

Functional integration of ecological networks through pathway proliferation

Stuart R. Borrett^{a,b,*}, Brian D. Fath^c, Bernard C. Patten^a

^a*Institute of Ecology, University of Georgia, Athens, GA 30606, USA*

^b*Skidaway Institute of Oceanography, Savannah, GA 31411, USA*

^c*Biology Department, Towson University, Towson, MD 21252, USA*

Received 11 April 2006; received in revised form 24 August 2006; accepted 22 September 2006

Available online 24 September 2006

Abstract

Large-scale structural patterns commonly occur in network models of complex systems including a skewed node degree distribution and small-world topology. These patterns suggest common organizational constraints and similar functional consequences. Here, we investigate a structural pattern termed pathway proliferation. Previous research enumerating pathways that link species determined that as pathway length increases, the number of pathways tends to increase without bound. We hypothesize that this pathway proliferation influences the flow of energy, matter, and information in ecosystems. In this paper, we clarify the pathway proliferation concept, introduce a measure of the node–node proliferation rate, describe factors influencing the rate, and characterize it in 17 large empirical food-webs. During this investigation, we uncovered a modular organization within these systems. Over half of the food-webs were composed of one or more subgroups that were strongly connected internally, but weakly connected to the rest of the system. Further, these modules had distinct proliferation rates. We conclude that pathway proliferation in ecological networks reveals subgroups of species that will be functionally integrated through cyclic indirect effects.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Cycles; Ecological network analysis; Eigenvalue; Food-web; Hierarchy; Indirect effects; Modularity; Network analysis; Network environment analysis; Spectral radius

1. Introduction and motivation

Large-scale structural patterns have been uncovered in network models of complex systems, suggesting the possibility of common organizational constraints and similar functional consequences. Network models are mathematical graphs composed of nodes and undirected edges or directed arcs that connect the nodes. For example, in a social network nodes might represent individuals in a community and the edges or links could represent a social relationship between the individuals such as collaboration (Newman, 2001a, b). In a model of the World Wide Web, web pages are nodes connected by hyperlinks (Albert et al.,

1999; Barabási and Albert, 1999). Ecologists use network models in many ways, including to represent trophic relations in food-webs and more generally energy–matter flux in ecosystems (Cohen et al., 1990; Higashi and Burns, 1991; Margalef, 1963; Pimm, 1982). In these networks, species or functional groups form the node set while the presence of energy and matter transfers and transformations are represented by links.

Traditionally, complex systems have been modeled using random graphs (Erdős and Rényi, 1959, 1960; Gardner and Ashby, 1970; May, 1972). However, ecologists have demonstrated that random graphs are inadequate models of ecological systems; food-web and ecosystem models often contain structures not commonly found in random graphs (Cohen et al., 1990; DeAngelis, 1975; Lawler, 1978; Pimm, 1979, 1982; Pimm et al., 1991; Ulanowicz and Wolff, 1991). In food-webs, these hypothesized structures include short food-chain lengths (Pimm and Lawton, 1977;

*Corresponding author. Current address: Computational Learning Laboratory, Center for the Study of Language and Information, Stanford University, Stanford, CA 94305, USA. Tel.: +1 650 723 1684.

E-mail address: sborrett@stanford.edu (S.R. Borrett).

Post, 2002) and little or no cycling (Cohen et al., 1990). In addition, several forms of modularity—hierarchical compartmentalization into subsystems—have been hypothesized for food-webs and ecosystems (Allen and Starr, 1982; May, 1972; Pimm, 1979; Pimm and Lawton, 1980; Yodzis, 1982). Furthermore, Ulanowicz and Wolff (1991) demonstrated that random networks (based on the Poisson, uniform, Gaussian, negative exponential, and log-normal probability distribution functions) failed to capture the distribution of connections in real ecosystems. Ecologists hypothesize that these structural differences exist because ecological systems are shaped and constrained by thermodynamic laws and natural history (Jørgensen et al., 1992; Lawler, 1978; Müller and Leupelt, 1998; Williams and Martinez, 2000).

Likewise, investigations of other types of complex systems have identified a number of distinctive patterns common in complex systems not found in purely random graphs (Albert and Barabási, 2002; Moreno and Jennings, 1938; Newman, 2002, 2003; Price, 1965). For example, the distribution of node degree (i.e. the number of edges (links) incident to a node) is often skewed in models of complex systems, following an exponential distribution or a power-law distribution rather than the Poisson distribution of random graphs. The power-law distribution was found in the World Wide Web (Barabási and Albert, 1999), metabolic networks (Jeong et al., 2000), and some but not all food-webs (Dunne et al., 2002b; Montoya and Sole, 2002). The power-law degree distribution implies that there are a large number of nodes with very few connections, while a few nodes have a large number of connections (Barabási, 2002). This topology tends to increase network robustness to random node or edge deletion, while making it more sensitive to targeted attacks (Albert et al., 2000; Dunne et al., 2002b). The small-world pattern is another commonly found topology (Watts, 1999; Watts and Strogatz, 1998). In small-world networks, the degree of node clustering is larger and the maximum distance (where distance is the shortest path between two nodes) is lower than expected from random graphs. This arrangement tends to increase the transmission speed of diseases, energy, matter, and information through networks. The largest distance in food-web graphs tends to be small, but the degree of clustering varies (Dunne et al., 2002a).

Pathway proliferation is another large-scale topological characteristic of networks, with implications for energy, matter, and information transmission. It is the tendency for the number of pathways in a network to increase geometrically without bound as pathway length increases. Patten and colleagues (Patten, 1985a, b; Patten et al., 1982) first observed this tendency in small, well-connected ecosystem models during the early development of ecosystem network analysis. More recently, Fath (1998) and Borrett and Patten (2003) showed that the rate of pathway proliferation is variable among networks. This is significant because the pathway proliferation rate characterizes how quickly the number of indirect pathways

increases, and thus, the number of pathways available for interactions. Food-web investigations often emphasize the shortest pathways, assuming that most significant interactions occur over these routes (e.g. Caldarelli et al., 1998; Post et al., 2000). However, previous results from Network Environ Analysis, an environmental application and extension of economic Input–Output Analysis, indicate that flows over longer indirect pathways can be significant or even dominant constituents of total system throughflow (Higashi and Patten, 1986, 1989; Patten, 1983), which is a measure of whole system activity. This result has important implications for trophodynamics (Burns et al., 1991; Patten et al., 1990; Whipple, 1998) and biogeochemical cycling in ecosystems (Borrett et al., 2006; Finn, 1980; Patten et al., 1976). Given the possible significance of indirect pathways in network models of conservative transport systems like ecosystems, it is critical to understand the network characteristics influencing the pathway proliferation rate.

In this paper, we clarify the pathway proliferation concept, describe factors influencing the proliferation rate, and characterize pathway proliferation rates in 17 large empirical food-webs. In Section 2 we review relevant mathematics to build a better understanding of pathway proliferation. In Section 3 we apply this understanding to 17 food-web models drawn from the literature. This analysis reveals a type of modularity in some of the food-webs, lending support to the hypothesis that food-webs have a modular structure (Allesina et al., 2005; Krause et al., 2003; May, 1972; Pimm and Lawton, 1980; Yodzis, 1982). We conclude by summarizing our findings, discussing their relevance for ecological systems, and suggesting next steps along this research path.

2. Pathway proliferation

Although Patten and colleagues (Patten, 1985a, b; Patten et al., 1982) introduced pathway proliferation into the ecological literature over two decades ago, it is not well understood. In this section, we synthesize mathematical results from graph theory and matrix algebra to determine a method for quantifying the node–node pathway proliferation rate, and to identify the bounds and expected value of the rate. In addition, we uncover the possibility of differing rates of pathway proliferation for modules within a network.

Network models of complex systems are mathematically graphs which can be directed or weighted (Bang-Jensen and Gutin, 2001; Ponstein, 1966). A graph G is specified by a set of n nodes and e unoriented edges ($0 \leq e \leq n(n-1)/2 + n$), where edges indicate an undirected relationship between two nodes. A directed graph (digraph) D is also specified by a set of n nodes, but instead of edges it has L oriented arcs or links ($0 \leq L \leq n^2$). Digraph structure is partially characterized by two connectivity measures, connectance $C = L/n^2$ and link density L/n , which are common metrics in the food-web literature (Cohen et al.,

1990; Martinez, 1991). Edges and links can be assigned weights to represent the relationship strength.

In this paper, we focus on simple unweighted digraphs, where simple implies no more than one link from one node to any other. We do this for two reasons. First, directed graphs are often appropriate for ecological applications as many ecological processes are oriented (e.g. predation and excretion generate energy and matter flows from one ecosystem element to another). Second, while the mathematics described in this paper may apply to non-simple and weighted graphs, our interest here is primarily network structure as it is a necessary element of understanding ecosystem organization.

2.1. Quantifying pathway proliferation

Here, we review relevant definitions and results from graph theory and linear algebra. We first show why the dominant eigenvalue of a strongly connected digraph is a good measure of the node–node proliferation rate. We conclude by describing how the dominant eigenvalue can distinguish between fundamentally different ecological networks.

In a directed graph, a pathway is an alternating sequence of nodes and links connecting a starting and a terminal node. Pathway length m is the number of links in the pathway. For example, in the directed graph D in Fig. 1a there is a pathway of length 2 from nodes 1 to 3 (e.g. $1 \rightarrow 2 \rightarrow 3$). Cycles are pathways with the same starting and terminal nodes, and a cycle of length 1 is a self-loop. In our example network, $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ is cycle of length 3, and $4 \rightarrow 4$ is a self-loop. Pathways with self-loops are termed walks, those without are paths (Patten, 1985a). Alternatively, we can represent D with its associated and isomorphic adjacency matrix $\mathbf{A}_{n \times n} = (a_{ij})$, where $a_{ij} = 1$ if and only if there is a link from j to i (note column to row orientation), otherwise $a_{ij} = 0$ (Fig. 1b). The number of direct links terminating or starting at a node is termed the in- and out-degree, respectively. These are calculated as $\mathbf{k}^{in} = (k_i^{in}) = \sum_{j=1}^n a_{ij}$ and $\mathbf{k}^{out} = (k_j^{out}) = \sum_{i=1}^n a_{ij}$, where \mathbf{k}^{in} and \mathbf{k}^{out} are $1 \times n$ and $n \times 1$ vectors, respectively. Average in- and out-degrees ($\langle k^{in} \rangle$ or $\langle k^{out} \rangle$), and degree distributions $P(k)$ are ways of characterizing network structure (Albert and Barabási, 2002; Newman, 2003; Ulanowicz and Wolff, 1991).

Indirect pathways ($m > 1$) are enumerated by raising the adjacency matrix to the m th power $\mathbf{A}^m = (a_{ij}^m)$ (Ponstein, 1966). Again, pathway proliferation is the tendency for the number of pathways in a network to increase without bound as a function of increasing pathway length (Fig. 1c). Borrett and Patten (2003) approximated whole-system pathway proliferation rate as

$$\frac{\log(\sum \sum a_{ij}^{(m)})}{\log(\sum \sum a_{ij}^{(m-1)})},$$

where $a_{ij}^{(m)}$ signifies the a_{ij} element of \mathbf{A}^m , but this aggregate estimate only holds for sufficiently large m . Therefore,

following Fath (1998), here we will define pathway proliferation in terms of the rate at which $a_{ij}^{(m)}$ changes as $m \rightarrow \infty$. This makes pathway proliferation a combinatorial property of \mathbf{A} (Seneta, 1973).

As m increases, the development of \mathbf{A}^m is determined by its characteristic polynomial, $\pi(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A})$, where λ is the variable in the polynomial and $\mathbf{I}_{n \times n}$ is the identity matrix (Godsil, 1993; Seneta, 1973). The eigenvalues (λ_i , $i = 1, \dots, n$) of \mathbf{A} are roots of the characteristic polynomial, determined as solutions to $\pi(\lambda) = 0$. The set of eigenvalues $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_i \geq \dots \geq \lambda_n\}$ is the spectrum of \mathbf{A} , which Cvetkoviâc et al. (1980) study to determine structural properties of graphs. In undirected graphs, \mathbf{A} is symmetric and all λ_i are real; in directed graphs λ_i may be a complex number.

In addition to being the roots of the characteristic polynomial, the eigenvalues must satisfy $\mathbf{A}\mathbf{W} = \mathbf{\Lambda}\mathbf{W}$, where $\mathbf{W}_{n \times n} = [\mathbf{w}_1 | \mathbf{w}_2 | \dots | \mathbf{w}_i | \dots | \mathbf{w}_n]$ is a composite matrix of the right eigenvectors \mathbf{w}_i , and $\mathbf{\Lambda}_{n \times n} = \text{diag}(\lambda_i)$ is a matrix with the eigenvalues of \mathbf{A} on the principle diagonal and zeros in all other positions. Further, the eigenvalues must satisfy $\mathbf{V}\mathbf{A} = \mathbf{V}\mathbf{\Lambda}$, where $\mathbf{V}_{n \times n} = [\mathbf{v}_1 | \mathbf{v}_2 | \dots | \mathbf{v}_i | \dots | \mathbf{v}_n]$ is a composite matrix of the left eigenvectors \mathbf{v}_i . \mathbf{A}^m is then determined as

$$\mathbf{A}^m = \mathbf{W} \times \mathbf{\Lambda}^m \times \mathbf{W}^{-1},$$

$$\mathbf{A}^m = \mathbf{W} \times \begin{bmatrix} \lambda_1^m & 0 & \dots & 0 \\ 0 & \lambda_2^m & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^m \end{bmatrix} \times \mathbf{W}^{-1}, \quad (1)$$

where \mathbf{W} is as before and \mathbf{W}^{-1} is the matrix inverse of \mathbf{W} , which are the transposed complex conjugates of the left eigenvectors \mathbf{V} (Caswell, 2001). If we let \mathbf{v}_i^* be the i th row of \mathbf{W}^{-1} , then we can rewrite Eq. (1) as

$$\mathbf{A}^m = \lambda_1^m \mathbf{w}_1 \mathbf{v}_1^* + \lambda_2^m \mathbf{w}_2 \mathbf{v}_2^* + \dots + \lambda_i^m \mathbf{w}_i \mathbf{v}_i^* + \lambda_n^m \mathbf{w}_n \mathbf{v}_n^*. \quad (2)$$

Eq. (2) is the pathway generating function of \mathbf{A} (Godsil, 1993), and illustrates how the development of \mathbf{A}^m depends on the spectrum of \mathbf{A} .

Next, we introduce two classification schemes, one for digraphs and one for matrices, because they are required to apply a theorem that will let us develop a succinct estimate of the pathway proliferation rate. In the first scheme, digraphs are classified as one of three types: strongly connected, weakly connected, and disconnected. A digraph is strongly connected (strong) if it is possible to reach every node from every other over a pathway of unspecified length. Bang-Jensen and Gutin (2001) define a (sub)digraph with only one node as strong, though this is trivial for our purposes. A weakly connected (weak) digraph is one in which it is possible to reach any node from any other node if we ignore link orientation, but it is impossible when following link orientation. Nodes of a weak digraph must have an in-degree or out-degree of at least 1.

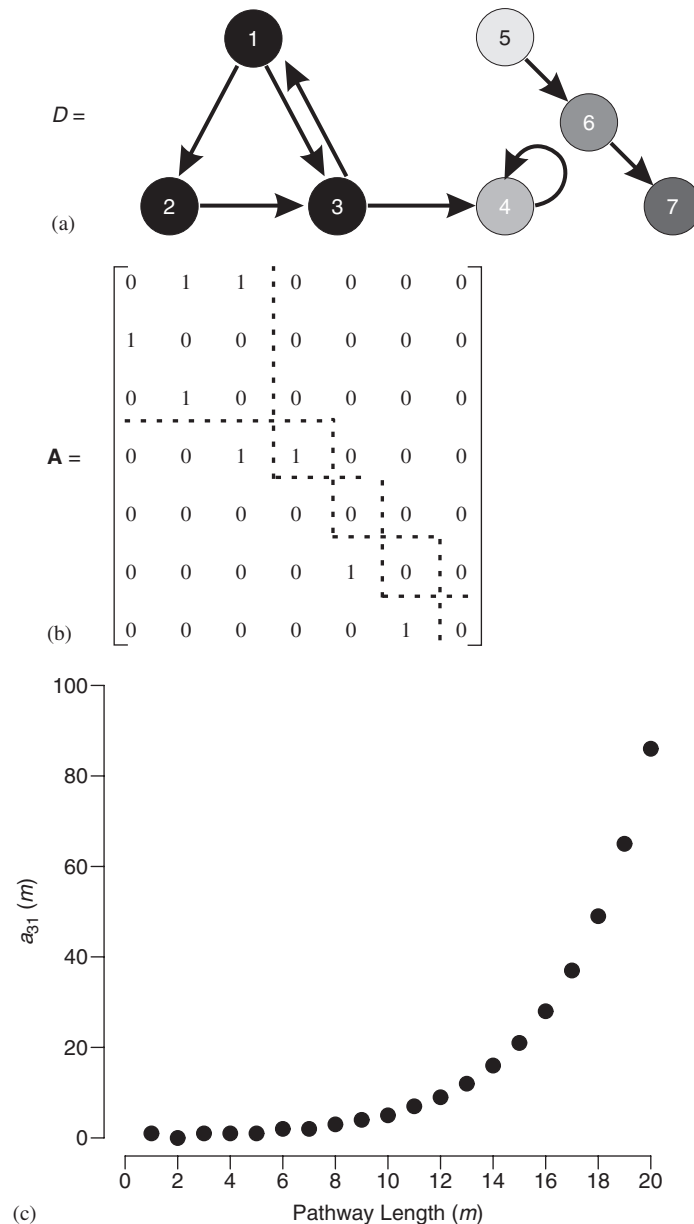


Fig. 1. Example digraph, its associated adjacency matrix, and pathway proliferation. (a) Digraph D composed of strongly connected components $\mathbf{K}_1 = \{1, 2, 3\}$, $\mathbf{K}_2 = \{4\}$, $\mathbf{K}_3 = \{5\}$, $\mathbf{K}_4 = \{6\}$, $\mathbf{K}_5 = \{7\}$, (b) the adjacency matrix \mathbf{A} associated with the digraph D (dotted lines denote strongly connected components), and (c) a plot demonstrating pathway proliferation from nodes 3 to 1: the number of pathways $a_{31}^{(m)}$ increases as pathway length m increases. Transient effects created by smaller eigenvalues are visible in the first 10 pathway lengths, but the rate of pathway proliferation has nearly converged to $\lambda_1(\mathbf{K}_1) = 1.32$ by a pathway length of 20.

A disconnected graph is one that contains one or more non-adjacent strong or weak components. The digraph in Fig. 1a is disconnected.

The second classification scheme distinguishes three types of adjacency matrices. A matrix is irreducible if and only if it is associated with a strong digraph, while one associated with a weakly connected or disconnected digraph is reducible (Berman and Plemmons, 1979). Irreducible matrices are further divided into two classes: primitive and imprimitive. A primitive matrix is an irreducible matrix that becomes positive ($a_{ij} > 0$, for all i, j) when raised to a sufficiently large power (Seneta, 1973).

Furthermore, all adjacency matrices are non-negative because all a_{ij} are greater than or equal to zero.

Weak and disconnected digraphs are decomposable into a unique set of maximally induced, strong subdigraphs (\mathbf{K}_i , $i = 1, \dots, \alpha$, where $\alpha \leq n$), which are termed *strongly connected components* (Bang-Jensen and Gutin, 2001). An induced subdigraph of D is a subset of nodes in D with all links that both start and terminate on the node subset, and a maximally induced subdigraph is the largest one that is strong. This implies that there is at least one simple cycle (no repeated medial nodes) that connects all nodes in a non-trivial strongly connected component. Further,

adjacency matrices associated with \mathbf{K}_i are irreducible. For example, the digraph in Fig. 1a contains two connected subdigraphs {1, 2, 3, 4} and {5, 6, 7}. In addition, it can be partitioned into five strongly connected components $\mathbf{K}_1 = \{1, 2, 3\}$, $\mathbf{K}_2 = \{4\}$, $\mathbf{K}_3 = \{5\}$, $\mathbf{K}_4 = \{6\}$, $\mathbf{K}_5 = \{7\}$, of which only \mathbf{K}_1 is non-trivial. The adjacency matrix associated with each strongly connected component is irreducible; the adjacency matrices associated with \mathbf{K}_1 and \mathbf{K}_2 are primitive.

Given these definitions, the Perron–Frobenius theorem guarantees there is one real eigenvalue equal to or larger than all other eigenvalues, $\lambda_1 \geq \lambda_i$ ($i = 2, \dots, n$), in irreducible matrices (Berman and Plemmons, 1979; Seneta, 1973). In the literature $\lambda_1(\mathbf{A})$ is alternately referred to as the dominant eigenvalue, the Perron eigenvalue, and the spectral radius. Next, we illustrate why $\lambda_1(\mathbf{A})$ is a good measure of the pathway proliferation rate.

As shown by Caswell (2001), we can divide both sides of Eq. (2) by λ_1 to obtain

$$\frac{\mathbf{A}^m}{\lambda_1^m} = \mathbf{w}_1 \mathbf{v}_1 + \frac{\lambda_2^m}{\lambda_1^m} \mathbf{w}_2 \mathbf{v}_2 + \frac{\lambda_3^m}{\lambda_1^m} \mathbf{w}_3 \mathbf{v}_3 + \dots + \frac{\lambda_n^m}{\lambda_1^m} \mathbf{w}_n \mathbf{v}_n. \quad (3)$$

If \mathbf{A} is primitive and irreducible, then λ_1 is strictly larger than $\|\lambda_i\|$ for all $i > 1$, where $\|\bullet\|$ is the norm of \bullet (this is necessary since λ_i may be complex). Taking the limit of both sides of Eq. (3) as pathway length increases, we find that

$$\lim_{m \rightarrow \infty} \frac{\mathbf{A}^m}{\lambda_1^m} = \mathbf{w}_1 \mathbf{v}_1. \quad (4)$$

Thus, smaller eigenvalues influence pathway proliferation over shorter path lengths (Fig. 1c), but as path length increases the pathway proliferation rate asymptotically becomes λ_1 . Consequently, $\mathbf{A}^{m+1}/\mathbf{A}^m \rightarrow \lambda_1$ as $m \rightarrow \infty$ (Hill, unpublished ms.), and λ_1 is the asymptotic rate of pathway proliferation in a strongly connected graph with a primitive adjacency matrix. In addition, the damping ratio

$$\rho = \frac{\lambda_1}{\|\lambda_2\|} \quad (5)$$

characterizes the rate of convergence to $\lambda_1(\mathbf{A})$ (Caswell, 2001).

In strong digraphs with an imprimitive and irreducible adjacency matrix there are $c \leq n$ eigenvalues with the same absolute magnitude, and one or more may be complex (Seneta, 1973). The Perron–Frobenius theorem then indicates that the common absolute magnitude of the c eigenvalues will be larger than the other $n - c$ eigenvalues. In this case, the dominant eigenvalue has a multiplicity of c , and as $m \rightarrow \infty$ only the c largest eigenvalues will influence pathway proliferation. Caswell (2001) reports that these digraphs generate oscillatory dynamics.

Given these mathematical results, each strong component \mathbf{K}_i of a weak or disconnected digraph will have an independent rate of pathway proliferation, $\lambda_1(\mathbf{K}_i)$ (read “ λ_1 of \mathbf{K}_i ”). The eigenvalues of a reducible matrix are the union set of the eigenvalues of the adjacency matrices associated

with strongly connected components (Jain and Krishna, 2003). Thus, the maximum dominant eigenvalue of the strongly connected components will be the dominant eigenvalue of the whole digraph. Further, trivial strongly connected components—those with only one node—will have a pathway proliferation rate of unity or 0 depending on whether or not it has a self-loop. Thus, if a digraph is composed of only trivial strongly connected components without self-loops, pathway proliferation will not occur. This is true of all acyclic digraphs, and suggests that we can use the dominant eigenvalue to detect the presence of cycles in digraphs (Jain and Krishna, 2003). For this application there are three cases:

- (1) if $\lambda_1(\mathbf{A}) = 0$, then \mathbf{A} has no cycles;
- (2) if $\lambda_1(\mathbf{A}) = 1$, then \mathbf{A} has at least one cycle and all cycles occur in strongly connected components that have only one simple cycle; and
- (3) if $\lambda_1(\mathbf{A}) > 1$, then \mathbf{A} has more than one simple cycle.

Based on an independent development, Fath (1998) interpreted similar results as three classes of feedback: (1) no feedback, (2) simple feedback, and (3) cyclic feedback in strongly connected networks. Notice that a graph with $\lambda_1(\mathbf{A}) \geq 1$ could have a reducible or irreducible adjacency matrix, while the adjacency matrix of a graph with $\lambda_1(\mathbf{A}) = 0$ is necessarily reducible with $\alpha = n$ trivial strongly connected components.

Similarly, as $m \rightarrow \infty$ we can summarize three possibilities for the dominant eigenvalue as a measure of pathway proliferation in digraphs:

- (1) if $\lambda_1(\mathbf{A}) = 0$, then the number of pathways between two nodes declines to zero;
- (2) if $\lambda_1(\mathbf{A}) = 1$, then the number of pathways between nodes in a strongly connected component remains constant; and
- (3) if $\lambda_1(\mathbf{A}) > 0$, then the number of pathways between nodes in at least one strongly connected component (\mathbf{K}_i) increases without bound at an asymptotic rate equal to $\lambda_1(\mathbf{K}_i)$ where $\max(\lambda_1(\mathbf{K}_i)) = \lambda_1(\mathbf{A})$.

2.2. Bounds and expected values of pathway proliferation

Given that $\lambda_1(\mathbf{A})$ is the asymptotic rate of pathway proliferation in strong digraphs, it would be useful to know its theoretical bounds and expected value. Again, existing mathematics provides us with some of these answers.

Matrix theory bounds the dominant eigenvalue of a non-negative matrix by the minimum and maximum column (row) sum, which in the context of directed graphs is the minimum and maximum out-degree (in-degree), where equality holds only if $k^{in} = k^{out}$ (Berman and Plemmons, 1979; Seneta, 1973). Thus, $\max(\min(k_j^{in}), \min(k_j^{out})) \leq \lambda_1(\mathbf{A}) \leq \min(\max(k_j^{in}), \max(k_j^{out}))$. In a strongly connected digraph with more than one node, all nodes must have at

least one and a maximum of n incoming and outgoing links. Therefore, $1 \leq \lambda_1(\mathbf{A}) \leq n$ for a strong digraph. As stated previously, a trivial component with no self-loops will have $\lambda_1(\mathbf{K}_i) = 0$, and a complete graph will have $\lambda_1(\mathbf{A}) = n$ (allowing self-loops). Notice that in the binary matrix \mathbf{A} , $\lambda_1(\mathbf{A})$ cannot take values between 0 and 1.

With these bounds, we can now examine the expected value of $\lambda(\mathbf{A})$. In undirected random graphs G with $\mathbf{A} = (a_{ij})$ where $a_{ij} = a_{ji} = 1$ with probability p ($0 < p < 1$) and $a_{ij} = a_{ji} = 0$ with probability $(1 - p)$, Juhász proved that

$$\lim_{n \rightarrow \infty} \frac{\lambda_1(\mathbf{A})}{n} = p \quad (6)$$

(Cvetkoviáč and Rowlinson, 1990). This implies that $\lambda_1(\mathbf{A}) \sim np$ in the limit of large n . Furthermore, given $\lambda_1(\mathbf{A})$ and n we can determine the approximate number of undirected edges in \mathbf{A} . A random graph is not necessarily connected, but Erdős and Rényi (1959, 1960) showed that the fraction of nodes connected in a single component increases rapidly when the average link density exceeds unity.

The expected value of $\lambda_1(\mathbf{A})$ is sensitive to the assumptions of random graphs. For example, Farkas et al. (2001) and Goh et al. (2001) found that in random graphs with a power-law distribution of node degrees rather than the standard Poisson distribution, $\lambda_1(\mathbf{A})$ increased with approximately $n^{1/4}$. Furthermore, de Aguiar and Bar-Yam (2005) demonstrated that the expected $\lambda_1(\mathbf{A})$ is further modified if the network topology displays a hierarchic modularity.

Random graphs are well studied, but properties of random digraphs are less well known. Some characteristics are similar to undirected graphs. For example, the in- and out-degree of random digraphs has a Poisson distribution, and when link density is greater than unity the expected size of the largest strongly connected component increases rapidly (Barbosa et al., 2003; Karp, 1990; Luczak, 1990). However, we are unaware of results regarding the spectra of random directed graphs. Therefore, we numerically verified that $\lambda_1(\mathbf{A}) \sim np$ remains plausible for random directed graphs by determining the largest eigenvalue in an ensemble of 99,000 random digraphs (50 from each combination of $n = \{2, 3, \dots, 100\}$ and $p = \{0.05, 0.10, 0.15, \dots, 1\}$). Our results indicate that $\lambda_1(\mathbf{A}) \sim nC = L/n$, where $C = L/n^2$ is an estimate of p (Fig. 2). As either L increases or n decreases the residual error decreases. We conclude that in random digraphs, as in undirected random graphs, $\lambda_1(\mathbf{A})$ is largely determined by the combination of the number of nodes and number of direct connections; pattern of connections has a minor influence. In digraphs with a more structured topology—such as those with power-law in-degree or out-degree distributions or modularity—we might expect $\lambda_1(\mathbf{A})$ to deviate from L/n as it does in undirected graphs, though this remains to be explored.

In ecological networks where n is the number of species (functional groups, etc.) and L is the number of direct

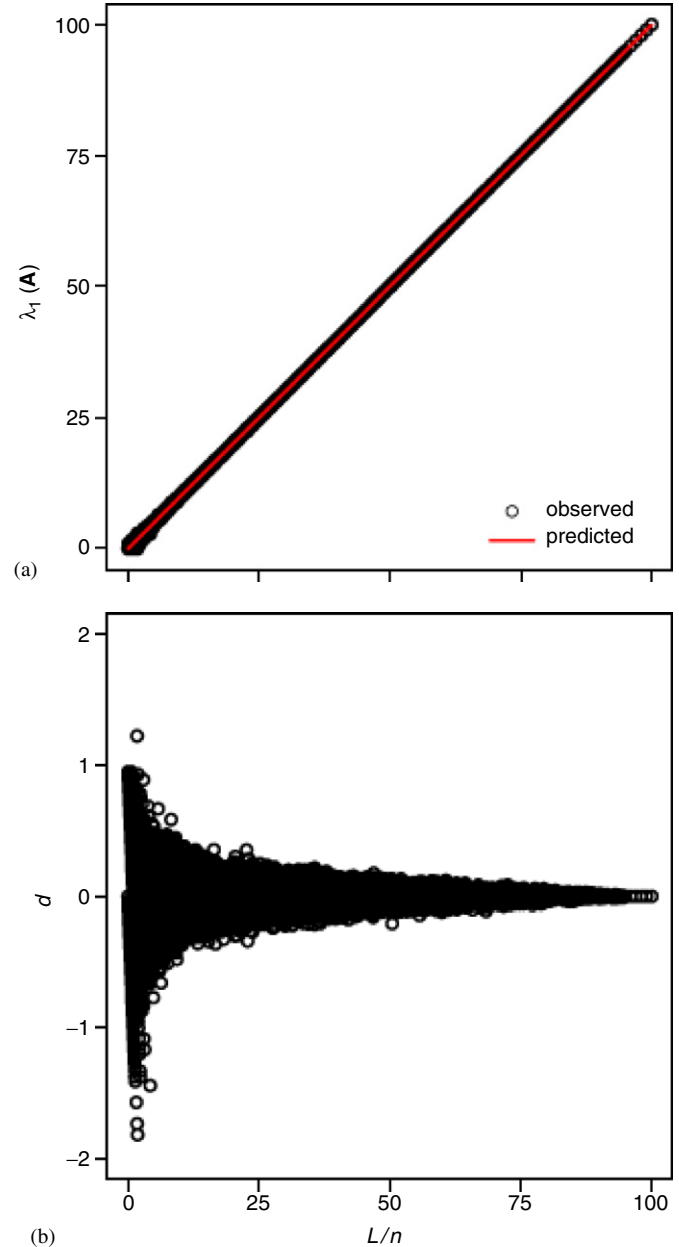


Fig. 2. Dominant eigenvalue in random digraphs. (a) Points show the relationship between dominant eigenvalues $\lambda_1(\mathbf{A})$ and link density L/n in 99,000 uniform random digraphs where $a_{ij} = 1$ with probability p and $a_{ij} = 0$ with probability $(1 - p)$ (50 replicates of each combination of $n = \{2, 3, \dots, 100\}$ and $p = \{0.05, 0.10, \dots, 1\}$). The line indicates the expected $\lambda_1(\mathbf{A}) = L/n$ relationship. (b) Distribution of residual values indicating that as L/n increases it becomes a better predictor of $\lambda(\mathbf{A})$.

transactions, the rate of pathway proliferation will be heavily influenced by species richness and direct link abundance. However, the results of Farkas et al. (2001) and Goh et al. (2001) suggest that if the degree distributions are skewed, as has been demonstrated for some foodwebs (Dunne et al., 2002b; Montoya and Sole, 2002; Williams et al., 2002), or the networks contain other types of order such as modularity, then the residual error $d = |\lambda_1(\mathbf{A}) - L/n|$ may be larger than expected from random graphs.

We have presented three key results in this section. First, pathway proliferation occurs only if there is more than one cycle or feedback in the graph (i.e. it does not occur in acyclic networks). Second, the dominant eigenvalue of strongly connected components $\lambda_1(\mathbf{K}_i)$ is the asymptotic proliferation rate for all nodes within that component, and this rate can vary between strongly connected components within a network. Further, the largest $\lambda_1(\mathbf{K}_i)$ of a network is the dominant eigenvalue for the whole network such that $\max(\lambda_1(\mathbf{K}_i)) = \lambda_1(\mathbf{A})$. Third, although topology can be a factor, the proliferation rate is heavily influenced by the number of nodes and number of direct links. Additionally, networks composed of α non-trivial strongly connected components \mathbf{K}_i ($i = 1, \dots, \alpha$) have a form of structural modularity that may be functionally significant to the complex systems being modeled.

3. Modularity and pathway proliferation in food-webs

In this section we build on our conceptual and mathematical developments by applying the theory to determine the pathway proliferation rates in 17 of the largest empirical food-webs currently available, which are identified in Table 1. Five of the food-webs are terrestrial, three are from freshwater habitats, and seven represent marine or oceanic habitats. Following accepted protocol, original food-webs were modified such that any species or functional groups with identical predators and prey were grouped as a “trophic species” or trophospecies to reduce methodological bias in the data (Cohen et al., 1990; Pimm

et al., 1991; Yodzis, 1982, 1998). These food-webs have been the subject several network analyses (Dunne et al., 2002a, b, 2004; Williams et al., 2002; Williams and Martinez, 2000) which previously reported their number of trophospecies or nodes n , connectance $C = L/n^2$, the proportion of basal species $\%B$ ($k_i^{in} = 0$), proportion of intermediate species $\%I$ ($k_i^{in}, k_i^{out} > 0$), proportion of top species $\%T$ ($k_i^{out} = 0$), and link density L/n . We reproduce this basic network information in Table 1 for comparison. In addition, Dunne et al. (2002a, b) showed that several have skewed degree distributions (i.e. power-law, exponential).

3.1. Methods

We first identified and characterized all strongly connected components (\mathbf{K}_i), including their rates of pathway proliferation and damping ratio, in these food-webs. We envisioned three possible outcomes. If food-webs were adequately modeled by random digraphs then we would expect each web to have one strongly connected component encompassing most if not all of the nodes with a single pathway proliferation rate close to link density $\lambda_1(\mathbf{A}) \sim L/n$. This seemed unlikely given the known skewed degree distributions and additional evidence that ecological processes construct non-random topologies (Cohen et al., 1990; Williams and Martinez, 2000), despite arguments to the contrary (Kenny and Loehle, 1991). A second possibility is based on the observation that most early food-webs were acyclic (Cohen et al., 1990). Thus, the 17

Table 1
Topological properties of 17 empirical food-webs

Habitat	Food-web	Original reference	Taxa ^a	n^a	C^a	$\%B^a$	$\%I^a$	$\%T^a$	L/n^a	$\lambda_1(\mathbf{A})$	d	$\text{Pr}(d)$	#K	$\%K$
Terrestrial	Coachella Valley	Polis, 1991	30	29	0.31	10	90	0	9.03	6.35	2.7	0.001*	2	59
	St. Martin Island	Goldwasser and Roughgarden, 1993	44	42	0.12	14	69	17	4.88	0.00	4.9	0.001*	0	0
	El Verde Rainforest	Waide and Reagan, 1996	156	155	0.06	18	69	13	9.74	10.25	0.5	0.001*	1	45
	UK Grassland	Martinez et al., 1999	75	61	0.03	18	69	13	1.59	0.00	1.6	0.001*	0	0
	Scotch Broom	Memmott et al., 2000	154	85	0.03	1	40	59	2.62	1.00	1.6	0.001*	0	0
Lake/Pond	Skipworth Pond	Warren, 1989	35	25	0.32	4	92	4	7.88	3.00	4.9	0.001*	2	20
	Bridge Brook Lake	Havens, 1992	75	25	0.17	32	68	0	4.28	2.00	2.3	0.001*	1	8
	Little Rock Lake	Martinez, 1991	181	92	0.12	13	86	1	10.84	6.20	4.6	0.001*	2	26
Stream	Canton Creek	Townsend et al., 1998	108	102	0.07	53	22	25	6.83	1.00	5.8	0.001*	0	0
	Stony Stream	Townsend et al., 1998	112	109	0.07	56	27	17	7.61	1.00	6.6	0.001*	0	0
Estuary	Chesapeake Bay	Baird and Ulanowicz, 1989	33	31	0.07	16	52	32	2.19	1.00	1.2	0.001*	0	0
	St. Marks Estuary	Christian and Luczkovich, 1999	48	48	0.10	10	80	10	4.60	1.00	3.6	0.001*	0	0
	Ythan Estuary, 1991	Hall and Raffaelli, 1991	92	83	0.06	9	54	37	4.76	1.62	3.1	0.001*	1	2
	Ythan Estuary, 1996	Huxham et al., 1996	134	124	0.04	4	56	40	4.67	1.62	3.1	0.001*	1	2
Marine	Benguela	Yodzis, 1998	29	29	0.24	7	93	0	7.00	3.00	4.0	0.001*	2	21
	Caribbean Reef, small	Opitz, 1996	50	50	0.22	6	94	0	11.12	8.63	2.5	0.001*	2	60
	NE US Shelf	Link, 2002	81	79	0.22	3	94	4	17.76	4.87	12.9	0.001*	2	39

Taxa refers to the original number of species; n is the number of nodes or trophospecies; $C = L/n^2$ is connectance; $\%B$, $\%I$, and $\%T$ are the proportions of basal (indegree = 0), intermediate (indegree and outdegree > 0), and top (outdegree = 0) trophospecies; L/n is link density; $\lambda_1(\mathbf{A})$ is the dominant eigenvalue of the entire foodweb; $d = |\lambda_1(\mathbf{A}) - L/n|$; $\text{Pr}(d)$ is the fraction of an ensemble of 1001 random digraphs in which d is greater than or equal to that observed in \mathbf{K}_i ; * indicates statistically significant $\text{Pr}(d)$ at $\alpha = 0.05$; #K is the number of non-trivial strongly connected components; $\%K$ is the percent of species in a non-trivial strongly connected component.

^aTopological properties previously reported for these food-webs (Dunne et al., 2002a, 2004; Williams and Martinez, 2000).

food-webs in our study, all developed since 1990, might also be acyclic digraphs with no non-trivial strongly connected components. In this case, pathway proliferation would not occur and $\lambda_1(\mathbf{A}) = 0$. A final possibility is that the food-webs would tend to have one or more strongly connected components and multiple pathway proliferation rates. This outcome would support the hypothesized modularity of ecological systems which is thought to increase system stability (Krause et al., 2003; May, 1972; Pimm and Lawton, 1980; Yodzis, 1982).

In Section 2 we hypothesized that the absolute difference between the dominant eigenvalue and its expected value in random digraphs (L/n) might be a useful indicator of the significance of network topology. To assess this hypothesis, we used Monte Carlo simulations to determine if $d = |\lambda_1(\mathbf{A}) - L/n|$ was larger than expected. We had two scales of analysis: whole food-web and non-trivial strongly connected components. For both, we constructed 1001 uniform random digraphs with n nodes, where each possible link was connected with probability p equal to the original network's connectivity ($p = C$). We assessed statistical significance by determining the fraction of random digraphs in which d was equal to or greater than observed in our network of interest, $\Pr(d)$. Assuming a significance level of $\alpha = 0.05$, $\Pr(d) < 0.05$ implies d is statistically significant.

When applied to the entire food-web, a significant difference with the null model implies that topological factors beyond species and link richness are significant in determining the whole system dominant eigenvalue. This could be the size or frequency of strongly connected components within the network, as suggested by the analysis in Section 2, or perhaps a skewed degree distribution. If the food-webs had more than one non-trivial component, then we expected the deviation to be large.

When applied to strongly connected components, a significant deviation of d also indicates the importance of network topology. However, given that they are irreducible by definition, a significant deviation of d within a component must indicate the significance of another element of topology, such as the degree distribution.

3.2. Results

Food-webs included in this study range in size from 25 to 155 trophospecies and 3% to 32% connectance (Table 1). They tend to have a large proportion of intermediate species (i.e. those with $k_j^{in} > 0$ and $k_j^{out} > 0$), although the two stream food-webs are notable exceptions. Ten of the 17 food-webs examined contained at least one non-trivial strongly connected component; six had two. Five of the remaining food-webs had a dominant eigenvalue of unity, implying that at least one node contained a self-loop. Our results reveal that the majority of these food-webs have at least one directed cycle, contrary to earlier food-web theory (Cohen et al., 1990).

While the majority of the food-webs have a modular organization that is based on strongly connected components, the proportion of species involved in the modules is variable. In food-webs that have a non-trivial strongly connected component, the proportion of the original nodes involved ranges from 2% in the two Ythan Estuary food-webs to 60% in the Caribbean Reef model. Notice that the definition of a strongly connected component bans nodes that have no inputs or no outputs, which excludes basal species ($k_j^{in} = 0$) and top consumer species ($k_j^{out} = 0$). Therefore, the total number of species in strongly connected components is limited by the number of intermediate species. This may be a factor in why the two stream food-webs and the Scotch Broom food-web contain no non-trivial components.

The absolute difference between the dominant eigenvalue of the entire food-web and its expected value based on random digraphs of the appropriate size and connectance ($d = |\lambda_1(\mathbf{A}) - L/n|$) ranged from a minimum of 0.5 for the El Verde rainforest to a maximum of 12.9 for the NE US shelf food-web (Table 1). In all cases, this difference was significantly different from the random digraph null model, indicating that topology is a significant factor in determining $\lambda_1(\mathbf{A})$. This result is consistent with the presence of one or more small non-trivial components and acyclic digraphs. Our results provide another line of evidence suggesting that the ecological processes that create food-webs lead to more ordered network topologies; random digraphs are not good models for these systems.

Inspection of the strongly connected components reveals a diversity of topologies as shown in Table 2. The largest strongly connected component, with 70 trophospecies, occurs in the El Verde rainforest model; although it is the least well connected (13%), it still has the largest rate of pathway proliferation ($\lambda_1(\mathbf{A}) = 10.25$). In contrast, 10 of the 16 strongly connected components have four or fewer species. Five of the strongly connected components contain only two trophospecies, requiring a single simple cycle of path length 2 (e.g. $j \rightarrow i \rightarrow j$). While the two strongly connected components in the Coachella Valley, Skipworth Pond, and Benguela are about the same size, one of the two components in Little Rock Lake, Caribbean Reef, and NE US shelf is substantially larger than the other. Table 3 lists the trophospecies in the two strongly connected components of the Coachella Valley.

The dominant eigenvalues of all strongly connected components have a multiplicity of one, so the adjacency matrices associated with the component subdigraphs are primitive. Therefore, the dominant eigenvalues represent the strongly connected component asymptotic rates of pathway proliferation. These range from 1.62 in strongly connected components of the two Ythan Estuary food-webs to 10.25 in the large El Verde rainforest strongly connected component and generally increase with link density as would be expected in random digraphs. However, half of the strongly connected components have a statistically significant difference between the dominant

Table 2
Topological properties of strongly connected components in 17 empirical food-webs

Model	\mathbf{K}_i	n	L	C	L/n	$\lambda_1(\mathbf{K}_i)$	$\text{mult}(\lambda_1)$	ρ	d	$\text{Pr}(d)$
Coachella Valley	1	11	71	0.59	6.45	6.35	1	3.40	0.1	0.001*
	2	6	22	0.61	3.67	3.56	1	3.56	0.1	0.001*
El Verde Rainforest	1	70	633	0.13	9.04	10.25	1	2.43	1.2	0.001*
	2	3	9	1.00	3.00	3.00	1	—	0.0	1.001
Skipworth Pond	1	3	9	1.00	3.00	3.00	1	—	0.0	1.001
	2	2	4	1.00	2.00	2.00	1	—	0.0	1.001
Bridge Brook Lake	1	2	4	1.00	2.00	2.00	1	—	0.0	1.001
Little Rock Lake	1	21	167	0.38	7.95	6.20	1	2.42	1.7	0.001*
	2	3	9	1.00	3.00	3.00	1	—	0.0	1.001
Ythan Estuary, 1991	1	2	3	0.75	1.50	1.62	1	2.62	0.1	0.212
Ythan Estuary, 1996	1	2	3	0.75	1.50	1.62	1	2.62	0.1	0.226
Benguela	1	3	9	1.00	3.00	3.00	1	—	0.0	1.001
	2	3	7	0.78	2.33	2.25	1	4.05	0.1	0.001*
Caribbean Reef, small	1	2	3	0.75	1.50	1.62	1	2.62	0.1	0.217
	2	28	244	0.31	8.71	8.63	1	4.11	0.1	0.001*
NE US Shelf	1	4	11	0.69	2.75	2.88	1	4.41	0.1	0.001*
	2	27	243	0.33	9.00	4.87	1	1.69	4.1	0.001*

\mathbf{K}_i indicates the non-trivial strongly connected component number; n is the number of nodes (trophospecies), L is the number of links; $C = L/n^2$ is connectance, L/n is the link density, $\lambda_1(\mathbf{K}_i)$ is the dominant eigenvalue of \mathbf{K}_i , $\text{mult}(\lambda_1)$ is the multiplicity of the dominant eigenvalue, $\rho = \lambda_1(\mathbf{K}_i)/\|\lambda_2(\mathbf{K}_i)\|$ is the damping ratio (— indicates ρ is undefined because $\|\lambda_2(\mathbf{K}_i)\| = 0$), $d = |\lambda_1(\mathbf{A}) - L/n|$ is the absolute difference between the dominant eigenvalue and link density, $\text{Pr}(d)$ is the fraction of an ensemble of 1001 random digraphs in which d is greater than or equal to that observed in \mathbf{K}_i , and * indicates statistically significant $\text{Pr}(d)$ at $\alpha = 0.05$.

Table 3
Trophospecies in the two non-trivial strongly connected components of the Coachella Valley food-web

\mathbf{K}_1	\mathbf{K}_2
Primarily herbivorous mammals and birds	Small arthropod predators
Small omnivorous mammals and birds	Medium arthropod predators
Primarily carnivorous lizards	Large arthropod predators
Primarily carnivorous snakes	Facultative arthropod predators
Large primarily predacious birds	Life-history arthropod omnivore
Large primarily predacious mammals	Spider parasitoids
	Primary parasitoids
	Hyperparasitoids
	Predacious mammals and birds
	Arthropodivorous snakes
	Primarily arthropodivorous lizards

eigenvalue and link density including Coachella Valley (\mathbf{K}_1 and \mathbf{K}_2), El Verde Rainforest (\mathbf{K}_1), Little Rock Lake (\mathbf{K}_1), Benguela (\mathbf{K}_2), Caribbean Reef (\mathbf{K}_2), and NE US shelf (\mathbf{K}_1 and \mathbf{K}_2). The topological arrangement of species and links in these three strongly connected components influences their rate of pathway proliferation; the others are largely determined by their species and link richness.

The damping ratio defined in Eq. (5) is an index of the speed of convergence to the asymptotic rate of pathway proliferation; a larger ratio indicates faster convergence.

Five of the strongly connected components are completely connected. They have a pathway proliferation rate equal to their trophospecies richness and an undefined damping ratio because their second eigenvalues are zero. In these cases the asymptotic rate of pathway proliferation is achieved instantaneously. The other damping ratios range from 1.69 in \mathbf{K}_2 of the NE US Shelf to 4.41 in \mathbf{K}_1 of the same food-web. Transient dynamics of the pathway proliferation rate, determined by the smaller eigenvalues, are more influential in NE US shelf (\mathbf{K}_2). Its pathway proliferation rate does not converge until a pathway length of nearly 22, while in NE US shelf (\mathbf{K}_1) the proliferation rate converges by a pathway length of 8.

In summary, the majority of the food-webs we examined contained at least one non-trivial strongly connected component. Six food-webs had two non-trivial strongly connected components; none had more than two. The proportion of species involved in a strongly connected component ranged from 2% to 60%. In all cases, the difference between the dominant eigenvalue of the food-web and the expected value (L/n) in a random network was significant. This difference occurs because the topology of food-webs is non-random; thermodynamic processes and species characteristics (e.g. the species niche) combine to form non-random structures (Chase and Leibold, 2003). Within the strongly connected components, the rate of pathway proliferation ranged from 1.62 to 10.25 and half were indistinguishable from random graphs based on the expected rate of pathway proliferation.

4. Discussion

As with any analysis of network models that reveals previously hidden structural patterns, we are left with two questions. First, what, if any, significance do these patterns hold for our systems of interest? Are strongly connected components and pathway proliferation simply another pretty pattern, another network or food-web statistic to report, or do they impart some functional significance? The second question cannot be divorced from the first; what system processes might create these structural patterns? Are there ecological processes or forces that might lead to the development of these structures? These are not easy questions to answer, but in this section we attempt to address them for the presence of strongly connected components and pathway proliferation in ecological networks.

Strongly connected components introduce a form of modularity into network models, where modularity is defined as a hierarchical system subdivision into more or less interacting subsystems. Several types of modularity have been proposed in ecological systems. Building on earlier ideas in general systems theory that linked hierarchical organization to system stability (e.g. Simon, 1962), May (1973) hypothesized that ecosystems have modular structures. He found that the Lyapunov stability of randomly assembled ecosystems tended to be greater when the species were partitioned into blocks of interacting species with few if any connections to other blocks. Pimm (1979, p. 145) termed these blocks of species “compartments”, stating that they are “...characterized by strong interactions within compartments, but weak interactions among compartments”. Pimm and Lawton (1980) concluded from a study of binary empirical food-webs that, while there was evidence species were grouped into subsystems largely by habitat, compartmentalization as defined by Pimm was an uncommon phenomenon. They noted, however, that a complete test of the hypothesis would require knowledge of the strength of interactions, which was absent in their food-web models. It is also possible that they were unable to identify compartmentalization in their food-webs because their models exclude detrital recycling. However, Neutel et al. (2002) showed that even without detrital recycling, long, weak links enhanced system stability.

Yodzis (1982) remarked that modular organization was an old idea in ecology, citing the guild concept (Root, 1967) as an example. He applied the dominant clique idea from graph theory to identify another type of modularity in food-webs. He defined “a clique as a set of species in a given ecosystem with the property that every pair in the set has some food resource in common, and ... a dominant clique as a clique which is contained in no other clique” (Root, 1967, p. 552). More recently, Krause et al. (2003) used a methodology developed to identify cohesive subgroups in social network analysis to classify another type of modularity in food-webs. They demonstrated that

this type of organization increased system stability to species deletion by localizing the effect within a module. In addition, Allesina et al. (2005) found four or more modules based on strongly connected components in 17 network models of carbon flux. The four modules usually grouped into four types: inputs, outputs, dissipation, and species and nutrient pools. Their work is the most similar to that presented in this paper, but there are two important differences. First, we examined different types of ecological networks. They examined ecosystem flow networks, while we restricted our analysis to food-webs. Second, we considered inputs, outputs, and dissipation to be external to the system and therefore we could not identify these as separate modules.

The dominant ecological hypothesis is that food-web or ecosystem modularity increases overall system stability by localizing interactions within modules. Given the static, binary, presence–absence information of food-webs in our study we were unable to meaningfully test this hypothesis; stability is inherently a dynamic concept. Known issues with food-web model construction further make this hypothesis difficult to resolve (Cohen et al., 1993; Polis and Strong, 1996). Empirical food-web models usually do not indicate interaction strength or the temporal and spatial variation of the interactions, as these details are expensive to acquire.

Despite the challenge of assessing their effect on system stability, strongly connected components in ecological networks appear to be important functional elements of system organization and provide new insights about species participating in them. By definition (Section 2), there is minimally one simple cycle that encircles all nodes in the strongly connected component. This provides at least one channel for cybernetic feedback (positive or negative) to propagate among species in the module (DeAngelis et al., 1986; Patten, 1959; Patten and Odum, 1981). Furthermore, in food-webs it is reasonable to assume that each predatory species directly benefits by its consumption of prey. This establishes an indirect mutualism that spans the strongly connected component, and provides the necessary conditions for the strongly connected components of food-webs to function as autocatalytic cycles—systems that catalyse their own production (Maynard Smith and Szathmáry, 1995; Ulanowicz, 1997). Autocatalytic cycles are an essential element of metabolism in chemical and living systems and may have played a role in the origin of life (Maynard Smith and Szathmáry, 1995). Maynard Smith and Szathmáry (1995) describe autocatalytic cycles as a force for cooperation among the member species and efficient information integrators.

Ulanowicz (1997) identifies several emergent properties that autocatalytic cycles may possess, including centripetality, persistence, and autonomy. Centripetality is the tendency of the cycle to attract more of the energy–matter flux of the system. If any member species becomes more efficient at using its resources or better able to acquire new resources such that its population increases, this positive

change tends to cascade through the module, collectively benefiting the populations of all species involved. Autocatalytic cycles tend to persist because their general form can be maintained in a system despite fluctuations in the interaction strengths and possible element replacement. In food-webs this implies that when a trophically similar species is introduced to the system, if it is more efficient or in some way ecologically more competitive, it may wholly replace an existing species in the autocatalytic cycle, but the cycle remains. Finally, autocatalytic cycles can establish a degree of autonomy because species in the cycle can actively influence at least a portion of their input environment. In this sense, species in strongly connected components of food-webs are involved in ecosystem engineering (Jones et al., 1997) and niche construction (Laland et al., 1999; Odling-Smee et al., 2003). Ulanowicz (1986, 1997) further argues that the autocatalytic nature of cycles in ecosystems makes them a principal agent in ecosystem growth and development.

Maynard Smith and Szathmáry (1995) remarked that autocatalytic cycles are sensitive to cheaters or parasites that feed off the strongly connected component without participating in the cycle. Top predators feeding on species in a strongly connected component or a downstream strongly connected component might function as parasites in this sense. Perhaps this is why strongly connected components do not occur in all the food-webs analysed. This and the tendency for centripetality may explain why there are fewer than three strongly connected components in these food-webs. However, we are unable to assess these possibilities with these data because the differences may simply reflect disparities in food-web modeling decisions.

Pathway proliferation rates of strongly connected components provide additional information about these modules. Each additional link in a strongly connected component beyond those that form the defining cycle introduces another embedded simple cycle. This lowers the maximal distance between nodes in the module, increases the potential pathways for energy–matter flux, tends to increase the rate of pathway proliferation, and leads to the unbounded growth of pathways as length increases. In some cases, the rate of pathway proliferation will not increase as expected with the number of links. For example, half of the strongly connected components identified in our food-webs had pathway proliferation rates that were significantly different from the expected rate based on random graphs with a Poisson degree distribution. This suggests that module topology differs from what we would expect from a random generating process.

As mentioned earlier, we expected the ecological processes forming food-webs to generate non-random structures. Species characteristics such as metabolic requirements, food preferences, capture ability, and handling time, as well as other niche requirements and natural history constraints should directly influence the choices of “who eats whom” and how much. In addition, the

emergent properties of autocatalytic cycles and ecosystems more generally may provide whole-system constraints.

Perhaps the interesting question is not why the eight strongly connected components did not match the random expectation, but why the other half did? Notice that the strongly connected components with non-random topologies were the largest modules, while the eight strongly connected components with topologies indistinguishable from random digraphs involved only two or three trophospecies. The universe of possible topologies is much smaller in these small and well-connected ($0.75 \leq C \leq 1$) modules, making the ecologically created topologies reflected in the food-webs more likely. Five of these modules were completely connected, generating only one possible arrangement. The eight strongly connected components with apparently non-random topologies were less well connected ($0.13 \leq C \leq 0.78$), generating a much larger universe of possible topologies. In six cases the pathway proliferation rate was significantly less than expected, but in the large strongly connected component of the El Verde rainforest and the smaller module of the Caribbean Reef, the pathway proliferation rate was more than expected. At this point, we are unable to provide a satisfactory explanation for these differences.

In our discussion thus far, we have been interchanging food-webs and ecosystems. It is important to recognize, however, that food-webs are a subset of a broader class of ecosystem models of energy–matter flow. Food-webs are generally defined by the relation “who eats whom” that is one process generating energy–matter flux, while ecosystem flow–storage models typically trace a conserved flow unit (e.g. energy, nitrogen, phosphorus) through the system, regardless of the process producing the flow. Thus, non-trophic ecological processes such as excretion and death are captured in flow–storage models, revealing a different picture of ecosystem organization with additional cycling.

Pathway proliferation influences the development and significance of indirect flows in ecosystem flow–storage models. Indirect flows are an important aspect of the ecological significance of the strongly connected components, so here we take a closer look. Indirect flows are derived from two fundamentally distinct types of pathways: chains (e.g. $5 \rightarrow 6 \rightarrow 7$ in Fig. 1a) and cycles (e.g. $1 \rightarrow 3 \rightarrow 1$ in Fig. 1a). Indirect flows in chains are limited by transfer efficiencies and chain length. In cycles, the number and length of pathways are unlimited such that indirect flows are only limited by transfer efficiencies reflecting energy–matter dissipation and export. As ecosystems are open thermodynamic systems, shorter indirect pathways individually will tend to carry larger indirect flows than longer indirect pathways. A faster rate of pathway proliferation $\lambda_1(\mathbf{A})$ implies that there will be more shorter indirect pathways, increasing the possibility that the magnitude of indirect flows will surpass that of direct flows. Thus, within a strongly connected component the large number of indirect pathways will tend to carry a large fraction of the flow between species (nodes).

More generally, $\lambda_1(\mathbf{A})$ indicates the potential for direct and indirect energy, matter, and information transmission between compartments in a strongly connected component. Realized transmission rates are dependent on the realized use of each pathway. Previous ecosystem network analyses reveal some of the system-level consequences of differential pathway use (Fath, 2004; Fath and Patten, 1998, 1999; Higashi and Patten, 1989; Patten, 1985b; Ulanowicz, 1986), but there is much left to learn about this subject. The interplay of this potential and realized network structure is an interesting, important topic for understanding the organization and transformation of complex adaptive systems like ecosystems.

We conclude that the strongly connected components and pathway proliferation are ecologically relevant phenomena because they provide novel insights about the system of interest. Without knowing the strength of interactions or energy–matter flux rates, the presence of these structural features suggests groups of species functionally integrated by indirect effects mediated by autocatalytic cycles. They portend the possibility of integral species relationships that are shifted toward more positive associations and the possibility of the dominance of indirect flows. In some cases, apparent negative interactions such as predation or competition may become more positive through indirect interactions mediated by the autocatalytic cycles of the strongly connected components.

Acknowledgments

We are grateful to Jennifer Dunne for providing the food-web data, and discussions with Stuart Whipple and the University of Georgia Systems and Engineering Ecology group that enhanced this work. The manuscript benefited from reviews by M.B. Beck, H.R. Pulliam, P.G. Verity, D.K. Gattie, S. Bata, and two anonymous peers. S.R.B. was supported by National Science Foundation Grant no. OPP-00-83381 and IIS-0326059.

References

- Albert, R., Barabási, A.L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.
- Albert, R., Jeong, H., Barabási, A.L., 1999. Diameter of the World-Wide Web. *Nature* 401 (6749), 130–131.
- Albert, R., Jeong, H., Barabási, A.L., 2000. Error and attack tolerance of complex networks. *Nature* 406 (6794), 378–382.
- Allen, T.F.H., Starr, T.B., 1982. *Hierarchy: Perspectives for Ecological Complexity*. University of Chicago Press, Chicago.
- Allesina, S., Bodini, A., Bondavalli, C., 2005. Ecological subsystems via graph theory: the role of strongly connected components. *Oikos* 110, 164–176.
- Baird, D., Ulanowicz, R.E., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* 59, 329–364.
- Bang-Jensen, J., Gutin, G., 2001. *Digraphs: Theory, Algorithms, and Applications*. Springer, London.
- Barabási, A.L., 2002. *Linked: The New Science of Networks*. Perseus, Cambridge, MA.
- Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
- Barbosa, V.C., Donangelo, R., Souza, S.R., 2003. Directed cycles and related structures in random graphs: I—static properties. *Physica A* 321, 381–397.
- Berman, A., Plemmons, R.J., 1979. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York.
- Borrett, S.R., Patten, B.C., 2003. Structure of pathways in ecological networks: relationships between length and number. *Ecol. Modelling* 170, 173–184.
- Borrett, S.R., Whipple, S.J., Patten, B.C., Christian, R.R., 2006. Indirect effects and distributed control in ecosystems. Temporal variability of indirect effects in a seven-compartment model of nitrogen flow in the Neuse River Estuary (USA): time series analysis. *Ecol. Modelling* 194, 178–188.
- Burns, T.P., Higashi, M., Wainright, S.C., Patten, B.C., 1991. Trophic unfolding of a continental shelf energy-flow network. *Ecol. Modelling* 55, 1–26.
- Caldarelli, G., Higgs, P.G., McKane, A.J., 1998. Modelling coevolution in multispecies communities. *J. Theor. Biol.* 193, 345–358.
- Caswell, H., 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*, second ed. Sinauer Associates, Sunderland, MA.
- Chase, J.M., Leibold, M.A., 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press, Chicago.
- Christian, R.R., Luczkovich, J.J., 1999. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecol. Modelling* 117, 99–124.
- Cohen, J.E., Briand, F., Newman, C.M., 1990. *Community Food Webs: Data and Theory*. Springer, New York.
- Cohen, J.E., Beaver, R.A., Cousins, S.H., Deangelis, D.L., Goldwasser, L., Heong, K.L., Holt, R.D., Kohn, A.J., Lawton, J.H., Martinez, N., Omalley, R., Page, L.M., Patten, B.C., Pimm, S.L., Polis, G.A., Rejmanek, M., Schoener, T.W., Schoenly, K., Sprules, W.G., Teal, J.M., Ulanowicz, R.E., Warren, P.H., Wilbur, H.M., Yodzis, P., 1993. Improving food webs. *Ecology* 74, 252–258.
- Cvetkovic, D.M., Rowlinson, P., 1990. The largest eigenvalue of a graph: a survey. *Linear Multilinear Algebra* 28, 3–33.
- Cvetkovic, D.M., Doob, M., Sachs, H., 1980. *Spectra of Graphs: Theory and Application*. Academic Press, New York.
- de Aguiar, M.A.M., Bar-Yam, Y., 2005. Spectral analysis and the dynamic response of complex networks. *Phys. Rev. E* 71, 016106.
- DeAngelis, D.L., 1975. Stability and connectance in food web models. *Ecology* 56, 238–243.
- DeAngelis, D.L., Post, W.M., Travis, C.C., 1986. *Positive Feedback in Natural Systems*. Springer, New York.
- Dunne, J.A., Williams, R.J., Martinez, N.D., 2002a. Food-web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* 99, 12917–12922.
- Dunne, J.A., Williams, R.J., Martinez, N.D., 2002b. Network topology and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* 5, 558–567.
- Dunne, J.A., Williams, R.J., Martinez, N.D., 2004. Network structure and robustness of marine food webs. *Mar. Ecol. Prog. Ser.* 273, 291–302.
- Erdős, P., Rényi, A., 1959. On random graphs. *Publicationes Math.* 6, 290–297.
- Erdős, P., Rényi, A., 1960. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5, 17–61.
- Farkas, I., Derényi, I., Barabási, A.L., Vicsek, T., 2001. Spectra of “real-world” graphs: beyond the semicircle law. *Phys. Rev. E* 64, 026704.
- Fath, B.D., 1998. *Network analysis: foundations, extensions, and applications of a systems theory of the environment*. Ph.D. Dissertation, University of Georgia.
- Fath, B.D., 2004. Network analysis applied to large-scale cyber-ecosystems. *Ecol. Modelling* 171, 329–337.
- Fath, B.D., Patten, B.C., 1998. Network synergism: emergence of positive relations in ecological systems. *Ecol. Modelling* 107, 127–143.
- Fath, B.D., Patten, B.C., 1999. Quantifying resource homogenization using network flow analysis. *Ecol. Modelling* 107, 193–205.
- Finn, J.T., 1980. Flow analysis of models of the Hubbard Brook ecosystem. *Ecology* 61, 562–571.

- Gardner, M.R., Ashby, W.R., 1970. Connectance of large dynamical (cybernetic) systems: critical values for stability. *Nature* 228, 784.
- Godsil, C.D., 1993. *Algebraic Combinatorics*. Chapman & Hall, New York.
- Goh, K.I., Kahng, B., Kim, D., 2001. Spectra and eigenvectors of scale-free networks. *Phys. Rev. E* 64, 051903.
- Goldwasser, L., Roughgarden, J., 1993. Construction and analysis of a large Caribbean food web. *Ecology* 74, 1216–1233.
- Hall, S.J., Raffaelli, D.G., 1991. Food web patterns: lessons from a species rich web. *J. Anim. Ecol.* 60, 823–841.
- Havens, K., 1992. Scale and structure in natural food webs. *Science* 257, 1107–1109.
- Higashi, M., Burns, T.P., 1991. *Theoretical Studies of Ecosystems: The Network Perspective*. Cambridge University Press, Cambridge.
- Higashi, M., Patten, B.C., 1986. Further aspects of the analysis of indirect effects in ecosystems. *Ecol. Modelling* 31, 69–77.
- Higashi, M., Patten, B.C., 1989. Dominance of indirect causality in ecosystems. *Am. Nat.* 133, 288–302.
- Hill, J.V., unpublished ms. On the largest real root of a polynomial.
- Huxham, M., Beaney, S., Raffaelli, D., 1996. Do parasites reduce the chances of triangulation in a real food web? *Oikos* 76, 284–300.
- Jain, S., Krishna, S., 2003. Graph theory and the evolution of autocatalytic networks. In: Bornholdt, S., Schuster, H.G. (Eds.), *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, Berlin, pp. 355–395.
- Jeong, H., Tomnor, B., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic networks. *Nature* 407, 651–654.
- Jones, C.G., Lawton, J.H., Shachak, M., 1997. Positive and negative effects of organisms as physical ecosystem engineers. *Ecology* 78, 1946–1957.
- Jørgensen, S.E., Patten, B.C., Strakraba, M., 1992. Ecosystems emerging: toward an ecology of complex systems in a complex future. *Ecol. Modelling* 62, 1–27.
- Karp, R.M., 1990. The transitive closure of a random digraph. *Random Struct. Algorithms* 1, 73–93.
- Kenny, D., Loehle, C., 1991. Are food webs randomly connected? *Ecology* 72, 1794–1799.
- Krause, A.E., Frank, K.A., Mason, D.M., Ulanowicz, R.E., Taylor, W.W., 2003. Compartments revealed in food-web structure. *Nature* 426 (6964), 282–285.
- Laland, K.N., Odling-Smee, F.J., Feldman, M.W., 1999. Evolutionary consequences of niche construction and their implications for ecology. *Proc. Natl Acad. Sci. USA* 96, 10242–10247.
- Lawler, L., 1978. A comment on randomly constructed model ecosystems. *Am. Nat.* 112, 445–447.
- Link, J., 2002. Does food web theory work for marine ecosystems? *Mar. Ecol. Prog. Ser.* 230, 1–9.
- Luczak, T., 1990. The phase transition in the evolution of random digraphs. *J. Graph Theory* 14, 217–223.
- Margalef, R., 1963. Certain unifying principles in ecology. *Am. Nat.* 97, 357–374.
- Martinez, N.D., 1991. Artifacts or attributes—effects of resolution on the little rock lake food web. *Ecol. Monogr.* 61, 367–392.
- Martinez, N.D., Hawkins, B.A., Dawah, H.A., Feifarek, B.P., 1999. Effects of sampling effort on characterization of food-web structure. *Ecology* 80, 1044–1055.
- May, R., 1972. Will a large complex system be stable? *Nature* 238, 413–414.
- May, R.M., 1973. *Stability and complexity in model ecosystems*. Princeton University Press, Princeton, NJ.
- Maynard Smith, J., Szathmáry, E., 1995. *The Major Transitions in Evolution*. W.H. Freeman Spektrum, New York.
- Memmott, J., Martinez, N.D., Cohen, J.E., 2000. Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. *J. Anim. Ecol.* 69, 1–15.
- Montoya, J.M., Sole, R.V., 2002. Small world patterns in food webs. *J. Theor. Biol.* 214, 405–412.
- Moreno, J.L., Jennings, H.H., 1938. Statistics of social configurations. *Sociometry* 1, 342–374.
- Müller, F., Leupelt, M., 1998. *Eco Targets, Goal Functions, and Orientors*. Springer, New York.
- Neutel, A.M., Heesterbeek, J.A.P., de Ruiter, P.C., 2002. Stability in real food webs: weak links in long loops. *Science* 296, 1120–1123.
- Newman, M.E.J., 2001a. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* 6401, 6131.
- Newman, M.E.J., 2001b. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA* 98, 404–409.
- Newman, M.E.J., 2002. Random graphs as models of networks. In: Bornholdt, S., Schuster, H.G. (Eds.), *Handbook of Graphs and Networks*. Wiley-VCH, Berlin.
- Newman, M.E.J., 2003. The structure and function of complex networks. *SIAM Rev.* 45, 167–256.
- Odling-Smee, F.J., Laland, K.N., Feldman, M.W., 2003. *Niche Construction: The Neglected Process in Evolution*. Princeton University Press, Princeton, NJ.
- Opitz, S., 1996. Trophic interactions in Caribbean coral reefs. Technical Report 43, ICLARM.
- Patten, B.C., 1959. An introduction to the cybernetics of the ecosystem: the trophic dynamic aspect. *Ecology* 40, 221–231.
- Patten, B.C., 1983. On the quantitative dominance of indirect effects in ecosystems. In: Lauenroth, W.K., Skogerboe, G.V., Flug, M. (Eds.), *Analysis of Ecological Systems: State-of-the-Art in Ecological Modelling*. Elsevier, Amsterdam, pp. 27–37.
- Patten, B.C., 1985a. Energy cycling in the ecosystem. *Ecol. Modelling* 28, 1–71.
- Patten, B.C., 1985b. Energy cycling, length of food chains, and direct versus indirect effects in ecosystems. *Can. Bull. Fish. Aquat. Sci.* 213, 119–138.
- Patten, B.C., Odum, E.P., 1981. The cybernetic nature of ecosystems. *Am. Nat.* 118, 886–895.
- Patten, B.C., Bosserman, R.W., Finn, J.T., Cale, W.G., 1976. Propagation of cause in ecosystems. In: Patten, B.C. (Ed.), *Systems Analysis and Simulation in Ecology*, vol. IV. Academic Press, New York, pp. 457–579.
- Patten, B.C., Richardson, T.H., Barber, M.C., 1982. Path analysis of a reservoir ecosystem model. *Can. Water Resour. J.* 7, 252–282.
- Patten, B.C., Higashi, M., Burns, T.P., 1990. Trophic dynamics in ecosystem networks: significance of cycles and storage. *Ecol. Modelling* 51, 1–28.
- Pimm, S.L., 1979. The structure of food webs. *Theor. Popul. Biol.* 16, 144–158.
- Pimm, S.L., 1982. *Food Webs*. Chapman & Hall, New York.
- Pimm, S.L., Lawton, J.H., 1977. On feeding on more than one trophic level. *Nature* 275, 542–544.
- Pimm, S.L., Lawton, J.H., 1980. Are food webs divided into compartments? *J. Anim. Ecol.* 49, 879–898.
- Pimm, S.L., Lawton, J.H., Cohen, J.E., 1991. Food web patterns and their consequences. *Nature* 350, 669–674.
- Polis, G.A., 1991. Complex trophic interactions in deserts: an empirical critique of food-web theory. *Am. Nat.* 138, 123–155.
- Polis, G.A., Strong, D.R., 1996. Food web complexity and community dynamics. *Am. Nat.* 157, 813–846.
- Ponstein, J., 1966. *Matrices in Graph and Network Theory*. Van Nostrand Reinhold, New York.
- Post, D.M., 2002. The long and short of food-chain length. *Trends Ecol. Evol.* 17, 269–277.
- Post, D.M., Pace, M.L., Hairston, N.G., 2000. Ecosystem size determines food-chain length in lakes. *Nature* 405, 1047–1049.
- Price, D., 1965. Networks of scientific papers. *Science* 149, 510–515.
- Root, R.B., 1967. The niche exploitation pattern of the blue-gray gnatcatcher. *Ecol. Monogr.* 37, 317–350.
- Seneta, E., 1973. *Non-negative Matrices: An Introduction to Theory and Applications*. Wiley, New York.
- Simon, H.A., 1962. The architecture of complexity. *Proc. Am. Philos. Soc.* 106, 467–482.

- Townsend, C.R., Thompson, R.M., McIntosh, A.R., Kilroy, C., Edwards, E., Scarsbrook, M.R., 1998. Disturbance, resource supply, and food-web architecture in streams. *Ecol. Lett.* 1, 200–209.
- Ulanowicz, R.E., 1986. *Growth and Development: Ecosystems Phenomenology*. Springer, New York.
- Ulanowicz, R.E., 1997. *Ecology, The Ascendent Perspective*. Columbia University Press, New York.
- Ulanowicz, R.E., Wolff, W.F., 1991. Ecosystem flow networks—loaded dice? *Math. Biosci.* 103, 45–68.
- Waide, R.B., Reagan, W.B., 1996. *The Food Web of a Tropical Rainforest*. University of Chicago Press, Chicago, IL.
- Warren, P.H., 1989. Spatial and temporal variation in the structure of a fresh-water food web. *Oikos* 55, 299–311.
- Watts, D.J., 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, Princeton, NJ.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of “small-world” networks. *Nature* 393, 440–442.
- Whipple, S.J., 1998. Path-based network unfolding: a solution for the problem of mixed trophic and non-trophic processes in trophic dynamic analysis. *J. Theor. Biol.* 190, 263–276.
- Williams, R.J., Berlow, E.L., Dunne, J.A., Barabási, A.-L., Martinez, N.D., 2002. Two degrees of separation in complex food webs. *Proc. Natl Acad. Sci. USA* 99, 12913–12916.
- Williams, R.J., Martinez, N.D., 2000. Simple rules yield complex food webs. *Nature* 404, 180–183.
- Yodzis, P., 1982. The compartmentation of real and assembled ecosystems. *Am. Nat.* 120, 551–570.
- Yodzis, P., 1998. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *J. Anim. Ecol.* 67, 635–658.