Review TRENDS in Ecology and Evolution Vol.20 No.6 June 2005

Full text provided by www.sciencedirect.com
SCIENCE DIRECT

Network thinking in ecology and evolution

Stephen R. Proulx[1], Daniel E.L. Promislow[2] and Patrick C. Phillips[1]

[1]Center for Ecology and Evolutionary Biology, 5289 University of Oregon, Eugene, OR 97403-5289, USA
[2]Department of Genetics, The University of Georgia, Athens, GA 30602-7223, USA

Although pairwise interactions have always had a key role in ecology and evolutionary biology, the recent increase in the amount and availability of biological data has placed a new focus on the complex networks embedded in biological systems. The increased availability of computational tools to store and retrieve biological data has facilitated wide access to these data, not just by biologists but also by specialists from the social sciences, computer science, physics and mathematics. This fusion of interests has led to a burst of research on the properties and consequences of network structure in biological systems. Although traditional measures of network structure and function have started us off on the right foot, an important next step is to create biologically realistic models of network formation, evolution, and function. Here, we review recent applications of network thinking to the evolution of networks at the gene and protein level and to the dynamics and stability of communities. These studies have provided new insights into the organization and function of biological systems by applying existing techniques of network analysis. The current challenge is to recognize the commonalities in evolutionary and ecological applications of network thinking to create a predictive science of biological networks.

Introduction

During the late 1970s, a flight attendant named Gaetan Dugas regularly visited gay bathhouses throughout North America and, over the course of the next several years, had sexual intercourse with hundreds of individuals. By 1981, a small group of men displaying an unusual array of symptoms, the likes of which were typically only seen in immuno-suppressed adults, began to appear in hospitals in New York, Los Angeles and San Francisco [1–4]. In the course of follow-up work with 40 of these patients, it was discovered that they were all part of a single network of sexual contacts, with Dugas, 'Patient Zero', at its center [1,3]. That all of these patients shared a common attribute (i.e. direct or indirect contact with Dugas within a network of sexual partners) contributed to the early conclusion that AIDS was caused by a sexually transmitted blood-borne agent [1,3,4]. This early application of network thinking to the spread of the AIDS epidemic was the first of many such attempts to use the existing mathematical theory of networks to understand epidemiological processes in general, and HIV in particular.

In human sexual contact networks, most individuals have just one sexual partner, whereas a small class of individuals might have tens or hundreds of partners per year. The frequency distribution of sexual partners typically follows a 'power-law' distribution (see Glossary). The specific structure of the network helps us to not only trace the origin and cause of diseases, but also predict the potential success of interventions, such as quarantine and vaccination [5,6]. In particular, the power-law distribution common to sexual contact networks allows the spread of diseases with extremely low transmission rates [7–11], but high levels of clustering (Box 1) can reduce the rate of

Glossary

Average path length/diameter: the mean shortest path between all nodes in the network.
Betweenness: the number of shortest paths that the focal node lies on.
Bipartite graph: a network with two distinct types of node.
Closeness: the mean shortest path between a focal node and all other nodes in the network.
Component: a group of nodes that are mutually interconnected.
Clustering coefficient: a measure of the proportion of neighboring nodes that can be reached through the nodes other neighbors; calculated as the proportion of a focal nodes neighbors who are themselves neighbors.
Degree/Connectivity: the number of edges that connect the focal node to other nodes.
Degree distribution: the frequency distribution of the individual node degree for an entire network.
Directed graph: nodes in a directed graph are connected by an asymmetric relationship, such as predation.
Edge: interacting nodes are connected by edges.
Graph theory: a branch of mathematics dealing primarily with the statistical description of static networks.
Long-tailed distribution: any degree distribution that decreases more slowly than exponentially over a portion of the range.
Motif: a small pattern within a network. For example, a feedback loop where two nodes each effect the other.
Node: an individual element within a network.
Poisson degree distribution: a network formed by randomly connecting a fixed number of nodes has a Poisson degree distribution. Such distributions are characterized by a modal hump at the mean degree with exponentially decreasing tails.
Power-law degree distribution: a network with a degree distribution described by $f(k) = \beta k^{-\alpha}$; also called scale-free distributions because there is no modal hump.
Preferential attachment model: the formation of a network by connecting nodes to nodes that already have many connections.
Shortest path: the path that traverses the minimum number of edges between the two nodes.
Small-world property: a network is said to have the small-world property if the diameter of the network is small relative to the size (total number of nodes) of the network.
Undirected graph: nodes in an undirected graph are connected by a symmetric relationship, such as physical interactions.

Corresponding author: Proulx, S.R. (proulx@proulxresearch.org).
Available online 16 April 2005

 0169-5347/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tree.2005.04.004

Box 1. What is a network?

A network is any collection of units potentially interacting as a system. In the most simple case, a network can be represented by a set of uniform nodes connected by undirected edges, as in Figure Ia. The nodes can represent units at most levels of the biological hierarchy, from genes and proteins to neurons and organs and limbs, and from individuals in a population to species in a community. Edges usually represent some kind of interaction between nodes, including transcriptional control, biochemical interaction, energy flow and species interactions.

Mathematical approaches can be used to understand the behavior of networks depending on the level of resolution and degree of quantification desired (Box 2). The recent explosion of work on biological networks has mostly utilized results from graph theory, in part because it enables the properties and behavior of a network to be quantified in a few simple statistics. In particular, the node degree distribution has been reported for a wide array of networks because it is easy to measure and can help to determine how the network was formed [14].

The degree distribution is an example of an aggregate statistic: it can be obtained solely by examining the properties of individual nodes within the network, the degree of the node in this case. For example, the degree distribution of a food web can be calculated by observing a species in the network at random and counting the species that it eats or is eaten by. We would then be able to plot the degree distribution, but would not be able to draw the network. The clustering coefficient, however, can only be calculated by observing larger fragments of the network [14]. We need to be able to observe both the connections from a focal node and the connections made by its neighbors. Recently, researchers have focused their attention above the level of individual nodes to look at the distribution of network motifs, which will be important if motif structure is related to modular functionality.

Whereas clustering coefficients and network motifs are local properties of networks, the shortest path between two nodes depends on the structure of the entire network (Figure Ia) Statistics based on measuring shortest path lengths, such as closeness and betweenness, can be perturbed by changes far removed from the focal node [14,32]. Network diameter, relative to the total number of nodes, is a global measure of how integrated the network is. Networks with small diameters relative to the number of nodes are said to have the small-world property and often have a few highly connected nodes that make it possible to traverse the network rapidly [14].

Several models of random network formation can be compared to the statistical features of observed networks. These can all be considered neutral models in that the network is built through a series of random steps that are not affected by the structure of the network. The oldest such model forms the network out of N nodes, where two nodes are connected with probability p. This yields a Poisson degree distribution in the limit of large N (Figure Ib) [14,65]. Biological networks, however, tend to have more asymmetry in node degree than appears in Poisson networks. There are many nodes with only one or a few connections, but there are also a few nodes with many more connections than the average degree, leading to a long-tailed degree distribution.

One way to achieve a long-tailed degree distribution is the preferential attachment model of Barabasi *et al.* [67,68]. In this model, the network is initialized with a few unconnected nodes. New nodes are added to the network and are more likely to have edges connecting them to highly connected nodes. This causes a 'rich get richer' phenomenon that produces networks with power-law degree distributions that appear linear on a log–log plot (these are also known as scale-free* networks) (Figure Ib) [67,68].

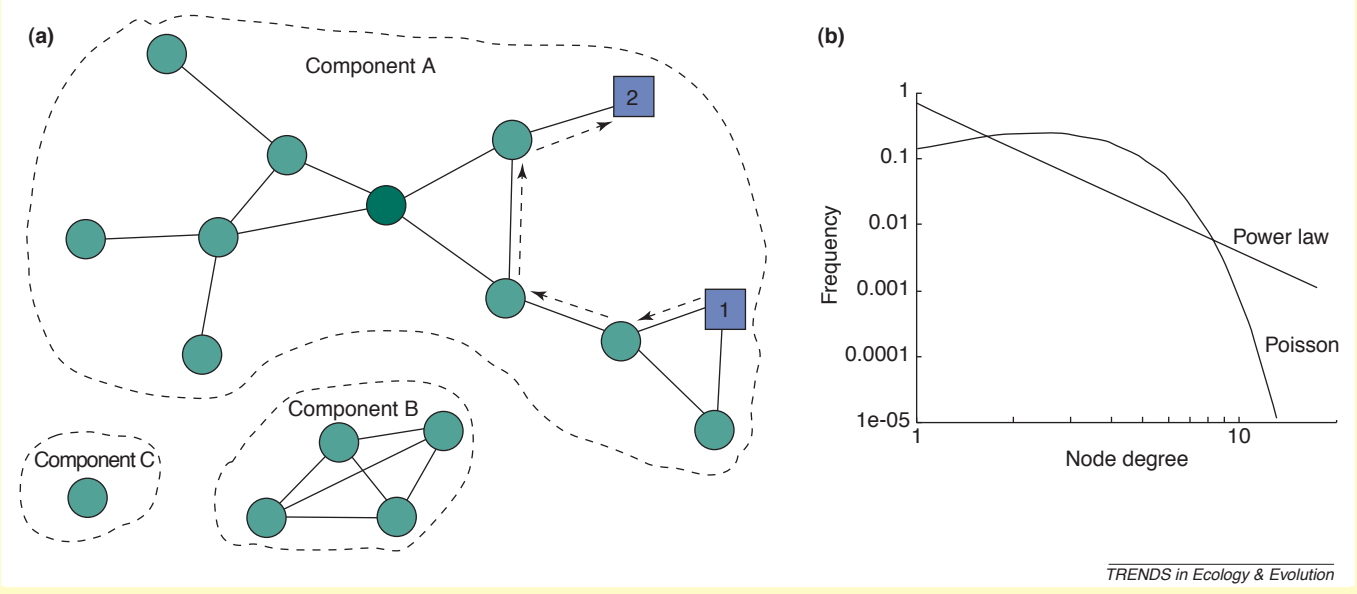


Figure I. Representation of a network. **(a)** The nodes are represented by circles and squares and vary in degree between zero and four. Edges are represented by solid black lines. The central node with degree 4 is shaded darker green. The network is divided into three components of varying size, outlined with a dashed line. The shortest path between square node 1 and square node 2, illustrated by dashed arrows, takes four steps. The central dark-green node has the highest betweenness, lying on 41 shortest paths (as an exercise, find the node on 35 shortest paths). Component B has a high clustering coefficient because all of sets of three nodes are connected within a triangle of edges. **(b)** Comparison of the degree distributions produced by models of network formation, each with mean connectivity of 3, shown on a log–log plot. Poisson random networks have degree distributions characterized by a modal degree, in this case set to 3. The preferential attachment model produces a power-law degree distribution that appears linear.

* The term 'scale-free' is something of a misnomer as it has nothing to do with any spatial scale or distance scale within the network. It refers only to the feature of the degree distribution where the probability of a node having k^*c edges is a fraction (dependant on only c) of the probability of a node having k edges.

spread [12,13]. These applications of the network approach show that taking note of higher order descriptions of network structure can yield unique insights into disease transmission and biological networks in general.

Here, we review and illustrate some of the basic concepts underlying network analysis, using case studies drawn from ecology and evolutionary biology (see [14] for an in-depth mathematical review and [15] for a review that includes dynamical systems). Networks exist at all scales of biological organization, from genes that interact through mutual regulation to interacting species in communities [16–19] (Figure 1). In spite of some recent major advances in network studies, we are still in the 'natural history' phase of this field, discovering the basic structure of biological networks at a variety of scales.

In its most useful form, the network approach focuses on components of network structure that cannot be recreated from observing individual nodes alone. It is in this area that the network approach has the most power to change the way that we understand biological processes, but it is also here that theoreticians face some of the most exciting and difficult challenges. Network studies in biology have now reached a turning point, where empirical studies must provide the motivating details for novel theory, and theoretical studies must provide a rigorous predictive framework in which to test hypotheses about network formation and network function.

A renewed focus on the nature of interactions within biological networks is being motivated by developments along two axes. First, in the past few years, ecological observatories, GIS systems, microarray studies and genome projects have provided an avalanche of data, highlighting the complex nature of biological systems. The second axis has developed as researchers have attempted to understand the nature and consequences of this complexity using theories developed to analyze network

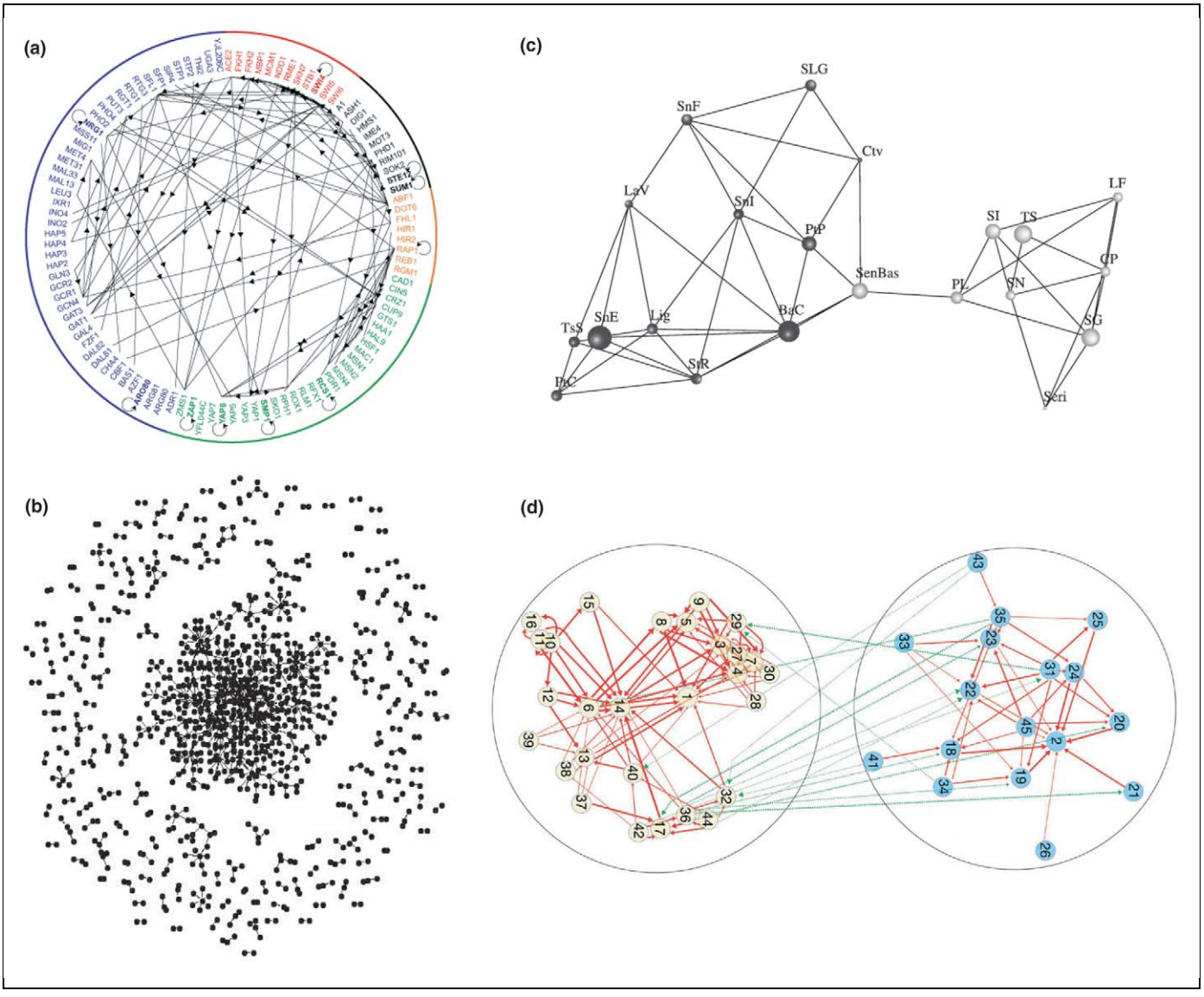


Figure 1. The use of network concepts to explore the structure and function of a variety of biological systems from genes **(a)** and proteins **(b)** to individuals within a population **(c)** and species within an ecosystem **(d)**. (a) The network of regulatory interactions in the yeast *Saccharomyces cerevisiae*, where genes encoding transcription factors interact by binding the regulatory regions of other regulatory genes [16]. (b) The protein interaction network in which proteins that physically interact are connected by edges [17]. (c) The genetic relationship of populations of the cactus *Lophocereus schottii* [18]. In this graph, edge length represents the fraction of the total genetic variation explained by the connected populations. (d) Predator–prey interactions in the Chesapeake Bay food web [19]. Reproduced, with permission from [16] (a), [17] (b), [18] (c) and [19] (d).

structure in other contexts (e.g. in physics, computer science and social science). Graph theoretic methods (Box 1) have received much attention, probably because the level of analysis is based on the presence or absence of interactions and ignores variation in the strength of interactions. This methodology fits well with our current understanding of large networks, but new theoretical techniques must be developed in parallel with advances in our ability to resolve biological networks at finer and more quantitative scales (Box 2).

Gene and protein networks

The first molecular networks were characterized over 50 years ago by Donald Nicholson [20]. With the advent of modern molecular tools, researchers can now characterize protein–protein interaction networks [21] and gene regulatory networks [16] with ever-increasing accuracy. These molecular networks have structures that are similar to one another and even bear similarity to non-biological networks. For example, metabolic, gene regulatory and protein–protein interaction networks all have an approximately power-law degree distribution (Box 1) [22–24], where most nodes in the network interact with just one other node, although a few can interact with tens or hundreds of others.

Protein–protein interaction networks

This similarity in network structure across different types of network is mirrored in the conservation of network features over millions of years. In a comparison of protein–protein interaction networks across yeast, worms and flies, a recent study found that many of the local structures within molecular networks have been conserved over deep evolutionary time [25]. Over the past few years, a series of studies has shown that the structure of nodes within a network has several evolutionarily important features. Jeong *et al.* [26] showed that the most highly connected proteins in the protein interaction network were three times more likely to be essential for survival than were weakly connected proteins. More highly connected proteins are also more pleiotropic [27], evolve more slowly [28], and are less likely to be lost over evolutionary time [29].

Evolutionary biologists now face two complementary challenges: understanding how evolutionary forces shape network structure and how network structure, in turn, determines organismal function. To this end, biologists have begun to expand their study of networks beyond connectivity to include other network attributes. Some of these descriptors are relatively local. For example, the connectivity of nodes that are connected to one another tends to be negatively correlated. Highly connected nodes are less likely to be connected to other highly connected nodes than one would expect by chance [30]. Maslov and Sneppen [30] suggest that this negative correlation prevents unwanted 'cross talk' between different functional modules within a network. Similarly, many networks appear to contain large numbers of small, local motifs. For example, in the yeast gene regulatory network, feed-forward loops are about seven times more frequent than expected [31].

Other attributes, such as 'betweenness' and modularity (Box 1), depend on broader measures of network structure. The essentiality of proteins is even more highly correlated with betweenness than it is with connectivity [32].

Metabolic pathways

Network analyses of metabolism have been used to understand the selective forces that act on pathway function. In metabolic or biosynthetic pathways, we might expect that upstream genes, which have the potential to influence a relatively large number of downstream processes, should be under stronger selection than are genes further downstream. Consistent with this expectation, in the anthocyanin pathway in *Ipomea*, genes that are upstream in the pathway tend to evolve relatively slowly in a variety of taxa [33]. Here again, the position of a gene within the larger network structure is the important unit of analysis, whereas local measures (e.g. connectivity) are insufficient. This pattern is, however, far from universal. Some enzymes at the top of the glycolytic pathway of *Drosophila* appear to be under weak or diversifying selection [34,35].

Robustness

One of the more interesting system-wide properties of networks is robustness [36,37]. Robustness refers to the relative insensitivity of a particular function or structure to change in the face of environmental or genetic perturbations. The *Escherichia coli* metabolic network, for example, is highly robust to damage. For most enzymes in the network, a change in concentration or complete loss has little effect on overall network function [38,39], both because of redundancy at the gene level and because the network itself encodes distributed robustness [40]. However, organisms are not adapted to constant laboratory conditions and genetic networks face strong selection to perform well under a variety of environmental conditions. Many yeast enzymes that appear to be redundant under typical laboratory conditions are essential under other growth conditions [41].

Gene-regulatory networks

Some of the most exciting results in the study of molecular networks have come from the analysis of gene interaction networks, from large-scale surveys of epistatic interactions between pairs of single-gene mutants [42] to analyses of how gene regulatory networks shape patterns of development. In the developing fly larva, the network of regulatory genes and proteins that gives rise to segment polarity is robust to substantial changes in initial conditions and kinetic rate constants [43]. Similarly, the regulatory network that determines sex in *Drosophila* can maintain its function in the face of a relatively large variety of different mutations [44]. Whereas some genes can be selected to preserve function under a range of conditions, others might be selected to vary in response to environmental changes. A recent study of phenotypic plasticity in yeast genes found that genes regulated by a relatively large number of upstream genes tend to show greater variation in expression levels across a range of environments [45].

Box 2. Alternative representations of networks

Real biological systems are unique and complex, and network representations will necessarily involve condensing some of that information. For example, the enzymes in a metabolic network can be thought of as nodes that are connected by the substrates that they metabolize. However, the activity of each enzyme will depend in different ways on many factors, such as the concentration of substrate, temperature and pH. By representing this network as enzyme nodes connected by reaction edges, we lose information about the rates of reaction, but gain the ability to calculate network statistics and to compare different networks. We can think of the amount of information that is retained as the level of resolution of a network (Figure I). This resolution will depend on how much functional information is encoded by the nodes and edges (i.e. how much of the variation in node and edge behavior is retained in the network representation).

In undirected graphs (Figure Ia), only the topology of the network is retained, and all further analysis will focus exclusively on that feature. For example, the protein–protein interaction network represents the presence or absence of physical interactions between proteins. All proteins are treated as identical in terms of their function in the network, and all interactions are represented by a single edge type. More information is retained in directed graphs (Figure Ib), where the direction of interactions is indicated by an arrow. In food-web networks, for example, one species is linked to another by predation, so edges must contain this directional information. Even more information will be preserved if information about the magnitude of the interaction effects is included (often as the width of the arrow).

In the preceding networks, all nodes are treated as identical, but, in some cases, it might be crucial to keep track of different types of node using a bipartite graph (Figure Ic). For example, sexual contact networks that involve males and females will have two distinct types of node. The distinct characteristics of these nodes might affect various processes, such as the way in which a sexually transmitted disease spreads through the network.

Finally, networks can contain functional information, where each node and the interactions between the nodes are unique (Figure Id). In this kind of network, as much information as possible is preserved while describing the network (e.g. through differential equations or logical operators). For example, a representation of a developmental network might include transcription rates and information about the functional integration of signals from other genes. Likewise, representations of food webs can include competition coefficients and functional responses. Although this kind of network is essential for studying how dynamic processes on networks act, there are currently few general mathematical results that can be broadly applied. The level of network resolution necessary to address specific problems in biological networks remains an open question.

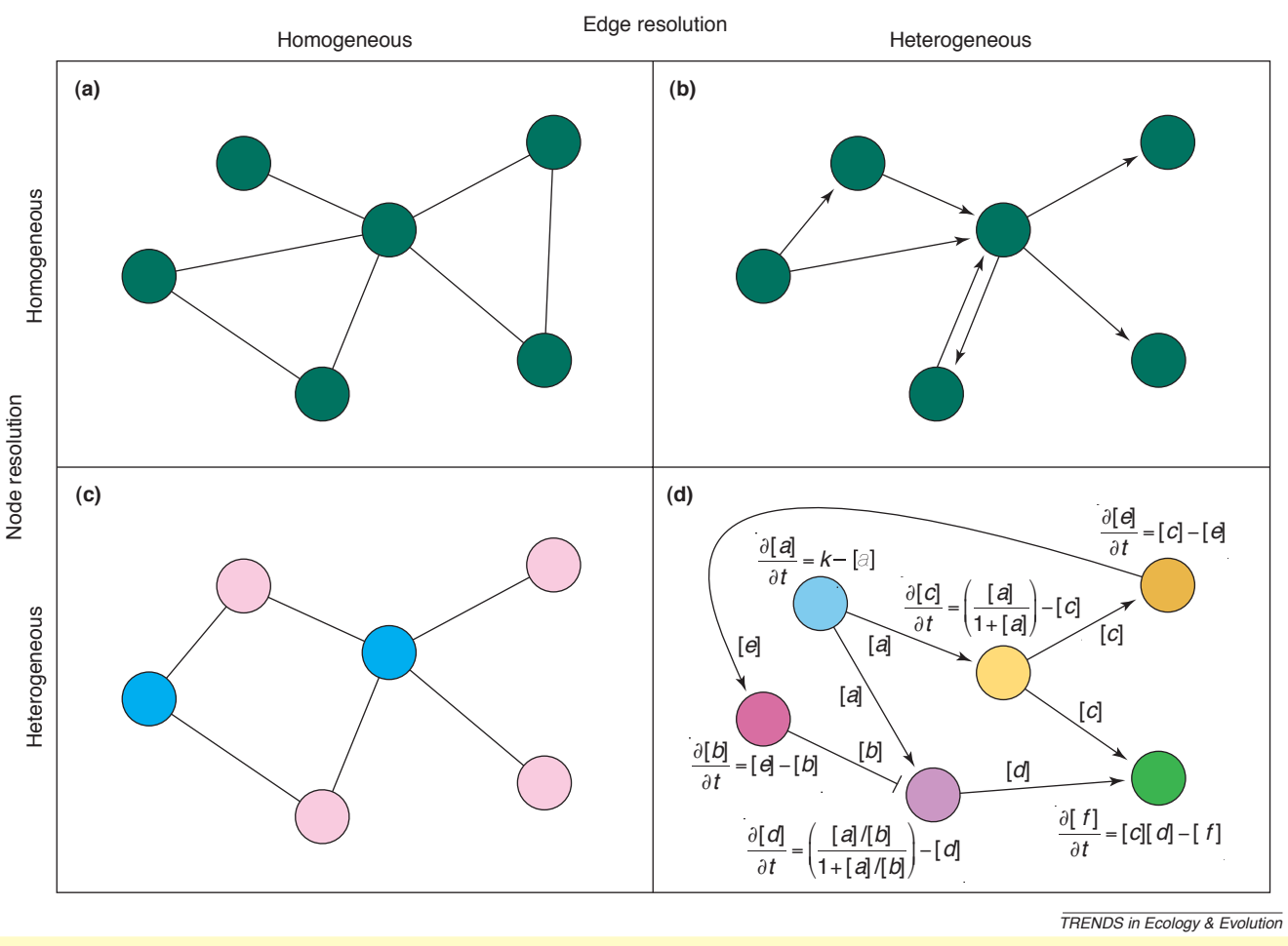


Figure I. Representations of biological systems at different levels of network resolution. For example, in **(a)**, the protein interaction network consists of a single type of node, protein, and a single kind of edge, physical interaction. **(b)** Representations of food webs often retain directional information showing the relationship between predators and prey. **(c)** In some networks, the state of individuals is variable. For example, in sexual contact networks, male and female individuals exhibit different behaviors. **(d)** Gene regulatory networks can include non-linear integration of transcription factors from several genes. This functional information can be retained to describe the network topologically and as a system of differential equations.

Whereas most studies of gene regulatory networks have focused on network function, one study illustrates the different ways in which networks can fail to function. In many species of ants, workers do not develop wings. In an analysis of the gene regulatory network responsible for wing development, Abouheif and Wray [46] show that, in ant castes where wings are produced, gene regulation for wing development has been conserved over several hundred million years. By contrast, among castes that do not produce wings, a different regulatory gene was disrupted in each of four species.

Food webs and community stability

A longstanding question in community ecology centers around the relationship between food-web structure and the stability of the community [47,48]. Mathematical ecologists have used dynamic models to explore how the size and connectivity of food webs determines how stable a community will be in the face of fluctuations in density [49], or invasion of new species [50], and the long-term persistence of the community under non-linear population dynamics [51]. In these studies, network parameters were assumed to be random (and uniformly distributed) whereas the network structure itself was typically assumed to be complete (all species interact with every other species). These assumptions produced networks that had interesting complex dynamics but had uniform network structure unlike real-world communities [52].

Degree distributions

Studies of network structure in food webs generally agree that mean path lengths are short and that the degree distribution is not Poisson (Box 1), rejecting the trivial hypothesis that food webs are assembled by randomly connecting an existing set of species [53–57]. These studies disagree, however, on whether the degree distributions are best fit by a power-law or by some other distribution. This disagreement stems from the lack of objective statistical tests to compare degree distributions and from the fact that food webs vary in their structure. Some questions remain about the way that data on community structure are converted to food webs, and it has been suggested that this process could bias our perception of food webs [58,59]. In particular, although many authors use trophic species (groups of species that both consume the same prey and are consumed by the same predators) to define nodes in food webs, in plant–animal interaction networks nodes are often resolved to the species level. Further study is needed to determine whether the discrepancies in observed degree distributions are best explained by the differences in the coding of network structure or by biological differences between food webs and plant–animal interaction networks.

Community stability and network robustness

Just as some of the most promising work on molecular networks has been on the problem of robustness (exploring what happens when we remove or alter nodes in a network) there is a related (and much deeper) vein running through the study of ecological networks. In particular, ecologists have set out to assess how removing or replacing native species with exotic invaders can alter food-web structure [19,60,61]. The effects of these perturbations throughout the rest of the network can be measured by the number of secondary extinctions and by the break-up of the network into smaller components. These studies show that, on average, removal of the most connected species causes more secondary extinctions than does the random removal of species [60,61]. However, in at least one study, removal of species with low connectivity sometimes had a large effect on the community [61], reinforcing the notion that keystone species are not strictly defined by high connectivity [52].

In particular, Allesina and Bodini [62] demonstrate that conceptualizing food webs as energy-flow networks can produce more striking results. They describe food webs based not on direct interaction, but on the notion of 'dominating' nodes that act as energy bottlenecks for resources flowing to other members of the food web. For example, the only herbivore in a tri-trophic community of plants, herbivores and predators would represent an energy bottleneck: all energy that reached the predator class must pass through the single herbivore. When the network is reformulated in this way, the species with large numbers of connections are species that act as energy bottlenecks and, thus, cause large numbers of secondary extinctions when removed. Therefore, highly connected species in the food web are expected to cause more secondary extinctions to the degree that connectivity is correlated with energy flow.

Beyond network topology

The network approach to food webs has primarily achieved general results by ignoring dynamical processes operating on such networks. However, the removal of a species can have subtle effects that extend beyond those caused by the direct effect it has on the species with which it directly interacts. Solé and Montoya [60] point out that fragmentation of a food web might destabilize the community because of dynamic effects within isolated compartments of the network (Box 1), in a similar way to the removal of a keystone species. Ongoing work aims at combining known topological features of food webs with quantitative variation in species interactions to predict community stability [63,64]. The challenge for future studies is to incorporate these quantitative features of networks without losing the simplifying power of analyses based on network topology alone.

Toward predictive theories of biological networks

The case studies that we have discussed so far demonstrate that the application of existing network approaches to biological systems can yield valuable insights. Future progress, however, will depend on a synergistic interplay between network thinking and biological processes.

The dynamics of biological network formation

The mathematical treatment of network formation has generally assumed that networks are formed by random processes involving homogeneous sets of nodes (Box 1) [65–68], but life is characterized by variation and change. To develop better models of biological networks, we must

account for three key features of biological networks: (i) networks can gain or lose nodes; (ii) the properties of nodes can change; and (iii) external forces can act on network structure. In particular, we need to incorporate the effect that natural selection might have not only on individual nodes (e.g. genes or species), but also on the whole network structure. The specific properties of the biological entities that comprise these networks will also influence the dynamics of network formation.

A realistic model of genetic network formation will need to include patterns of gene duplication and loss, the way that evolution of genes changes existing interactions, and the effect of natural selection on network structure. Toward this goal, the preferential attachment model (Box 1) [67,68] has been extended to include gene duplication and loss [69–71]. In particular, Wagner [70] has shown that a roughly power-law distribution can emerge simply through a neutral model of the repeated duplication and loss of genes, which creates a kind of preferential attachment. This does not, however, demonstrate that the preferential attachment model is the hypothesis with the best fit; other models have produced similar degree distributions. For example, Kunin *et al.* [72] create a network with a power-law distribution by assuming that proteins are most likely to connect with other proteins in the same functional class. In addition, as researchers look more closely at network structure, higher order network features emerge. Networks with power-law degree distributions can still differ in terms of clustering [13], motif frequency [31,73], nestedness [74] and fractal structure [75]. Finally, the role that natural selection has in the evolution of network structure remains unknown. Even if neutral models are sufficient to explain the patterns in networks at their current state of resolution, alternative hypotheses are needed to direct future empirical work.

In contrast to gene networks, we need to incorporate a different array of biological phenomena, including local extinction, immigration, behavioral changes in foraging, and evolution of species interactions, to understand the formation of food webs [54,76–79]. Several models of community formation are based on niche subdivision, where species are arrayed on a linear niche axis, and can produce many of the features of observed food webs [54,76]. The addition of spatial and phylogenetic effects can alter the predicted food-web topology and sometimes produce a better fit with observed communities [78,79]. A major challenge is to include evolutionary and spatial dynamics in a single framework to predict not only the structure, but also the functional dynamics of food webs.

The emergence of network robustness

In network studies, we often equate robustness with a structural phenomenon: the preservation of a single giant component in the network. But robustness is also related to important functional concepts in biology, including canalization in evolutionary theory [80] and community stability in ecology [48]. Networks that have significant variance in node connectivity, regardless of whether they are best described by power-law degree distributions, are robust to random removal of nodes [5,60,61,81]. As a result of this finding, several authors have suggested that high variance in connectivity is an evolved feature of network topology [23,35]. However, other more local network features, such as feedback regulation, might also be responsible for robustness, as has been suggested for developmental networks [43,82,83] and has been reported to evolve in computer simulations [84].

The relationship between network form and function

There are only a few examples where we have a good understanding of how network topology determines the way in which networks function and respond to change [43,49,82,83,85–88]. For example, working from the set of logical functions in the segment polarity network in *Drosophila*, developmental patterns can be generated in a robust way [82,83]. Likewise, the predator–prey relationships in food webs are sometimes sufficient to predict food-web stability [49,87]. Unfortunately, there is no single function that networks perform in all biological situations. This heterogeneity in network function creates a major stumbling block for our ability to predict function from structure. For instance, in studies of protein–protein interaction networks, we can relate a host of biologically interesting phenomena to network structure. However, we have yet to explain adequately the relationship between these network measures, such as connectivity or betweenness, and biologically relevant measures, such as fitness. The current explanation for these relationships is that rapid information transfer within genetic networks is adaptive. However, genetic networks do not function in the same way as information exchanging nodes on the internet; selection acts on the dynamic output of the network, which is unlikely to be optimized solely by maintaining the flow of information between genetic states. Whereas rapid progress has been made in identifying the existence of connections between genes, proteins and species, it might be some time before large-scale techniques are developed to map the quantitative and functional relationships that define biological networks (Box 2). To this end, a qualitative theory of quantitative networks must be developed to determine the connections between the structure of networks and their adaptive value.

Conclusion

Many features and processes of ecological and evolutionary systems can be well represented by networks of interacting elements. The real usefulness of network thinking becomes most apparent in cases where the novel insights that we gain when we observe the entire network far outstrip what we learn from an analysis of its parts. As biologists, we still face several impediments as we try to create a science of networks that begins with the necessary formalisms of mathematical network theory, but which incorporates the complex realities of evolutionary biology and ecology. Currently, most measures of network structure strongly depend on how the network is defined, especially in terms of relationships between network parameters and network function. Furthermore, to test adequately hypotheses about network structure, we

need to make sure that we are using null models that reflect biological reality.

As ecologists and evolutionary biologists develop new approaches to these problems, they will no doubt benefit from increased crosstalk, as their problems are likely to have more common features than either have with classic mathematical network models. The hope is that network approaches will free us from the Gordian knot of accumulated data to reveal the global patterns behind large-scale ecological and evolutionary processes. The fear is that all of the fine structure will still matter in the end, leaving us tangled in detail. The next decade should show us how the knot is undone.

Acknowledgements

We thank Sally Otto for numerous comments and discussion, several reviewers for their valuable comments, and the authors who kindly made figures available. Funding support was provided by an NIH fellowship to S.R.P., an Ellison Medical Foundation award to D.P., and a NIH and NSF grants to P.C.P.

References

1 Auerbach, D.M. *et al*. (1984) Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact. *Am. J. Med.* 76, 487–492
2 Grmek, M.D. (1990) *History of AIDS: Emergence and Origin of a Modern Pandemic*, Princeton University Press
3 Klovdahl, A.S. (1985) Social networks and the spread of infectious diseases: the AIDS example. *Soc. Sci. Med.* 21, 1203–1216
4 CDC. (1982) A cluster of Kaposi's sarcoma and *Pneumocystis carinii* pneumonia among homosexual male residents of Los Angeles and range counties, California. *Morb. Mortal. Weekly* 31, 305–307
5 Albert, R. *et al*. (2000) Error and attack tolerance of complex networks. *Nature* 406, 378–382
6 Ancel Meyers, L. *et al*. (2003) Applying network theory to epidemics: control measures for outbreaks of *Mycoplasma pneumoniae*. *Emerg. Infect. Dis.* 9, 204–210
7 Anderson, R.M. and May, R.M. (1988) Epidemiological parameters of HIV transmission. *Nature* 333, 514–519
8 May, R.M. and Lloyd, A.L. (2001) Infection dynamics on scale-free networks. *Phys. Rev. E* 65, 066112
9 Lloyd, A.L. and May, R.M. (2001) Epidemiology – how viruses spread among computers and people. *Science* 292, 1316–1317
10 Liljeros, F. *et al*. (2001) The web of human sexual contacts. *Nature* 411, 907–908
11 Pastor-Satorras, R. and Vespignani, A. (2001) Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200–3203
12 Szendroi, B. and Csanyi, G. (2004) Polynomial epidemics and clustering in contact networks. *Proc. R. Soc. Lond. B. Biol. Sci.* 271, S364–S366
13 Jones, J.H. and Handcock, M.S. (2003) An assessment of preferential attachment as a mechanism for human sexual network formation. *Proc. R. Soc. Lond. B. Biol. Sci.* 270, 1123–1128
14 Newman, M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.* 45, 167–256
15 Strogatz, S.H. (2001) Exploring complex networks. *Nature* 410, 276
16 Lee, T.I. *et al*. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804
17 Wagner, A. (2003) Does selection mold molecular networks? *Sci. STKE* 2003, PE41
18 Dyer, R.J. and Nason, J.D. (2004) Population graphs: the graph theoretic shape of genetic structure. *Mol. Ecol.* 13, 1713–1727
19 Krause, A.E. *et al*. (2003) Compartments revealed in food-web structure. *Nature* 426, 282–285
20 Dagley, S. and Nicholson, D.E. (1970) *An Introduction to Metabolic Pathways*, John Wiley & Sons
21 Uetz, P. *et al*. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627
22 Guelzim, N. *et al*. (2002) Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* 31, 60–63
23 Wagner, A. and Fell, D.A. (2001) The small world inside large metabolic networks. *Proc. R. Soc. Lond. B. Biol. Sci.* 268, 1803–1810
24 Wagner, A. (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* 18, 1283–1292
25 Sharan, R. *et al*. (2005) Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1974–1979
26 Jeong, H. *et al*. (2001) Lethality and centrality in protein networks. *Nature* 411, 41–42
27 Promislow, D.E.L. (2004) Protein networks, pleiotropy and the evolution of senescence. *Proc. R. Soc. Lond. B. Biol. Sci.* 271, 1225–1234
28 Fraser, H.B. *et al*. (2002) Evolutionary rate in the protein interaction network. *Science* 296, 750–752
29 Krylov, D.M. *et al*. (2003) Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13, 2229–2235
30 Maslov, S. and Sneppen, K. (2002) Specificity and stability in topology of protein networks. *Science* 296, 910–913
31 Milo, R. *et al*. (2002) Network motifs: simple building blocks of complex networks. *Science* 298, 824–827
32 Hahn, M.W. and Kern, A.D. (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22, 803–806
33 Rausher, M.D. *et al*. (1999) Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.* 16, 266–274
34 Eanes, W.F. (1999) Analysis of selection on enzyme polymorphisms. *Annu. Rev. Ecol. Syst.* 30, 301–326
35 Cork, J.M. and Purugganan, M.D. (2004) The evolution of molecular genetic pathways and networks. *Bioessays* 26, 479–484
36 de Visser, J.A.G.M. *et al*. (2003) Perspective: evolution and detection of genetic robustness. *Evolution* 57, 1959–1972
37 Proulx, S.R. and Phillips, P.C. (2005) The opportunity for canalization and the evolution of genetic networks. *Am. Nat.* 165, 147–162
38 Lemke, N. *et al*. (2004) Essentiality and damage in metabolic networks. *Bioinformatics* 20, 115–119
39 Edwards, J.S. and Palsson, B.O. (2000) Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 16, 927–939
40 Wagner, A. (2005) Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* 27, 176–188
41 Papp, B. *et al*. (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429, 661–664
42 Tong, A.H. *et al*. (2004) Global mapping of the yeast genetic interaction network. *Science* 303, 808–813
43 von Dassow, G. *et al*. (2000) The segment polarity network is a robust developmental module. *Nature* 406, 188–192
44 MacCarthy, T. *et al*. (2003) The evolutionary potential of the *Drosophila* sex determination gene network. *J. Theor. Biol.* 225, 461–468
45 Promislow, D.E.L. A regulatory network analysis of phenotypic plasticity in yeast. *Am. Nat*. (in press)
46 Abouheif, E. and Wray, G.A. (2002) Evolution of the gene network underlying wing polyphenism in ants. *Science* 297, 249–252
47 Elton, C.S. (1958) *The Ecology of Invasions by Animals and Plants*, Methuen
48 MacArthur, R.H. (1955) Fluctuations of animal populations and a measure of community stability. *Ecology* 36, 533–536
49 May, R.M. (1973) *Stability and Complexity in Model Ecosystems*, Princeton University Press
50 Case, T.J. (1990) Invasion resistance arises in strongly interacting species-rich model competition communities. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9610–9614
51 Hastings, A. and Powell, T. (1991) Chaos in a 3-species food-chain. *Ecology* 72, 896–903
52 Berlow, E.L. *et al*. (2004) Interaction strengths in food webs: issues and opportunities. *J. Anim. Ecol.* 73, 585–598
53 Montoya, J.M. and Sole, R.V. (2002) Small world patterns in food webs. *J. Theor. Biol.* 214, 405–412
54 Camacho, J. *et al*. (2002) Robust patterns in food web structure. *Phys. Rev. Lett.* 88, 228102
55 Williams, R.J. *et al*. (2002) Two degrees of separation in complex food webs. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12913–12916

56 Dunne, J.A. *et al*. (2002) Food-web structure and network theory: the role of connectance and size. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12917–12922
57 Jordano, P. *et al*. (2003) Invariant properties in coevolutionary networks of plant-animal interactions. *Ecol. Lett.* 6, 69–81
58 Martinez, N.D. (1993) Effects of resolution on food web structure. *Oikos* 66, 403–412
59 Solow, A.R. and Beet, A.R. (1998) On lumping species in food webs. *Ecology* 79, 2013–2018
60 Sole, R.V. and Montoya, J.M. (2001) Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B. Biol. Sci.* 268, 2039–2045
61 Dunne, J.A. *et al*. (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* 5, 558–567
62 Allesina, S. and Bodini, A. (2004) Who dominates whom in the ecosystem? Energy flow and bottlenecks and cascading extinctions. *J. Theor. Biol.* 230, 351–358
63 Ives, A.R. and Cardinale, B.J. (2004) Food-web interactions govern the resistance of communities after non-random extinctions. *Nature* 429, 174–177
64 Williams, R.J. and Martinez, N.D. (2004) *Diversity, Complexity, and Persistence in Large Model Ecosystems*, SFI Working Paper (http://www.santafe.edu/research/publications/wpabstract/200407022)
65 Erdos, P. and Renyi, A. (1959) On random graphs. *Publ. Math. Debrecen* 6, 290–297
66 Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature* 393, 440–442
67 Barabasi, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509–512
68 Barabasi, A.L. *et al*. (1999) Mean-field theory for scale-free random networks. *Physica A* 272, 173–187
69 Sole, R.V. *et al*. (2002) A model of large-scale proteome evolution. *Adv. Comp. Syst.* 5, 43–54
70 Wagner, A. (2003) How the global structure of protein interaction networks evolves. *Proc. R. Soc. Lond. B. Biol. Sci.* 270, 457–466
71 Berg, J. *et al*. (2004) Structure and evolution of protein networks: a statistical model for link dynamics and gene duplications. *BMC Evol. Biol.* 4, 51
72 Kunin, V. *et al*. (2004) Functional evolution of the yeast protein interaction network. *Mol. Biol. Evol.* 21, 1171–1176
73 Wuchty, S. *et al*. (2003) Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat. Genet.* 35, 176–179
74 Bascompte, J. *et al*. (2003) The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9383–9387
75 Song, C. *et al*. (2005) Self-similarity of complex networks. *Nature* 433, 392–395
76 Williams, R.J. and Martinez, N.D. (2000) Simple rules yield complex food webs. *Nature* 404, 180–183
77 Kondoh, M. (2003) Foraging adaptation and the relationship between food-web complexity and stability. *Science* 299, 1388–1391
78 Cattin, M.F. *et al*. (2004) Phylogenetic constraints and adaptation explain food-web structure. *Nature* 427, 835–839
79 Arii, K. and Parrott, L. (2004) Emergence of non-random structure in local food webs generated from randomly structured regional webs. *J. Theor. Biol.* 227, 327–333
80 Waddington, C.H. (1942) The canalization of development and genetic assimilation of acquired characters. *Nature* 150, 1008–1023
81 Albert, R. *et al*. (2001) correction: error and attack tolerance of complex networks. *Nature* 409, 542
82 Albert, R. and Othmer, H.G. (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223, 1–18
83 Ingolia, N.T. (2004) Topology and robustness in the *Drosophila* segment polarity network. *PLoS Biol.* 2, e123
84 Siegal, M.L. and Bergman, A. (2002) Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10528–10532
85 Barkai, N. and Leibler, S. (1997) Robustness in simple biochemical networks. *Nature* 387, 913–917
86 Ideker, T. *et al*. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
87 Dambacher, J.M. *et al*. (2003) Qualitative stability and ambiguity in model ecosystems. *Am. Nat.* 161, 876–888
88 Price, N.D. *et al*. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–897

Free journals for developing countries

The WHO and six medical journal publishers have launched the Access to Research Initiative, which enables nearly 70 of the world's poorest countries to gain free access to biomedical literature through the Internet.

The science publishers, Blackwell, Elsevier, the Harcourt Worldwide STM group, Wolters Kluwer International Health and Science, Springer-Verlag and John Wiley, were approached by the WHO and the *British Medical Journal* in 2001. Initially, more than 1000 journals will be available for free or at significantly reduced prices to universities, medical schools, research and public institutions in developing countries. The second stage involves extending this initiative to institutions in other countries.

Gro Harlem Brundtland, director-general for the WHO, said that this initiative was 'perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries'.

See http://www.healthinternetwork.net for more information.