3

Network models

In this chapter we present a review of network models that will be used to study dynamical processes in the context of computational approaches. These different models will also help to determine the influence that specific network features have on the various phenomena that will be considered in the next chapters. To this end, we will discuss the different modeling approaches and put the activity focused on each of the different dynamical models into the proper perspective. Particular emphasis will be devoted to models which have been developed as theoretical examples of the different specific classes of real-world networks empirically observed.

3.1 Randomness and network models

Static random graph models and topology generators such as the paradigmatic Erdős–Rényi model (Erdős and Rényi, 1959; 1960; 1961) and the network generation algorithm of Molloy and Reed (1995) are the simplest network models to include stochasticity as an essential element. They are characterized by an absolute lack of knowledge of the principles that guide the creation of edges between nodes. Lacking any information, the simplest assumption one can make is to connect pairs of nodes at random with a given connection probability p. In its original formulation, an Erdős–Rényi graph $G_{N,E}$ is constructed starting from a set of N different vertices which are joined by E edges whose ends are selected at random among the N vertices. A variation of this model (Gilbert, 1959) is the graph $G_{N,p}$ constructed from a set of N different vertices in which each of the $N(N-1)/2$ possible edges is present with probability p (the *connection probability*) and absent with probability $1-p$.

The relation between them is straightforward: in the latter case, the probability, when constructing a graph $G_{N,p}$, of obtaining a particular graph $G_{N,E}$ with N vertices and exactly E edges is

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

$$P(G_{N,E}) = p^E (1-p)^{\frac{1}{2}N(N-1)-E}. \tag{3.1}$$

Since each edge can be either present or absent, the ensemble of graphs $G_{N,p}$ contains $2^{N(N-1)/2}$ elements, and many of their properties can be easily derived. For example, in order to compute the average degree, we observe that the average number of edges generated in the construction of the graph is $\langle E \rangle = \frac{1}{2}N(N-1)p$. Since each edge contributes to the degree of two vertices, we obtain

$$\langle k \rangle = \frac{2\langle E \rangle}{N} = (N-1)p \simeq Np, \tag{3.2}$$

where the last equality is valid for large N. The two ensembles of graphs $G_{N,p}$ and $G_{N,E}$ are in fact statistically equivalent when N goes to infinity with

$$pN(N-1)/2 = E.$$

From the previous equation we observe that, for any finite p, the average degree diverges with the number of vertices in the graph. Since real-world graphs are most often characterized by a finite average degree, in many cases it is a natural choice to consider the behavior of the model for a wiring probability that decreases with N; i.e. $p(N) = \langle k \rangle / N$. The average degree of the random graph is also a determinant parameter in establishing the connectivity structure of the resulting network. If $\langle k \rangle < 1$ the network is composed of many small subgraphs that are not interconnected. For $\langle k \rangle = 1$, we observe a phase transition equivalent to the percolation transition in infinite dimension and for $\langle k \rangle > 1$ a giant component emerges with size proportional to the number of vertices in the network. A more detailed account of the component structure of random graphs can be found in Chapter 6.

In order to obtain the degree distribution $P(k)$, we notice that, in a graph with wiring probability p, the probability of creating a vertex of degree k is equal to the probability that it is connected to k other vertices and not connected to the remaining $N-1-k$ vertices. Since the establishment of each edge is an independent event, this probability is simply given by the binomial distribution

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \tag{3.3}$$

In the limit of large N and for $pN = \langle k \rangle$ constant, the binomial distribution can be approximated by the Poisson distribution (Gnedenko, 1962)

$$P(k) = \mathrm{e}^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \tag{3.4}$$

recovering the result obtained from more rigorous arguments by Bollobás (1981). The most characteristic trait of the degree distribution of the Erdős–Rényi model is that it decays *exponentially* for large k, allowing only very small degree fluctuations. The Erdős–Rényi model represents, in this sense, the prototypical example of

a statistically *homogeneous* random graph, in which, for the purpose of the large-scale characterization of the network, the degree of the different vertices can be approximately considered as uniform and equal to the average degree, $k \simeq \langle k \rangle$.

In the Erdős–Rényi model it is also possible to derive easily the clustering and small-world properties. The clustering coefficient $\langle C \rangle$ of the Erdős–Rényi model follows from the independence of the connections. For any vertex, the probability that any two of its neighbors are also connected to each other is given by the connection probability p. Therefore the average clustering coefficient is equal to

$$\langle C \rangle = p = \frac{\langle k \rangle}{N}. \tag{3.5}$$

From the previous expression it is easy to conclude that the clustering coefficient of the Erdős–Rényi model, at fixed $\langle k \rangle$, *decreases* with the graph size, and approaches zero in the limit of an infinitely large network. In other words, the absence of cohesive ordering is implicit in the complete randomness of the model. For a connected network of average degree $\langle k \rangle$, the average number of neighbors at distance 1 of any vertex i is $\langle k \rangle$. If the position of the edges is completely random and the effect of cycles is neglected, the number of neighbors at a distance d can be approximated by $\langle k \rangle^d$. Let us define r_G such that $\langle k \rangle^{r_G} \simeq N$. Since the quantity $\langle k \rangle^d$ grows exponentially fast with d, an overwhelming majority of vertices are at a distance of order r_G from the vertex i. We can thus approximate the average shortest path length $\langle \ell \rangle$ by r_G and we obtain[1]

$$\langle \ell \rangle \simeq \frac{\log N}{\log \langle k \rangle}. \tag{3.6}$$

This approximate estimate can be proved rigorously (Bollobás, 1981), showing that the Erdős–Rényi model exhibits an average shortest path length $\langle \ell \rangle$ that scales logarithmically with the graph size N. This scaling behavior is the signature of the *small-world* effect observed in many complex networks.

3.1.1 Generalized random graphs

The random graph paradigm of the Erdős–Rényi model can be extended to accommodate the construction of generalized random graphs with a predefined degree distribution – not necessarily Poisson – that are otherwise random in the assignment of the edges' end-points. This procedure, first proposed by Bender and Canfield

[1] It must be noted that the present result is not valid if one considers averages over a structure that is wholly tree-like. In this case, the existence of one-dimensional-like trees leads to the result that $\langle \ell \rangle$ usually scales as a power of N (Burda, Correia and Krzywicki, 2001).

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

(1978), and later developed in several works (Molloy and Reed, 1995; Molloy and Reed, 1998; Aiello, Chung and Lu, 2001), consists of assigning the graph a fixed degree sequence $\{k_i\}$, $i = 1, \ldots, N$, such that the ith vertex has degree k_i, then distributing the end-points of the edges among the vertices according to their respective degrees. This procedure generates graphs which are in all respects random,[2] with the imposed degree distribution $P(k)$. The inherent randomness of these networks allows the analytical calculation of properties such as the clustering coefficient and the average shortest path in the general case of random uncorrelated graphs with given degree distribution $P(k)$, as detailed in Appendix 1. For instance, the general expression for the average clustering coefficient reads as

$$\langle C \rangle = \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}. \tag{3.7}$$

This implies that the clustering properties have an intrinsic dependence on the degree distribution moments. In the limit of infinite size $N \to \infty$, the clustering coefficient is null as expected in a completely random graph. However, in finite size networks with large fluctuations of the degree distribution, the clustering coefficient could have much larger values than in the case of the Poissonian Erdős–Rényi graph because of the large values attained by $\langle k^2 \rangle$.

A generalization of Equation (3.6) for the scaling of the average shortest path of the graph can similarly be obtained and reads (Appendix 1)

$$\langle \ell \rangle \approx 1 + \frac{\log[N/\langle k \rangle]}{\log[(\langle k^2 \rangle - \langle k \rangle)/\langle k \rangle]}. \tag{3.8}$$

Small-world properties are thus confirmed for general degree distribution $P(k)$, and exist simply because of the randomness of the graph.

As we have just seen, the random graph framework readily explains the presence of small-world properties and provides a convenient first order approximation for the modeling of a wide range of networks. It is therefore not surprising that, in view of their simplicity and elegance, these graphs have been used as the central paradigm of network modeling for almost four decades.

3.1.2 Fitness or "hidden variables" models

In the Erdős–Rényi and the generalized random graph models, the probability of connecting two nodes is independent of the nodes themselves. In certain real situations, it is however reasonable to think that two nodes will be connected depending on some of their intrinsic properties such as social status, information content, or

[2] Correlations are generated because of finite size effects and may not be negligible in some cases (see Section 3.4).

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

friendship. In order to explore this idea, Söderberg (2002), and Caldarelli *et al.* (2002), have introduced a network model in which each node i ($i = 1, \ldots, N$) is assigned a "fitness" described by a random real variable x_i distributed according to a certain probability distribution $\rho(x)$, or by a discrete "type" (Söderberg, 2002). Each pair of nodes (i, j) is then connected with a probability depending on the fitnesses or types of the two nodes, $p_{ij} = f(x_i, x_j)$, where f is a given function. In the case where $f = \text{const.}$, the present model is equivalent to an Erdős–Rényi graph. The expected degree of a node with fitness x is given by

$$k(x) = N \int_0^\infty f(x, y)\rho(y)\mathrm{d}y \equiv NF(x), \tag{3.9}$$

and the degree distribution can then be deduced by

$$\begin{aligned} P(k) &= \int \mathrm{d}x\rho(x)\delta[k - k(x)] \\ &= \rho\left[F^{-1}\left(\frac{k}{N}\right)\right]\frac{\mathrm{d}}{\mathrm{d}k}F^{-1}\left(\frac{k}{N}\right), \end{aligned} \tag{3.10}$$

where we have supposed that $F(x)$ is a monotonic function of x. From this last equation, one can see that if the fitness distribution is a power law and F is for example linear the resulting network will be scale-free. This property, however, does not yield an explanation of the presence of scale-free degree distributions in real complex networks since the use of a power-law $\rho(x)$ is not a priori justified. A more surprising result appears if we choose a peaked fitness distribution of the exponential type ($\rho(x) \sim \mathrm{e}^{-x}$) and for the function f a threshold function of the form

$$f(x_i, x_j) = \theta\left[x_i + x_j - z(N)\right], \tag{3.11}$$

where θ is the Heaviside function and $z(N)$ is a threshold depending in general on the size N. Caldarelli *et al.* (2002) have shown that the degree distribution is then a power law of the form $P(k) \sim k^{-2}$. This result implies that for static networks a peaked distribution of fitnesses can generate scale-free networks.[3] Moreover, both the assortativity $k_{\mathrm{nn}}(k)$ and the clustering spectrum $C(k)$ behave as power laws. Interestingly, Boguñá, Pastor-Satorras (2003) show that this framework can be generalized into a class of models in which nodes are tagged by "hidden" variables that completely determine the topological structure of the network through their probability distribution and the probability of connecting pairs of vertices.

[3] Geographical constraints may easily be introduced as well, see e.g. Masuda, Miwa and Konno (2005)

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

3.1.3 The Watts–Strogatz model

In random graph models the clustering coefficient is determined by the imposed degree distribution and vanishes in the limit of very large graphs. The empirical observation of a very large clustering coefficient in many real-world networks is therefore a conceptual challenge that spurred the definition of models in which it is possible to tune $\langle C \rangle$ to any desired value. Inspired by the fact that many social networks (Milgram, 1967; Wasserman and Faust, 1994) are highly clustered while at the same time exhibiting a small average distance between vertices, Watts and Strogatz (1998) have proposed a model that interpolates between ordered lattices (which have a large clustering coefficient) and purely random networks (which possess a small average path length).

The original Watts and Strogatz model starts with a ring of N vertices in which each vertex (i_1, i_2, etc.) is symmetrically connected to its $2m$ nearest neighbors (m vertices clockwise and counterclockwise of i_1, then i_2, etc. as shown in Figure 3.1 for $p = 0$). Then, for every vertex, each edge connected to a clockwise neighbor is rewired with probability p, and preserved with probability $1 - p$. The rewiring connects the edge's end-point to a randomly chosen vertex, avoiding self-connections, and typically creates *shortcuts* between distant parts of the ring. The parameter p therefore tunes the level of randomness present in the graph, keeping the number of edges constant. With this construction, after the rewiring process, a graph with average degree $\langle k \rangle = 2m$ is obtained. It is however worth noting that even in the limit $p \to 1$, the fact that each vertex keeps a minimum degree m induces some memory of the generating procedure so that the network is not locally equivalent to an Erdős–Rényi graph (Barrat and Weigt, 2000). The degree distribution of the Watts–Strogatz model can be computed analytically (Barrat and Weigt, 2000) and reads

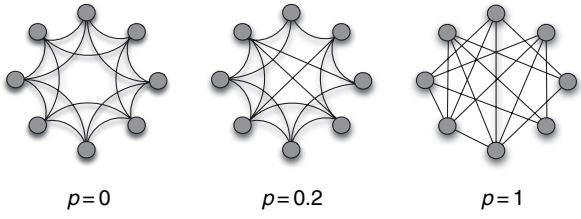

Fig. 3.1. Construction leading to the Watts–Strogatz model. We start with $N = 8$ nodes, each one connected to its four nearest neighbors. By increasing p, an increasing number of edges is rewired. Rewired edges are represented as straight arcs. At $p = 1$ all edges have been rewired. Adapted from Watts and Strogatz (1998).

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

$$P(k) = \sum_{n=0}^{\min(k-m,m)} \binom{m}{n} (1-p)^n p^{m-n} \frac{(pm)^{k-m-n}}{(k-m-n)!} \mathrm{e}^{-pm}, \quad \text{for } k \geq m. \quad (3.12)$$

In the limit of $p \to 1$ the above expression reduces to

$$P(k) = \frac{m^{k-m}}{(k-m)!} \mathrm{e}^{-m}, \quad (3.13)$$

which is a Poisson distribution for the variable $k' = k - m$, with average value $\langle k' \rangle = m$.

While the degree distribution has essentially the same features as a homogeneous random graph, the parameter p has strong effects on the clustering coefficient and the average shortest path length. When $p = 0$ the number of connections among the neighbors of each node is $3m(m-1)/2$, while the total possible number of connections is $2m(2m-1)/2$. This yields a clustering coefficient $\langle C \rangle = 3(m-1)/2(2m-1)$. At the same time, the shortest path length scales as in the one-dimensional regular grid, i.e. $\langle \ell \rangle \sim N$. This picture changes dramatically as soon as the rewiring probability is switched on. For very small p the resulting network has a full memory of a regular lattice and consequently a high $\langle C \rangle$. In particular, Barrat and Weigt (2000) derived the dependence of the clustering coefficient defined as the fraction of transitive triples, obtaining

$$\langle C(p) \rangle \simeq \frac{3(m-1)}{2(2m-1)} (1-p)^3. \quad (3.14)$$

On the other hand, even at small p, the appearance of shortcuts between distant vertices in the lattice dramatically reduces the average shortest path length. For $p \to 1$ the network eventually becomes a randomized graph, with a logarithmically small $\langle \ell \rangle$ and a vanishing clustering coefficient. Watts and Strogatz (1998) focused on the transition between these two regimes (shown in Figure 3.2), noting that in a wide range of $p \ll 1$, the shortest path length, after decreasing abruptly, almost reaches the value corresponding to a random graph, while the clustering coefficient remains almost constant and close to that of the original ordered lattice. Therefore, a broad region of the parameter space exists in which it is possible to find graphs with a large $\langle C \rangle$ and a small $\langle \ell \rangle$, as observed in most natural networks.

Interestingly, the smallest value of p at which the small-world behavior sets in is related to the size of the network, as can be seen from the following scaling form for the average shortest path (Barthélemy and Amaral, 1999b)

$$\langle \ell \rangle \sim N^* F\left(\frac{N}{N^*}\right), \quad (3.15)$$

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

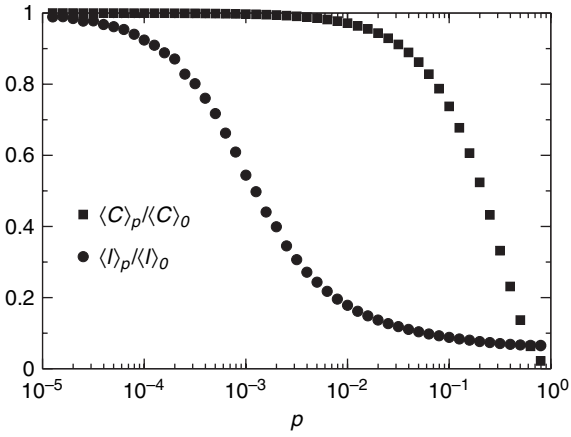

Fig. 3.2. Normalized clustering coefficient $\langle C\rangle_p/\langle C\rangle_0$ (squares) and average shortest path length $\langle \ell\rangle_p/\langle \ell\rangle_0$ (circles) as a function of the rewiring probability p for the Watts–Strogatz model. The results correspond to networks of size $N = 1000$ and average degree $\langle k\rangle = 10$, and are averaged over 1000 different realizations.

where the scaling function has the limit behaviors $F(x \ll 1) \sim x$ and $F(x \gg 1) \sim \log x$. We thus have a crossover from a lattice-like behavior to a small-world one and the crossover size N^* scales with p as $N^* \sim 1/p$ (Barthélemy and Amaral, 1999a; Barrat and Weigt, 2000). This behavior can easily be understood qualitatively. The typical size N^* of the regions between shortcuts is given by the total number of vertices N divided by the average number pN of shortcuts present in the graph, leading to $N^* \sim 1/p$. If the characteristic size N^* of these regions is much smaller than the size of the graph, enough shortcuts connect distant regions on the ring, producing the desired small-world effect. This immediately tells us that if $p \gg 1/N$ the average shortest path is going to be very small. Since, on the other hand, a large clustering coefficient is obtained for $p \ll 1$, the whole parameter region $N^{-1} \ll p \ll 1$ yields networks sharing both small-world properties and a high clustering coefficient. Notably, in the case of very large graphs, $N \to \infty$, even a very small amount of randomness is then sufficient to produce the small-world effect.

The Watts–Strogatz model represents an important development in the modeling of social networks and many other systems (Strogatz, 2000a) since it allows the tuning of the clustering coefficient within the framework of static random graph theory. In addition it can explain the high clustering coefficients observed in real networks as the memory of an initial ordered structure that has been reshaped by some stochastic element. Finally, as the original model displays a Poisson degree distribution, several variations have been proposed in the literature in order to make room within the Watts–Strogatz construction for arbitrary

degree distributions. Several models have therefore been defined in order to obtain highly clustered graphs with heavy-tailed degree distributions and non-trivial correlation properties (Dorogovtsev, Mendes and Samukhin, 2001b; Davidsen, Ebel and Bornholdt, 2002; Holme and Kim, 2002a; Warren, Sander and Sokolov, 2002; Rozenfeld *et al.*, 2002; Chung and Lu, 2004; Volz, 2004; Andersen, Chung and Lu, 2005; Serrano and Boguñá, 2005).

3.2 Exponential random graphs

By looking at the basic definition of the Erdős–Rényi graph, it is possible to see that, by drawing edges at random with a given probability, there is a total of $\binom{N(N-1)/2}{E}$ possible different graphs, which form a probability ensemble in which each graph has the same likelihood. In this respect, the previous construction resembles the *microcanonical ensemble* in classical equilibrium statistical mechanics (Pathria, 1996; Burda *et al.*, 2001).

This idea can be better formalized by the well-established group of models represented by the exponential random graph family[4] largely studied in social network analysis (Holland and Leinhardt, 1981; Frank and Strauss, 1986), and more recently cast in the general framework of the equilibrium statistical mechanics of networks. This modeling approach considers the adjacency matrix – also called the sociomatrix in the social network literature – $\mathbf{X} = \{x_{ij}\}$ characterizing the graph of size N as a random matrix whose realization occurs with a probability $P(\mathbf{X})$ defined in the sample space of all possible graphs. In this hyper-dimensional space, each coordinate represents the possible values that each degree of freedom (in the case of the graph the variables x_{ij}) may have. Each point in this space thus specifies the values of the microscopic variables defining a realization of the network. The exponential random graphs family defined in social network analysis assumes that the general probability distribution in the sample space has the form

$$P(\mathbf{X}) = \frac{\exp\left[\sum_i \theta_i z_i(\mathbf{X})\right]}{\kappa(\{\theta_i\})}, \tag{3.16}$$

where $\{\theta_i\}$ is a set of model parameters and $z_i(\mathbf{X})$ is a set of the network's statistical observables. The statistics $z_i(\mathbf{X})$ are numerous and range from the very simple average degree of the graph $\langle k \rangle$ (equivalently the total number of edges $E = N\langle k \rangle/2$) to complete degree sequences and even probability distributions of attributes. The function $\kappa(\{\theta_i\})$ ensures the correct normalization (the sum of $P(\mathbf{X})$ over all possible graphs $\mathbf{X}$ allowed in the sample space is equal to 1). Once the relevant statistics

[4] In the statistical and social science literature these models are also referred to as Logit models, p^*-models, and Markov random graphs.

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

and assumptions are included in the model, the parameters θ_i have to be estimated by comparison with the real data. This has spurred the development of a wide array of techniques such as pseudo-likelihood estimation and Monte Carlo maximum likelihood estimation (Strauss and Ikeda, 1990; Wasserman and Pattison, 1996). This framework has been adapted to introduce the dynamical evolution of edges, given a fixed number of nodes (Sanil, Banks and Carley, 1995; Banks and Carley, 1996), and numerical techniques for the estimation of the modeling distribution parameters have been put forward in computational tools (Snijders, 2001).

The form of Equation (3.16) has deep connections with the basic principles of equilibrium statistical physics. It is indeed possible to show in a very simple way that exponential random graph models are equivalent to the statistical mechanics of Boltzmann and Gibbs for networks (Burda *et al.*, 2001; Berg and Lässig, 2002; Burda and Krzywicki, 2003; Dorogovtsev, Mendes and Samukhin, 2003; Farkas *et al.*, 2004; Park and Newman, 2004). Equilibrium statistical physics assumes that the probability for the system (in our case the network) to be in a specific configuration $\mathbf{X}$ is given by the distribution $P(\mathbf{X})$ that maximizes the Gibbs entropy

$$S[P] = -\sum_{\mathbf{X}} P(\mathbf{X}) \ln P(\mathbf{X}), \tag{3.17}$$

where the sum is over all possible stochastic realizations allowed. The entropy is a measure of the disorder encoded in the probability distribution and it is reasonable to expect that, at equilibrium, statistical disorder is maximal. It is worth stressing that equilibrium here does not imply a static vision and only refers to a situation in which the probability distribution describing the possible states is not biased or constrained to be on a specific subset of the allowed states. In the context of physical systems, this equilibrium assumption can be more formally stated and related to their microscopic dynamics. The maximization of the entropy is moreover constrained by a certain number of statistical observables $z_i(\mathbf{X})$ for which one assumes one has statistical estimates

$$\langle z_i \rangle = \sum_{\mathbf{X}} P(\mathbf{X}) z_i(\mathbf{X}), \tag{3.18}$$

and by the normalization condition $\sum_{\mathbf{X}} P(\mathbf{X}) = 1$. The maximization of the entropy function is made by introducing a Lagrange multiplier α_i for each constraint $\langle z_i \rangle$, with α_0 being the multiplier relative to the normalization condition. The distribution has therefore to satisfy the equation

$$\frac{\delta}{\delta P(\mathbf{X})}\left[S[P] + \alpha_0\left(1 - \sum_{\mathbf{Y}} P(\mathbf{Y})\right) + \sum_i \alpha_i \left(\langle z_i \rangle - \sum_{\mathbf{Y}} P(\mathbf{Y}) z_i(\mathbf{Y})\right)\right] = 0. \tag{3.19}$$

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

This functional derivative yields the condition, valid for all possible realizations $\mathbf{X}$,

$$\ln P(\mathbf{X}) + 1 + \alpha_0 + \sum_i \alpha_i z_i(\mathbf{X}) = 0, \tag{3.20}$$

which leads to the solution

$$P(\mathbf{X}) = \frac{\exp\left[-\sum_i \alpha_i z_i(\mathbf{X})\right]}{Z(\{\alpha_i\})}, \tag{3.21}$$

where the normalization condition imposes

$$Z(\{\alpha_i\}) = \mathrm{e}^{\alpha_0+1} = \sum_{\mathbf{X}} \mathrm{e}^{-\sum_i \alpha_i z_i(\mathbf{X})}. \tag{3.22}$$

Finally, the explicit values of the parameters α_i are found by imposing the self-consistent condition on the statistical observables for all the observables z_i used in the model construction:

$$\langle z_i \rangle = \sum_{\mathbf{X}} z_i(\mathbf{X}) \frac{\exp\left[-\sum_j \alpha_j z_j(\mathbf{X})\right]}{Z(\{\alpha_j\})}. \tag{3.23}$$

The simple substitution $\theta_i = -\alpha_i$ and $\kappa(\{\theta_i\}) = Z(\{-\theta_i\})$ readily yields a probability distribution $P(\mathbf{X})$ identical to the distribution of the exponential random graph family. Generally, the function $H(\mathbf{X}) = \sum_i \alpha_i z_i(\mathbf{X})$ defining the statistical weight is called the Hamiltonian of the system and the function Z defines the partition function. Analogies can be pushed further with different statistical constraints corresponding to different statistical ensembles in the statistical mechanics definition. Moreover, it is possible to show that this formalism also contains random graphs such as the Erdős–Rényi one (Park and Newman, 2004). For instance, the random graph family $G_{N,p}$ can be recovered by imposing, as a constraint, the corresponding value of the number of links. The exponential random graph family used in statistics is therefore the distribution form corresponding to the equilibrium ensembles of statistical mechanics developed in physics and can be considered as the statistical mechanics of Gibbs for networks.

3.3 Evolving networks and the non-equilibrium approach

The modeling approaches introduced in the previous section are focused on the stationary properties of the network for which they derive the probability distribution in phase space. However, many physical, social, and biological systems are the result of microscopic dynamical processes determining the occurrence of the various configurations. The creation of a social relation, the introduction of a hyperlink to a web page, and the peering of two Internet service providers are dynamical events based on local interactions among individuals that shape the evolution of

the network. Non-equilibrium systems may still have a stationary state in which the probability distribution is time-independent, but in this case we find an overwhelming number of constraints that must be considered for the entropy maximization, rendering its computation infeasible. In the non-equilibrium situation, the system might favor a particular region of the phase space depending on the initial conditions or on a specific dynamical drift. In this case, unless we know exactly all the constraints on the system evolution, for instance the exact partition of accessible configurations depending on the initial conditions, it is impossible to find a meaningful solution. This is also the case for networks with a continuously increasing number of nodes and edges, whose phase space dimensionality is continuously enlarging. In many of these cases, it is more convenient to rely on approaches dealing directly with the dynamical evolution of the network. To this end, we have to introduce the time variable and the probability of a particular network realization $\mathbf{X}$ at time t given by the distribution $P(\mathbf{X}, t)$. The temporal evolution of the probability distribution is generally expressed in the form of a master equation (Dorogovtsev, Mendes and Samukhin, 2000; Krapivsky and Redner, 2001). This is a linear differential equation for the probability that any network, owing to the microscopic dynamics, is in the configuration $\mathbf{X}$. In this case a stochastic description is applied by introducing the rates $r_{\mathbf{X}\to\mathbf{Y}}$ that express the transition from the realization $\mathbf{X}$ to the realization $\mathbf{Y}$.[5] If we assume that the process has no time memory, we have a Markovian process, and the temporal change of $P(\mathbf{X}, t)$ obeys a master equation of the form

$$\partial_t P(\mathbf{X}, t) = \sum_{\mathbf{Y}\neq\mathbf{X}} \left[P(\mathbf{Y}, t) r_{\mathbf{Y}\to\mathbf{X}} - P(\mathbf{X}, t) r_{\mathbf{X}\to\mathbf{Y}}\right]. \tag{3.24}$$

The transition rates have to satisfy the relation $\sum_{\mathbf{Y}} r_{\mathbf{X}\to\mathbf{Y}} \mathrm{d}t = 1$ since summing the rates on all possible configurations must give a probability equal to 1. The probability distribution $P(\mathbf{X}, t)$ is also normalized and it is trivial to see that the master equation (3.24) preserves this normalization.

In the master equation approach, it is crucial to consider transition rates or probabilities r reflecting the actual dynamics of the system under consideration. In the modeling, the attention thus shifts from the statistical quantities describing the system to the dynamical laws governing its evolution. In the absence of any details on such laws, the dynamical approach is often a difficult exercise in which rough assumptions and often uncontrolled approximations have to be made. On the other hand, it has the advantage of being more intuitive and suitable to large-scale computer simulations and theoretical discussions. In practice, the master equation can

[5] In the case of dynamics in discrete time $t \to t + 1$, one assigns transition probabilities $p_{\mathbf{X}\to\mathbf{Y}}$ instead of transition rates.

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

be exactly solved in only a small number of specific cases and it is more practical to work with a specific projection of the probability distribution, depending on the quantity of interest, such as the degree distribution or any other statistical observables in the network.

To provide a specific example, let us consider a continuously growing network in which new nodes appear and wiring processes such as edge adding, removal and rewiring take place (Krapivsky and Redner, 2003b). Each time a new node enters the network, it establishes m new reciprocal edges with nodes already existing in the network. For the sake of simplicity let us assume that once an edge is established it will not be rewired. In this case, the simplest quantity we might be interested in is the degree distribution specified by the number N_k of vertices with degree k. The master equation for such a growing scheme is simply given by

$$\partial_t N_k = r_{k-1\to k} N_{k-1} - r_{k\to k+1} N_k + \delta_{k,m}. \tag{3.25}$$

The first term on the right corresponds to processes in which a vertex with $k-1$ links is connected to the new vertex, thus entering the class of degree k and yielding an increase of the number N_k. The second term corresponds to vertices with degree k that acquire a new edge, with, as a consequence, a decrease of N_k. Finally the last term, given by a Kronecker symbol with values $\delta_{k,m} = 1$ if $k = m$, and $\delta_{k,m} = 0$ otherwise, corresponds to the entry of the new vertex into the ensemble of vertices with degree m. The solution of this evolution equation therefore depends on the rates $r_{k-1\to k}$ and $r_{k\to k+1}$ that specify the network dynamics. In various specific cases, it can be obtained explicitly, allowing the calculation of many quantities of interest.

In the case of continuously growing networks the natural time scale for the network's evolution is given by its size N. In this way, the time is measured with respect to the number of vertices added to the graph, resulting in the definition $t = N - m_0$, m_0 being the size of the initial core of vertices from which the growth process starts. Therefore, each time step corresponds to the addition of a new vertex that establishes a number of connections (edges) with already existing vertices following a given set of dynamical rules. A full description of the system is achieved through the probability $p(k, s, t)$ that a vertex introduced at time s has degree k at the time $t \geq s$. Once the probability $p(k, s, t)$ is known, we can obtain the degree distribution at time t (i.e. for a network of size $N = t + m_0$) using the expression

$$P(k, t) = \frac{1}{t + m_0} \sum_{s=0}^{t} p(k, s, t). \tag{3.26}$$

In this case a simple projection is provided by the study of the average degree value $k_s(t) = \sum_{k=0}^{\infty} k p(k, s, t)$ of the sth vertex at time t. Let us consider for the sake

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

of simplicity that the only relevant characteristic of a node which will determine its further evolution is given by its degree (see Section 3.4 for references on more involved approaches). The dynamical rate equation governing the evolution of $k_s(t)$ can then be formally obtained by considering that the degree growth rate of the sth vertex will increase proportionally to the probability $\Pi[k_s(t)]$ that an edge is attached to it.[6]

In the simple case that edges only come from the newborn vertices, the rate equation reads

$$\frac{\partial k_s(t)}{\partial t} = m\Pi[k_s(t)], \tag{3.27}$$

where the proportionality factor m indicates the number of edges emanating from every new vertex. This equation is moreover constrained by the boundary condition $k_s(s) = m$, meaning that at the time of its introduction, each vertex has an initial degree m. In this formulation, all the dynamical information is contained in the probability or growth kernel $\Pi[k_s(t)]$, which defines the properties of each particular model.

The projection could consider other quantities such as the number of vertices $N(k|\ell)$ with degree k which share an edge with a vertex of degree ℓ, etc. Higher order statistics clearly yield increasingly involved evolution equations. Similarly, the dynamics might be complicated further by the introduction of more refined processes such as edge removal, rewiring, inheritance, and vertices' disappearance. Space and other attributes for the network formation have also been considered in various forms. For a review of dynamical models we refer the reader to Albert and Barabási (2002); Dorogovtsev and Mendes (2003); Pastor-Satorras and Vespignani (2004); Newman (2003b); Boccaletti *et al.* ([2006]); Caldarelli (2007) and references therein.

It is clear from the previous discussion that the dynamical approach is potentially risky and unless we have precise experimental information on the dynamics, it will not allow quantitatively accurate predictions. Moreover it does not provide any systematic theoretical framework, with each model focusing on some specific features of the system of interest. On the other hand, the study of the dynamics is well suited to identify general growing mechanisms from seemingly very different dynamical rules. In this perspective, the dynamical approach appears as the right tool to capture the emergent behavior and complex features of networks. The emphasis is indeed on the evolutionary mechanisms that generate the observed topological properties, which become a by-product of the system's dynamics. The similarities are evident with the statistical physics approach to complex phenomena

[6] For the sake of analytical simplicity one often considers the degree k and the time t as continuous variables. (Barabási, Albert and Jeong, 1999; Dorogovtsev *et al.*, 2000).

that aims to predict the large-scale emergent properties of a system by studying the collective dynamics of its constituents.

3.3.1 The preferential attachment class of models

The dynamical approach to networks is exemplified by the Barabási–Albert class of models (Barabási and Albert, 1999) which provides an example of the emergence of networks with heavy-tailed degree distributions in terms of the elementary process governing the wiring of new vertices joining the network. The insight behind this approach is the fact that in most real networks, new edges are not located at random but tend to connect to vertices which already have a large number of connections (a large degree). For example, very popular web pages are likely to receive new incoming edges from new websites. Similarly, a person with a large number of social relations is more likely to acquire new friends than somebody who has just a close and small circle of friends. In the Internet, new service providers aim at optimizing their centrality by peering with well-connected existing providers. This class of systems might therefore be described by models based on such a preferential attachment mechanism – also known as the rich-get-richer phenomenon, the Matthew effect (Merton, 1968), the Gibrat principle (Simon, 1955), or cumulative advantage (de Solla Price, 1976).

Barabási and Albert (1999) have combined the preferential attachment condition with the growing nature of many networks by defining a simple model based on the following two rules:

Growth: The network starts with a small core of m_0 connected vertices. At every time step we add a new vertex, with m edges ($m < m_0$) connected to old vertices in the system.

Preferential attachment: Each new edge is connected to the old sth vertex with a probability proportional to its degree k_s.

These rules are used to define a class of dynamical algorithms that, starting from a connected initial core, generates connected graphs with fixed average degree $\langle k \rangle = 2m$ (Barabási and Albert, 1999; Barabási *et al.*, 1999; Bollobás *et al.*, 2001). These algorithms can be easily implemented in computer simulations, and Figure 3.3 represents a typical graph of size $N = 200$ and average degree 6 ($m = 3$).

The numerical simulations indicate that the graphs generated with the Barabási–Albert algorithms spontaneously evolve into a stationary power-law degree distribution with the form $P(k) \sim k^{-3}$ (see Figure 3.4). The model can be further explored by using the dynamical approach outlined in the previous section. The preferential attachment mechanism can be easily cast in mathematical form since

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

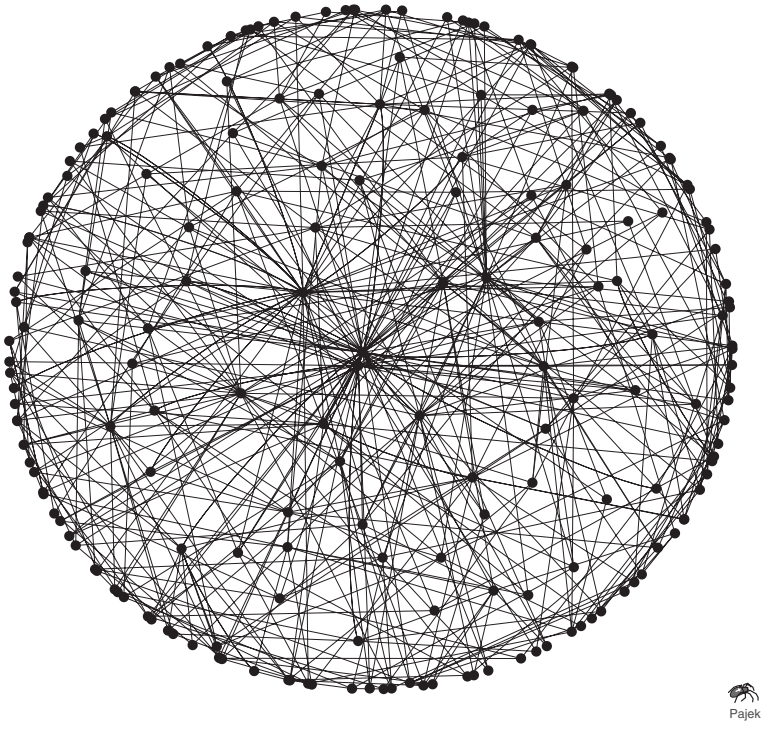

Fig. 3.3. Typical Barabási–Albert network of size $N = 200$ and average degree $\langle k \rangle = 6$. Higher degree nodes are at the center of the graph. The figure is generated with the Pajek package for large network analysis, `http://vlado.fmf.uni-lj.si/pub/networks/pajek/`.

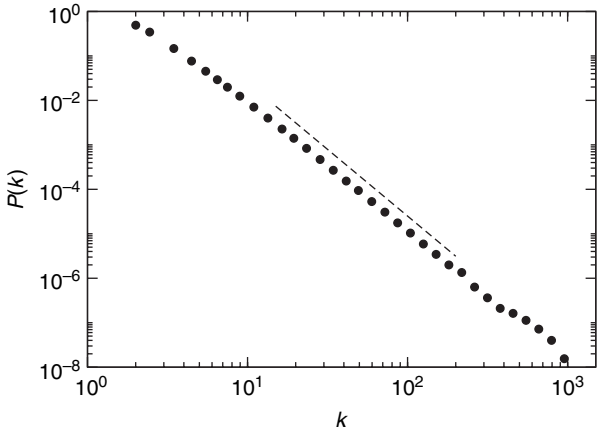

Fig. 3.4. Degree distribution of a Barabási–Albert network of size $N = 10^5$ with average degree $\langle k \rangle = 4$ in double logarithmic scale. The dashed line corresponds to a power-law behavior $P(k) \sim k^{-\gamma}$ with exponent $\gamma = 3$.

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

it states that the probability that the vertex s acquires a new edge is proportional to its degree, obtaining the explicit form of the growth rate $\Pi[k_s(t)]$

$$\Pi[k_s(t)] = \frac{k_s(t)}{\sum_j k_j(t)}, \tag{3.28}$$

where the denominator is the required normalization factor; i.e. the sum of all the degrees of the vertices in the network. Since each new edge contributes with a factor 2 to the total degree, and at time t we have added tm edges, the evolution equation (3.27) for k_s takes the form, within the continuous k approximation,

$$\frac{\partial k_s(t)}{\partial t} = \frac{m k_s(t)}{2mt + m_0\langle k\rangle_0}, \tag{3.29}$$

where $\langle k\rangle_0$ is the average connectivity of the initial core of m_0 vertices. This differential equation with the boundary condition $k_s(s) = m$ can be readily solved, yielding in the limit of large networks ($t, s \gg m\langle k\rangle_0$),

$$k_s(t) \simeq m\left(\frac{t}{s}\right)^{1/2}. \tag{3.30}$$

By considering the continuum limit, the degree distribution is obtained as

$$P(k,t) = \frac{1}{t+m_0}\int_0^t \delta[k - k_s(t)]\,\mathrm{d}s \equiv -\frac{1}{t+m_0}\left(\frac{\partial k_s(t)}{\partial s}\right)^{-1}\Bigg|_{s=s(k,t)}, \tag{3.31}$$

where $\delta(k - k_s(t))$ is the Dirac delta function and $s(k,t)$ is the solution of the implicit equation $k = k_s(t)$. In the Barabási–Albert model the solution of the previous expression yields

$$P(k,t) = 2m^2\frac{t + (m_0/2m)\langle k\rangle_0}{t+m_0}k^{-3}, \tag{3.32}$$

and gives in the limit of large sizes $t \to \infty$ the solution (valid for any $\langle k\rangle_0$),

$$P(k) = 2m^2k^{-3}. \tag{3.33}$$

For the Barabási–Albert class of models it is also possible to obtain analytic expressions providing the clustering coefficient and the average shortest path scaling. More precisely, the average clustering coefficient reads (Klemm and Eguíluz, 2002a; Szabó, Alava and Kertész, 2003; Barrat and Pastor-Satorras, 2005)

$$\langle C\rangle_N = \frac{m}{8N}(\ln N)^2, \tag{3.34}$$

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

and the average shortest path length scales as (Bollobás and Riordan, 2003; Cohen and Havlin, 2003)

$$\langle \ell \rangle \sim \frac{\log(N)}{\log\log(N)}. \tag{3.35}$$

In some situations, it can be reasonable to assume that besides degrees, nodes have other attributes which can render them more "attractive." Bianconi and Barabási (2001) have proposed that for these cases, all the properties (other than the degree) that modify the preferential attachment probability can be encoded in a *fitness* η chosen randomly according to some distribution $\rho(\eta)$. Note that fitness is considered here in a dynamical perspective in contrast with the quantities used in the static fitness model described in Section 3.1.2. In such a modified preferential attachment framework, the probability that a new edge will connect to node s is then given by

$$\Pi_s = \frac{\eta_s k_s}{\sum_j \eta_j k_j}. \tag{3.36}$$

Very fit nodes can thus attract new edges despite a moderate value of their degree and this new "fittest-get-richer" process is thus superimposed on the degree-driven attachment mechanism. Analytical insights into this model can be obtained by using an approximation for the sum $\sum_j \eta_j k_j$, and in the continuous time and degree limits, the degree distribution of the resulting network reads as

$$P(k) = \frac{c}{m} \int \mathrm{d}\eta \frac{\rho(\eta)}{\eta} \left(\frac{k}{m}\right)^{-(1+c/\eta)}, \tag{3.37}$$

where the constant c depends on the fitness distribution (Bianconi and Barabási, 2001). In the case of a uniform distribution of fitness in the interval $[0, 1]$, this leads to $P(k) \sim k^{-\gamma}$ for $k \to \infty$ with $\gamma \approx 2.255$. Remarkably enough, this result demonstrates that a small amount of randomness in fitness can lead to growing scale-free networks with a non-trivial exponent for the degree distribution.

The success of the models based on the preferential attachment mechanism resides in the simple explanation, through a basic dynamical principle, of the emergence of graphs with a power-law degree distribution and small-world properties. The importance of the preferential attachment has even been reinforced recently. Fortunato, Flammini and Menczer (2006a) discuss the attachment mechanism for different quantities such as degree, age, etc., and argue that the relative values of nodes' properties matter more than their absolute values. In other words, the relevant quantity in a preferential attachment mechanism is the rank of each node according to a certain attribute. Such information is often easier to estimate than the absolute value of a node's attribute. For example, even if we are not usually able to quantify precisely the wealth of different people, we can propose a ranking

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

of the richest individuals. Following this idea, Fortunato *et al.* (2006a) study the general case of the attachment probability of a new node to an old vertex s given by the form

$$\Pi_s = \frac{R_s^{-\alpha}}{\sum_j R_j^{-\alpha}}, \tag{3.38}$$

where R_s denotes the rank of the node s for some specific attribute and where α is a positive parameter. In this case, the degree distribution is given by $P(k) \sim k^{-\gamma}$ with

$$\gamma = 1 + \frac{1}{\alpha}. \tag{3.39}$$

This result points to the fact that preferential attachment mechanisms in growing networks lead to scale-free networks, even in the absence of a complete knowledge of the values of the nodes' attributes.

The Barabási–Albert class of models represents a very simple implementation of the preferential attachment mechanism and is not intended to be a realistic model of any real-world network. Rather, it is a zeroth order conceptual model which can be used as the paradigm for much more realistic models taking into account the particular processes in the system under consideration. For this reason, after the introduction of the Barabási–Albert model, a large number of other network models, considering a wide array of degree-based growth mechanisms, have been proposed, incorporating various ingredients in order to account for power-law degree distributions with a connectivity exponent $2 < \gamma < 3$, local geographical factors, rewiring among existing nodes, or age effects. We refer the reader to the reviews by Dorogovtsev and Mendes (2003) and Newman (2003b), and to the references therein to get a flavor of the work done in this direction.

3.3.2 Copy and duplication models

One of the strengths of the dynamical approach is that the presence of similarities in the dynamical microscopic mechanisms allows the explanation of shared properties and similarities in very different systems. In particular, there is a large class of models that at first sight seem completely unrelated to the preferential attachment mechanism, but in which a closer look reveals the existence of this mechanism in disguise. Such a result represents an important step in the basic understanding of the microscopic origin of the preferential mechanism in many real-world systems.

The class of copying models relies on the plausible assumption that new elements arriving in a system have the tendency to copy the choice of already present elements. This mechanism was first proposed in the context of WWW simulations for the generation of skewed degree distributions (Kleinberg *et al.*, 1999; Kumar

et al., 2000). The chief consideration for this mechanism is that new pages dedicated to a certain thematic area copy hyperlinks from existing pages with similar content. This has been translated into a directed growing model in which, at each time step, a new vertex (web page) is added to the network and a corresponding *prototype* vertex is selected at random among those already existing. Each new vertex emits $m \geq 1$ new outgoing edges initially pointing towards vertices pointed by the prototype vertex. At this point a *copy factor* α (constant for all new vertices) is introduced. With probability $1 - \alpha$ each new edge is retained as it is; with probability α it is rewired towards a randomly chosen vertex of the network. The copy factor introduces the possibility that the new vertex does not just copy all its edges from the prototype vertex, since the web page author might find other interesting pages in the network by a random exploration. A pictorial illustration of the copying model dynamics is provided in Figure 3.5.

This model can be easily studied analytically and numerically: the in-degree distribution is power-law distributed and the copy factor acts as a tuning parameter for the degree exponent of the model. These results can be obtained by writing the basic evolution equation for the model. Let us focus here on a generic vertex of the network and calculate its probability of receiving an edge during the addition of a new vertex. For each of the m edges of the new vertex, there is a probability α to rewire it to another vertex chosen at random (uniformly for example). Thus any vertex has a probability α/N of receiving an edge, where N is the size of the network. With probability $1 - \alpha$, on the other hand, the vertex which is pointed to by one of the edges of the prototype vertex is selected. The probability that any given vertex s is pointed to by this edge is given by the ratio between the number of incoming edges of that vertex and the total number of edges, i.e. $k_{\text{in},s}/(mt)$. This second process increases the probability that high degree vertices will receive new incoming edges, and in the limit of large network sizes the mean-field evolution for the copying model can be written in the form of the usual growth rate equation as

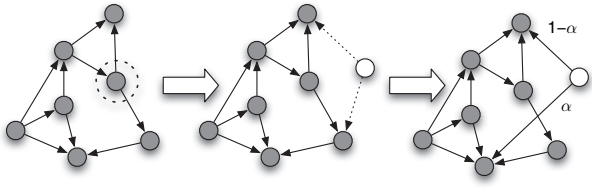

Fig. 3.5. Illustration of the rules of the copying model. A prototype vertex is selected at random (circled by a dashed line) and a new vertex is created with edges pointing to the out-neighbors of the prototype. Each new edge is kept with probability $1 - \alpha$, and rewired to a randomly chosen vertex with probability α.

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

$$\frac{\partial k_{\text{in},s}(t)}{\partial t} = m\left[\frac{\alpha}{t} + (1-\alpha)\frac{k_{\text{in},s}(t)}{mt}\right], \tag{3.40}$$

where $N \simeq t$ for large linearly growing networks. Therefore, through its local dynamical rules, the copying model produces effective preferential attachment growth dynamics. This is a striking result, since the model is defined on the very simple assumption of selecting a prototype vertex, without any knowledge of the popularity or the degree importance of the vertex. The copying model thus offers a microscopic explanation for the preferential attachment mechanism that was just used as a phenomenological law in other models. The solution of the above equation, with the boundary condition $k_{\text{in},s}(s) = 0$, yields the in-degree distribution

$$P(k_{\text{in}}) \sim (k_0 + k_{\text{in}})^{-(2-\alpha)/(1-\alpha)}, \tag{3.41}$$

where $k_0 = \alpha m/(1-\alpha)$ is an offset constant, confirming the presence of heavy tails when a preferential attachment mechanism is governing the network growth.

Although the copy model was first used in the context of the World Wide Web, similar copying mechanisms have also been used in models for the citation network (Krapivsky and Redner, 2005) and the evolution of gene and protein interaction networks (Solé *et al.*, 2002; Vázquez *et al.*, 2003; Wagner, 2003). Indeed, the genomes of most organisms are generally thought to evolve through the duplication of genes and subsequent diversification in order to perform different biochemical roles. Genome evolution by duplication/divergence corresponds to the evolution of the protein–protein interaction network (interactome) whose nodes represent the proteins expressed by the genes. The process of gene duplication can be translated, in terms of protein–protein interaction networks, into the duplication of a node (protein) sharing the same interacting partners as its ancestor, while divergence mechanisms lead to the loss or gain of interactions (see also Chapter 12 on biological networks).

3.3.3 Trade-off and optimization models

Despite the complexity and evolving characters of most networks, it is clear that a series of technical constraints and engineering principles are acting at least on the local properties of networks. Just to give an example, it is clear that within the administrative boundaries, Internet service providers decide upon the topology structure of the network and in addition experience the constraints imposed by the capacity of industrial routers. From this perspective, the network topology might be viewed as emerging from deterministic design decisions that seek to optimize certain domain-specific and network characteristics. Models adopting this perspective are based on the competition between a global optimization of the

https://doi.org/10.1017/CBO9780511791383.004 Published online by Cambridge University Press

system and some local constraints. In other cases, non-trivial collective behavior arises from conflict among the various local expectations. This is, for instance, the case of the heuristically optimized trade-off (HOT) model introduced by Fabrikant, Koutsoupias and Papadimitriou (2002), successively applied to Internet router level topology (Li *et al.*, 2004b). This model suggests that the network emerges through the optimization of conflicting objectives pursued in its set-up. As a practical implementation of these ideas, the HOT Internet model is a growing model in which at every time step a new vertex is added to the network and placed in a random position on the unit square. The new vertex i is connected with an edge to the vertex j that minimizes the function $\Psi(i, j) = \alpha(N)d_{\mathrm{E}}(i, j) + \phi(j)$, where $d_{\mathrm{E}}(i, j)$ is the Euclidean distance between vertices i and j, $\alpha(N)$ is a constant that depends on the final size of the network, and $\phi(j)$ is a measure of the centrality of the vertex j. In the original paper, several centrality measures based on the shortest path length (in terms of hops on the network) were used. It is clear that in this model each new element seeks to lower the costs of establishing the physical connection by reducing as much as possible the Euclidean distance, while trying at the same time to be "centrally located" in the network, thus reducing the hop distance to other vertices. Several other variations of this model have been proposed, some reproducing specific network architecture, others accounting for the degree heterogeneity (Li *et al.*, 2004b; Alvarez-Hamelin and Schabanel, 2004; Mahadevan *et al.*, 2006). Another important aspect of these models is that randomness plays only a marginal role, since the choice of the connection (i, j) is deterministic once i has been added to the network. In principle this is different from stochastic evolution models where randomness models a set of external factors that acts as a noise on the system. The network then emerges as a result of the contrast between such randomness and the preference function which is encoded in the form of the attachment probability. On closer examination, however, HOT models and preferential-attachment-like models are found to lie on common ground. For instance, the original HOT model of Fabrikant *et al.* (2002) minimizes the average distance from the attachment node to the rest of the network. Since this distance directly depends on the degree of the node (Dorogovtsev, Mendes and Oliveira, 2006; Hołyst *et al.*, 2005), the model can actually be reduced to a form of the preferential attachment model. In other words, the HOT model somehow considers the same ingredients that stand at the basis of degree-driven models, although cast in a different dynamical rule. Analogously, the introduction of more involved generating rules in stochastic evolving networks may effectively account for design principles or constraints of increasing complexity and often competing amongst themselves. Such considerations clearly lead to a convergence of modeling perspectives between the stochastic evolution and the optimization approaches.

While the trade-off paradigm balances an optimization principle with some opposing local constraints, a wide class of models just considers the emergence of networks from purely *global* optimization principles. For instance, variational approaches have been used in practical problems by road traffic engineers (Wardrop, 1952) and problems both of optimal traffic on a network (Ahuja, Magnanti and Orlin, 1993) and of optimal networks (Jungnickel, 2004) have a long tradition in mathematics and physics. Recently, it has been shown that optimal networks are relevant in a large variety of systems such as the mammalian circulatory system (Mahon and Bonner, 1983), food webs (Garlaschelli, Caldarelli and Pietronero, 2003), general transportation networks (Banavar, Maritan and Rinaldo, 1999), metabolic rates (West, Brown and Enquist, 1997), river networks (Maritan *et al.*, 1996), gas pipelines or train tracks (Gastner and Newman, 2006), and the air-travel network (Barthélemy and Flammini, 2006). In these studies, the nodes of the network are embedded in a Euclidean space so that the degree is almost always limited and the connections restricted to "neighbors" only. A second recently investigated broad class of optimal networks where spatial constraints are absent has shown that optimizing both the average shortest path and the total length can lead to small-world networks (Mathias and Gopal, 2001). More generally, degree correlations (Berg and Lässig, 2002) or scale-free features (Valverde, Ferrer i Cancho and Solé, 2002) can emerge from an optimization process. Valverde *et al.* (2002) have shown that the minimization of the average shortest path and the link density leads to a variety of networks including exponential-like graphs and scale-free networks. The minimization of search costs leads, on the other hand, either to star-like or to homogeneous networks (Guimerà *et al.*, 2002b). Finally, the interplay between obtaining short routes and little congestion leads to a variety of network structures when the imposed average degree is varied (Colizza *et al.*, 2004). Even if the philosophy behind these studies differs from the usual approach in complex systems, namely that a complex behavior emerges from a local mechanism and interactions between a large set of units, they show that global optimization could in certain cases be a relevant mechanism in the formation of complex networks.

3.4 Modeling higher order statistics and other attributes

As we have seen in the previous chapter, real-world networks are not just characterized by the degree distribution or the small-world properties. Non-trivial association properties, degree correlations, and other complex features all contribute to define the network structure. In many cases, such features are precisely the elements that allow a meaningful model validation. In the study of the dynamical processes occurring on networks, on the other hand, we might be interested in generating networks with specific correlations or clustering properties whose

effects we want to study. In addition, we are generally interested in associating these properties with a specific degree distribution or growth behavior. This still constitutes an open field of study (Boguñá and Pastor-Satorras, 2003; Serrano and Boguñá, 2005; Catanzaro, Boguñá and Pastor-Satorras, 2005), since it is not completely known how the various topological properties are related, or if it is possible to single out a family of metrics defining all others (Mahadevan *et al.*, 2006). On the other hand, basic degree correlation properties are inherent even in simple random and evolving models just because of the constraints imposed by the generating algorithms. For instance, it is possible to show that disassortative correlations are inherent in heavy-tailed random graphs unless very specific precautions are taken in the network's construction. The origin of the problem lies in the fact that the completely random wiring process may generate multiple edges among the same nodes. If the degree distribution imposed by construction has a finite second moment $\langle k^2 \rangle$, the fraction of multiple edges resulting from the construction process vanishes in the thermodynamic limit and, as a consequence, they can be neglected. In contrast, in scale-free degree distributions with exponent $2 < \gamma \leq 3$, the weight of these multiple edges with respect to the overall number of edges cannot be ignored since they are not evenly distributed among all the degree classes. In the thermodynamic limit, a finite fraction of multiple edges will remain among high degree vertices, and imposing restrictions on the algorithm to avoid those occurrences induces disassortative correlations (Boguñá *et al.*, 2004; Maslov, Sneppen and Zaliznyak, 2004; Catanzaro *et al.*, 2005). Analogously, dynamically evolving networks with scale-free degree distributions also spontaneously generate disassortative correlations. The first theoretical derivation of this result (Krapivsky and Redner, 2001) 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 exist (Dorogovtsev and Mendes, 2002; Park and Newman, 2003; 2004; Szabó *et al.*, 2003; Barrat and Pastor-Satorras, 2005). For the average degree of nearest neighbors, it is found that in the large k limit

$$k_{\mathrm{nn}}(k) \sim N^{(3-\gamma)/(\gamma-1)} k^{-(3-\gamma)}, \tag{3.42}$$

for $\gamma < 3$. Correlations between two vertices are therefore disassortative by construction, and characterized by a power-law decay. It is also interesting to note that for $\gamma = 3$, the $k_{\mathrm{nn}}(k)$ function converges to a constant value independent of k and proportional to $\ln N$: the Barabási-Albert model lacks appreciable correlations between degrees of neighboring vertices.

While in this chapter we have mainly discussed the modeling of the topological properties of networks, the need for more specific and data-driven models has led to the formulation of models in which many other features characterizing real-world structures are introduced. Among those, spatial characteristics

and vertex attributes have been plugged into both static and growing network models (Medina, Matt and Byers, 2000; Manna and Sen, 2002; Xulvi-Brunet and Sokolov, 2002; Yook, Jeong and Barabási, 2002; Barthélemy, 2003; Barrat *et al.*, 2005). Weighted properties have also been introduced in modeling attempts in order to understand the interplay and feedback between topology and traffic in complex networks (Yook *et al.*, 2001; Barrat, Barthélemy and Vespignani, 2004b; Barrat, Barthélemy and Vespignani, 2004c; Wang and Zhang, 2004; Barrat *et al.*, 2005; Bianconi, 2005; Dorogovtsev and Mendes, 2005; Wang *et al.*, 2005; Almaas, Krapivsky and Redner, 2005). Finally, various social, economic, and demographic factors have been considered in realistic approaches to the Internet and large-scale infrastructures (Chang, Jamin and Willinger, 2003; Alderson, 2004; Guimerà and Amaral, 2004; Ramasco *et al.*, 2004; Serrano, Boguñá and Díaz-Guilera, 2005; Dimitropoulos *et al.*, 2005; Chang, Jamin and Willinger, 2006; Mahadevan *et al.*, 2006). These works pair with the important observation that many networks have a considerable level of engineering acting at the local and intermediate scale level (Li *et al.*, 2004b; Doyle *et al.*, 2005). This corresponds to particular structural arrangements and connectivity properties requiring specific modeling approaches that cannot be captured with general models, which generally work on simple statistical assumptions. While these elements are very important in the faithful representation of networked systems, they introduce a level of specificity that in most cases does not allow general conclusions to be drawn. For this reason, in the rest of the book we will mainly use models that abstract very general properties whose effect on dynamical processes can be considered as general results, valid in a wide range of systems. On the other hand it is worth stressing that any realistic or detailed study of dynamical processes in specific networks cannot neglect the introduction of domain-specific features and details that might play an important role in the system's behavior.

3.5 Modeling frameworks and model validation

The considerations of the previous sections provide the proper context for a discussion of the various modeling frameworks. The exponential random models, which are akin to the statistical mechanics of networks, are built on very solid statistical foundations and have been mathematically and conceptually developed for many years. On the other hand, they are less intuitive and in many practical instances present technical problems out of our reach. This is the case for a network whose size is rapidly changing and for many non-equilibrium situations. In these cases, the dynamical approach, even if based on uncontrolled assumptions, is the only viable option. This is particularly true if we want to study very large-scale networks, for which it is nowadays possible to rely on large-scale computer

simulations based on the microscopic dynamics of the system's elements. In many ways, the recent explosion in dynamical modeling approaches is a consequence of the informatics revolution over recent years. The advent of high-throughput biological experiments, and the possibility of gathering and handling massive data sets on large information structures, and of tracking the relations and behaviors of millions of individuals, have challenged the community to characterize and to model networks of unprecedented sizes. For instance, at the moment, internet data contain more than 10^4 service providers and 10^5 routers, and keeps track of the behavior of 10^7 to 10^8 users, and available data set sizes are continuously increasing. WWW crawls offer maps of the Web with more than 10^8 nodes. In addition, networks with similar sizes and dynamical characteristics are gathered every day for communication infrastructure such as mobile telephone and ad hoc networks, transportation networks, digital documents, etc. Finally, in biology we are witnessing a change of paradigm with an increasing focus on the so-called system's biology and the many large interaction networks that may be measured by taking advantage of high-throughput experiments. Most importantly, the dynamical features of these systems cannot be neglected since we are dealing in general with networks growing exponentially in time because of their intrinsic dynamics. In this context, dynamical modeling offers an intuitive grasp of the main mechanisms at play in network formation and allows the construction of basic generators that capture the rapidly evolving dynamics of these systems. In other words, the dynamical modeling approach suits the need to understand the evolution and non-equilibrium properties of these networks and might be easily plugged into large-scale numerical simulations which allow the generation of synthetic networks of 10^6 nodes with reasonable computational effort.

The availability of large-scale data sets and the feasibility of large-scale simulations also result in new conceptual questions. While networks may be extremely different from each other in their functions and attributes, the large-scale analysis of their fabric has provided evidence for the ubiquity of several asymptotic properties (see Chapter 2) and raised the issue of the emergence of general and common self-organizing principles. This question is akin to the issue of "universality" addressed in statistical physics and in particular in phase transitions. Universality refers to the fact that very different physical systems (such as a ferromagnet or a liquid), are governed by the same set of statistical laws for their large-scale properties at specific points of their phase diagram. While the nature of these systems might be very different, the shared similarities of the microscopic interactions give rise to the same macroscopic behavior. This behavior is an emergent phenomenon and does not reveal any characteristic scale, i.e., the large-scale properties look the same for all systems. The statistical physics approach has been exploited as a very convenient tool because of the possibility of characterizing emergent macroscopic

phenomena in terms of the dynamical evolution of the microstates of the various systems.

At this point, however, it is worth remarking that universality does not imply equivalence. Universality refers only to large-scale statistical properties or correlation functions. Naturally, at the local level each system will be described by different properties and details. Universality focuses on specific large-scale properties and looks for possible similarities in the dynamical mechanisms responsible for creating these features across different systems. It is related to the identification of general classes of complex networks whose macroscopic properties are dictated by suitable evolution rules for the corresponding microstates, but this identification is representing a scientific challenge.

In such a context, the issue of model validation, the falsification of assumptions, and the reliability of the statistical models are major questions for which it is difficult to provide a general answer. First of all, models always contain hypotheses which are mainly dictated by our interests. While dynamic modeling is well suited for large-scale and evolutionary properties, statistical modeling based on maximum entropy considerations is more suitable to take fully into account the statistical observables at hand. Indeed, the data set properties play a different role in the two approaches. By using the exponential random graph modeling, the statistical measures obtained from the data are used to obtain the parameters of the model via fitting techniques or via the exact solution. The obtained distribution will then provide statistical predictions that can be tested through new measurements on the data set. In the dynamical approach, the network is defined by local dynamical rules that in general are not related to the statistical observables. The data set properties are not an input of the model and are instead used in the model validation process.

In most cases one needs to identify suitable approximations, since it is rarely possible to consider all the properties of a network. Depending on which question one wants to tackle, different models will be considered and different quantities will become prominent or neglected. In this sense, most models are incomplete and address only a limited set of questions. In particular, it is important to distinguish between models aimed at predicting the behavior of a system in a realistic way and models whose goals are to understand the basic mechanisms governing the behavior of the system. This discussion is the subject of many debates in the complex systems community, and as we will see in the next chapter, many of these considerations can be exported to the discussion of models concerning the dynamics occurring on networks.