

12

Networks in biology: from the cell to ecosystems

Networks have long been recognized as having a central role in biological sciences. They are the natural underlying structures for the description of a wide array of biological processes across scales varying from molecular processes to species interactions. Especially at smaller scales, most genes and proteins do not have a function on their own; rather they acquire a specific role through the complex web of interactions with other proteins and genes. In recent years this perspective, largely fostered by the recent abundance of high-throughput experiments and the availability of entire genome sequences and gene co-expression patterns, has led to a stream of activities focusing on the architecture of biological networks.

The abundance of large-scale data sets on biological networks has revealed that their topological properties in many cases depart considerably from the random homogeneous paradigm. This evidence has spurred intense research activity aimed at understanding the origin of these properties as well as their biological relevance. The problem amounts to linking structure and function, in most cases, by understanding the interplay of topology and dynamical processes defined on the network. Empirical observations of heterogeneities have also revamped several areas and landmark problems such as Boolean network models and the issue of stability and complexity in ecosystems.

While concepts and methods of complex network analysis are nowadays standard tools in network biology, it is clear that a discussion of their relevance and roles has to be critically examined by taking into account the specific nature of the biological problem. The biological arena is incredibly vast (see the books of Palsson [2006] and Alon [2007a]), and this chapter is just a bird's eye view of the possible applications of dynamical processes on networks to relevant problems in biology. The material presented here should be considered an appetizer to possibly the most exciting and promising area for the application of network science methods and tools.

12.1 Cell biology and networks

Cell functioning is the consequence of a complex network of interactions between constituents such as DNA, RNA, proteins, and other molecules. A wealth of information on the cell is encoded in the DNA which includes both genes and non-coding sequences. The genes preside over the production of proteins and other molecules, which themselves interact with the genes on several levels. For instance, transcription factors, proteins that activate or inhibit the transcription of genes into mRNA, are the product of genes that may be active or not. Ultimately, genes regulate each other via a complex interaction pattern forming the genetic regulatory network. Proteins in their turn perform functions by forming interacting complexes. The map of physical interactions among proteins defines the protein–protein interaction network. Cell metabolism can also be thought of as a network whose fluxes are regulated by enzymes catalyzing the metabolic reactions. In general, all of these interaction networks are connected through a cascade of biochemical reactions when stimulated by a change in the cell environment, activating an appropriate transcriptional regulation that, in its turn, triggers the metabolic reaction enabling the cell to survive the change.

Researchers have long been constrained to focus on individual cellular constituents and their functions, but the development of high-throughput data collection techniques now provides the possibility of simultaneous investigation of many components and their interactions. Microarrays, protein chips, and yeast two-hybrid screens are some of the techniques that allow the gathering of data on gene, protein, and molecule interactions. These data can be used to obtain a system description usually mapped into interaction networks, which in turn form the network of networks that regulates cell life. These experiments are by construction error-prone, and many false positive and negative signals are usually found in the resulting data sets. This evidence has stimulated lively debate on how far these data sets can be considered reliable, and it is clear that continuous checking with specific high-confidence experiments is needed. On the other hand, for the first time it is possible to gather global information on cell functioning that challenges the research community to develop a systematic program to map and understand the many cell networks.

A first relevant result derived from the systematic analysis of a variety of cellular networks is that their topology is far from random, homogeneous graphs. On the contrary, the degree distribution of most cellular networks is skewed and heavy-tailed. In addition, it is generally accepted that these networks are fragmented into groups of diverse molecules or modules, each one being responsible for different cellular functions. One of the first pieces of evidence comes from the analysis of protein interaction networks (PIN) where nodes are single

proteins and a link among them represents the possibility of binding interactions. PINs have been constructed for several organisms including viruses (McCraith *et al.*, 2000), prokaryotes (Rain *et al.*, 2001), eukaryotes such as the yeast *Saccharomyces cerevisiae* (Schwikowski, Uetz and Fields, 2000; Jeong *et al.*, 2001; Ito *et al.*, 2001; Gavin *et al.*, 2002; Ho *et al.*, 2002), and *Drosophila melanogaster* (Giot *et al.*, 2003). In all cases the topological analysis shows that the PINs have a giant connected component with small-world properties and a degree distribution which strongly differs from the Poisson distribution of homogeneous random graphs. In the case of yeast (see Figure 12.1), the network is characterized by hubs and presents non-trivial correlations, as measured by a slightly disassortative behavior of the degree-degree correlations, and by the clustering coefficient, which is around 10 times larger than that of an Erdős–Rényi random graph with



Fig. 12.1. Map of the protein–protein interaction network of yeast. Figure courtesy of H. Jeong.

the same number of nodes and links. Interestingly, in the case of PINs the network structure can be modeled by taking into account two mechanisms which are thought to be crucial for the understanding of the evolution of organisms, namely gene duplication and divergence (Ohno, 1970). Duplication means that, occasionally, an organism transmits two copies of one or more genes to an offspring. The divergence is then the result of possible mutations of one of the duplicated genes, which will survive and be further transmitted if the resulting organism is viable. This divergence process can lead to differences in the genes' functions and, at the proteome level, to different proteins with slightly different interaction patterns and functionalities. By using these ideas, a series of dynamical network models (see also Chapter 3) aimed at capturing the structure of PINs has been developed (Solé *et al.*, 2002; Vázquez *et al.*, 2003; Wagner, 2003). In addition to the issue of modeling, the representation of protein's interactions as a network provides interesting insights into the biological role and importance of the various proteins. For instance, it turns out that the lethality of a protein (i.e. the likelihood that the removal of a protein is lethal for the organism) is indeed correlated with its degree in the PIN (Jeong *et al.*, 2001). Hubs are about three times more likely to be essential than less connected proteins. This result can be related to the resilience of the PIN under damage and provides a vivid illustration of the relevance of targeted attacks in networks (see Chapter 6).

Further evidence for the complexity of biological networks is found in transcription regulatory networks, in which the nodes correspond to genes and transcription factors, and the directed links to transcriptional interactions. These networks are available for *E. coli* (Shen-Orr *et al.*, 2002) and *S. cerevisiae* (Guelzim *et al.*, 2002; Lee *et al.*, 2002). In this case the corresponding networks again exhibit heavy-tailed out-degree distributions while an exponential behavior is found for the in-degree distribution. These properties provide valuable information on the mechanisms presiding over gene regulation processes. The broad out-degree distribution signals that there is an appreciable probability that some transcription factors interact with a large number of genes. In contrast, each gene is regulated only by a small number of transcriptional factors as indicated by the exponential in-degree distribution. A detailed inspection of the basic transcriptional regulation patterns has been performed in Uri Alon's laboratory by looking at particular subgraphs, called motifs, which are much more recurrent than in the null hypothesis of a random graph (Milo *et al.*, 2002; Shen-Orr *et al.*, 2002; Alon, 2007b). Again, the relation between structure and function constrains the topology of the network and induces deviations from the simple random construction.

Like the networks previously described, the metabolic networks of various organisms display heterogeneous topologies and non-trivial correlations as revealed for instance by the clustering spectrum. Metabolic networks are directed

and have various levels of representations that can consider up to three types of nodes (Jeong *et al.*, 2000; Wagner and Fell, 2001; Lemke *et al.*, 2004). Several simplified descriptions, however, have been proposed. In the most general representation, the metabolites (or substrates) constitute the nodes of the network and are connected through links which represent chemical reactions. Links are therefore directed, and for each metabolite the number of incoming and outgoing links corresponds to the number of reactions in which it participates as a product and ingredient, respectively. Along with the degree heterogeneity, the study of 43 different organisms (Jeong *et al.*, 2000) showed that the ranking (according to the total degree) of the most connected substrates is practically identical for all these organisms: species differences appear at the level of less connected substrates. The global picture which emerges from these results is the existence of a species-independent set of highly connected substrates which connect the other less connected nodes organized in modules serving species-specific enzymatic activities (Jeong *et al.*, 2000). Furthermore, the presence of modular and hierarchical network structures has been clearly documented in the case of metabolic networks (Ravasz *et al.*, 2002).

12.2 Flux-balance approaches and the metabolic activity

Although topological characterization of the cell functional networks can be extremely useful in achieving a system-level picture of the cell's machinery, it is clear that a complete understanding necessitates a dynamical description. In other words, the intensity and temporal aspects of the interactions have to be considered. The case for the importance of the dynamical description is illustrated by metabolic reactions. The chemical reactions defining metabolic networks may carry very heterogeneous fluxes: in fact, these reactions form *directed weighted networks*, and the characterization of the weights, given by the reaction fluxes, may provide crucial information in their study.

The metabolic flux-balance approaches can generate quantitative predictions for metabolic fluxes at equilibrium and provide testable hypotheses on the importance and weight of each reaction (Fell, 1996; Varma and Palsson, 1993; Segrè, Vitkup and Church, 2002). Let us consider a metabolite A_i ; the variation in its concentration $[A_i]$ is given by the balance of the reactions in which it participates,

$$\frac{d}{dt}[A_i] = \sum_j S_{ij} v_j, \quad (12.1)$$

where S_{ij} is the stoichiometric coefficient of metabolite A_i in the reaction j , and v_j is the flux of reaction j . The metabolite A_i is a product in this reaction if $S_{ij} > 0$ and an ingredient in the opposite case $S_{ij} < 0$. In this context, Almaas *et al.* (2004)

have analyzed on the global scale the metabolic fluxes of the bacterium *E. coli*, whose metabolic network contains 537 metabolites and 739 reactions, by assuming a steady state ($d[A_i]/dt = 0$ for all A_i) and that the metabolism is optimized for a maximal growth rate. These assumptions make it possible to compute the flux of each reaction, leading to the conclusion that the distribution of fluxes for *E. coli* can be fitted by the form

$$P(\nu) \sim (\nu + \nu_0)^{-\alpha} \quad (12.2)$$

with an exponent of the order $\alpha \approx 1.5$, as shown in Figure 12.2. The experimental measure of metabolic fluxes is clearly very delicate, but the strong heterogeneity revealed by the numerical approach is confirmed by the experimental data (Emmerling *et al.*, 2002) which yield an exponent of order 1 (Figure 12.2). The metabolic network thus displays both topological heterogeneity and broadly distributed fluxes. Most metabolic reactions have a small flux, and the metabolism activity is dominated by a relatively small number of reactions with very high fluxes.

As presented in Chapter 1, the heterogeneity of weights at a global level may correspond to very different situations at the individual node level. The fluxes can be locally homogeneous or heterogeneous around a given metabolite: in the first case, the metabolite participates in similar proportions in the various reactions, while in the second case, it participates mostly in one or few reactions. The

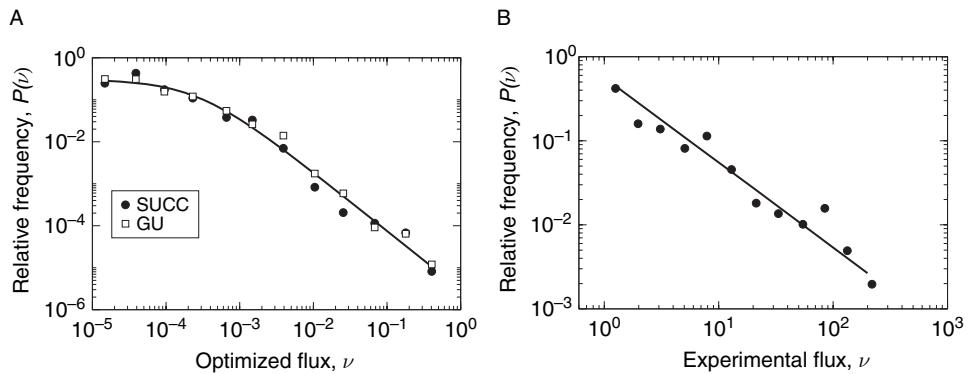


Fig. 12.2. Flux distribution for the metabolism of *E. coli*. A, Flux distribution for optimized biomass production when *E. coli* is put on succinate-rich (circles) and glutamate-rich (squares) uptake substrates. The solid line corresponds to the power law fit $(\nu + \nu_0)^{-\alpha}$ with $\nu_0 = 0.0003$ and $\alpha = 1.5$. B, The distribution of experimentally determined fluxes (Emmerling *et al.*, 2002) from the central metabolism of *E. coli* also displays a power-law behavior with an exponent value $\alpha \approx 1$. Data from Almaas *et al.* (2004).

disparity Y_2 distinguishes between these situations, and is defined for a metabolite i as (see also Chapter 1):

$$Y_2(i) = \sum_{j=1}^{k_i} \left(\frac{\hat{v}_{ij}}{\sum_{\ell=1}^{k_i} \hat{v}_{i\ell}} \right)^2, \quad (12.3)$$

where k_i is the total number of reactions in which i participates, and \hat{v}_{ij} is the mass carried by the reaction j which produces or consumes the metabolite i . If all the reactions producing (or consuming) the metabolite i have comparable fluxes, we will observe the scaling $Y_2(i) \sim 1/k_i$ (for large k_i). In contrast, Almaas *et al.* (2004) obtain for *E. coli* $Y_2(i) \sim k_i^{-\theta}$ with $\theta \approx 0.2$, showing that a relatively small number of reactions have dominant fluxes (see Figure 12.3).

The inhomogeneity of the local fluxes suggests that for most metabolites there is a single reaction dominating both its production and consumption. It is then possible to identify a high flux backbone (HFB) defined as the structure formed by metabolites linked by their dominant reactions. This backbone, which encompasses the subset of reactions that dominate the whole metabolism, forms a giant component including most of the metabolites. Surprisingly, only reactions with high fluxes are sensitive to changes in the environment: some are even completely inactive in one environment and extremely active in another. Changes

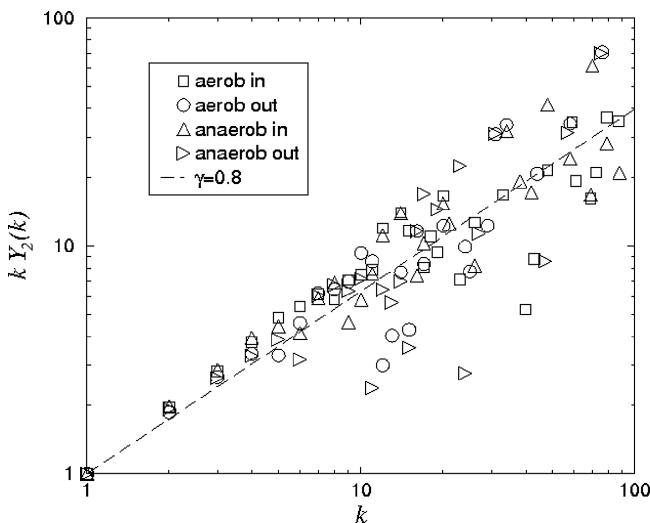


Fig. 12.3. Characterization of the local inhomogeneity of the metabolic flux distribution for the metabolic network of *E. coli*. The measured quantity $kY_2(k)$ calculated in various environments (with aerobic or anaerobic conditions) is averaged over all metabolites and displays a behavior characterized by an exponent $\gamma = 1 - \theta \approx 0.8$ (straight line). Similar behaviors are observed when k is the in- or the out-degree. Figure courtesy of E. Almaas.

in the external conditions therefore lead to (de)activation of certain metabolic reactions, without altering the fundamental identity and function of the major pathways that form the backbone (Almaas *et al.*, 2004). Flux balance analysis combined with network theory thus provides new and interesting insights into the metabolic network. This type of analysis, if done on a growing number of metabolic networks of prokaryotic and eukaryotic organisms, will give access to more information about the interplay between the topology of the network and the reaction dynamics.

12.3 Boolean networks and gene regulation

Modeling the dynamical activity is also crucial in understanding the various genetic regulatory interactions and co-expression patterns. As we have previously explained, transcription factors are themselves produced by genes, and transcriptional interaction can effectively be described as an interaction between the genes. In addition, at a very coarse level, the gene can be represented as having just two possible states: active or inactive. For this reason, although detailed experimental data have only recently started to become available, the modeling of gene regulatory networks can be traced back to the pioneering works of Kauffman (1969), who put forward a simple model of interacting genes, each characterized by a binary on/off variable which describes their activity: an active gene is “on” while an inactive gene is “off”. In the corresponding *random Boolean network model*, the genes are considered as the nodes of the network, and each gene receives inputs from a fixed number K of randomly selected genes: each node has in-degree K , its in-neighbors being chosen at random. For each gene i , the value of its Boolean variable σ_i is updated deterministically at each time step according to

$$\sigma_i(t+1) = F_i(\{\sigma_j(t)\}), \quad (12.4)$$

where $\{\sigma_j(t)\}$ is the set of K variables on the randomly chosen in-neighbors of the node i , and F_i is a Boolean function of its K inputs which is chosen at random for each node i . More precisely, for each set of values $\{\sigma_j\}$, $F_i(\{\sigma_j\})$ is equal to 1 with probability p and to zero with probability $1 - p$. Many studies have been devoted to this model and its variations, and we refer the reader to the book by Kauffman (1993) and to the recent reviews by Bornholdt (2001), Aldana, Coppersmith and Kadanoff (2003), Albert (2004) and Drossel (2007) for more details.

The random Boolean network model is of interest both as a pure dynamical system and as an applied model for biological processes. In both cases, a crucial point concerns the type of dynamical behavior these networks display. In a finite system of N genes (nodes), the configuration space of the set of binary

variables $\{\sigma_i, i = 1, \dots, N\}$ has a finite size 2^N . This ensures that, after a time of at most 2^N steps, the system reaches a state that it has already encountered. Since the dynamics is deterministic, the system will then repeatedly follow the same steps, which define an *attractor*. In fact, the system can either reach a fixed point (point attractor), or go successively and repeatedly through a certain set of configurations, which then form a cycle attractor. Different initial configurations lead to different attractors. In the biological context, the fact that the system goes repeatedly through the same set of configurations can be thought of as defining a set of functionalities. In this framework, each attractor would correspond to a different type of cell in an organism. Each cell contains the same genetic material, so that they may differ only by the pattern of gene activity. The computation of the number and lengths of the attractors is thus of particular interest (see for example Albert and Barabási [2000]; Kaufman, Mihaljev and Drossel [2005]; Correale *et al.* [2006]). Early numerical simulations by Kauffman with relatively small network sizes had led to the result that the mean number of attractors scales as \sqrt{N} for $K = 2$ (and linearly with N for $K > 2$), in agreement with biological data available at the same time, under the assumption that the number of genes in a cell was proportional to the DNA mass. This apparent agreement has spurred much excitement since it seemed to give support to the existence of universal laws, which could be reproduced by very simple models. More recent results, however, have shown that both early empirical data and computer simulations have led to incorrect conclusions. On the one hand, the number of attractors in random Boolean networks with $K = 2$ increases faster than any power law as a function of the number of genes, as uncovered by numerical simulations with large enough sizes and analytical calculations (Drossel, 2007; Kaufman *et al.*, 2005). On the other hand, the sequencing of genomes has revealed that the number of genes is not proportional to the mass of the DNA. The existence of universal laws concerning the number of attractors therefore remains an open issue.

Another, and maybe more essential, characteristic of a Boolean network is that such a system can be either ordered or chaotic: in the ordered phase, most nodes are frozen, i.e. their Boolean internal variable does not change value, while most nodes are continuously changing state in the chaotic phase. The transition from order to chaos (“the edge of chaos”) is characterized by the percolation of clusters of non-frozen nodes. The difference between ordered and chaotic states lies also in the response of the system to a perturbation: in the chaotic phase, a perturbation of a single element can propagate to the whole system and make it jump from one cycle attractor to another, while perturbations die out in the ordered or “robust” phase. This difference is crucial as robust features are clearly needed for a biological system to perform its tasks.

The original model considers that each gene receives inputs from exactly K in-neighbors. If this assumption is relaxed, and the genes form an Erdős–Rényi graph with average in-degree K , it can be shown (Derrida and Pomeau, 1986) that there exists a critical value K_c such that for $K > K_c$ the system is in the chaotic regime:

$$K_c = \frac{1}{2p(1-p)}, \quad (12.5)$$

where p is the probability that the output of a Boolean function is 1 in Equation (12.4). In particular, for $K \leq 2$ the network is always in the ordered phase, while for larger average degrees the ordered phase is obtained only for $p < p_c$ or $p > 1 - p_c$, with $p_c = 1/2(1 - \sqrt{1 - 2/K})$. This implies a narrow range of p values as soon as K is not small. For example, for $K=3$, the system is robust with respect to perturbations only for $p \geq 0.788$ or $p \leq 0.21$. In other words, a fine-tuning of the model is needed in order to build ordered systems, robust with respect to perturbations: the number of inputs has to be small or the bias in the choice of the Boolean functions has to be large.

These results, however, lead to a contradiction with recent experimental evidence. On the one hand, indeed, real cell cycle networks appear to be robust (Fox and Hill, 2001; Li *et al.*, 2004a; Shmulevich, Kauffman and Aldana, 2005), with either fixed points or periodic behaviors. On the other hand, the average number of inputs per element may be larger than 2, and some elements in particular may present a very large in-degree. In fact, broad degree distributions are observed (Guelzim *et al.*, 2002; Tong *et al.*, 2004), and the impact of such topological heterogeneities has to be taken into account. Fox and Hill (2001) therefore consider Boolean networks with various distributions of in-degrees $P(k_{\text{in}})$: the initial model corresponds to $P(k_{\text{in}}) = \delta(k_{\text{in}} - K)$, but Poisson or even scale-free distributions $P(k_{\text{in}}) \sim k_{\text{in}}^{-\gamma}$ may lead to different behaviors. Interestingly, the stability criterion turns out to involve only the first moment of the distribution. More precisely, in the thermodynamic limit, the critical point for the transition from order to chaos depends only on the average degree of the network, and not on the degree fluctuations (Fox and Hill, 2001; Aldana, 2003). Despite this result, more detailed investigations show that scale-free topology leads to more stable systems, with larger fractions of frozen nodes (Fox and Hill, 2001). This enhanced stability stems from the fact that scale-free networks contain many nodes of small degree, which may easily become frozen. Moreover, hubs have many inputs of small degree, and therefore can themselves become frozen. In other words, the order introduced by the large number of nodes with

small in-degree may outweigh the disorder effect due to nodes with a large number of inputs.¹

Aldana (2003) has built on this result by noting that the exponent of the degree distribution is often a more relevant quantity than the average degree in scale-free topologies. In particular, if the in-degree follows the distribution $P(k_{\text{in}}) \sim k_{\text{in}}^{-\gamma}$ with $1 < \gamma < 3$, for $k_{\text{in}} \geq 1$, the average degree is given by $\zeta(\gamma - 1)/\zeta(\gamma)$, where $\zeta(x) = \sum_{k=1}^{\infty} k^{-x}$ is the Riemann zeta function. The transition to chaos therefore takes place at a critical value γ_c given by

$$\frac{\zeta(\gamma_c - 1)}{\zeta(\gamma_c)} = \frac{1}{2p(1-p)}. \quad (12.6)$$

This equation can be solved numerically and the result, displayed in Figure 12.4, shows that scale-free topologies lead to a wide range of parameter values for which the network has a robust behavior. In particular, one observes that the maximum value reached is $\gamma_c \approx 2.48$ for $p = 0.5$. This is an interesting result since many networks seem to present an exponent in the range [2, 2.5], meaning that they are close to the edge of chaos. To reconcile this result with the fact that the critical

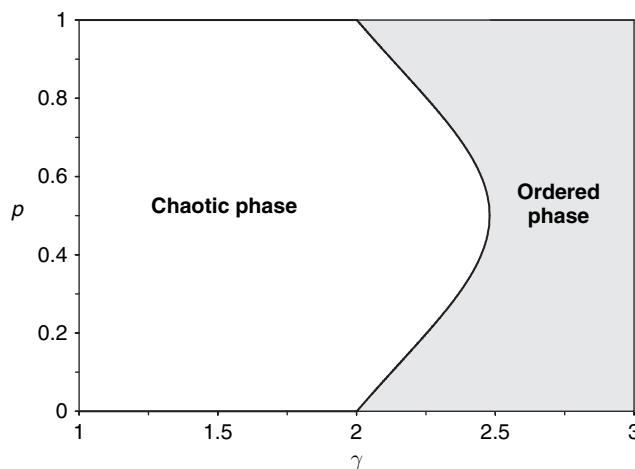


Fig. 12.4. Phase diagram for scale-free networks with power-law in-degree distribution $P(k_{\text{in}}) \propto k_{\text{in}}^{-\gamma}$ with exponent γ and minimum in-degree $k_{\text{in},\min} = 1$. The quantity p describes the probability that each Boolean function has an output equal to one for a given input. From Aldana (2003).

¹ Oosawa and Savageau (2002) also consider various distributions for the out-degree, for a network with uniform in-degree, and show that the percentage of active sites is smaller for a broad out-degree distribution. On the other hand, the system exhibits a high degree of dependence upon the few nodes with large out-degree.

point is independent of the topology (Fox and Hill, 2001), it is important to emphasize that Equation (12.6) is valid under the condition that the minimum in-degree is $k_{\text{in},\min} = 1$: when γ increases, an increasing number of nodes have degree 1, leading to increased order (as $\gamma \rightarrow \infty$ for instance, all nodes have degree 1). For networks with $k_{\text{in},\min} > 1$ the range of stability is strongly reduced.² When perturbations are applied, the hubs turn out to play a crucial role, as usual in complex networks: the network is sensitive to perturbations applied to the most connected elements even in the ordered phase (Aldana and Cluzel, 2003).³

While taking into account the heterogeneous topology of real regulation networks is certainly an important step in their understanding, the interplay between dynamics and topology also depends on the fact that neither the biological networks nor the dynamics of single nodes are random (for example, some particular subgraphs are over-represented with respect to random networks, see Milo *et al.* [2002]). The validation of Boolean network models against empirical data therefore requires these aspects to be taken into account. Kauffman *et al.* (2003) consider, for example, the known architecture of the yeast transcriptional network, and show that the system is marginally stable for randomly chosen Boolean functions, but that the robustness is greatly enhanced if the Boolean functions are *nested canalizing* (Kauffman, 1993). A Boolean function F_i is said to be canalizing if there exists one input j_0 such that fixing σ_{j_0} to a certain value a (0 or 1) completely determines the value of the function: the Boolean variable σ_i becomes then independent of i 's other in-neighbors. A function is nested canalizing if, when σ_{j_0} is not equal to a , there exists another input which canalizes F_i , and so on. It turns out that such rules are widely present in empirically observed networks (Harris *et al.*, 2002), and that such nested canalizing rules lead to stable systems for a wide range of possible topologies (Kauffman *et al.*, 2004). Albert and Othmer (2003) (see also Albert [2004]) study in detail a Boolean model where both the structure of the network and the Boolean rules are taken from empirical data for the fruit fly *D. melanogaster*, and show that this model is able to reproduce the various observed gene expression patterns, even if it neglects the existence of various time scales for the chemical reactions, and abstracts the genes and their products as being either “on” or “off” while their concentration can instead take a continuum of values (see also Shmulevich *et al.* [2005] and Li *et al.* [2004a] for the study of the robustness of other regulatory networks).

² Few studies also take into account that nodes with no input, i.e. 0 in-degree, may exist. These nodes then correspond to the propagation of external variables (Mahmoudi *et al.*, 2007).

³ In this study, a system is perturbed by setting one node's Boolean variable to randomly chosen values instead of the one it should take according to its Boolean function F_i . The overlap between the perturbed and an unperturbed system is then computed and averaged over time.

Such results may seem puzzling, since the modeling of regulatory networks by simple on/off “switches” can appear to be too crude a representation of the biological complexity. Moreover, the evolution of Boolean networks assumes a synchronous updating: in other words, all units are updated together in discrete time steps. In this respect, Klemm and Bornholdt (2005) have shown that, if this particular updating rule is slightly modified by introducing small shifts of update events forward or backward in time, many attractors become unstable. Since a certain amount of noise is unavoidable in real biological systems, this seems to contradict the reproduction of biological results by the Boolean network modeling approach. Zhang *et al.* (2006) and Braunewell and Bornholdt (2007) have therefore studied how the introduction of noise affects the update dynamics of the Boolean network of yeast genes. Very interestingly, both studies (which introduce two different kinds of noise) reach the conclusion that both the biologically relevant attractor and its basin of attraction (the set of initial configurations which lead the system to this attractor) are stable against such noise in the updating rule. All these results suggest that the precise kinetic details of the activation or inhibition of genes might not be essential to understand the system level functioning of regulatory networks (Bornholdt, 2005).

12.4 The brain as a network

While protein interaction networks and metabolic networks operate inside cells, and the food-webs described in the next section concern large-scale phenomena, biological networks can also be encountered at intermediate scales. A prime example is one of the most complex objects in existence: the human brain. Trying to understand its functioning, and how the microscopic biochemical and biophysical processes at work inside and between neurons can give rise to high level phenomena such as perception, memory, or language, is perhaps the most challenging open problem of modern science. Despite much progress made in neuroscience, in particular at the level of molecular and genetic mechanisms, most questions remain largely open.

It is no surprise that the set of neurons and their interconnections through synapses can be considered as a network. In fact, the networked structure of the brain has inspired a whole field of research about *artificial neural networks*. In such systems, the nodes are elements connected together in order to perform certain tasks, such as pattern recognition, function approximation, or data processing. These networks, however, are not “complex” in the sense that they are engineered and put together according to definite patterns. We will therefore not deal with this huge research area and refer the interested reader to the numerous recent books, for example by Dreyfus (2005) and Haykin (2007). On the other hand, the recent

advances and tools developed in the area of complex networks have also been applied to the field of neuroscience (Sporns, 2003; Sporns *et al.*, 2004; Stam and Reijneveld, 2007).

The human cortex contains about 10^{10} neurons with about 10^{14} connections, whose topography, forming the human “connectome,” remains largely unknown (Bota, Dong and Swanson, 2003; Sporns, Tononi and Kötter, 2005). More information is available on some animal brains, and various projects have started to collect data and create repositories for neuroscience researchers in order to foster the efforts of this community. For instance, the Brain Architecture Management System (<http://brancusi.usc.edu/bkms/>) gathers data on brain connectivity structures from different species, collated from the literature. A graph theory toolbox can be found on the website <http://www.indiana.edu/~cortex/connectivity.html> to analyze neural connectivity patterns. The Blue Brain project (see Markram [2006] and <http://bluebrain.epfl.ch/>) aims at creating a biologically accurate, functional model of the brain in a supercomputer, starting from realistic models of neurons connected in a biologically accurate manner.

When considering the brain as a network, a first aspect concerns the *anatomical* connectivity, which describes the biological connections between neurons. Data on this network can unfortunately be obtained only through invasive techniques (injection of markers), so that they are available only for certain animals (mostly the cat or primates such as the macaque). The connectivity patterns reveal a structure formed of densely connected clusters linked together in a globally small-world network (Sporns and Zwi, 2004).⁴ Such a structure turns out to be essential for the brain since it combines two fundamental aspects of its functioning. The first one is *segregation*: similar specialized neuronal units are organized in densely connected groups called “columns”. In order to obtain globally coherent states, these groups need to be interconnected, leading to the second necessary aspect: the functional *integration* which allows the collaboration of a large number of neurons to build particular cognitive states (Sporns, 2003).

A fundamental question concerns the impact of the particular observed structure on the dynamics of the brain. A natural way to tackle this problem consists of simulating the dynamics of neurons interacting on various topologies (see Stam and Reijneveld [2007] for a recent review). For example, Kötter and Sommer (2000) have used the real (experimentally known) anatomical connectivity of the cat cortex to simulate the propagation of neuronal activity and compare it with a randomly connected network in order to demonstrate the existence of a relationship between

⁴ Interestingly, the seminal paper of Watts and Strogatz (1998) which introduced the Watts–Strogatz model showed that the neural network of *Caenorhabditis elegans* is a small world.

structure and function. Other studies have compared various network structures and shown that the combined strong local clustering and short topological distances favor the interplay between segregation and integration (Sporns and Tononi, 2002), and allow faster learning or information transmission than regular or fully random topologies (Simard, Nadeau and Kröger, 2005). Many works have also focused on the effect of particular model topologies (small-world, scale-free...) on synchronization phenomena on complex networks, as described in Chapter 7, showing for example how predefined dynamical patterns of firing neurons can be obtained by specifically designing the corresponding network of connections between neurons (Memmesheimer and Timme, 2006). In this framework, Zhou *et al.* (2006) study the synchronization dynamics in a realistic anatomical network of the cat cortex (see also Zhou *et al.* [2007]). This approach considers a multilevel network in which each node is a region of the cortex. Since detailed anatomical information about the connectivity at the neuronal level is lacking, these regions are modeled by small-world subnetworks of interacting neurons. The regions (nodes) are themselves connected using the anatomically known connections. Biologically plausible dynamics emerge as a result of this architecture, and the synchronization patterns can be analyzed in detail, providing insights into the relationship between the hierarchical organization of the network structure and its dynamics.

Interestingly, complex network approaches have recently been used to study a different kind of “brain network.” The dynamics of the neurons give rise to a (virtual) network, which describes the *functional* connectivity of the brain. More precisely, in this representation the nodes of the functional network are millimetric regions of the brain, and a link is considered to exist between two nodes if the corresponding regions display a correlated activity. In this construction, the links are therefore not necessarily physically present, but they describe the existence of correlations in the dynamics. Data on these correlations can be obtained through various non-invasive techniques such as electroencephalography (EEG), magnetoencephalography (MEG) or functional magnetic resonance imaging (fMRI), and therefore even the human brain can be examined in this way (see Bassett and Bullmore [2006]; Stam and Reijneveld [2007] for reviews on recent experiments). The activity of brain areas of millimetric size (called “voxels”) is measured either by the electromagnetic field generated locally by neuronal activity (EEG or MEG) or by the level of oxygen in the blood (fMRI), which also depends on the local activity of neurons. Time-series of the activity of the voxels are recorded, and a link between two voxels is drawn whenever the linear correlation coefficient of their time-series exceeds a given threshold. The obtained network, which defines the functional connectivity, is therefore the result of the activity of the brain itself, and different states of the brain (for instance, rest, or certain predefined tasks such as finger tapping) give rise to different networks. Various questions then

naturally arise about the structure of these networks, and the relationship between anatomical and functional connectivity (Sporns, 2004). The application of graph theoretical techniques to the analysis of functional connectivity is in fact quite recent (Dodel, Herrmann and Geisel, 2002), and has allowed interesting properties to be uncovered, such as small-world features with large clustering coefficient and small average shortest path length (Stam, 2004), and scale-free degree distributions (Eguíluz *et al.*, 2005a), possibly with exponential cut-offs (Achard *et al.*, 2006). Strikingly, recent studies of the functional connectivity of patients suffering various types of brain diseases (Alzheimer's disease, schizophrenia, epilepsy) have provided evidence of an association of these diseases with deviations of the functional network topology from the small-world pattern, with an increase of the average shortest path length (see e.g. Stam *et al.* [2007] and the review by Stam and Reijneveld [2007]).

All these experimental approaches are still in their initial stages, and show how the integration of graph theoretical analysis and the use of complex network theory opens the door to many further investigations on the links between anatomical and functional connectivity patterns in the brain. For instance, one can wonder how the anatomical structure evolves during growth and development, possibly through feedback effects in which function shapes the anatomical structure. Finally, the impact of genetic or environmental factors on both anatomical and functional properties remains to be asserted, and the possible relationship between pathologies and particular network properties appears to be a very exciting area of research (Stam and Reijneveld, 2007).

12.5 Ecosystems and food webs

The stability of ecosystems and the preservation of biodiversity are very real and critical topics. Biodiversity reflects the number, variety, and variability of living organisms which is, at present, strongly decreasing. One of many possible examples is the 30% decrease in the number of vertebrate species in the last 30 years. In this context, the theoretical approach to model ecosystems provides valuable insights into issues such as the fragility and stability of populations and species interactions.

An ecosystem can be abstracted as a set of species interacting with the environment. From a global point of view, it is a system that, through the food chain, allows the flow of energy across various levels of the ecosystem, from the basic nutrients to higher order vertebrates. This energy flow is encoded in the ecosystem food web which is specified by the set of relations identifying, at the level of species, "who eats whom". This set of relations can be naturally represented as a directed network where the nodes are the species, and any directed link from a

species A to another species B implies that the species B feeds on the species A. More precisely, a directed link is present between the “prey” or “resource” and its “predator” or “consumer”. Links are therefore in the direction of the energy flow as the prey is the energy source for the predator.⁵ While the nodes represent species in most studies, they can also represent a whole group of taxonomically related species or another type of grouping (in some studies, the term “taxa” is used instead of “species”). Food webs are also organized in *trophic levels* which are sets of species sharing the same set of predators (out-neighbors in the directed network) and prey (in-neighbors) and therefore play equivalent roles in the ecosystem. Food webs usually have a small number (typically about 3) of trophic levels. Figure 12.5 displays a typical example of such a chain of trophic levels: green plants form the first trophic level, herbivores the second, while carnivores form the third and usually the last trophic level. The flow of energy in the system goes from the lowest trophic level to the highest. Several catalogs of food webs have been proposed in the last three decades, and we refer the interested reader to the book of Pimm (2002) for an updated review. The recent availability of data sets with increasing accuracy and the interest in complex networks have also stimulated the activity in the field, as testified in the reviews of Drossel and McKane (2003) and Montoya *et al.* (2006).

Food webs, as other networks, can be characterized by simple global parameters such as the number of links E , the number of species (nodes) N , and average clustering coefficient. More local information can be gained from quantities such as the degree distribution $P(k)$. Since the network is directed (see Chapter 1), the number of links per species E/N is equal to the average in-degree – and also to the average out-degree –, i.e., to the average number of prey or predators. The connectance in its turn measures the “density” of links and is given for a directed network by⁶

$$\mathcal{D} = \frac{E}{N(N - 1)}, \quad (12.7)$$

which is the ratio between the actual number of directed links and the maximum possible number of links, self-loops being excluded. As reviewed by Dunne (2006), there is a lively debate in the research community about the values of E/N and \mathcal{D} and their dependence with the food-web size N . In particular, the average number of prey or predators of a species, E/N , has been considered for a long time to be almost constant, with values in the interval [1, 2]. This conjecture, known as the “link-species relation,” means that species have the same constant number of prey regardless of the global size of the food web. It has, however, been argued

⁵ In this representation omnivorous species will have a large in-degree and cannibal species present self-loops.

⁶ Note that, in undirected networks, the connectance is $2E/N(N - 1)$ (see Chapter 1).

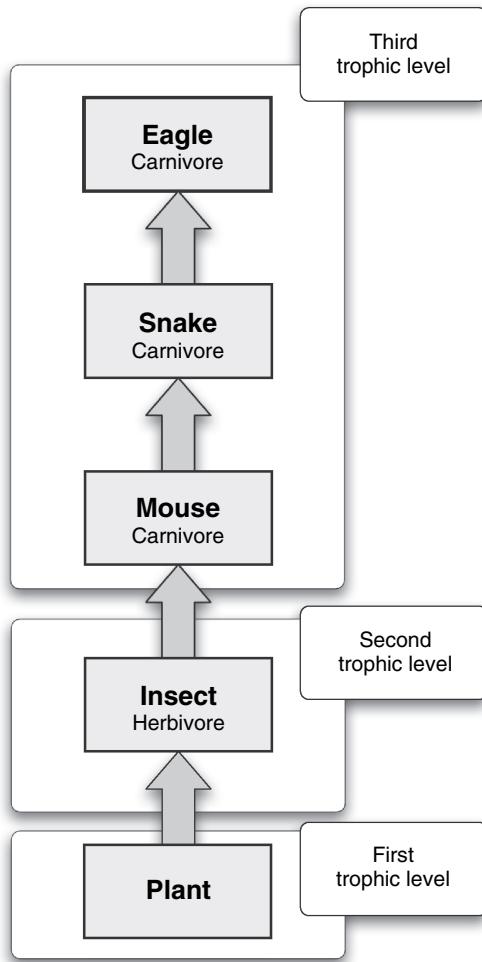


Fig. 12.5. Example of a terrestrial food chain illustrating the concept of trophic levels. The energy flow goes from the lower level usually constituted by plants to the higher order levels such as carnivores.

that the number of different prey does not in general increase with N , in contrast to the number of predators against which a species has to protect itself. The recent availability of larger data sets supports the idea that E/N does depend on N (Dunne, 2006), with empirical studies showing that in general

$$E \sim N^b \quad (12.8)$$

where the exponent b can vary from $b = 1$ (which corresponds to the link-species relation) to $b = 2$. The case $b = 1$ corresponds to a sparse graph with a finite average degree and with a small connectance. In contrast, the case $b = 2$ corresponds

to a very dense graph whose average degree increases with size and where the connectance is constant. The empirical results demonstrate that the connectance can reach values as large as 10% and that the value of b is not universal, varying from one ecosystem to the other.

Various empirical studies have allowed the characterization of other topological properties of food webs. Montoya and Solé (2002) and Williams *et al.* (2002) have shown on several examples of food webs that the diameter is small (with typically “two degrees of separation”) with a large clustering coefficient. A more extensive analysis has been performed by Dunne *et al.* (2002a) who analyze 16 different ecosystems with different sizes (the number of species ranging from 25 to 172) but also different connectances (from 0.026 to 0.315) and average degrees (from about 3 to 50). Larger connectance corresponds to larger average degree, larger clustering coefficient and smaller diameter, all simple consequences of a larger density of links. Interestingly, some particular subgraphs (motifs) have also been shown to be over-represented with respect to a random reshuffling of links (Milo *et al.* 2002). Figure 12.6 shows two particular regular patterns which are often found in food webs. For instance, the abundance of the bi-parallel motif shown in Figure 12.6B suggests that species that share the same predator also tend to share the same prey.

A number of empirical studies have tackled the analysis of degree distributions and other statistical properties. It is worth remarking that the limited size of the available data sets is an obvious limitation to this type of analysis. In general, however, the degree distribution does not exhibit a remarkable level of universality (see also Figure 12.7). While the degree distribution is in some cases approximated by homogeneous and Poisson-like behavior, it is skewed and possibly heavy-tailed in other instances (Camacho, Guimerà and Amaral, 2002b; Montoya and Solé, 2002; Dunne *et al.*, 2002a). In fact, the study of Dunne *et al.* (2002a) shows that $P(k)$ evolves with the connectance of the food web; the larger the connectance, the more homogeneous the degree distribution.

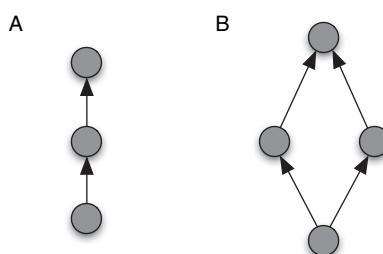


Fig. 12.6. Motifs over-represented in seven food webs (Milo *et al.*, 2002). A, Observed in five food webs. B, Observed in all seven food webs.

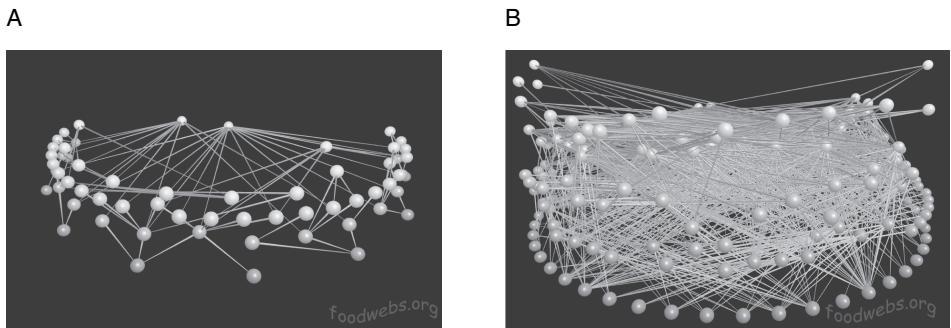


Fig. 12.7. Two examples of food webs (considered here as undirected). A, Example of a common food web with a skewed degree distribution (grassland). B, A food web with an exponentially decaying degree distribution (El Verde rainforest). Figures courtesy of Pacific Ecoinformatics and Computational Ecology Laboratory, www.foodwebs.org.

A review of topological models for food webs is beyond the scope of this book, and we mention only a few of the interesting models put forward by the research community. The most basic model simply considers that a directed link is drawn from one species to another with uniform probability p . It is therefore simply a directed version of the Erdős–Rényi model that does not take into account biological constraints, and reproduces simply the connectance (and the average degree), the average number of links being $p \times N(N - 1)$ (Cohen, Briand and Newman, 1990). Despite its simplicity, it can be used as a benchmark in order to observe and identify deviations from a purely random model. To restore some structure between the trophic levels, the cascade model (Cohen *et al.*, 1990) assigns to each species i a random number n_i drawn uniformly from the interval $[0, 1]$. A species i can prey on another species j only if $n_i > n_j$ and with probability p . For two randomly chosen species i and j , the probability that $n_i > n_j$ is simply $1/2$, so the probability that there is a link from i to j is $p/2$, and the average number of links is therefore $p \times N(N - 1)/2$. Taking $p = 2\mathcal{D}N/(N - 1) \simeq 2\mathcal{D}$ (for large $N \gg 1$), where \mathcal{D} is the desired connectance, one obtains the correct number of links on average. This model, however, has a certain number of drawbacks, including the overestimation of the food-chain length.

A more elaborate framework is the niche model proposed by Williams and Martinez (2000). In this model, analogously to the cascade model, species are described by a “niche parameter” n_i but a species i can prey on all species contained in an interval of width r_i centered at a random position c_i drawn uniformly in the interval $[r_i/2, n_i]$ (see Figure 12.8). Species may remain isolated and are thus removed from the web. Species with the same list of prey and predators are instead merged.

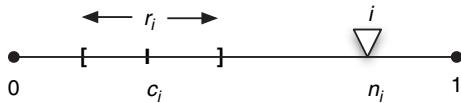


Fig. 12.8. Niche model (Williams and Martinez, 2000): Each species i preys on all species in the interval of length r_i centered on $c_i \in [r_i/2, n_i]$.

The length of the niche interval is chosen as $r_i = xn_i$ where x is a random variable (independent from n_i) drawn in the interval $[0, 1]$ according to the distribution $p(x) = b(1 - x)^{b-1}$, this functional form being chosen for the sake of simplicity. In contrast with the cascade model, a species can prey here also on species with larger values of n_i (up to $n_i + r_i/2$ if $c_i = n_i$). The average of n is $1/2$ and thus $\bar{n} = \bar{x}/2 = 1/[2(1 + b)]$. The values of the parameters b and N then determine the total average degree and the directed connectance: $\langle k \rangle = 2N\bar{n}$ and $\mathcal{D} = \bar{n}$. The niche model yields a good agreement with empirical data for various food-web characteristic quantities such as the food-chain length, the fraction of top species (having no predators), and the fraction of basal species (having no prey). They therefore provide an interesting starting point for a realistic modelling of food webs (Williams and Martinez, 2000). Interestingly, Camacho, Guimerà and Amaral (2002a) have obtained analytically the degree distributions characterizing the niche model. The results show that both the in-degree and the out-degree of a species i decay very rapidly (at least as an exponential) with a typical cut-off of the order of the average degree. In addition there is only one characteristic parameter $\langle k_{in} \rangle = \langle k_{out} \rangle$ governing these distributions. Even if these results are in relative good agreement with empirical data for some ecosystems, they do not explain why some ecosystems with a small connectance have a distribution decaying much slower than an exponential. On the other hand, many factors are not taken into account in the niche model, including the predator–prey dynamics occurring on the network.

12.5.1 Dynamics and stability of ecosystems

Although much effort has been devoted to the study of the static and topological properties of food webs, it is clear that their descriptions must include the dynamics of species: in particular, the number of individuals per species varies, and the number of species itself evolves, on a larger time scale, through extinction, mutation, and invasions. The interaction strength between species in its turn may depend upon the species populations and on external resources. The basic building block of dynamical ecosystem models is the prey–predator interaction. Population dynamics models for general webs are usually considering the Lotka–Volterra scheme. If

we define by N_i the population size of the species i , the Lotka–Volterra equations for a web of N interacting species can be written as

$$\frac{dN_i}{dt} = N_i \left(b_i + \sum_j a_{ij} N_j \right). \quad (12.9)$$

In this set of equations the parameters b_i and a_{ij} are usually independent of the population size and encode the birth/death rates of species and their interactions. In general, b_i is a positive birth rate for the basal species, and a negative death rate for the other species. The parameters a_{ij} describe the rate at which predators convert prey (food) into predator births and conversely the rate of prey consumed by each predator. Therefore, if i is a predator and j is prey, a_{ij} is positive. If i is prey and j is a predator a_{ij} is instead negative. In general, given the predator-prey relation (i, j) , the two interactions are related as $a_{ij} = -\lambda a_{ji}$. The factor λ , called the ecological efficiency, expresses the fraction of consumed prey that is actually transformed into the birth of predators. Some general variations of the model consist of including a saturation term by allowing $a_{ii} < 0$ for basal species and a competition term $a_{ji} = a_{ij} < 0$ if two predators i and j feed on the same species. The food web structure and topology enters the dynamics of the ecosystem by imposing the structure of the interaction matrix a_{ij} and therefore defining the set of Lotka–Volterra equations. The general study of this form of equations has, however, led to models where random numbers are used for the interactions a_{ij} and in some cases the constraints on the opposite sign of a_{ij} and a_{ji} are relaxed.

One of the most important issues in the study of food webs concerns their stability with respect to perturbations, and the effect of their structure on this stability (recent reviews can be found for example in Pimm [2002] and Dunne [2006]). Early considerations lead to the argument that a large number of paths through each species is necessary in order to limit overpopulation effects, since each species is regulated by many others (MacArthur, 1955). More complex food webs, i.e. food webs with more species and more links, would therefore be more stable.⁷ In this context, May (1973), following the work of Gardner and Ashby (1970), produced an argument in the opposite direction, showing that the stability is *decreased* by increasing the number of species or the connectance of the network. May used a linear stability analysis by considering small perturbations around a supposed equilibrium point N_i^* of the Lotka–Volterra equations so that $N_i = N_i^* + \delta N_i$. In this case it is possible to

⁷ We note here that the definition of complexity in terms of connectance is not analogous to the modern definition of complex systems we have provided in this book. In this context we are more inclined to talk about the intricacy of the web.

write the linearized Lotka–Volterra equations for the deviations of the population size as

$$\frac{d}{dt}(\delta N_i) = \sum_j w_{ij} \delta N_j, \quad (12.10)$$

where w_{ij} is the community matrix obtained by the Taylor expansion of each basic Equation (12.9) whose elements characterize the interaction among species near the equilibrium point. The diagonal elements w_{ii} represent the relaxation time to equilibrium in the absence of any interactions and for the sake of simplicity are all set to be $w_{ii} = -1$ by a proper rescaling on the time scale. The choice of the w_{ij} corresponds to the choice of the food-web structure and in order to be as general as possible May (1973) considers that the non-diagonal matrix elements are randomly set equal to zero with probability $1 - \mathcal{D}$ (the connectance), and chosen from a distribution of average 0 and width w with probability \mathcal{D} . The issue of stability is then studied by looking at the behavior of the set of differential equations. Stability implies that the real part of all eigenvectors of w_{ij} are negative so that the system goes back to equilibrium after a perturbation. By using general results from random matrix theory (Mehta, 1991), it is possible to reach the general conclusion that in the space of all possible random community matrices the system will be almost certainly unstable (with probability 1) if

$$w\sqrt{N\mathcal{D}} > 1, \quad (12.11)$$

and almost certainly stable otherwise. In other words, the stability decreases for increasing values of $w\sqrt{N\mathcal{D}}$ as the negative eigenvalues move to zero. According to this result, food webs should become unstable as the complexity (measured by the average strength, the number of species, or the connectance) increases. While this could be accepted in the framework of constant connectivity networks ($E \propto N$, i.e. finite $N\mathcal{D}$ as $N \rightarrow \infty$), it is, however, hardly compatible with the evidence of superlinear behavior $E \propto N^b$ with $b > 1$, which implies that many food webs should be unstable.

Although the seminal work of May sets the framework for the discussion of ecosystem stability, the general results concerning random community matrices are obtained at the price of introducing several modeling assumptions. While one issue is the structure of the Lotka–Volterra equations which can be certainly generalized to more complicate functional form, possibly the weakest point is the assumption of random matrix elements. Real food-web networks are far from random graphs and the structure used for the community matrix should contain elements of biological realism. Since the 1970s, work has been done in constructing more realistic topologies and studying their stability (De Angelis, 1975; Lawlor, 1978; Yodzis, 1981).

Another important issue is the definition of stability itself, which can be generalized to include processes that go beyond small perturbation to the equilibrium state, such as the complete deletion of a species (Pimm, 1979). In this context, various authors investigate the link between the topology of a food web and its stability (Jordán and Molnar, 1999; Solé and Montoya, 2001; Dunne, Williams and Martinez, 2002b; Memmott, Waser and Price, 2004; Allesina, Bodini and Bondavalli, 2005). In particular, considering the results concerning the resilience and robustness of complex heterogeneous networks presented in Chapter 6, several studies deal with the effect on the food-web topological structure of the removal of a certain fraction of species (Solé and Montoya, 2001; Dunne *et al.*, 2002b; Memmott *et al.*, 2004) and the corresponding interactions. Species (nodes) can be removed either at random or by following selective procedures such as targeting species with largest degree. Since food webs are directed, the removal of species with largest out-degree, i.e. those with most predators, has been studied, to model for instance the extinction of large herbivores (Solé and Montoya, 2001). The robustness with respect to the removal of the most connected nodes can also be investigated (Dunne *et al.*, 2002b), with the exception of basal species whose disappearance generally leads to the immediate collapse of the whole ecosystem. As customary in the study of complex networks (see Chapter 6), the robustness of a food web can be measured by the size of the largest connected component after a fraction f of nodes has been removed. In the ecological context, however, another key assessment of the importance of species removal is the possibility of *secondary extinctions*, i.e. the disappearance of species caused by the removal of others. In particular, species that lose all their prey become extinct: a node whose in-degree vanishes because of the removal of other nodes is therefore removed from the network. The studies of Solé and Montoya (2001) and Dunne *et al.* (2002b) show that food webs are robust with respect to random removals but highly fragile and vulnerable if species with large degree are removed, with a large number of secondary extinctions in the latter case.⁸ Notably, both in- and out-degree play an important role: a high degree corresponds to a high potential to affect the community, whether this degree is due to a large number of prey or predators.⁹ Food webs with heterogeneous topologies appear to be the most vulnerable, but the difference between the various removal procedures is evident also in food webs with exponentially decaying degree distributions. A slightly different definition of robustness considers the fraction of removed species that triggers a total loss of at least 50% of the

⁸ The protection of basal species leads in certain cases to a vulnerability decrease (Dunne *et al.*, 2002b).

⁹ The extinction of a particular species with relatively few links may also lead to important secondary extinctions, in a way linked to detailed, precise and non-general connectivity patterns.

species. By using this definition it is found that the robustness of the web increases with its connectance (Dunne *et al.*, 2002b).

Purely structural studies of robustness completely ignore the population dynamics, while the linear stability analysis neglects the importance of network structure. The integration of both topological and dynamical aspects therefore appears necessary in order to provide a more complete picture. Recent attempts in this direction include the work of Borrrell, Ebenman and Jonsson (2000), who consider various fixed structures in which species are divided into functional groups, all species inside a functional group having the same prey and predators, and evolve according to Lotka–Volterra dynamics. The risk of cascading extinction events is then shown to be smaller when the number of species per functional group increases. Chen and Cohen (2001) also consider web structures obtained from the cascade model together with Lotka–Volterra dynamics, but on small number of species. Quince, Higgs and McKane (2005) finally build on the model of Caldarelli, Higgs and MacKane (1998) and Drossel, Higgs and MacKane (2001) to study community responses to species deletions using realistic global dynamics with realistic network structures. In this last study, no evidence was found that complexity destabilizes food webs. We finally refer the reader to Jordán and Scheuring (2004), Martinez, Williams and Dunne (2006), Dunne (2006) and Montoya *et al.* (2006) for reviews on these promising new research directions.

12.5.2 Coupling topology and dynamics

The investigation of stability and dynamical features of food webs naturally raises the question of the effect of population dynamics in the shaping of food-web topology. In other words, a more fundamental approach to the modeling of the formation and structure of food webs should consider the coevolution of the populations and their interaction patterns. Generally, evolutionary models separate the dynamics on short and long time scales. On the short time scale, species are fixed and their populations evolve according to the predator–prey dynamics. On long time scales species mutate or change their predation links because of competition effects. For instance, Caldarelli *et al.* (1998) consider a short time scale evolution where a certain percentage of each species is eaten by its predators, while Drossel *et al.* (2001) and Lässig *et al.* (2001) use more detailed Lotka–Volterra-like dynamics. These models also allow for mutations and changes of predator–prey relations at large time scales. The resulting network structure and the strength of the interactions are therefore the outcome of the dynamical evolution at the various time scales (Caldarelli *et al.*, 1998; Drossel *et al.*, 2001; Lässig *et al.*, 2001; Quince *et al.*, 2005).

Evolutionary models are capable of generating artificial food webs whose structures are compatible with real data. For instance, a quantity of interest is the number of species present in different trophic levels. This function – the shape of the food web – is represented in Figure 12.9 for different real food webs, and the model of Lässig *et al.* (2001) is able to reproduce such a shape and to explain the existence of a maximum. At low trophic levels, species feed on, and compete for, limited external resources (large open symbols in Figure 12.9A), while resisting predation. The model of Lässig *et al.* (2001) predicts that, as the trophic level l increases, the number of species increases thanks to the increasing prey diversity, but that the population of each species decreases. For still increasing trophic level, therefore, more and more species have populations too small to support predation, and the number of species starts to decrease.

Other ecological ingredients such as predator–predator competition or the limited number of resources are obviously crucial ingredients for a realistic description of food webs (Chen and Cohen, 2001; Jordán and Scheuring, 2004; Martinez *et al.*, 2006; Dunne, 2006). Although no definitive model exists at this time, the recent modeling approaches suggest that it is crucial to allow the population dynamics to affect and coevolve with the food-web structure.

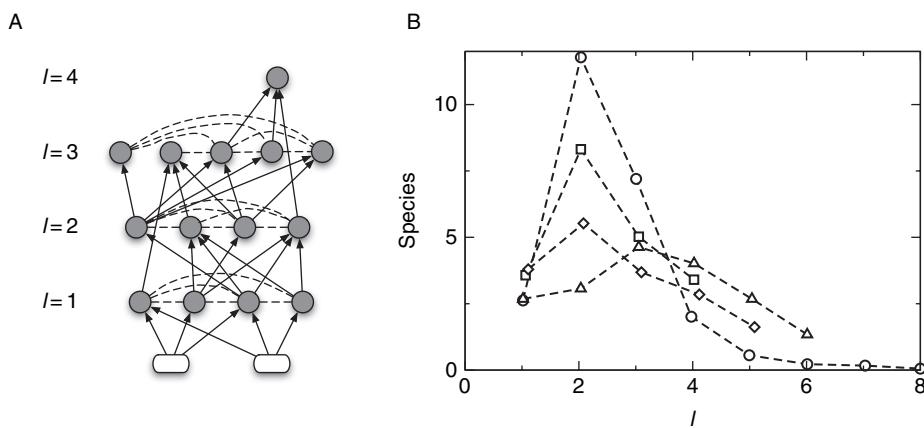


Fig. 12.9. A, Example of a food web (Pamlico estuary food web with 14 species represented by filled circles) with four trophic levels starting from detritus ($l = 1$) which feed on external resources (large open symbols) and finishing with predatory fishes ($l = 4$). Arrows point from prey to predator and the dashed lines connect species with some common prey or predators. B, Average number of species versus the trophic rank for a set of natural ecosystems (different symbols correspond to different data sets). Data from Lässig *et al.* (2001).

12.6 Future directions

Without any doubt, biological research is producing terrific advances at an unprecedented pace. At the same time, however, many areas are still in their infancy as the research community is facing for the first time such an abundance and continuous stream of data. While data reliability has to be assessed and its precision improved, a large part of our understanding depends on the possibility of organizing these data in a meaningful way. Multiscale modeling of the dynamics occurring on networks unavoidably forms one of the key issues in our understanding of biological systems. Although in the biological realm a certain degree of universality appears to be present in the large-scale organization of networks and the ensuing dynamics, it is also clear that it is important to introduce specific features characterizing the different scales and phenomena of the system at hand. At the same time we can see that new challenges lie ahead, in which networks and the dynamics of networks are important. Applications ranging from drug design and pathogen–host interactions to disease networks (Uetz *et al.*, 2006; Goh *et al.*, 2007) are just starting to be explored.