ELSEVIER

ECOLOGICAL COMPLEXITY
An International Journal on
BIOCOMPLEXITY IN THE ENVIRONMENT
AND THEORETICAL ECOLOGY

Editor-in-Chief
Bai-Lian (Larry) Li

Characterization of topological keystone species Local, global and "meso-scale" centralities in food webs

Ernesto Estrada*

Complex Systems Research Group, X-rays Unit, RIAIDT, Edificio CACTUS, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

ARTICLE INFO

ABSTRACT

An important question in the network representation of ecological systems is to determine how direct and indirect interactions between species determine the positional importance of species in the ecosystem. Here we present a quantitative analysis of the similarities and differences of six different topological centrality measures as indicators of keystone species in 17 food webs. These indicators account for local, global and "meso-scale" – intermediate between local and global – topological information about species in the food webs. Using factor analysis we shown that most of these centrality indices share a great deal of topological information, which range from 75% to 96%. A generalized keystone indicator is then proposed by considering the factor loadings of the six-centrality measures, which contains most of the information encoded by these indices. However, the individual ordering of species according to these criteria display significant differences in most food webs. We simulate the effects of species extinction by removing species ranked according to a local and a "meso-scale" centrality indicator. The differences observed on three network characteristics – size, average distance and clustering coefficient of the largest component – after the removal of the most central nodes indicate that the consideration of these indices have different impacts for the ranking of species with conservational biology purposes. The "meso-scale" indicator appears to play an important role in determining the relative importance of species in epidemic spread and parasitism rates.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Food webs are complex networks representing the trophic relationships of species in a natural ecosystem (Pimm, 1982; Proulx et al., 2005; Jordán and Scheuring, 2004). In these networks species are represented by nodes and trophic relationships are represented by network links. It has been believed for long time that all species are not equally important in an ecosystem. In fact, there are some species "whose impact on its community or ecosystem is large, and disproportionately large relative to its abundance" (Power et al., 1996). These species receive the denomination of keystones, and their extinction can lead to community collapses which make their investigation of great importance in conservation ecology (Paine, 1966; Estes and Palmisano, 1974; Mills et al., 1993; Menge et al., 1994; Power et al., 1996; Springer et al., 2003). The identification of keystones species in a food web is both an important and elusive task for ecologists. Consequently, several measures for *a priori* identification of keystones have been proposed (Power et al., 1996; Christianou and Ebenman, 2005; Jordán et al., 1999, 2006; Jordán and Scheuring, 2002).

In recent years there have been important discoveries of certain network topological properties which are common to complex systems in a wide range of disciplines ranging from

Internet to cellular networks (Strogatz, 2001; Albert and Barabási, 2002; Newman, 2003). These results have also impacted the study of keystones identification. For instance, it has been established that the loss of highly connected species in a food web has the most serious consequences for the persistence of the network (Solé and Montoya, 2001; Dunne et al., 2002a). This has conducted to the acceptance of node degree as the most widely used index of positional importance in a food web (Solé and Montoya, 2001; Dunne et al., 2002a). Node degree is the number of species having trophic relationships with a particular species. However, node degree represents the most local index characterizing the centrality of a node in a network which excludes the long-range interactions arising in the network (Jordán et al., 2006). These indirect effects have been recognized as relevant for understanding the functioning of ecological systems (Wooton, 1994). In general the notion of centrality comes from its use in social networks (Wasserman and Faust, 1994). Intuitively, it is related to the ability of a node to communicate directly with other nodes, or to its closeness to many other nodes or to the quantity of pairs of nodes which need a specific node as intermediary in their communications (Gómez et al., 2003). These ideas have materialised in some well-known centrality measures such as degree centrality (DC), closeness centrality (CC), and betweenness centrality (BC) (Wasserman and Faust, 1994; Freeman, 1979). Other centrality measures, known as eigenvector centrality (EC) and information centrality (IC), were developed by Bonacich (1972, 1987) and Stephenson and Zelen (1989), respectively.

On the other hand, some attempts have been made to account for all indirect effects in a food web by studying some global properties of these networks, such as their "small-worldness", or "scale-freeness" (Camacho et al., 2002; Montoya and Solé, 2002; Williams et al., 2002; Dunne et al., 2002b). The first is based on the global characteristics of a network of being highly clustered, as measured by means of the clustering coefficient, and having small average distance among every pair of species (Watts and Strogatz, 1998). The second is the global statistical property characterizing the link distribution of nodes (Barabási and Albert, 1999).

An intermediate approach – neither local nor global – has been suggested, which is based on a "meso-scale" view of the positional importance of species (Jordán and Scheuring, 2002; Jordán et al., 2006). According to these authors "considering either only direct partners or the whole community is not the key to assessing the importance of species". The "meso-scale" view should takes into account that the strength of indirect effects decreases with the length of the pathway in such a way that an index accounting for species importance should be between the local and the global scales reflecting "far reaching effects but only to smaller and smaller extent". This "meso-scale" view has been considered in several approaches for the study of topological characteristics of food webs (Melián and Bascompte, 2002; Jain and Krishna, 2002; Garlaschelli et al., 2003; Luczkovich et al., 2003; Allesina and Bondini, 2005). We have recently introduced a centrality index in the context of complex networks which fulfill the requirements to be considered as a "meso-scale" index of positional importance of a node (Estrada and Rodríguez-Velázquez, 2005a). This index has resulted very efficient in identifying relevant nodes

in different networks, such as authors in collaboration networks (Estrada and Rodríguez-Velázquez, 2005a) or essential proteins in protein interaction networks (Estrada, 2006).

Here we study several centrality measures which can be grouped as local or global as well as our "meso-scale" centrality index. These centrality indices characterize the positional importance of a species in a food web according to different topological characteristics. We study these indices in ranking species in 17 food webs representing a wide range of species numbers, linkage densities, taxa, and habitat types, which include lakes, ponds, streams, estuaries, marine and terrestrial ecosystems (temperate, desert and subtropical). We show some of the similarities and differences among these indices as well as their possibilities for the a priori identification of keystone species. A generalized centrality index is derived by means of factor analysis which resumes the information contained in the six-centrality measures analyzed here. Finally, we analyze the effects of removing keystone species identified by our "meso-scale" centrality index and by the degree centrality.

2.   Local and global centralities

We will consider two local centrality measures, the degree centrality and betweenness centrality. Degree centrality, $DC(i)$, is simply the number of ties incident upon a node i, i.e., the number of species that have trophic relationships with species i. Betweenness centrality (BC) is defined as the fraction of shortest paths going through a given node. If $\rho(i, j)$ is the number of shortest paths from species i to j, and $\rho(i, k, j)$ is the number of these shortest paths that pass through species k in the food web, then the betweenness centrality of node k is given by (Wasserman and Faust, 1994; Freeman, 1979):

$$BC(k) = \sum_i \sum_j \frac{\rho(i, k, j)}{\rho(i, j)}, \quad i \neq j \neq k \qquad (1)$$

There are three global centrality measures that will be considered here. The closeness centrality of species i, $CC(i)$, is the sum of topological distances from all other species in the food web, where the distance $d(i, j)$ from one species i to another j is defined as the number of links in the shortest path from one to the other. The closeness centrality of species i in a food web is given by the following expression (Wasserman and Faust, 1994; Freeman, 1979):

$$CC(i) = \frac{N - 1}{\sum_j d(i, j)} \qquad (2)$$

The eigenvector centrality (EC) introduced by Bonacich (1972, 1987) is defined as the principal eigenvector of the adjacency matrix $\mathbf{A}$ defining the network. The defining equation of an eigenvector is $\lambda \mathbf{e} = \mathbf{A}\mathbf{e}$, where $\mathbf{A}$ is the adjacency matrix of the food web, λ is an eigenvalue and $\mathbf{e}$ is the eigenvector. Thus, EC of species i is defined as the ith component of the eigenvector $\mathbf{e}_1$, $\mathbf{e}_1(i)$, that corresponds to the largest eigenvalue of $\mathbf{A}$, λ_1 (principal eigenvalue or index):

$$EC(i) = \mathbf{e}_1(i) \qquad (3)$$

Another global structural measure of centrality in a network was introduced by Stephenson and Zelen (1989) and is known as information centrality (IC). It is based on the information that can be transmitted between any two points in a connected network. If $\mathbf{A}$ is the adjacency matrix of a network, $\mathbf{D}$ a diagonal matrix of the degree of each node and $\mathbf{J}$ is a matrix with all its elements equal to one, then IC is defined by inverting the matrix $\mathbf{B}$ defined as $\mathbf{B} = \mathbf{D} - \mathbf{A} + \mathbf{J}$ in order to obtain the matrix $\mathbf{C} = (c_{ij}) = \mathbf{B}^{-1}$ from which the information matrix is obtained as follows:

$$I_{ij} = (c_{ii} + c_{jj} - c_{ij})^{-1} \tag{4}$$

The information centrality of the species i is then defined by using the harmonic average:

$$IC(i) = \left[\frac{1}{N} \sum_j \frac{1}{I_{ij}} \right]^{-1} \tag{5}$$

Stephenson and Zelen (1989) proposed to define I_{ii} as infinite for computational purposes, which makes $1/I_{ii} = 0$.

3. A "meso-scale" centrality measure

We have recently introduced an index characterizing the importance of a node in a network which fulfills the requirements for being considered a "meso-scale" index of positional importance of trophic groups in a food web. This index is named the "subgraph centrality" of a node because it characterizes the participation of such node in all subgraphs existing in the network. The "meso-scale" nature of this index is guarantied by the fact that smaller subgraphs receive more weight than larger subgraphs. The participation of a node in a subgraph is quantified by means of the so-called closed-walks (CWs). A walk of length r is a sequence of (not necessarily different) vertices $v_1, v_2, \cdots, v_r, v_{r+1}$ such that for each $i = 1, 2, \ldots, r$ there is a link from v_i to v_{i+1}. A closed walk (CW) is a walk in which $v_{r+1} = v_1$. A particular case of CW is the cycle, in which all vertices in $v_1, v_2, \cdots, v_r, v_{r+1}$ are different. Every CW has a subgraph associated to it. For instance, the CW of length two v_1, v_2, v_1 is associated with the link v_1, v_2 and the CW of length three v_1, v_2, v_3, v_1 is associated with the triangle v_1, v_2, v_3. CWs of length four can be associated with different subgraphs. For instance, v_1, v_2, v_1, v_2, v_1 is associated with the link v_1, v_2; v_1, v_2, v_3, v_2, v_1 is associated with the triple v_1, v_2, v_3 and v_1, v_2, v_3, v_4, v_1 is associated with the square v_1, v_2, v_3, v_4.

If we consider a particular node of the food web, i, the total number of CWs of length r starting (and ending) at i is designated by $\mu_r(i)$. In order to guaranty the "meso-scale" characteristics of our measure we give a weight to $\mu_r(i)$ in decreasing order of their length. For instance, a CW of length two is weighted by a factor of 1/2, a CW of length three is weighted by a factor of 1/6, and in general a CW of length r is weighted by a factor of $1/r!$ Then, the weighted participation of the node i in CWs of all lengths in the food web is given by

$$SC(i) = \sum_{r=0}^{\infty} \frac{\mu_r(i)}{r!} \tag{6}$$

Using spectral graph theory we have shown (Estrada and Rodríguez-Velázquez, 2005a) that SC(i) can be obtained from the eigenvalues and eigenvectors of the adjacency matrix of the network:

$$SC(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 e^{\lambda_j} \tag{7}$$

where $\gamma_j(i)$ is the ith component of the jth eigenvector of the adjacency matrix $\mathbf{A}$ and λ_j is the corresponding jth eigenvalue (Estrada and Rodríguez-Velázquez, 2005a). It should be noted that SC(i) counts all CWs in the network, which can be of even or odd length. CWs of even length can move back and forth in acyclic subgraphs, i.e., those that do not contain cycles, while odd CWs do not contain contributions from acyclic subgraphs. Consequently, SC(i) can be divided into two terms accounting for even and odd CWs in the food web (Estrada and Rodríguez-Velázquez, 2005b):

$$SC(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 \cosh \lambda_j + \sum_{j=1}^{N} [\gamma_j(i)]^2 \sinh \lambda_j$$
$$= SC_{even}(i) + SC_{odd}(i) \tag{8}$$

where the term $SC_{odd}(i)$ only accounts for subgraphs containing at least one odd cycle and can be considered as a meso-scale property of order in networks that characterize the odd-cyclic wiring of a typical neighborhood, which will be named the odd subgraph centrality of the ith species in the food web:

$$SC_{odd}(i) = \sum_{j=1}^{N} [\gamma_j(i)]^2 \sinh \lambda_j \tag{9}$$

Of course, $SC_{odd}(i)$ can be expressed in terms of the number of CWs of odd length using the following expression:

$$SC_{odd}(i) = \sum_{r=1}^{\infty} \frac{\mu_{2r}(i)}{(2r)!} \tag{10}$$

which is a clear sign of the meso-scale nature of this index because indicates that the shortest length CWs receive higher weights that the longest ones.

4. Food web datasets

The following are the names and description of the food webs studied including the number of species (network nodes), N, and trophic relationships (network links), E. Benguela ($N = 29$, $E = 191$) represents a marine ecosystem of Benguela off the southwest coast of South Africa (Yodzis, 1998). Bridge Brook Lake ($N = 75$, $E = 542$) represents pelagic species from the largest of a set of 50 New York Adirondack lake food webs (Havens, 1992). Canton Creek ($N = 108$, $E = 707$) is formed primarily by invertebrates and algae in a tributary of the Taieri River, surrounded by pasture, in the South Island of New Zealand (Townsend et al., 1998). Chesapeake Bay ($N = 33$, $E = 71$) represents the pelagic portion of an eastern US estuary, with an emphasis on larger fishes (Baird and Ulanowicz, 1989). Coachella Valley ($N = 30$, $E = 241$) is formed by a wide range of

highly aggregated taxa from the Coachella Valley desert in southern California (Polis, 1991). El Verde ($N = 156$, $E = 1439$) represents insects, spiders, birds, reptiles and amphibians in a rainforest in Puerto Rico (Waide and Reagan, 1996). Grassland ($N = 75$, $E = 113$) represents all vascular plants and all insects and trophic interactions found inside stems of plants collected from 24 sites distributed within England and Wales (Martinez et al., 1999). Little Rock Lake ($N = 181$, $E = 2318$) is formed by pelagic and benthic species, particularly fishes, zooplankton, macro-invertebrates, and algae of the Little Rock Lake, Wisconsin, US (Martinez, 1991). Reef Small ($N = 50$, $E = 503$) is a Caribbean coral reef ecosystem from the Puerto Rico-Virgin Island shelf complex (Opitz, 1996). Scotch Broom ($N = 154$, $E = 366$) contains trophic interactions between the herbivores, parasitoids, predators and pathogens associated with broom, *Cytisus scoparius*, collected in Silwood Park, Berkshire, England, UK (Memmott et al., 2000). Shelf ($N = 81$, $E = 1451$), which is a marine ecosystem on the northeast US shelf (Link, 2002). Skipwith Pond ($N = 35$, $E = 353$) represents invertebrates in an English pond (Warren, 1989). St. Marks Seagrass ($N = 48$, $E = 218$) is formed mostly by macroinvertebrates, fishes, and birds associated with an estuarine seagrass community, *Halodule wrightii*, at St. Marks Refuge in Florida (Christian and Luczkovich, 1999). St. Martin Island ($N = 44$, $E = 218$) represents interactions between birds and predators and arthropod prey of *Anolis* lizards on the island of St. Martin, which is located in the northern Lesser Antilles (Goldwasser and Roughgarden, 1993). Stony Stream ($N = 112$, $E = 830$) is similar to Canton Creek but in native tussock habitat (Townsend et al., 1998). Ythan1 ($N = 134$, $E = 593$) is formed mostly by birds, fishes, invertebrates, and metazoan parasites in a Scottish Estuary (Huxham et al., 1996) and Ythan2 ($N = 92$, $E = 416$) is a reduced version of Ythan1 with no parasites (Hall and Raffaelli, 1991).

5. Results and discussion

We calculated six different centrality indices for each of the nodes in the 17 food webs studied. In general, we observed that there are significant differences in the ranking of species by these different centralities. These divergences might be produced by the differences of scale used in the definition of centrality measures. In other words, these centrality indices characterize the positional importance of a species at different scales, e.g., based on local, global or "meso-scale" characteristics. On the one hand, degree and betweenness centralities rank nodes in a network according to their local interactions. DC considers the immediate effects of the closest neighbors to the species analyzed, and BC characterizes the degree of influence a species has in "communicating" between species pairs. On the other hand, CC, EC and IC take into account global network properties in ranking species in a network. According to the definition of EC this index simulates a mechanism in which each node affects all of its neighbours simultaneously (Borgatti, 2005). Accordingly, a species is considered central if it is adjacent to other species that themselves have high scores. In such a way, this centrality measure is a sort of degree which takes into account the "importance" of the other nodes in the network. In CC the global effect of all nodes in the network is accounted for by

means of the sum of all distances separating a node from the rest. In a similar way IC has been recognized as another closeness measure, which in essence measures the harmonic mean lengths of paths ending at a node i, which is smaller if i has many short paths connecting it to other nodes. In all these global indices the short- and long-range interactions receive the same weight, which make them substantially different from SC_{odd}. In this centrality measure both the short- and long-range effects are taken into account using different scales for accounting them, which merits the denomination of "meso-scale" for this index. Consequently, these different rankings should not be considered as a drawback of centrality measures but as the consideration of different topological aspects of the positional importance of species in a food web. These different rankings can be relevant for the analysis of distinct goals in theoretical ecology and conservation biology.

It was previously shown that strong correlations exist among different centrality measures in complex networks (Wutchy and Stadler, 2003). This is not surprising because these measures are defined so as to account for the notion of centrality of the nodes in the graph. In consequence, nodes with large degrees show in general short average distance to the other nodes in the network, which produces high correlations between node degrees and various measures of centrality. Nodes with large degrees are also expected to participate in large number of subgraphs, such as simply connected triplets, triangles, squares and so forth. Here we have observed that, in general, odd subgraph centrality yields the highest rank orders for those nodes of largest degrees in the network, despite the fact that both measures disagree very significantly for the majority of other nodes. This effect produces high correlations between SC_{odd} and DC for most of the food webs studied here. There are also high intercorrelations among the other pairs of centrality measures for these food webs. In Fig. 1 we illustrate these correlations for all pairs of centralities in the Chesapeake Bay food web. The exception in the majority of food webs is given by the betweenness centrality, which shows poor correlations with all the other indices. The best correlations obtained for BC are, in general, obtained with the degree, which agrees with our previous classification of these indices as "local-scale" ones.

We have further investigated the intercorrelation between these centrality measures for the 17 food webs studied here by using factor analysis with the principal component method (Rummel, 1970). The principal goal of this factor analysis is to simplify the structure of multidimensional data of centrality measures by reducing the dimensionality of the centrality space in order to see what basic features are behind these measures. In general, only one factor was necessary for explaining most of the variance contained in the six-centrality measures, which is a direct consequence of their collinearity. Most of the variables show high "factor loadings" in this principal component (PC1) indicating that centralities are highly "loaded" in this unique component. In Table 1 we give the values of the factor loading of each centrality, and the percentage of variance explained by the component extracted. It can be seen that the lowest loading is always obtained for BC, which is the least collinear variable of all centrality measures. This principal component is dominated by the DC, which shows the highest loadings in most of food webs. The average
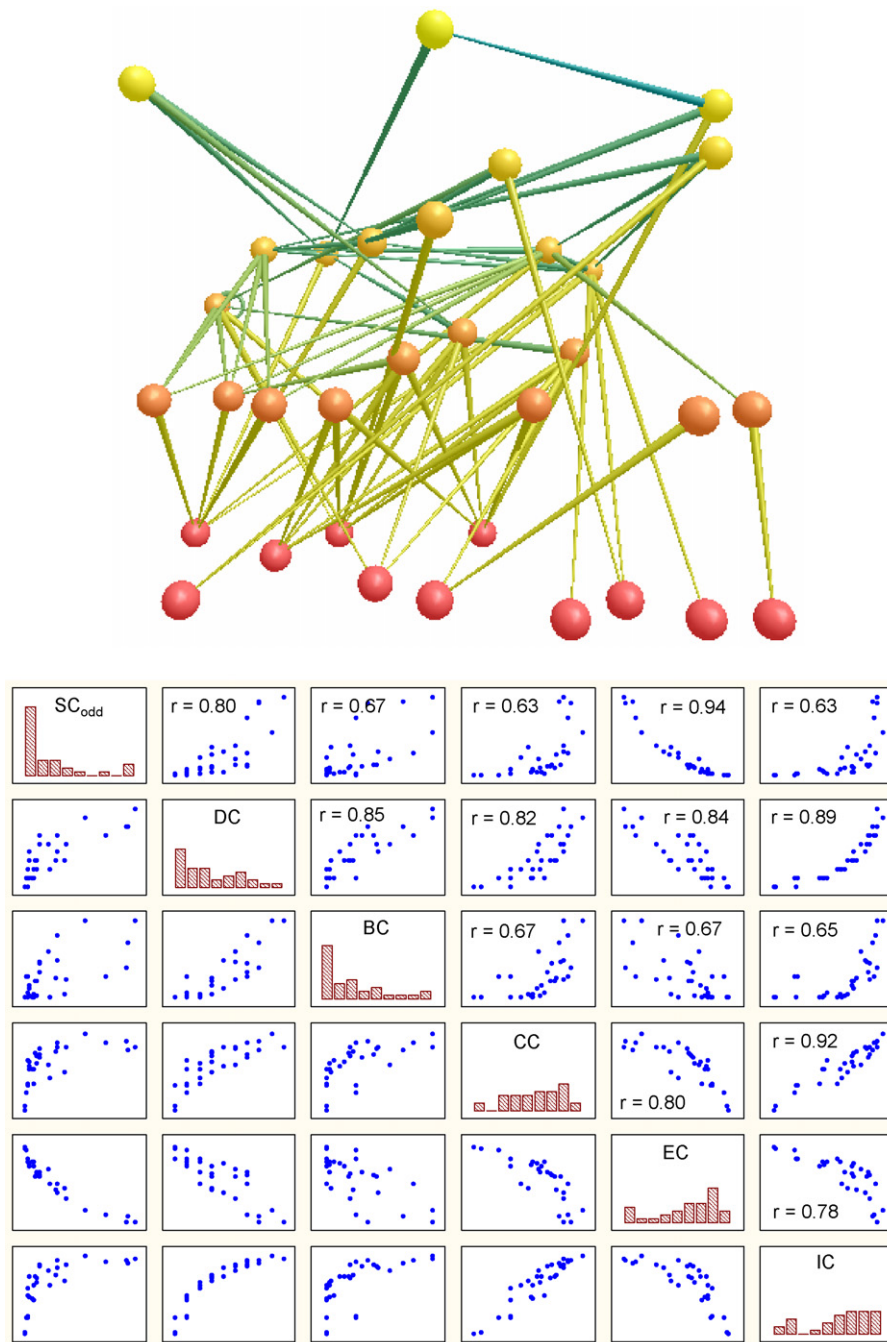
Fig. 1 – Intercorrelation between the six-centrality measures studied in this work for the Chesapeake Bay food web shown at the top. Correlation coefficients for every pair of variables are also given. The image of the food web was produced with FoodWeb3D, written by R.J. Williams and provided by the Pacific Ecoinformatics and Computational Ecology Lab (http://www.foodwebs.org).

correlation coefficients between all centrality measures are also illustrated in Table 1. It can be seen the existence of two different groups of food webs. On the one hand, there are networks in which the multicollinearity between all centrality measures is high, e.g., average correlation coefficient >0.80. On the other hand, there are networks with low intercorrelation between centrality measures, e.g. average correlation coefficient <0.80. In the first group of networks there is a general coincidence in the rankings produced by the different centrality criteria studied here, which in some way give certain homogeneity to these food webs. However, food webs in the second group are heterogeneous respect to the positional importance displayed by species according to different criteria. In general, in these networks high (low) connectivity of a node does not necessarily implies high (low) cliquishness, betweenness or closeness to other species. This means that only a partial ordering (Jensen et al., 2003) of the species is possible in these food webs.

Table 1 – Results of the factor analysis carried out to the six-centrality measures in the 17 food webs studied					
	Benguela	Bridge brook	Canton	Chesapeake	Coachella	El Verde
SC_{odd}	−0.9529	−0.9260	−0.9362	−0.8643	−0.9703	−0.9344
DC	−0.9965	−0.9794	−0.9912	−0.9632	−0.9980	−0.9869
BC	−0.5505	−0.7907	−0.7582	−0.8329	−0.5908	−0.7505
CC	−0.9874	−0.9442	−0.9487	−0.9000	−0.9913	−0.9466
EC	−0.9775	−0.9482	−0.9750	−0.9324	−0.9796	−0.9752
IC	−0.9423	−0.9292	−0.8737	−0.9060	−0.9711	−0.8519
% Var	83.7	84.9	84.1	81.1	86.2	83.1
Average r	0.78	0.82	0.80	0.77	0.81	0.79
	Grassland	Little rock	Reef Small	Scotch broom	Shelf	Skipwith
SC_{odd}	−0.8995	−0.9193	−0.9793	−0.8997	−0.9888	−0.9828
DC	−0.9121	−0.9872	−0.9977	−0.9489	−0.9982	−0.9992
BC	−0.8459	−0.6523	−0.8533	−0.6091	−0.8930	−0.9317
CC	−0.8669	−0.9602	−0.9979	−0.9200	−0.9960	−0.9975
EC	−0.9350	−0.9639	−0.9863	−0.9456	−0.9862	−0.9789
IC	−0.8667	−0.8501	−0.9454	−0.8516	−0.9598	−0.9866
% Var	78.9	80.3	92.4	75.8	94.3	96.0
Average r	0.75	0.75	0.91	0.69	0.93	0.95
	St. Marks		St. Martin	Stony	Ythan 1	Ythan 2
SC_{odd}	−0.9577		−0.9620	−0.9643	−0.9509	−0.9565
DC	−0.9975		−0.9923	−0.9906	−0.9871	−0.9889
BC	−0.9072		−0.8163	−0.8265	−0.8262	−0.8513
CC	−0.9871		−0.9739	−0.9447	−0.9333	−0.9507
EC	−0.9833		−0.9838	−0.9804	−0.9755	−0.9815
IC	−0.9552		−0.8739	−0.8936	−0.8316	−0.8542
% Var	93.2		87.6	87.4	84.6	86.9
Average r	0.92		0.85	0.85	0.81	0.84

The values of the factor loading of each centrality, the percentage of variance explained by the component extracted, and the average correlation coefficient between all centralities are given. Factor loadings ≥0.80 are boldfaced.

The use of factor analysis permits to transform the multidimensional space of centrality measures into a reduced space of principal components, which are linear combinations of the centrality indices. Thus, the "factor scores" obtained from the factor analysis represent a new variable which contains the information encoded in the six-centrality measures. These factor scores are linear combinations of the centrality measures previously normalized to mean zero and standard deviation of one. For instance, for Chesapeake Bay the factor score of each species can be obtained by means of the following equation:

$$PC1 = -0.1775SC_{odd} - 0.1978DC - 0.1711BC - 0.1849CC$$
$$- 0.1915EC - 0.1861IC$$

This new variable constitutes a sort of generalized centrality and can be useful for characterizing the positional importance of species in a food web. We will call this measure the generalized keystone index and will be used to rank species in a food web. In Fig. 2 we illustrate the ordering of species produced by the six-centrality measures in the Chesapeake Bay food web. Each species is represented by a different number whose names are given in the caption of this figure. For instance, species 32 (Summer flounder) is ranked as the most central one by SC_{odd}, DC, BC and EC, while species 15 (Rag worm) is ranked as number one by CC and IC. Species 31

(Weakfish) and 12 (Softshell clam) are ranked among the top four species by all centralities. The ranking produced by PC1 is illustrated at the top of this graphic, where it can be seen that species 32 is the most central, followed by 15, 31 and 12, in this order.

At the top of Fig. 2 we illustrate the distribution of species in Chesapeake Bay food web according to their positional importance determined by two centrality measures, SC_{odd} and DC, and by the generalized keystone index. The larger the area of the circle representing the node the most central the species. It was previously reported (Dunne et al., 2002b) that this food web displays an exponential distribution of node degrees but in Fig. 3 we show that there is a perfect fit with a uniform distribution of node degrees. However, the cumulative distribution of SC_{odd} clearly shows a power-law dependence in which only very few species participate in a very large number of subgraphs while most of species participate in only a few subgraphs.

In general, there is great coincidence in the distributions of node degrees and odd subgraph centralities for these food webs. For instance, Benguela, Bridge Brook, Coachella, Reef Small, Shelf, Skipwith, and Stony stream all display uniform distributions for both centrality measures. On the hand, Canton Creek, El Verde rainforest, Little Rock lake, St. Martin island, and Ythan estuary 1 all display exponential distributions for both degree and odd subgraph centralities. These
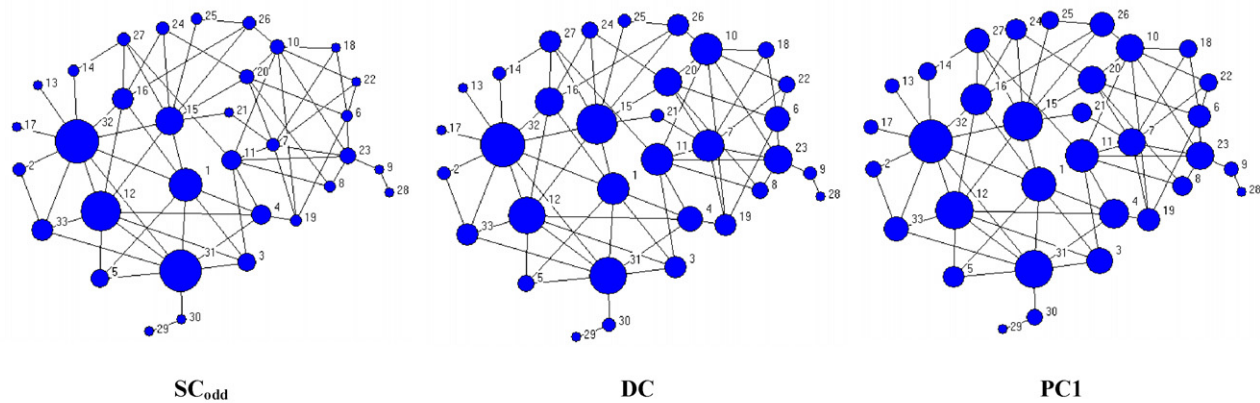
PC1	32	15	31	12	1	11	16	4	10	20	7	23	3	27	26	33	19	6	5	24	8	21	14	25	18	22	2	30	9	13	17	29	28
IC	15	32	12	11	1	31	16	10	7	20	4	23	3	6	27	19	26	33	24	5	8	18	22	21	14	25	2	30	13	17	9	29	28
EC	32	12	31	1	15	16	4	3	33	11	5	20	27	10	26	7	24	23	19	2	14	6	25	8	21	13	17	18	22	30	9	29	28
CC	15	12	11	1	32	4	16	20	31	3	27	10	26	7	21	19	24	23	6	8	25	14	33	5	18	22	2	30	13	17	9	29	28
BC	32	15	11	31	23	4	20	10	12	7	1	30	9	3	16	26	27	19	6	21	33	8	24	14	18	22	5	2	25	13	17	29	28
DC	32	15	31	12	1	11	10	7	16	23	20	4	6	33	3	26	27	19	5	24	8	18	22	2	25	14	30	21	9	13	17	29	28
SC_{odd}	32	31	12	1	15	16	33	4	11	3	5	23	10	20	7	26	24	27	2	8	25	6	14	19	30	13	17	21	18	22	9	29	28

Fig. 2 – Ranking of species in the Chesapeake Bay food web according to the six different centrality measures used in this work. At the top the relative importance of every species in the web according to three criteria are represented in graphical form. The names of species are as follows: 1 = Phytoplankton, 2 = bacteria attached to suspended POM, 3 = bacteria attached to sediment POM, 4 = Benthic Diatoms, 5 = free bacteria in water column/DOC, 6 = heterotrophic microflagellates, 7 = microzooplankton, 8 = zooplankton, 9 = Ctenophores, 10 = Sea Nettles, 11 = other suspension feeders, 12 = Softshell Clam, 13 = American Oyster, 14 = other Polychaetes, 15 = Nereis sp. (Rag worm), 16 = Macoma spp. (bivalves), 17 = Meiofauna, 18 = Crustacean deposit feeders, 19 = Blue Crab, 20 = fish larvae, 21 = Alewife and blue herring, 22 = Bay Anchovy, 23 = Atlantic menhaden, 24 = American shad, 25 = Atlantic Croaker, 26 = Hogchoker, 27 = Spot, 28 = White perch, 29 = Sea Catfish, 30 = Bluefish, 31 = Weakfish, 32 = Summer Flounder, 33 = Striped Bass.

coincidences are extended to Scotch Broom and Grassland which displays partial power-law and power-law distribution for both centralities. These networks have been identified as "source webs" constructed by following food chains upward from one or a few basal species and ignoring co-occurring generalists, which explain the power-law distributions observed (Dunne et al., 2002b). The exceptions for this rule apart from Chesapeake Bay are St. Marks, which displays uniform distribution for degree and exponential for SC_{odd}, and Ythan estuary 2, which display exponential distribution for node degrees and a power-law distribution for the odd subgraph centrality.

From the perspective of the analysis of the positional importance of species in a food web the differences in the ranking of species introduced by the different centrality indices appear to be the most relevant ones. In those food webs with large intercorrelation between the centrality indices it is clear and effective to select the most central species according to any centrality as the most relevant keystones in the web. However, in those food webs with relatively low collinearity between the different centrality measures the situation is substantially different. If several centrality measures point to different species as the most central one the question of which species is absolutely the most central arises. A possible solution has been given here by considering a generalized centrality index that encompass all

the measures analyzed by means of the extraction of a principal component containing most of the information encoded in the separated indices. Another possible solution is to use partial ordering rankings (Jensen et al., 2003) to determine the partial order of species in the food web. In any case we have to take into account that different rankings not necessarily are an indication of failures in the theoretical approaches used to rank the species in the food web. These differences might clearly indicate important structural differences between species, which will affect in different ways the structure and function of the ecosystem. In other words, the extinction of one species classified as the most central by one indicator will affect the ecosystem in a different way that the extinction of another species ranked as the most central by another measure. As an illustrative example we have selected the Little Rock lake food web, which is the largest one of all analyzed here and displays a relatively low intercorrelation between centralities (see Table 1). We have ranked all the species in this food web according to their values of DC, CC and SC_{odd}. These centralities represent examples of local, global and "meso-scale" keystone indicators. Then, we have removed the top 5, 10, 15, and so forth species according to each ranking. We have analyzed the effects of species extinction on the whole ecosystem by considering three global indicators of the robustness of the web: the size of the largest component, the average distance between species in
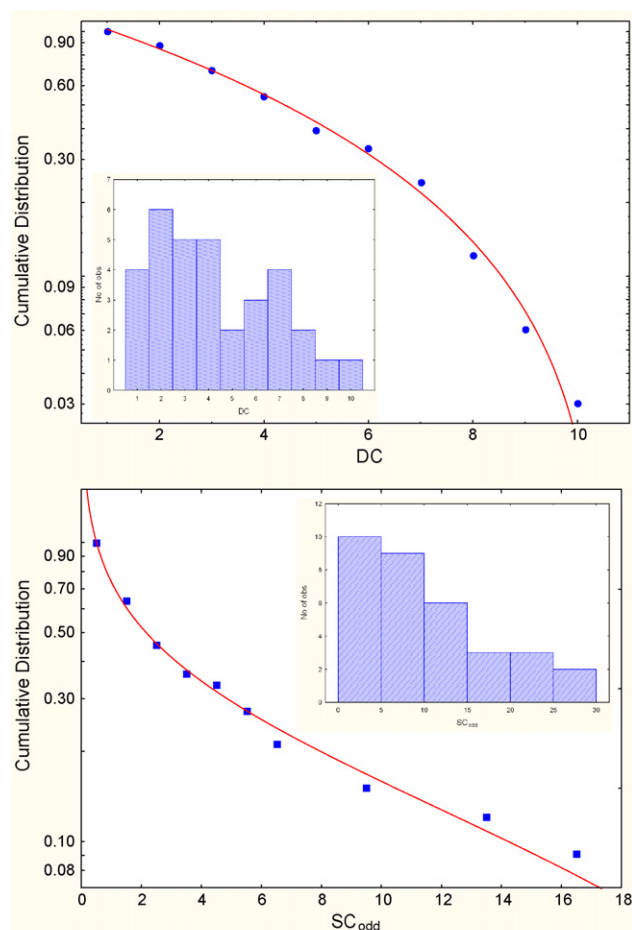
Fig. 3 – Illustration of the linear-log plots of the cumulative distribution of links per species (DC) and participation in weighted subgraphs per species (SC_{odd}). DC displays a uniform distribution characterized by a downward curved line. SC_{odd} displays a typical power-law distribution characterized by an upward curved line.
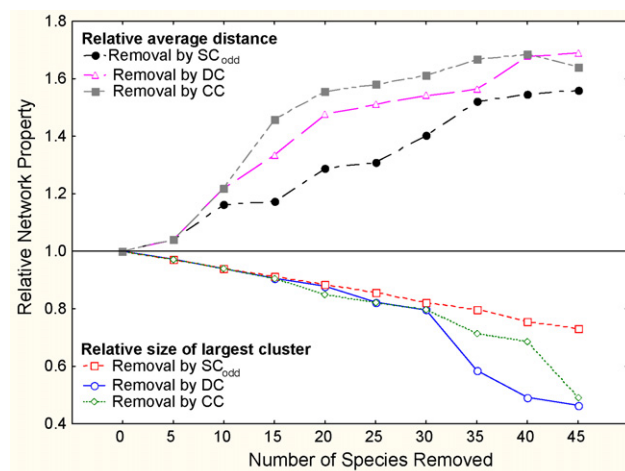

Fig. 4 – Simulation of the effects of species extinction in the Little Rock Lake food web on two different network properties: size of the largest remaining component and the average distance between species in such component. The removal of species is carried out by selecting the top 5, 10, 15, etc. species ranked according to three centrality measures: DC, CC and SC_{odd}.

such component and the clustering coefficient of the nodes in this cluster. The removal of the most connected species changes more significantly the relative size of the largest component as well as the average distance between species in such component, followed very closely by CC. In Fig. 4 we illustrate these changes on network topology produced by the removal of species according to these three criteria. However, when another network parameter – clustering coefficient – is considered the situation is reversed. The clustering coefficient of a species measures the proportion of the neighbors of that species which are adjacent to each other (Watts and Strogatz, 1998). In Fig. 5 we illustrate how the clustering coefficient drops dramatically after the removal of both the most connected species and those having the largest values of SC_{odd}. After the removal of the first 30 nodes the components obtained by removing species according to SC_{odd} show about the half clustering that those obtained by removing nodes according to DC. The clustering coefficient drops from 0.352 to about 0.01 when the nodes are removed according to their rankings based on SC_{odd}. This is a clustering 10 times larger

than the one obtained for the component remaining when species are removed according to CC. The consequences of these two types of network damages can be significantly different. For instance, it was found that the average path length is positively correlated with the time to reach an epidemic size as well as with the fastest rate of infection (Shirley and Rushton, 2005). This means that when species are removed by any of the three criteria analyzed here an epidemic can spread slowest and reaching shortest size than
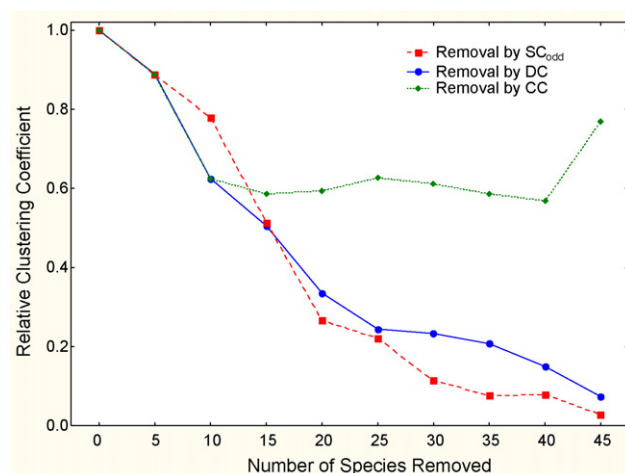

Fig. 5 – Simulation of the effects of species extinction in the Little Rock Lake food web on the cliquishness of the remaining largest component of the web. The removal of species is carried out by selecting the top 5, 10, 15, etc. species ranked according to three centrality measures: DC, CC and SC_{odd}, and the cliquishness is measured by the Watts–Strogatz clustering coefficient.

in the original network, which has lowest average path length. However, we can see in Fig. 4 that removal of species by SC_{odd} ordering produces networks with lower average path lengths than those obtained by removal according to DC or CC. Consequently, an epidemic spreads fastest and reach largest size in the component remaining after the removal of the most central nodes ranked according to SC_{odd}. Shirley and Rushton (2005) also observed that the clustering coefficient is significantly correlated with both disease spread parameters but unfortunately the correlations were not shown to extract any conclusion here (Shirley and Rushton, 2005). On the other hand, the total parasitism rates correlates negatively with the clustering coefficient of the food web (Montoya et al., 2003). In other words, a decrease of the clustering coefficient increases the percentage of parasitism rate in the web. In consequence, the components generated by the removal of nodes according to SC_{odd} are more and more susceptible to parasitism and this should happen in a greater proportion than in the components generated by DC and CC removal criteria. As can be seen at least on the basis of these two criteria ranking of species by the "meso-scale" index SC_{odd} can play an important role for understanding theoretical aspects of food web functioning as well as for conservationist purposes.

Finally, it is important to remark some characteristics of the index SC_{odd} that might be relevant for the study of food webs. First, this index accounts for the weighted participation of a species in all sub-structures of the web. This makes SC_{odd} more appropriated than other clustering indices for measuring food web complexity. While the clustering coefficient only measures the implication of a node in triangles, SC_{odd} takes into account all other sub-structures. In other words, while the clustering coefficient is a local index, SC_{odd} is a meso-scale parameter. Furthermore, SC_{odd} can also be applied to weighted networks (Estrada et al., 2006). This means that we can use this centrality measure when species and/or trophic interactions are weighted by some kind of numeric parameter. This permits to obtain information about more realistic situations in which interaction frequency, carbon flow, interaction strength, or other ecosystem parameters are considered. This situation is difficult or even impossible to be considered when other centrality measures are taken into account as some of the ones analyzed in this work.

Acknowledgements

REFERENCES

Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev Mod. Phys. 74, 47–97.

Allesina, S., Bondini, A., 2005. Food web networks: scaling relation revisited. Ecol. Complex. 2, 323–338.

Baird, D., Ulanowicz, R.E., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. Ecol. Monogr. 59, 329–364.

Barabási, A.-L., Albert, R., 1999. Emergente of scaling in random networks. Science 286, 509–512.

Bonacich, P., 1972. Factoring and weighting approach to clique identification. J. Math. Sociol. 2, 113–120.

Bonacich, P., 1987. Power and centrality: a family of measures. Am. J. Sociol. 92, 1170–1182.

Borgatti, S.P., 2005. Centrality and network flow. Social Networks 27, 55–71.

Camacho, J., Guimerá, R., Amaral, L.A.N., 2002. Robust patterns in food web structure. Phys. Rev. Lett. 88, 228102.

Christian, R.R., Luczkovich, J.J., 1999. Organization and understanding a winter's seagrass foodweb network through effective trophic levels. Ecol. Mod. 117, 99–124.

Christianou, M., Ebenman, B., 2005. Keystone species and vulnerable species in ecological communities: string and weak interactors? J. Theor. Biol. 235, 95–103.

Dunne, J.A., Williams, R.J., Martinez, N.D., 2002a. Network robustness and biodiversity loss in food webs: robustness increases with connectance. Ecol. Lett. 5, 558–567.

Dunne, J.A., Williams, R.J., Martinez, N.D., 2002b. Food-web structure and network theory: the role of connectance and size. Proc. Natl. Acad. Sci. U.S.A. 99, 12917–12922.

Estes, J.A., Palmisano, J.F., 1974. Sea otters: their role in structuring nearshore communities. Science 185, 1058–1060.

Estrada, E., Rodríguez-Velázquez, J.A., 2005a. Subgraph centrality in complex networks. Phys. Rev. E 71, 056103.

Estrada, E., Rodríguez-Velázquez, J.A., 2005b. Spectral measures of bipartivity in complex networks. Phys. Rev. E 72, 046105.

Estrada, E., 2006. Virtual identification of essential proteins within the protein interaction network of yeast. Proteomics 6, 35–40.

Estrada, E., Rodríguez-Velázquez, J.A., Randić, M., 2006. Atomic branching in molecules. Int. J. Quant. Chem. 106, 823–832.

Freeman, L.C., 1979. Centrality in social networks. I. Conceptual clarification. Social Networks 1, 215–239.

Garlaschelli, D., Caldarelli, G., Pietronero, L., 2003. Universal scaling relations in food webs. Nature 423, 165–168.

Goldwasser, L., Roughgarden, J.A., 1993. Construction of a large Caribbean food web. Ecology 74, 1216–1233.

Gómez, D., González-Arangüena, E., Manuel, C., Owen, G., del Pozo, M., Tejada, J., 2003. Centrality and power in social networks: a game theoretic approach. Math. Soc. Sci. 46, 27–54.

Hall, S.J., Raffaelli, D., 1991. Food web patterns: lessons from a species-rich web. J Anim. Ecol. 60, 823–842.

Havens, K., 1992. Scale and structure in natural food webs. Science 257, 1107–1109.

Huxham, M., Beany, S., Raffaelli, D., 1996. Do parasites reduce the changes of triangulation in a real food web? Oikos 76, 284–300.

Jain, S., Krishna, S., 2002. Large extinctions in an evolutionary model: the role of innovation and keystone species. Proc. Natl. Acad. Sci. U.S.A. 99, 2055–2060.

Jensen, T.S., Lerche, D.B., Sørensen, P.B., 2003. Ranking contaminated sites using a partial ordering method. Environ. Toxicol. Chem. 22, 776–783.

Jordán, F., Liu, W.-C., Davis, A.J., 2006. Topological keystone species: measures of positional important in food webs. Oikos 112, 535–546.

Jordán, F., Scheuring, I., 2002. Searching for keystones in ecological networks. Oikos 99, 607–612.

Jordán, F., Scheuring, I., 2004. Network ecology: topological constraints on ecosystem dynamics. Phys. Life Rev. 1, 139–172.

Jordán, F., Takács-Sánta, A., Molnár, I., 1999. A reliability for theoretical quest for keystones. Oikos 86, 453–462.

Link, J., 2002. Does food web theory work for marine ecosystems? Mater. Ecol. Prog. Ser. 230, 1–9.

Luczkovich, J.J., Borgatti, S.P., Johnson, J.C., Everett, M.G., 2003. Defining and measuring trophic role similarity in food webs using regular equivalence. J. Theor. Biol. 220, 303–321.

Martinez, N.D., 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. Ecol. Monogr. 61, 533–536.

Martinez, N.D., Hawkins, B.A., Dawah, H.A., Feifarek, B.P., 1999. Effects of sampling efforts on characterization of food web structure. Ecology 80, 1044–1055.

Melián, C.J., Bascompte, J., 2002. Complex networks: two ways to be robust? Ecol. Lett. 5, 705–708.

Menge, B.A., Berlow, E.L., Balchette, C.A., Navarrete, S.A., Yamada, S.B., 1994. The keystone species concept: variation in interaction strength in a rocky intertidal habitat. Ecol. Monogr. 64, 249–286.

Memmott, J., Martinnez, N.D., Cohen, J.E., 2000. Predators, parasitoids and pathogens: species richness, trophic generality and body size in a natural food web. J. Anim. Ecol. 69, 1–15.

Mills, L.S., Soulé, M.E., Doak, D.F., 1993. The keystone-species concept in ecology and conservation. BioScience 43, 219–224.

Montoya, J.M., Rodríguez, M.A., Hawkins, B.A., 2003. Food web complexity and higher-level ecosystem services. Ecol. Lett. 6, 587–593.

Montoya, J.M., Solé, R.V., 2002. Small world patterns in food webs. J. Theor. Biol. 214, 405–412.

Newman, M.J.E., 2003. The structure and function of complex networks. SIAM Rev. 45, 167–256.

Opitz, S., 1996. Trophic interactions in Caribbean coral reefs. ICLARM Tech. Rep. 43, 341.

Paine, R.T., 1966. Food web complexity and species diversity. Am. Nat. 100, 65–75.

Polis, G.A., 1991. Complex desert food webs: an empirical critique of food web theory. Am. Nat. 138, 123–155.

Pimm, S.L., 1982. Food Webs. Chapman and Hall, London.

Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S., Daily, G., Castilla, J.C., Lubchenco, J., Paine, R.T., 1996. Challenges in the quest for keystones. BioScience 46, 609–620.

Proulx, S.R., Promislow, D.E.L., Phillips, P.C., 2005. Network thinking in ecology and evolution. Trends Ecol. Evol. 20, 345–353.

Rummel, R.J., 1970. Applied Factor Analysis. Northwestern University Press, Evanston, IL.

Shirley, M.D.F., Rushton, S.P., 2005. the impacts of network topology on disease spread. Ecol. Complex. 2, 287–299.

Solé, R.V., Montoya, J.M., 2001. Complexity and fragility in ecological networks. Proc. R. Soc. B 268, 2039–2045.

Springer, A.M., Estes, J.A., van Vliet, G.B., Williams, T.M., Doak, D.F., Danner, E.M., Forney, K.A., Pfister, B., 2003. Sequential megafaunal collapse in the North Pacific Ocean: an ongoing legacy of industrial whaling? Proc. Natl. Acad. Sci. U.S.A. 100, 12223–12228.

Stephenson, K., Zelen, M., 1989. Rethinking centrality: methods and examples. Social Networks 11, 1–37.

Strogatz, S.H., 2001. Exploring complex networks. Nature 410, 268–275.

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.

Waide, R.B., Reagan, W.B. (Eds.), 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 freshwater food web. Oikos 55, 299–311.

Wasserman, S., Faust, K., 1994. Social Network Analysis. Cambridge University Press, Cambridge.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

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. U.S.A. 99, 12913–12916.

Wooton, J.T., 1994. The nature and consequences of indirect effects in ecological communities. Annu. Rev. Ecol. Syst. 25, 443–466.

Wutchy, S., Stadler, P.F., 2003. Centers of complex networks. J. Theor. Biol. 223, 45–53.

Yodzis, P., 1998. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. J. Anim. Ecol. 67, 635–658.