rajagopalan, D. Sivakumar and A. Tomkins, in *Proceedings of the* 27^{th} *VLDB Conference*, Morgan Kaufmann (2001), pp. 69-78.
209. M. Adler and M. Mitzenmacher, Tech. Rep. 00-39, University of Massachussets (2000).
210. P. Boldi and S. Vigna, in *WWW '04: Proceedings of the 13th International Conference on World Wide Web*, ACM Press (2004), pp. 595-602.
211. I. H. Witten, A. Moffat and T. C. Bell, *Managing gigabytes (2nd ed.): compressing and indexing documents and images*, Morgan Kaufmann Publishers Inc. (1999).
212. J. Abello, P. M. Pardalos and M. G. C. Resende, *Handbook of massive data sets*, Kluwer Academic Publishers (2002).
213. F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA (1969).
214. M. Mitzenmacher, *Internet Mathematics* **1**, 2 (2003).
215. J. Cho and H. Garcia-Molina, Parallel crawlers, in *Proc. of the 11th International World–Wide Web Conference* (2002).
216. P. Boldi, B. Codenotti, M. Santini and S. Vigna, Ubicrawler: A scalable fully distributed web crawler (2002).
217. J. Vitter and E. Shriver, *Algorithmica* **12**, 107-114 (1994).

218. J. Vitter and E. Shriver, *Algorithmica* **12**, 148-169 (1994).
219. The Stanford webbase project, http://www-diglib.stanford.edu/∼testbed/doc2/WebBase
220. Cyvellance, 2000, http://www.cyvellance.com.
221. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, in *Proc. of the 8th WWW Conference*, Elsevier North-Holland, Inc. (1999), pp. 1481-1493.
222. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, S. Stata, A. Tomkins and J. Wiener, *Computer Networks* **33**, 309-320 (2000).
223. G. Pandurangan, P. Raghavan and E. Upfal, in *Proc. of the 8th Annual International Conference on Combinatorics and Computing (COCOON)*, Lecture Notes in Computer Science, **2387**, Springer-Verlag (2002), pp. 330-339.
224. S. Raghavan, Personal communication (2002).
225. J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, in *Proc. International Conference on Combinatorics and Computing*, Lecture Notes in Computer Science, **1627**, Springer-Verlag (1999), pp. 1-18.
226. L. Laura, S. Leonardi, G. Caldarelli and P. De Los Rios, A multi-layer model for the webgraph, *On-line Proceedings of the 2nd International Workshop on Web Dynamics* (2002). http://www.dcs.bbk.ac.uk/webDyn2/onlineProceedings.html.
227. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, in *Proc. of 41st Symposium on Foundations of Computer Science (FOCS)*, IEEE Computer Society (2000), pp. 57-65.
228. D. Pennock, G. Flake, S. Lawrence, E. Glover and C. Giles, *Proceedings of the National Academy of Sciences (USA)* **99**, 5207-5211 (2002).
229. B. Bollobas and O. Riordan, *Internet Mathematics* **1**, 1-35 (2003).
230. J. Sibeyn, J. Abello and U. Meyer, Heuristics for semi-external depth first search on directed graphs. in *Proceedings of the fourteenth annual ACM symposium on Parallel algorithms and architectures* (2002).
231. D. Donato, L. Laura, S. Leonardi and S. Millozzi, Tech. Rep. D13, COSIN European Research Project, 2004. http://www.cosin.org.
232. A. Walker, *ACM Trans. Mathematical Software* **3**, 253-256 (1977).
233. D. E. Knuth, *Seminumerical Algorithms*, third ed., Vol. 2 of *The Art of Computer Programming*, Addison-Wesley, Reading, Massachusetts (1997).
234. T. H. Cormen, C. E. Leiserson and R. L. Rivest, *Introduction to algorithms*, 6th ed., MIT Press and McGraw-Hill Book Company (1992).
235. R. E. Tarjan, *SIAM Journal on Computing* **1**, 146-160 (1972).
236. M. Garey and D. Johnson, *Computers and Intractability*, W. H. Freeman (1979).
237. T. H. Haveliwala, *Efficient computation of PageRank*, Tech. rep., Stanford University (1999).
238. R. Kraft, E. Hastor and R. Stata, Timelinks: Exploring the link structure of the evolving web., in *Second Workshop on Algorithms and Models for the Web-Graph (WAW2003)* (2003).
239. R. Govindan and H. Tangmunarunkit, *IEEE INFOCOM 2000*, 1371–1380, Tel Aviv, Israel, June 2000, IEEE.

240. University of Oregon Route Views Project, http://www.routeviews.org/.
241. Router-Level Topology Measurements: "Cooperative Association for Internet Data Analysis". http://www.caida.org/tools/measurement/skitter/router_topology/.
242. "Distributed Internet Measurements and Simulations". http://www.netdimes.org.
243. Y. Shavitt and E. Shir, Dimes: Let the internet measure itself, Preprint cs.NI/0506099 (2005).
244. B. W. Bush, C. R. Files and D. R. Thompson, Empirical characterization of infrastructure networks, Technical Report LA-UR-01-5784, Los Alamos National Laboratory (2001).
245. D. J. Watts, *Small worlds: the dynamics of networks between order and randomness*, Princeton University Press, New Jersey (1999).
246. A. Broido and K. Claffy, Internet topology: connectivity of IP graphs, in *SPIE International symposium on Convergence of IT and Communication*, Denver, CO (2001).
247. S. Mossa, M. Barthélemy, H. E. Stanley and L. A. N. Amaral, *Physical Review Letters* **88**, 138701 (2002).
248. M. Molloy and B. Reed, A Critical Point for Random Graphs with a Given Degree Sequence, *Random Structures and Algorithms* **6**, 161-180 (1995); R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, *Physical Review Letters* **85**, 4626-4628 (2000); D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Physical Review Letters* **85**, 5468-5471 (2000).
249. R. Pastor-Satorras and A. Vespignani, *Physical Review Letters* **86**, 3200 (2001); R. Pastor-Satorras and A. Vespignani, *Physical Review E* **63**, 066117 (2001).
250. B. Bollobás, *Random Graphs*, Academic Press, London (1985).
251. H. Tangmunarunkit, J. Doyle, R. Govindan, S. Jamin, S. Shenker and W. Willinger, *Computer Communication Review* **31**, 7-10 (2001).
252. R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Physical Review Letters* **87**, 258701 (2001).
253. D. Magoni and J.-J. Pansiot, *ACM SIGCOMM Computer Communication Review* **31**, 26-37 (2001).
254. H. Chang, S. Jamin and W. Willinger, Inferring AS-level internet topology from router-level path traces, in *Proceedings of SPIE ITCom 2001*, Denver, CO, August 2001.
255. A. Vázquez, R. Pastor-Satorras and A. Vespignani, *Physical Review E* **65**, 066130 (2002).
256. H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker and W. Willinger, *Computer Communication Review* **32**, 76 (2002).
257. W. Willinger, R. Govindan, S. Jamin, V. Paxson and S. Shenker, *Proceedings of the National Academy Science (USA)* **99**, 2573-2580 (2002).
258. Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker and W. Willinger, in *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, Volume 2*, IEEE Computer Society Press (2002), pp. 608-617.

259. T. Bu and D. Towsley, On distinguishing between Internet power law topology generators, in *Proceedings of INFOCOM* (2002).
260. H. Burch and B. Cheswick, *IEEE computer* **32**, 97-98 (1999).
261. L. Li, D. Alderson, W. Willinger and J. C. Doyle, A First-Principles Approach to Understanding the Internet's Router-level Topology, in *Proceedings of ACM Sigcomm*, Portland (2004).
262. A. Lakhina, J. W. Byers, M. Crovella and P. Xie, Sampling biases in IP topology measurements, *Proc. IEEE INFOCOM* (2003).
263. T. Petermann and P. De Los Rios, *Eur. Physics J. B* **38**, 201-204 (2004).
264. J.-L. Guillaume and M. Latapy, Relevance of Massively Distributed Explorations of the Internet Topology: Simulation Results, *IEEE* 24^{th} *Infocom'05*, Miami, USA (2005).
265. A. Clauset and C. Moore, *Physical Review Letters* **94**, 018701 (2005).
266. P. Barford, A. Bestavros, J. Byers and M. Crovella, On the marginal utility of deploying measurement infrastructure, in *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001*, CA (2001).
267. L. Dall'Asta, I. Alvarez Hamelin, A. Barrat, A. Vázquez and A. Vespignani, analysis. *Lecture Notes in Computer Science* **3405**, 140-153 (2005).
268. L. Dall'Asta, I. Alvarez Hamelin, A. Barrat, A. Vázquez and A. Vespignani, *Physical Review E* **71**, 036135 (2005).
269. R. Govindan and A. Reddy, in *Proceedings of IEEE INFOCOM'97*, April 1997, p. 850.
270. H. Chang, R. Govindan, S. Jamin, S. Shenker and W. Willinger, ACM SIGMETRICS (2002), pp. 280-281.
271. J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat and A. Vespignani, k-core decomposition: a tool for the analysis of large scale Internet graphs, Preprint cs.NI/0511007 (2005).
272. S. N. Dorogovtsev, A. V. Goltsev and J. F. F. Mendes, k-core percolation and k-core organization of complex networks, cond-mat/0509102 (2005).
273. S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt and E. Shir (2005). `http://www.cs.huji.ac.il/~kirk/Jellyfish_Dimes.ppt`.
274. J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat and A. Vespignani, k-core decomposition: a tool for the visualization of large scale networks, Preprint cs.NI/0504107 (2005).
275. LArge NETwork VIsualization tool. `http://xavier.informatics.indiana.edu/lanet-vi/`
276. S. H. Strogatz, *Nature* **410**, 268 (2001).
277. J. H. Lawton, in *Ecological Concepts*, Ed. J. M. Cherret, Blackwell Scientific, Oxford, 43 (1989).
278. S. L. Pimm, *Food Webs*, Chapman & Hall, London (1982).
279. J. E. Cohen, F. Briand and C. M. Newman, *Community Food Webs: Data and Theory*, Springer, Berlin (1990).
280. D. Garlaschelli, G. Caldarelli and L. Pietronero, *Nature* **423**, 165 (2003).
281. D. Garlaschelli, *European Physical Journal B* **38**(2), 277 (2004).

282. C. Jeffry, *An Introduction to Plant Taxonomy*, Cambridge University Press, New York (1982).
283. R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy*, Freeman, San Francisco (1963).
284. J. C. Willis, *Age and Area*, Cambridge University Press, Cambridge (1922).
285. B. Burlando, *Journal of Theoretical Biology* **146**, 99 (1990).
286. B. Burlando, *Journal of Theoretical Biology* **163**, 161 (1993).
287. C. Caretta Cartozo, D. Garlaschelli, C. Ricotta, M. Barthelemy and G. Caldarelli, preprint (2004).
288. North-American *florae* are publicly available from the *Plants National Database* (`http://plants.usda.gov`).
289. C. S. Elton, *Animal Ecology*, Sidgwick & Jackson, London (1927).
290. J. M. Montoya and R. J. Solé, *Journal of Theoretical Biology* **214**, 405 (2002).
291. J. Camacho, R. Guimerà and L. A. N. Amaral, *Physical Review Letters* **88**, 228102 (2002).
292. J. A. Dunne, R. J. Williams and N. D. Martinez, *Proceedings of the National Academy of Science (USA)* **99**, 12917 (2002).
293. I. Rodriguez-Iturbe and A. Rinaldo, *Fractal River Basins: Chance and Self-Organization*, Cambridge University Press, Cambridge (1996).
294. J. Banavar, A. Maritan and A. Rinaldo, *Nature* **399**, 130 (1999).
295. G. B. West, J. H. Brown and B. J. Enquist, *Science* **276**, 122 (1997).
296. G. B. West, J. H. Brown and B. J. Enquist, *Science* **284**, 1677 (1999).
297. J. Memmott, N. D. Martinez and J. E. Cohen, *J. Anim. Ecol.* **69**, 1 (2000).
298. S. J. Hall and D. Raffaelli, *J. Anim. Ecol.* **60**, 823 (1991).
299. N. D. Martinez, *Ecol. Monogr.* **61**, 367 (1991).
300. M. Huxham, S. Beaney and D. Raffaelli, *Oikos* **76**, 284 (1996).
301. G. Caldarelli, D. Garlaschelli and L. Pietronero, in *Statistical Mechanics of Complex Networks*, Lecture Notes in Physics, Vol. 625, Ed. R. Pastor-Satorras, M. Rubi and A. Díaz-Guilera, Springer-Verlag (2003), p. 148.
302. A. Cronquist, *The Evolution and Classification of Flowering Plants*, The New York Botanical Garden, New York (1998).
303. J. Stiglitz and B. Greenwald, *Towards a New Paradigm in Monetary Economics, Raffaele Mattioli Lectures*, Cambridge (2003).
304. J. Scott, *Social Network Analysis: A Handbooks*, Sage Publications, London (2000).
305. N. E. Friedkin, *American Journal of Sociology* **96**, 1478-1504 (2001).
306. J. L. Moreno, *Who shall survive?*, Beacon House, Beacon, NY (1934).
307. A. Rapoport and W. J. Horvath, *Behavioral Sci.* **6**, 279-291 (1961).
308. T. J. Fararo and M. Sunshine, *A Study of a biased friendship network*, Syracuse University Press, Syracuse, NY (1964).
309. A. Davis, B. B. Gardner and M. R. Gardner, *Deep South*, University of Chicago Press, Chicago (1941).
310. F. J. Roethlisberger and W. J. Dickson, *Management and the Worker*, Harvard University Press, Cambridge, MA (1939).
311. P. S. Bearman, J. Moody and K. Stovel, *Am. Jour. Soc.* **110**, 44-91 (2004).

312. A. S. Klovdhal, J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muthand and W. W. Darrow, *Soc. Sci. Med.* **38**, 79-88 (1994).
313. J. F. Padgett and C. K. Ansell, *American Journal of Sociology* **98**, 1259-1319 (1993).
314. A. Rapoport, *Bull. Math. Biophys.* **19**, 145-157 (1957).
315. J. Travers and S. Milgram, *Sociometry* **32**, 425-443 (1969).
316. P. V. Marsden, *Ann. Review Soc.* **16**, 435-463 (1990).
317. L. Adamic and B. A. Huberman, *Science* **287**, 2115 (2000).
318. V. Batgelj and A. Mrvar, *Social Networks* **22**, 173-186 (2000).
319. M. Bordens and I. Gomez, in *The Web of Knowledge*, B. Cronin and H. B. Atkins (eds.), Information Today, Medson, NJ (2000).
320. R. De Castro and J.W. Grossman, *Math. Intelligencer* **21**, 51-63 (1999).
321. J. W. Grossman and P. D. F. Ion, *Congressus Numerantium* **108**, 129-131 (1995).
322. G. Mellin and O. Persson, *Scientometric* **36**, 363-377 (1996).
323. J. Moody, *American Society Review* **69**, 213-238 (2004).
324. M. E. J. Newman, *Proceedings of the National Academy of Science (USA)* **98**, 404-409 (2001).
325. G. F. Davis and H. R. Greve, *American Journal of Sociology* **103**, 1-37 (1996).
326. G. F. Davis, M. Yoo and W. E. Baker, *Strategic Organization* **1**, 301-326 (2003).
327. P. Mariolis, *Social Sci. Quart.* **56**, 425-439 (1975).
328. M. S. Mizruchi, *The American Corporate Network, 1904-1974*, Sage, Beverly Hills, CA (1982).
329. G. Chartrand and L. Lesniak, *Graphs and Digraphs*, Wadsworth and Brooks/Cole, Menlo Park (1986).
330. J. J. Ramasco, S. N. Dorogovtsev and R. Pastor-Satorras, *Physical Review E* **70**, 036106 (2004).
331. M. E. J. Newman, S. Forrest and J. Balthrop, *Physical Review E* **66**, 035101(R) (2002).
332. R. D. Smith, cond-mat/0206378 (2002).
333. W. Aiello, F. Chung and L. Lu, in *Proc. of the 32nd Annual ACm Symposium on Theory and Computing* (2000), pp. 171-180.
334. W. Aiello, F. Chung and L. Lu, in *Handbook of Massive Data Sets*, J. Abello, P. M. Pardalos and M. G. C. Resende (eds.), Kluwer Academic, Dordrecht (2002), pp. 97-122.
335. S. Garfinkel, *PGP: Pretty Good Privacy*, O'Reilly and Associates, Cambridge (1994).
336. W. Stallings, *BYTE* **20**, 161 (1995).
337. X. Guardiola, R. Guimera, A. Arenas, A. Díaz-Guilera, D. Streib and L. A. N. Amaral, cond-mat/0206240 (2002).
338. J. H. Jones and M. S. Handcock, *Proceedings of the Royal Society B* **270**, 1123-1128 (2003).
339. F. Liljeros, R. Edling, L. A. N. Amaral, H. E. Stanley and Y. Aberg, *Nature* **411**, 907 (2001).

340. R. K. Merton, *Science* **159**, 56-63 (1968).
341. S. Maslov, K. Sneppen and A. Zaliznyak, cond-mat/0205379.
342. J.-L. Guillaume and M. Latapy, cond-mat/0307095.
343. M. Girvan and M. E. J. Newman, *Proceedings National Academy Sciences (USA)* **99**, 7821 (2002).
344. L. C. Freeman, *Sociometry* **40**, 35 (1977).
345. J. R. Tyler, D. M. Wilkinson and B. A. Huberman, *The Information Society* **21**, 133-141 (2005)
346. G. Caldarelli, C. Caretta Cartozo, P. De Los Rios and V. D. P. Servedio, *Physical Review E* **69**, 035101 (2004).
347. P. De Los Rios, *Europhysics Letters* **56**, 898 (2001).
348. A. Maritan, A. Rinaldo, R. Rigon, A. Giacometti and I. Rodriguez-Iturbe, *Physical Review E* **53**, 1510 (1996).
349. A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vazquez and L. R. Bras, *Physical Review Letters* **70**, 822 (1993).
350. S. Kramer and M. Marder, *Physical Review Letters* **68**, 205 (1992).
351. K. Sinclair and R. C. Ball, *Physical Review Letters* **76**, 3360 (1996).
352. M. O. Jackson and A. Wolinsky, *Journal of Economic Theory* **71**, 44-74 (1996).
353. A. Kirman, *Economic Journal* **99**, 126-39 (1989).
354. J. Galaskiewicz, *Social Organization of the Urban Grant Economy*, Academic Press, New York (1985).
355. J. Galaskiewicz and P. V. Marsden, *Social Sci. Res.* **7**, 89-107 (1978).
356. S. Battiston, G. Caldarelli and D. Garlaschelli, preprint.
357. D. Garlaschelli, S. Battiston, M. Castri, V. D. P. Servedio and G. Caldarelli, *Physica A* **350**, 491-499 (2004).
358. M. A. Serrano and Boguñá, *Physical Review E* **68**, 015101(R) (2003).
359. Mediobanca, *Calepino dell'Azionista*, Milano (1986).
360. Banca Nazionale del Lavoro, *La meridiana dell'investitore 2002*, ClassEditori, Milano (2002).
361. *Fortune 1000* (Fortune 1000, data concerning the first 1000 US companies, raked by revenues. Data kindly provided by Gerald Davis, Michigan University).
362. B. Mintz, and M. Schwartz, *The Power Structure of American Business*, University of Chicago Press (1985).
363. L. D. Brandeis, *Other People's Money: And How the Bankers Use It*, Frederick A. Stokes, New York (1914).
364. S. Battiston, E. Bonabeau and G. Weisbuch, *Physica A* **322**, 567 (2003).
365. M. E. J. Newman and Juyong Park, *Physical Review E* **68**, 036122 (2003).
366. S. Battiston, G. Weisbuch and E. Bonabeau, *Advances in Complex Systems* **6**, 4 (2003).
367. P. R. Haunschild, *Administrative Science Quarterly* **38**, 564-592 (1993).
368. H. Follmer, *Journal of Mathematical Economics* **1**, 51-62 (1974).
369. S. Galam, Y. Gefen and Y. Shapir, *Journal of Sociology* **9**, 1-13 (1982).
370. M. Leone, A. Vazquez , A. Vespignani and R. Zecchina, *European Physical Journal B* **28**, 191 (2002).

Index

affinity, 71
airport network, 74
  clustering, 75
  disparity, 76
  strength distribution, 74
analytic layouts, 124
assortative generator, 57
average nearest neighbors degree, 40

betweenness, 98
  current-flow, 98
biological network, 83
bipartite core, 136
board of directors, 219
  assortativity, 225
  degree distribution, 224
  statistical properties, 222
breadth first search, 136

centrality, 97
centrality layouts, 125
circuit, 10
clique, 9, 95
clustered layouts, 126
clustering
  coefficient, 12
  random walks, 108
connected component, 136
  strongly, 136
  weakly, 136
connectivity, 10
correlations
  detailed balance, 37
  three vertices, 45
  two vertices, 41
coupling of traffic and topology, 91
cycle, 10

degree, 8, 135
  distribution, 12
  heterogeneity, 14
  in-, 8, 135
  out-, 8, 135
  sequence, 8
  weighted, 8
disassortative correlations, 55

edge, 5
  multiple, 7
end vertices, 6
Erdös-Rényi model, 168

finite size effects, 170
food web, 195
  spanning tree, 191
forest, 10

Gaussian distribution, 13
Girvan and Newman algorithm, 98
graph, 5
  bipartite, 9
  complete, 6
  directed, 7
  empty, 6
  order, 6
  oriented, 7
  r-partite, 9
  size, 6
  weighted, 7

Internet
  autonomous systems, 52
  average nearest neighbors degree, 175
  correlations, 173
  degree distribution, 168
  maps, 164
  small-world, 166

joint degree distribution, 38

L-shells, 101
link clustering, 100
links efficiency, 190
local heterogeneity, 72, 88
loop, 7

model
  Barabási-Albert, 25
    degree distribution, 27
    degree exponents, 27
    rate equation, 25
  clustered networks, 59
  configuration, 55
  copying, 144
  duplication divergence
    rate equation, 29
  duplication-divergence, 27
  fitness, 21, 62
    degree distribution, 22
  fitness and preferential attachment, 64
  growing, 56
  growing weighted networks, 30
    degree distribution, 31
  hidden color, 61
  multi-layer, 145
  preferential attachment, 23
  Q-potts, 110
  random graphs, 18, 168
    degree distribution, 19
  small-world, 32
    shortcuts, 32
  weighted network, 84
  world wide web, 142
modularity, 96
  maximization, 102
multigraph, 7

network
  assortative, 41
  collaboration, 207
  communication, 207
  disassortative, 41
  Markovian, 39
  mutual recognition, 208
  protein interaction, 47
  resistors, 99
    approximate, 106
  scale-free, 14
  scientific collaboration, 50
  social, 206
    assortativity, 210
    clustering, 213
  world airport, 53
network efficiency, 99
networks
  evolving, 18
  static, 18

optimisation
  constrained, 106
  extremal, 103
  greedy, 102

PageRank, 140
path, 9, 136
  Eulerian, 10
  Hamiltonian, 10
Poisson's distribution, 13
preferential attachment and duplication, 29
preprint archive, 17

random graph, 18
rich-club phenomenon, 76

scientific collaboration network, 80
  clustering, 82
  degree distribution, 80
  disparity, 82
  topology weight correlation, 80

six degrees of separation, 11
small world, 11
spectral analysis, 104
strength, 69
subgraph, 8
  induced, 136

taxonomy, 192
tier, 174
traversal, 136
tree, 10
  leaves, 11
  root, 11
trees
  spanning, 188
trophic interactions, 186

urban network, 77
  clustering, 78
  degree distribution, 77
  disparity, 78
  strength, 78

vertex, 5
  adjacent, 6
  incident, 6
visualization
  spectral layout, 121
  spring embedder methods, 117

walk, 136
web of trust, 51
weight, 68
  clustering coefficient, 69
  distribution, 68
world trade web, 53
world wide web
  cliques, 142
  crawl, 137
  degree distributions, 138
  WebBase, 138