
Quantitative patterns in the structure of model and empirical food webs

J. Camacho^{1,2}, R. Guimerà¹, D. B. Stouffer¹, and L. A. N. Amaral¹

¹ *Department of Chemical and Biological Engineering, Northwestern University,
Evanston, IL 60208, USA*

² *Departament de Física (Física Estadística), Universitat Autònoma de Barcelona,
E-08193 Bellaterra, Catalonia, Spain*

RUNNING TITLE: Patterns in food web structure

¹ **Email:** Juan.Camacho@uab.es, rguimera@northwestern.edu, d-stouffer@northwestern.edu, and amaral@northwestern.edu

Abstract

We analyze the properties of model food webs and of fifteen community food webs from a variety of environments—including freshwater, marine-freshwater interfaces and terrestrial environments. We first perform a theoretical analysis of a recently proposed model for food webs—the niche model of Williams and Martinez (2000). We derive analytical expressions for the distributions of species' number of prey, number of predators, and total number of trophic links and find that they follow universal functional forms. We also derive expressions for a number of other biologically relevant parameters which depend on these distributions. These include the fraction of top, intermediate, basal, and cannibal species, the standard deviations of generality and vulnerability, the correlation coefficient between species' number of prey and number of predators, and assortativity. We show that our findings are robust under rather general conditions; a result which could not have been demonstrated without treating the problem analytically. We then use our analytical predictions as a guide to the analysis of fifteen of the most complete empirical food webs available. We uncover quantitative unifying patterns that describe the properties of the model food webs and most of the trophic webs considered. Our results support a strong new hypothesis that the empirical distributions of number of prey and number of predators follow *universal functional forms that, without free parameters, match our analytical predictions*. Further, we find that the empirically observed correlation coefficient, assortativity, and fraction of cannibal species are consistent with our analytical expressions and simulations of the niche model. Finally, we show that two quantities typically used to characterize complex networks, the average distance between nodes and the average clustering coefficient of the nodes, show a high degree of regularity for both the empirical data and simulations of the niche model. Our findings suggest that statistical physics concepts such as scaling and universality may be useful in the description of natural ecosystems.

Key words: Food webs, complex networks, network structure, scaling, universality, patterns.

1 Introduction

In ecosystems, species are connected through intricate trophic relationships which define complex food webs (Briand and Cohen, 1984; Cohen et al., 1990; Williams and Martinez, 2000). Understanding the structure and mechanisms underlying the formation of these complex webs is of great importance in ecology (Rejmanek and Stry, 1979; Briand and Cohen, 1987; Cohen et al., 1990; Polis, 1991). In particular, food web structure provides insights into the behavior of ecosystems under perturbations such as the introduction of new species or the extinction of existing ones. It is thought that the nonlinear response of the elements composing the network may lead to possibly catastrophic outcomes for even small perturbations (Berlow, 1999; Chapin et al., 2000; McKann, 2000). Thus, an understanding of the topology and assembly mechanisms of food webs may be of great practical interest with regard to crop selection, management of fishing stocks, preservation of threatened ecosystems, and maintenance of biodiversity (Pimm et al., 1995; Stone, 1999; Covington, 2000; Hutchings, 2000).

Due to the importance of food webs, much effort has been put into collecting empirical data and uncovering unifying patterns which describe their structure (Rejmanek and Stry, 1979; Briand and Cohen, 1984, 1987; Lawton and Warren, 1988; Cohen et al., 1990; Hall and Raffaelli, 1991; Pimm et al., 1991; Polis, 1991; Martinez, 1993). However, in the last decade the construction of larger and more complete food webs has clearly demonstrated that the previously reported unifying patterns do not hold for the new webs (Hall and Raffaelli, 1991, 1993; Polis, 1991). Indeed, the complexity of the new webs has rendered it quite challenging to obtain quantitative patterns with which to “substitute” the old ones (Dunne et al., 2002).

The purpose of this work is twofold. First, to perform a theoretical analysis of a recently proposed model for food webs—the niche model of Williams and Martinez (2000). Second, to use the predictions of the model as a guide to a systematic statistical analysis of community food webs from a variety of environments.

Remarkably, we uncover quantitative unifying patterns that describe the properties of most of the diverse trophic webs considered and capture the random and non-random aspects of their structure. Specifically, we find that several quantities—such as the distributions of number of prey, number of predators, and number of trophic links—characterizing these diverse food webs obey robust functional forms that, as predicted by our analytical results for the niche model, depend on a single parameter—the linkage density of the food web.

The organization of the paper is as follows. In Section 2 we study the niche model proposed by Williams and Martinez (2000) analytically and numerically. In Section 3 we analyze the empirical food webs and show the existence of robust quantitative patterns. Finally, in Section 4 we present some concluding remarks.

2 Analytical solution of the niche model

Recently, Williams and Martinez (2000) have proposed a model for food web structure—the niche model—that with just a couple of ingredients appears to successfully predict key structural properties of the most comprehensive food webs in the literature. Numerical simulations of the niche model predicted values for many quantities typically used to characterize empirical food webs that are consistent with measured values for seven empirical webs.

In this Section, we systematically investigate the niche model from a *theoretical* perspective. We obtain analytical expressions for a number of quantities characterizing the structure of food webs in the limit of sparse food webs, i.e., webs with $L \ll S^2$, where L is the number of trophic interactions between species and S is the number of species in the web. This limit is the one of interest in ecology because (i) for most food webs reported in the literature the directed connectance, defined as $C \equiv L/S^2$, take values much smaller than one, and (ii) it corresponds to the limit of large web sizes S (Briand and Cohen, 1984, 1987; Solé et al., 2002), which is the interesting limit if one surmises that geographically separate ecosystems are in fact connected.

We first calculate the probability distributions of number of prey and of number of predators and find that for $C \ll 1$ they depend only on one parameter of the model—the linkage density $z \equiv L/S$, i.e. the average number of prey or predators. These distributions give valuable information about the structure of the network (Albert et al., 2000) and enable us to calculate other interesting quantities such as the fraction of “top,” “intermediate,” and “basal” species, and the standard deviation of the “vulnerability” and “generality” of the species in the food web (Williams and Martinez, 2000). We also calculate the correlation coefficient between number of prey and number of predators, the fraction of cannibals present, and two additional properties of interest in the characterization of complex networks: the average “distance” between species and the local redundancy of the connection between species (Watts and Strogatz, 1998).

2.1 The niche model

Consider an ecosystem with S species and L trophic interactions between these species. These species and interactions define a network with S nodes and L directed links. In the niche model, one first randomly assigns S species to “trophic niches” with niche values n_i mapped uniformly onto the interval $[0,1]$. For convenience, we will assume that the species are ordered according to their niche number, i.e., $n_1 < n_2 < \dots < n_S$.

A species i is characterized by its niche parameter n_i and by its list of prey. Prey are chosen for all species according to the following rule: A species i preys on all species j with niche parameter n_j inside a segment of length a_i centered in a position chosen randomly inside the interval $[a_i/2, n_i]$, with $a_i = xn_i$ and $0 \leq x \leq 1$ a random variable with probability density function

$$p_x(x) = b(1-x)^{(b-1)}. \quad (1)$$

Williams and Martinez (2000) appeared to have chosen this functional form for convenience, but, we will show later that the predictions of the model are mostly robust to changes in the specific form of p_x .

The values of the parameters b and S determine the linkage density $z = L/S$ of the food web, and the directed connectance $C = L/S^2$. Since the species are uniformly distributed along the segment $[0, 1]$, one can express the average number of prey per species as $S\bar{a}$, where the bar indicates an average over all species in the web. It then follows that the linkage density is

$$z = S\bar{a} \quad (2)$$

and the connectance is

$$C = \bar{a}. \quad (3)$$

One can also obtain these expressions in terms of b by noting that n and x are independent variables, namely

$$\bar{a} = \bar{n}\bar{x} = \frac{1}{2(1+b)}, \quad (4)$$

where $\bar{n} = 1/2$ and $\bar{x} = 1/(1+b)$.

In the niche model, isolated species—that is, species with no prey or predators—are eliminated, and species with the same list of prey and predators—that is, trophically-identical species—are “merged.” For $C = \bar{a} \ll 1$, the probability that two species are trophically identical is very small. This suggests that taxonomic and trophic classifications of species lead to similar results in this limit (Williams and Martinez, 2000).

In the following subsections we derive analytical expressions that enable us to predict the properties of the webs generated by model.

2.2 Distribution of number of prey

For large S , the number of prey of species i is $k_i = Sa_i$, so that the probability of having k prey p_{prey} is given directly by the distribution of a . Specifically,

$$p_{\text{prey}}(k) = p(a)/S. \quad (5)$$

In order to evaluate $p(a)$, let us note that a is the product of two independent stochastic variables, n and x , both ranging between 0 and 1. As illustrated in Fig. 1, it then follows that the cumulative probability $P(a' > a) = \int_a^1 da' p(a')$ is the probability that a pair of random values (n, x) fall

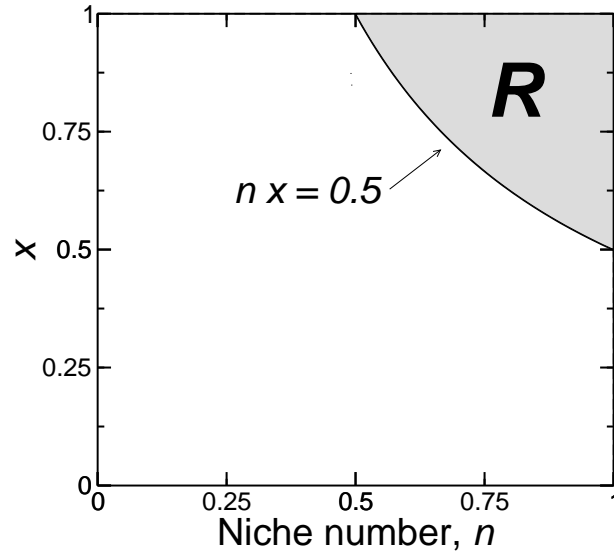


Fig. 1. Calculation of the distribution of segment lengths a . Each point inside the $n - x$ diagram yields a value $a = nx$. The hyperbole $nx = \text{const}$ divides the $n - x$ plane in two regions, which we color grey and white. The region R , colored in grey, contains all the points for which $a > 0.5$. Therefore, the probability of drawing a pair (n, x) belonging to region R can be calculated using Eq. (6).

in the region R of the $n - x$ diagram bounded by the lines $x = 1$, $n = 1$, and the hyperbole $a = nx$,

$$\begin{aligned}
 P(a' > a) &= \int_a^1 da' p(a') \\
 &= \int_R dx dn p_n(n) p_x(x) \\
 &= \int_a^1 dx \int_{a/x}^1 dn p_n(n) p_x(x), \tag{6}
 \end{aligned}$$

where $p_n(n) = 1$ is the probability density function for n . The integration of Eq. (6) gives rise to a closed form involving hypergeometric functions (Gradshteyn and Ryzhik, 2000).

However, we can obtain a simpler analytical solution for $p(a)$ than that given by the hypergeometric functions as follows. In the limit $C \ll 1$, which implies $b \gg 1$, $p_x(x)$ is negligible except when $x \ll 1$. We can then approximate p_x as

$$p_x(x) = b(1 - x)^{(b-1)} \simeq be^{-bx}. \tag{7}$$

in the entire x -range and expect the results to remain unchanged in the limit $C \ll 1$. Under this approximation, the solution of Eq. (6) yields

$$p(a) = bE_1(ba), \tag{8}$$

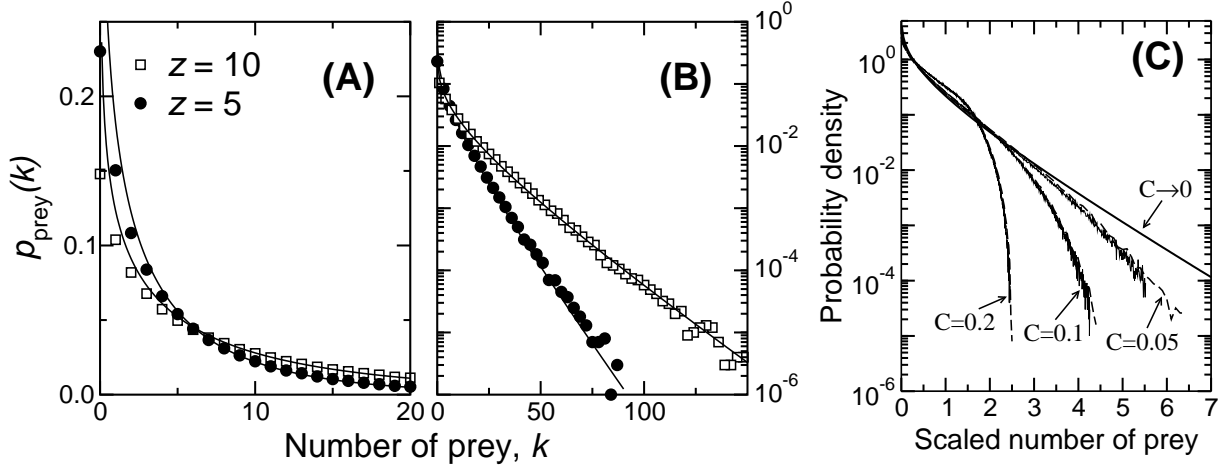


Fig. 2. (A) Linear and (B) log-linear plots of the distribution of number of prey for 1000 simulations of food webs with $S = 1000$. We show results for $z = 5$ and 10 and the corresponding theoretical predictions. As predicted by Eq. (9), we find an exponential decay of the distributions. (C) Log-linear plot of the probability densities for the scaled number of prey $k/2z$ for finite C . We show results for $C = 0.05, 0.1$ and 0.2 , for 50000 webs generated using the niche model. Dashed lines (solid lines) correspond to simulations with $S = 100$ ($S = 1000$) for the appropriate values of z (namely $z = 5, 10$, and 20 , for $S = 100$, and $z = 50, 100$, and 200 for $S = 1000$). The thick solid line is the theoretical prediction in the limit $C \rightarrow 0$. One observes the collapse of distributions having the same C but different S , similar to what we derived analytically for $C \rightarrow 0$. Note that the decay is faster than exponential for finite C .

where $E_1(x) = \int_x^\infty dt t^{-1} \exp(-t)$ is the exponential-integral function (Gradshteyn and Ryzhik, 2000). The probability distribution $p_{\text{prey}}(k)$ is obtained from Eq. (8) making the substitutions $a = k/S$ and $b = S/2z$ —which are valid in the limit $C \ll 1$ —yielding

$$p_{\text{prey}}(k) = (1/2z) E_1(k/2z). \quad (9)$$

In Figs. 2(A) and (B) we compare the predictions of Eq. (9) with numerical simulations of the model. We find close agreement between our analytical expression and the numerical results. In particular, p_{prey} shows an exponential decay for large k . The deviations observed for small values of k are due to the fact that $k_j = S a_j$ is a good approximation only when the fluctuations of k_j are small, which is not true for small k .

Equation (9) depends only on the scaled number of prey $k/2z$. Thus, for *any* value of z , the scaled variable $\tilde{k} = k/2z$ obeys the same probability density function,

$$p_{\text{prey}}(\tilde{k}) = E_1(\tilde{k}). \quad (10)$$

This probability density function is therefore universal, i.e., it is identical for any values of S and z provided that C is small. For finite C , p_{prey} has a truncation of the exponential decay at a value of \tilde{k} that is a decreasing function of C . Remarkably, even for finite C , p_{prey} only depends on C ; cf. Fig. 3(A).

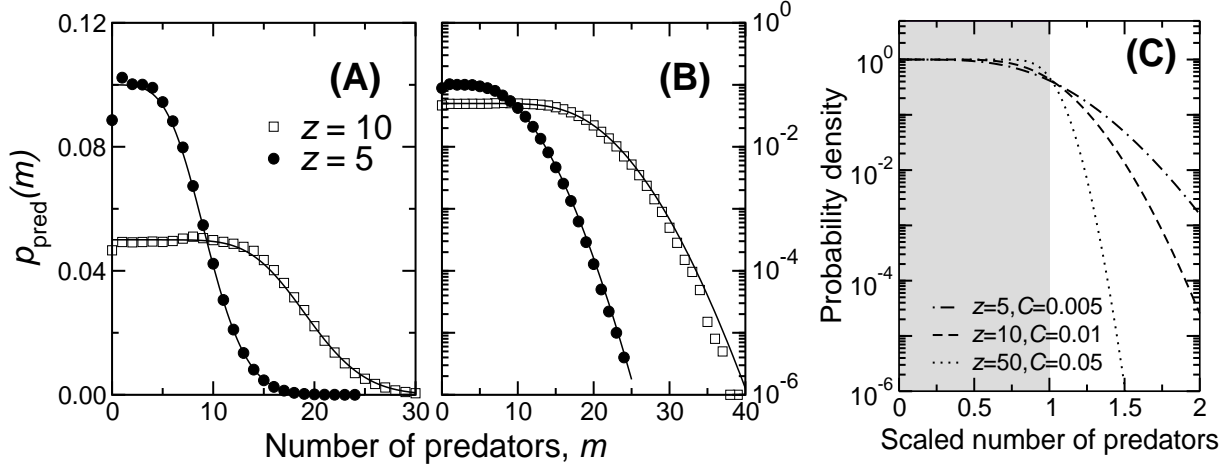


Fig. 3. (A) Linear and (B) log-linear plots of the distribution of number of predators for 1000 simulations of food webs with $S = 1000$. We show results for $z = 5$ and 10 and the corresponding theoretical predictions. As predicted by Eq. (14), we find a regime where the distribution is approximately uniform followed by a Gaussian decay. (C) Log-linear plot of the analytical expressions for the probability densities for the scaled number of predators $m/2z$ for three values of z . For all values of z , we find an approximately constant value of p_{pred} for $m/2z < 1$ (shaded region) and a fast Gaussian-like decay for $m/2z > 1$.

2.3 Distribution of number of predators

For $C \ll 1$ the predators of species i have niche values $n_j > n_i$, and the segment a_j is placed with equal probability in the interval $[0, n_j]$. Therefore, the probability that a species j preys on i , provided $n_j > n_i$, is

$$a_j/n_j = x_j n_j/n_j = x_j, \quad (11)$$

implying that the average probability that a species with $n_j > n_i$ preys on species i is \bar{x} .

If one notes that species i has $(S - i)$ potential predators—those species with $n_j > n_i$ —then it follows that, in the limit $C \ll 1$, the total number of predators of i is the result of $(S - i)$ independent “coin-throws” with probability \bar{x} of a given species j , with $n_j > n_i$, being a predator. This fact implies that the probability of species i having m predators is given by the binomial distribution,

$$p_{\text{pred}}^i(m) = \binom{S-i}{m} \bar{x}^m (1 - \bar{x})^{S-m-i} \quad (12)$$

It then follows that the distribution of number of predators for the set of all species in the web is an average over the binomials for the different values of i

$$p_{\text{pred}}(m) = \frac{1}{S} \sum_{i=1}^{S-m} p_{\text{pred}}^i(m) = \frac{1}{S} \sum_{i=1}^{S-m} \binom{S-i}{m} \bar{x}^m (1 - \bar{x})^{S-m-i} \quad (13)$$

In the limit of interest, $S \gg 1$, and with $S\bar{x} = 2z$, one can approximate the binomial distribution by a Poisson and the sum by an integral, yielding

$$p_{\text{pred}}(m) = \frac{1}{2z} \int_0^{2z} dt \frac{t^m e^{-t}}{m!} = \frac{1}{2z} \gamma(m+1, 2z), \quad (14)$$

where γ is the ‘‘incomplete gamma function’’ (Gradshteyn and Ryzhik, 2000). For $m < 2z$, p_{pred} is approximately constant because $\gamma(m+1, 2z) \approx 1$, while for $m > 2z$ p_{pred} decays with a Gaussian tail. In Figs. 2(A) and (B), we compare the predictions of Eq. (14) with numerical simulations and find good agreement.

Unlike the scaling seen for the distribution of number of prey, Eq. (14) is not simply a function of the scaled variable $m/2z$. However, for small values of $m/2z$, γ is a constant and thus it does not depend on m or z . So, the probability density for the scaled variable $\tilde{m} = m/2z$,

$$p_{\text{pred}}(\tilde{m}) = \gamma(2z\tilde{m} + 1, 2z) \approx 1 \quad \tilde{m} < 1 \quad (15)$$

for *any* z . For $\tilde{m} > 1$, $p_{\text{pred}}(\tilde{m})$ decays as a Gaussian. In Fig. 3(B) we show that the decay rate of p_{pred} increases with z , becoming a sharp step drop at $\tilde{m} = 1$ as $z \rightarrow \infty$.

2.4 Distribution of number of trophic links

The number r of trophic links of a species is the sum of the number of prey and number of predators of that species. In Fig. 4, we display the probability densities $p_{\text{links}}(\tilde{r} = r/2z)$ obtained from numerical simulations for $S = 100$ and $S = 1000$. Both distributions show a maximum around $\tilde{r} = 1$ and the same decay in the tail as the distribution of number of prey, i.e. an exponential decay for small C , and faster than exponential decay for larger C values. This similarity is expected since the distribution of number of predators decays very rapidly, so that large values of r must arise due to large values of k .

Obtaining an analytical solution for $p_{\text{links}}(r)$ is considerably more difficult than for p_{prey} or p_{pred} because it requires one to estimate the correlations between the number of prey and the number of predators. The rules underlying the niche model imply that the number of predators is negatively correlated with the number of prey. That is to say, a species having a large number of predators is likely to have a low niche number, and thus it will likely predate upon few or no species. This attribute of the niche model is discussed in more detail in Section 2.7.

If one neglects, for the moment, correlations between the distribution of number of prey and number of predators, the probability density for the scaled total number of links, $\tilde{r} = r/2z$, is found simply by the convolution of both distributions. In the limit $C \ll 1$, one has

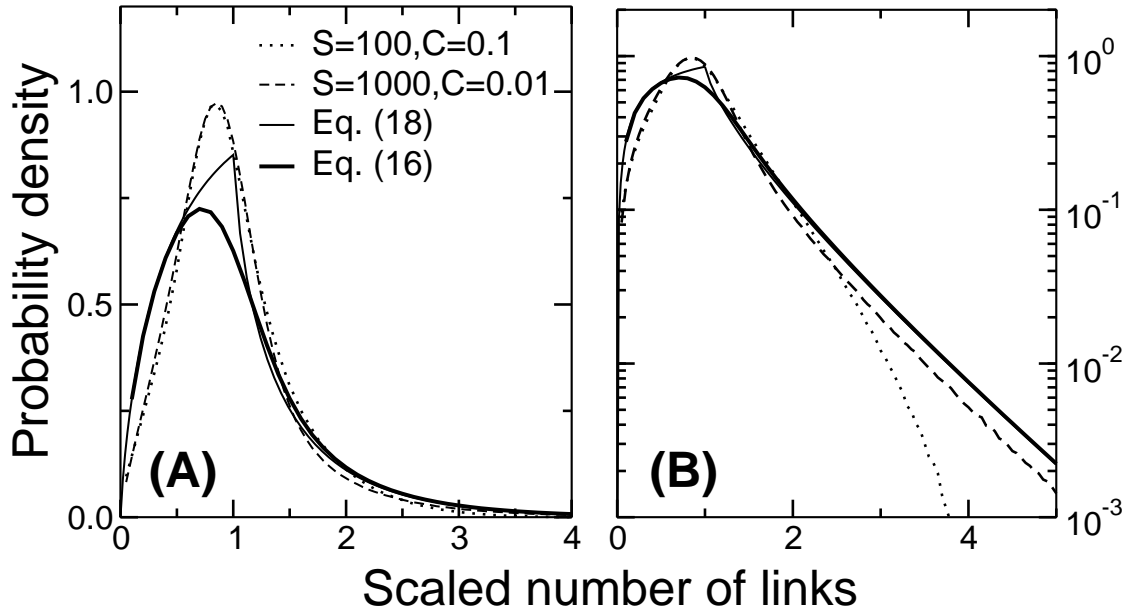


Fig. 4. (A) Linear and (B) log-linear plots of the the probability densities for the scaled number of links, $\tilde{r} = r/2z$. We show results for $z = 10$ and web sizes $S = 100$ (dotted line) and $S = 1000$ (dashed line). The two curves show a similar behavior: A maximum around $\tilde{r} = 1$, and a tail dominated by the distribution of number of prey. For $S = 1000$, one has $C = 0.01$ and the decay is exponential. For $S = 100$, $C = 0.1$ and the decay is faster than exponential (see Fig. 2). We also show two analytical curves in the limit $C \rightarrow 0$: the convolution, Eq. (16), for $z = 10$ (thick solid line) and for $z = \infty$ (thin solid line). As with the simulations, the theoretical curves also display a maximum at $\tilde{r} \approx 1$. The decay for both analytical curves is exponential, as is the case for the distribution of number of prey for $C \rightarrow 0$.

$$\begin{aligned}
 p_{\text{links}}(\tilde{r}) &= \int_0^{\tilde{r}} p_{\text{prey}}(t) p_{\text{pred}}(\tilde{r} - t) dt \\
 &= \int_0^{\tilde{r}} E_1(t) \gamma(2z(\tilde{r} - t) + 1, 2z) dt
 \end{aligned} \tag{16}$$

This integral cannot be expressed in terms of recognizable functions, but can be calculated numerically. In Fig. 4 we plot the prediction of Eq. (16) for $z = 10$. As we obtained for the simulations of the model, we find a maximum of p_{links} around $\tilde{r} = 1$ and an exponential decay in the tail. The latter arises from the exponential tail found for the distribution of number of prey, as previously mentioned.

One can also obtain an explicit expression for p_{links} when $z \rightarrow \infty$. One expects this solution not to differ much from that for a finite z , since the only difference is that a smooth step in the distribution of the number of predators around $\tilde{m} = 1$ is substituted by a sharp one. As discussed in the previous section, the limits of $z \rightarrow \infty$ and $C \ll 1$ yield

$$p_{\text{pred}}(m) = \begin{cases} 1 & \text{if } m < 2z, \\ 0 & \text{if } m > 2z. \end{cases} \quad (17)$$

Integration of Eq. (16) then yields

$$p_{\text{links}}(\tilde{r}) = \begin{cases} 1 - E_2(\tilde{r}) & \text{if } \tilde{r} < 1, \\ E_2(\tilde{r} - 1) - E_2(\tilde{r}) & \text{if } \tilde{r} > 1. \end{cases} \quad (18)$$

with $E_2(x) = 1 - \exp(-x) + xE_1(x)$ the exponential-integral of order 2 (Gradshteyn and Ryzhik, 2000). Figure 4 also displays Eq. (18), showing a curve rather similar to the convolution for $z = 10$ except close to $\tilde{r} = 1$.

From the analytical solution, one can demonstrate that the tail of p_{links} decays exponentially. For large \tilde{r} , Eq. (18) can be approximated by

$$\begin{aligned} E_2(\tilde{r} - 1) - E_2(\tilde{r}) &\simeq -E_2'(\tilde{r} - 1/2) \\ &= E_1(\tilde{r} - 1/2) \\ &\simeq \frac{\exp(-\tilde{r} + 1/2)}{\tilde{r} - 1/2}, \end{aligned} \quad (19)$$

where $E_2' = -E_1$ is the derivative of function E_2 and the last expression is the dominant term in the expansion of exponential-integral functions (Gradshteyn and Ryzhik, 2000). Note that Eq. (19) is a good approximation of Eq. (18) for $\tilde{r} > 2$.

2.5 Fraction of top, intermediate, basal and cannibal species

As the names indicate, top species have no predators and basal species have no prey, while intermediate species are those with both prey and predators. The fraction T of top species is, by definition,

$$T \equiv p_{\text{pred}}(0) = \frac{1 - \exp(-2z)}{2z}, \quad (20)$$

as directly seen from Eq. (14). Typically empirical webs have $2 < 2z < 20$, a region for which Eq. (20) can be approximated simply as $T = 1/2z$.

To calculate the fraction B of basal species, we note that a species has no prey only if its range a falls in a region with no species. In the limit of large S , the probability density for finding an empty interval of length δ is $Se^{-S\delta}$, as predicted by the canonical distribution (Pathria, 1972). Thus, the probability of finding a species-free segment of length larger than a is e^{-Sa} , which

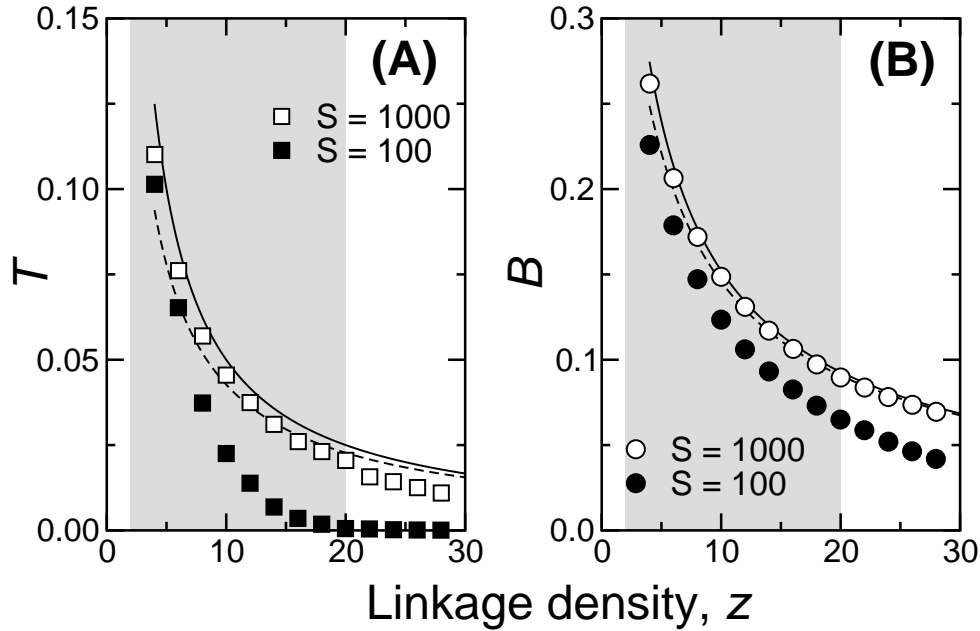


Fig. 5. Fraction of top and basal species as a function of the linkage density z . The shaded region corresponds to the interval of z typically observed in empirical food webs (Dunne et al., 2002). (A) Comparison of the results for 1000 simulations of food webs for which isolated species were removed with our theoretical prediction Eqs. (20), solid line, and (22), dashed line, for fraction of top species. (B) Comparison of the results of 1000 simulations of food webs for which isolated species were removed with our theoretical prediction Eqs. (21), solid line, and (22), dashed line, for fraction of basal species. Note that for both T and B there is good agreement between the analytical expressions and the numerical results when $C \ll 1$. Also note that the theoretical predictions provide narrow bounds for the numerical results in this limit.

yields the probability for a species with a range a not to prey on other species. Using Eq. (8), it follows that the fraction of bottom species is:

$$B = \int_0^1 da e^{-S_a} p(a) = \frac{\ln(1 + 2z)}{2z}. \quad (21)$$

Note that we must take this approach to the calculation because the definition of the fraction B of basal species, $p_{\text{prey}}(0)$, leads, in the continuous analysis, to a divergence as the exponential-integral function $E_1(x)$ diverges in the limit $x \rightarrow 0$.

In the niche model, isolated species are eliminated, so they are not counted toward top and basal species. To correct Eqs. (20) and (21) for this effect, we must account for the isolated species in our calculations. We estimate the number of isolated species to first order by assuming that having no prey is statistically independent of having no predators, implying that the fraction of isolated species is just the product of the fractions of top and basal species. This assumption does not take into account the (negative) correlations between the number of prey of a species and its number of predators. Nonetheless, this simple approximation provides an upper bound

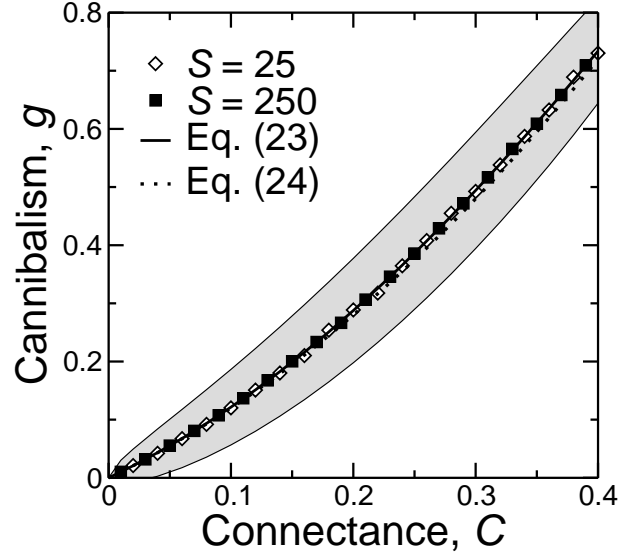


Fig. 6. Fraction of cannibals as a function of the connectance C . We plot the fraction of cannibal species averaged over 5000 realizations of the model for two system sizes, and the predictions of Eqs. (23) and (24). It is visually apparent that the results are insensitive to changes in S and that they are in good agreement with our theoretical predictions. The shaded area indicates one standard deviation from the average for $S = 25$; the upper and lower curves have been obtained using Eqs. (26) and (24). Our results demonstrate that in the niche model the fraction of cannibals depends only upon the connectance.

for the number of isolated species which leads to lower bounds on T and B ,

$$T' = \frac{T - TB}{1 - TB}, \quad B' = \frac{B - TB}{1 - TB} \quad (22)$$

In Fig. 5, we compare our analytical predictions for the fraction of top and basal species with numerical simulations of the model. As expected, Eqs. (20–22) provide bounds for the numerical results in the limit $C \ll 1$.

The fraction of intermediate species is just $I = 1 - (T + B)$.

Another quantity of ecological interest is the fraction of cannibal species. Consider a species i with niche number n_i . Its preys are the species inside an interval of length $a_i = xn_i$, chosen according to Eq. (1), and centered with uniform probability in the interval $[a_i/2, n_i]$. For i to predate on itself, it is necessary that the segment of length a_i contains species i . This is equivalent to the distance between the center of the segment and n_i being smaller than $a_i/2$, which occurs with probability $\frac{a_i}{2}/(n_i - \frac{a_i}{2})$. Therefore, the average probability g for a species to be a cannibal is given by

$$g = \int_0^1 b(1-x)^{b-1} \frac{x/2}{1-x/2} dx = {}_2F_1(1, 1; 2C; 1/2) - 1 \quad (23)$$

where ${}_2F_1$ is the Gauss hypergeometric function (Gradshteyn and Ryzhik, 2000). One thus finds that, for any value of S , the fraction of cannibals depends only upon C . Equation (23) can be

expanded for small C as

$$g \simeq C + 2C^2 + O(C^3). \quad (24)$$

Since the largest value of C reported in the literature is 0.3, we expect Eq. (24) to provide a good approximation for empirically relevant values of C (Fig. 6).

Next, we evaluate the standard deviation of g . If ℓ denotes the number of cannibal species in a single realization of the model with S species, the corresponding fraction of cannibals is simply $g_\ell = \ell/S$. Therefore, the standard deviation of g_ℓ is $\sigma_g = \Delta\ell/S$. Since g is the probability for a species to be a cannibal, the probability of having exactly ℓ cannibal species is given by a binomial distribution

$$p_{can}(\ell) = \binom{S}{\ell} g^\ell (1-g)^{S-\ell}. \quad (25)$$

Therefore, $\Delta\ell^2 = Sg(1-g)$, and the standard deviation for the fraction of cannibalism is

$$\sigma_g = \sqrt{\frac{g(1-g)}{S}}. \quad (26)$$

This expression implies that σ_g decreases with increasing S .

2.6 Standard deviation of generality and vulnerability

The vulnerability of a prey is defined as its number m of predators, and the generality of a predator as its number k of prey. Following Williams and Martinez (2000), we define the normalized standard deviations of the vulnerability as $\sigma_V^2 = \overline{m^2}/\overline{m}^2 - 1$ and of the generality as $\sigma_G^2 = \overline{k^2}/\overline{k}^2 - 1$. By definition, one has $\overline{m} = \overline{k} = z$.

To evaluate σ_V , we first calculate $\overline{m^2}$. Equation (14) yields $\overline{m^2} = 4z^2/3 + z$, so that

$$\sigma_V = \sqrt{\frac{1}{3} + \frac{1}{z}}, \quad (27)$$

which approaches $\sqrt{1/3}$ as $z \rightarrow \infty$ (Fig. 7). Since Eq. (14) is valid only in the limits $S \gg 1$ and $C \ll 1$, these are also the limits of validity for Eq. (27). We present these results, as well as simulations of the niche model in Fig. 7(A).

We next calculate σ_G by direct evaluation of $\overline{k^2}$. If $S \gg 1$, the number of prey of a species

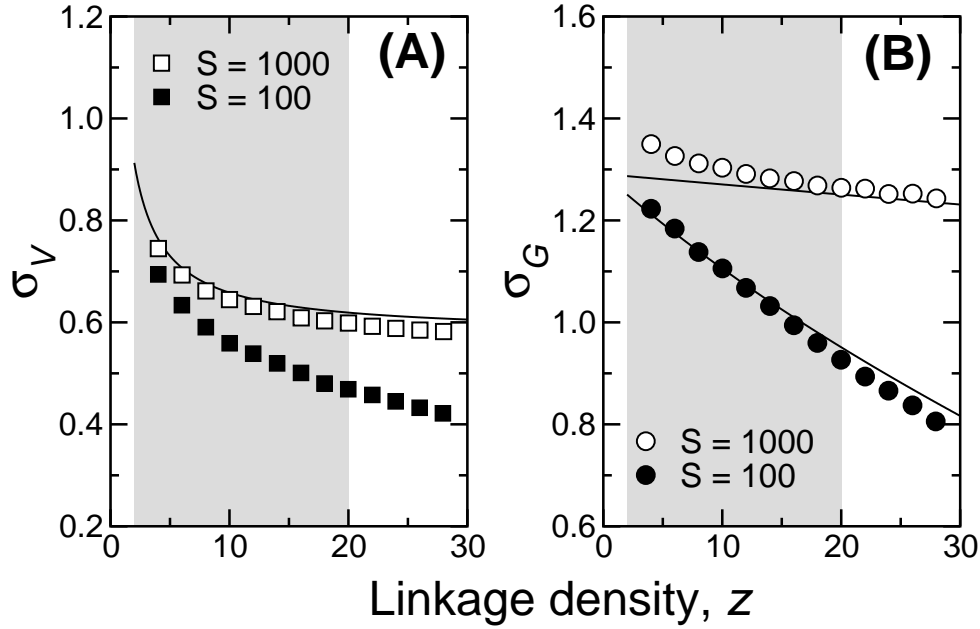


Fig. 7. Normalized standard deviations of generality and vulnerability as a function of the average linkage density z . The shaded region corresponds to the interval of z typically observed in empirical food webs (Dunne et al., 2002). (A) Comparison of the results for 1000 simulations of food webs for which isolated species were removed with our theoretical prediction Eq. (27) for the standard deviation of the vulnerability. (B) Comparison of the results for 1000 simulations of food webs for which isolated species were removed with our theoretical prediction Eq. (29) for the standard deviation of the generality. Note that Eq. (27) is only valid in the limits of $S \gg 1$ and $C \ll 1$, which is confirmed in (A) by the deviations found for $S = 100$. Equation (29) was derived independent of these limits and thus holds for any value of C . The small underestimation of our expression for σ_G relates to the fact that $k_j = Sr_j$ is a good approximation only when the fluctuations of k_j are small, which is no longer true for small k .

having a range a is $k = Sa$, yielding

$$\frac{\overline{k^2}}{\overline{k}^2} = \frac{\overline{a^2}}{\overline{a}^2} = \frac{\overline{n^2 x^2}}{\overline{n x^2}} = \frac{8(b+1)}{3(b+2)}. \quad (28)$$

Thus, σ_G is

$$\sigma_G = \sqrt{\frac{8}{3} \frac{1}{1+2C} - 1}. \quad (29)$$

For $C \ll 1$, $\sigma_G = \sqrt{5/3}$, a result that can also be obtained from Eq. (9). Unlike Eq. (27), Eq. (29) depends upon C and thus does not require the same limits considered in the derivation of Eq. (27). We show these results as well as those for numerical simulations of the niche model in Fig. 7(B).

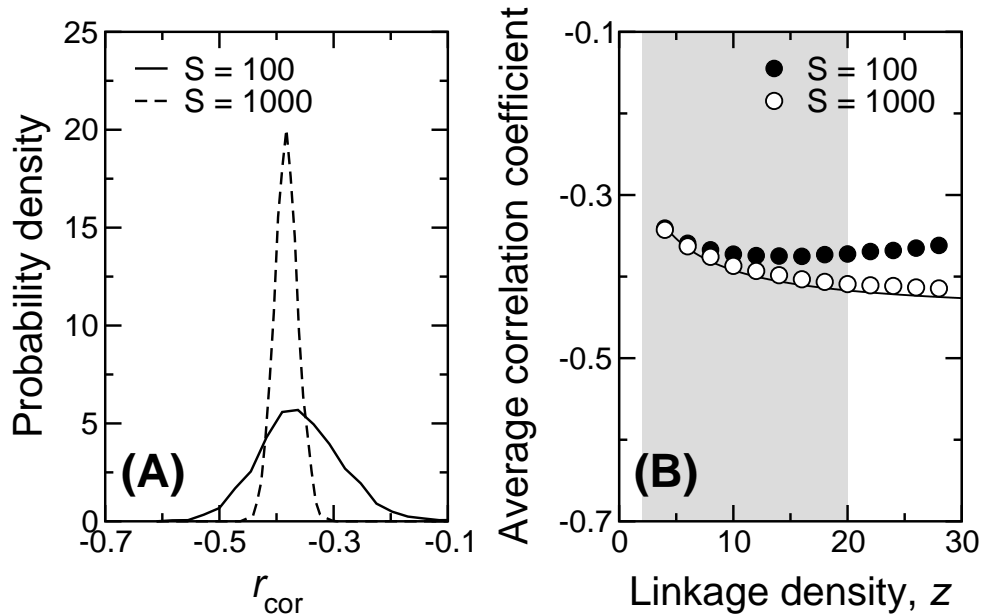


Fig. 8. (A) Distribution of r_{cor} , the correlation coefficient between number of prey and number of predators obtained from 5000 realizations of the niche model. We show simulation results for $z = 10$. (B) Comparison of the average correlation coefficient obtained from 5000 simulations of the model (for which isolated species were removed) with our theoretical prediction, Eq. (39). The negative correlations observed are inherent to the model, as a species having a large number of predators is likely to have a low niche number, and consequently will predate upon few or no species.

2.7 Correlation coefficient

The mechanisms by which webs are constructed in the niche model imply that a species with many predators is likely to have a low niche value. Furthermore, a species with a low niche value probably feeds on few prey. As a consequence, the more predators a species has, the smaller the number of prey will be, and vice versa. Thus, the number of prey of a species and its number of predators must be negatively correlated.

In order to quantify the correlations between k and m we define the correlation coefficient between the number of prey and the number of predators,

$$r_{\text{cor}} = \frac{\overline{km} - \bar{k}\bar{m}}{\sqrt{\overline{k^2} - \bar{k}^2} \sqrt{\overline{m^2} - \bar{m}^2}} \quad (30)$$

where, as before, the bar over a variable denotes an average over all species in a web.

We evaluate the average correlation coefficient, r_{cor} , for the niche model in the limits $S \gg 1$ and $C \ll 1$. Equation (30) can be rearranged to yield

$$r_{\text{cor}} = \frac{\overline{km}/z^2 - 1}{\sigma_G \sigma_V}, \quad (31)$$

where σ_G and σ_V are the normalized standard deviations of the generality and the vulnerability respectively. Thus, in order to obtain a closed form for Eq. (31), we need only to calculate

$$\overline{km} \equiv \sum_{k,m=0}^S km p(k, m), \quad (32)$$

where $p(k, m)$ is the probability that a species has k prey and m predators. If k and m were independent, $p(k, m)$ would be simply the product $p_{\text{prey}}(k)p_{\text{pred}}(m)$, yielding

$$\overline{km} = \sum_{k=0}^S kp_{\text{prey}}(k) \sum_{m=0}^S mp_{\text{pred}}(m) = \overline{k}\overline{m} = z^2 \quad (33)$$

and $r_{\text{corr}} = 0$. However, k and m are not independent.

Let us define $p(k, m, n)$ as the probability that a species has k prey, m predators, and niche number n . One then has

$$p(k, m, n) = p(k|m, n)p(m|n)p_n(n). \quad (34)$$

where $p(\cdot|\cdot)$ refers to the conditional probability.

Remembering that $p_n(n) = 1$ and noticing that $p(k|m, n) = p(k|n)$ because the number of prey in the niche model is determined solely by the niche value n , one can write

$$p(k, m) \equiv \int dn p(k, m, n) = \int dn p(k|n)p(m|n). \quad (35)$$

Using Eq. (35) in Eq. (32) and rearranging terms, yields

$$\overline{km} = \int_0^1 dn \overline{k}_n \overline{m}_n, \quad (36)$$

with \overline{k}_n and \overline{m}_n the average number of prey and of predators for a species with niche number n . One can calculate \overline{k}_n by realizing that $k_n = Snx$ and $S\overline{x} = 2z$, so that

$$\overline{k}_n = Sn\overline{x} = 2zn. \quad (37)$$

In the limits $S \gg 1$ and $C \ll 1$, the potential predators of a species with niche value n are those species with niche values larger than n . Each of these species has a probability \overline{x} of preying on it. Therefore, the average number of predators for a species with niche value n is

$$\overline{m}_n = S(1 - n)\overline{x} = 2z(1 - n). \quad (38)$$

Substituting Eqs. (37) and (38) into Eq. (36), one obtains $\overline{km} = 2z^2/3$. Finally, using Eq.(27) and Eq.(29) in the limit $C \rightarrow 0$ one obtains

$$r_{\text{cor}} = -\frac{1}{\sqrt{5(1 + 3/z)}}. \quad (39)$$

In Fig. 8(A), we show the probability density of model realized r_{cor} . The figure demonstrates that fluctuations from the average value are not negligible for S as large as 100. In Fig. 8(B) we compare Eq.(39) with numerical simulations of the niche model, finding good agreement in the limits of interest, $S \gg 1$ and $C \ll 1$. The two figures confirm our initial hypothesis that the niche model generates webs with negative correlations between k and m .

2.8 Assortativity

The assortativity of a food web quantifies the correlation between the number of links—trophic interactions—of a species and the number of links of its prey and predators (Newman, 2002). In a food web with positive assortativity—an assortative food web—highly connected species tend to be connected to one another. Conversely, in a food web with negative assortativity—a disassortative food web—highly connected species tend to connect to species with a small number of links. Prior work by Newman (2002) suggests that assortativity tends to be positive for social networks and negative for technological and biological networks. In this subsection we evaluate the assortativity of the food webs generated by the niche model in the limit of large S and small C .

Let us consider a trophic interaction in the network, in which the predator has r trophic links and the prey has r' trophic links. The assortativity A is defined as (Newman, 2002)

$$A = \frac{\langle rr' \rangle - \langle \frac{r+r'}{2} \rangle^2}{\langle \frac{r^2+r'^2}{2} \rangle - \langle \frac{r+r'}{2} \rangle^2}, \quad (40)$$

where $\langle \dots \rangle$ denotes the average over all trophic links in the food web.

We first calculate $\langle rr' \rangle$. In section 2.7 we obtained the average number of prey and the average number of predators for a species with niche number n , namely $\overline{k_n} = 2zn$, and $\overline{m_n} = 2z(1-n)$. Thus the average number of links of a species with niche number n is $\overline{r_n} = \overline{k_n} + \overline{m_n} = 2z$, which is independent of n . It follows then that the average number of links is the same for all species, independently of the niche number. Therefore, r and r' are independent variables, so that $\langle rr' \rangle = \langle r \rangle \langle r' \rangle$.

In order to evaluate these averages, we transform the averages over links, $\langle \dots \rangle$, to averages over

species, $\overline{\cdot}$. Let us call r_i the number of links of the predator in link i , by definition $\langle r \rangle$ is

$$\langle r \rangle = \frac{\sum_{\text{links}} r_i}{L}, \quad (41)$$

where L is the number of links in the network. The predator from link i —denoted as species j —has k_j prey, so that it appears as the predator in k_j links. Therefore, one can write

$$\langle r \rangle = \frac{\sum_{\text{species}} k_j r_j}{L}. \quad (42)$$

Since $r_j = k_j + m_j$, the numerator can be rewritten as $S(\overline{k^2} + \overline{k m})$, with the overbar denoting an average over species, as used in previous sections. Finally, one has

$$\langle r \rangle = \frac{\overline{k^2} + \overline{k m}}{z}, \quad (43)$$

Analogously, one can obtain

$$\langle r' \rangle = \frac{\overline{m^2} + \overline{k m}}{z}, \quad (44)$$

$$\langle r^2 \rangle = \frac{\overline{k^3} + \overline{k m^2} + 3 \overline{k^2 m}}{z}, \quad (45)$$

$$\langle r'^2 \rangle = \frac{\overline{m^3} + \overline{k^2 m} + 3 \overline{k m^2}}{z}. \quad (46)$$

These averages can be evaluated as in section 2.7 (see Appendix A). Substituting Eqs. (43)–(46) into Eq. (40), one obtains

$$A = -\frac{(2z/3 - 1/2)^2}{26z^2/9 + z/3 + 1/4} \leq 0. \quad (47)$$

In Fig. 9, we compare this expression with numerical simulations and find good agreement in the limits for which our derivation is valid.

2.9 Robustness of the analytical results

In Table 1 we summarize the analytical results obtained for the niche model. An important question is that of the robustness of our predictions to changes in the particular formulation of the *details* of the model. The approximations used in the derivations of the expressions for the distributions of the number of prey and predators, Eqs. (9) and (14), allow us to conclude that:

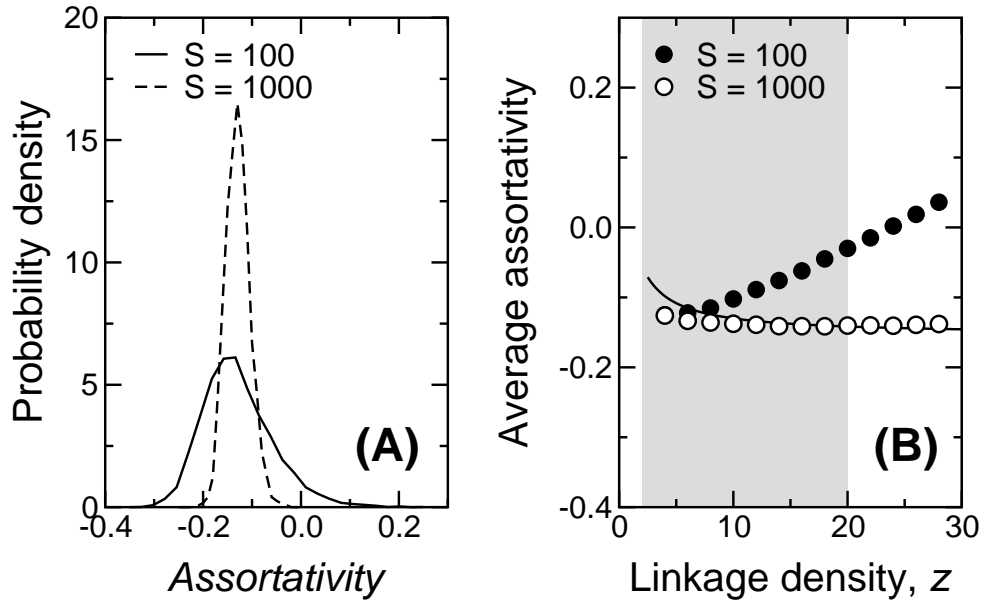


Fig. 9. (A) Distribution of the assortativity A obtained from 5000 realizations of the niche model. We show simulation results for $z = 10$. (B) Comparison of the average assortativity obtained from 5000 simulations of the model (for which isolated species were removed) with our theoretical prediction, Eq. (47). As predicted by Eq. (47), the niche model yields webs that are slightly disassortative in the limit of large S . For small S , food webs become slightly assortative at large linkage densities.

Property	Expression
Distribution of number of prey	$p_{\text{prey}}(k) = (1/2z) E_1(k/2z)$
Distribution of number of predators	$p_{\text{pred}}(m) = (1/2z) \gamma(m+1, 2z)$
Fraction of top species	$T = \frac{1 - \exp(-2z)}{2z}$
Fraction of basal species	$B = \frac{\ln(1+2z)}{2z}$
Fraction of cannibals	$g = {}_2F_1(1, 1; 2C; 1/2) - 1$ $\approx C + 2C^2 + O(C^3)$
Standard deviation of the vulnerability	$\sigma_V = \sqrt{1/3 + 1/z}$
Standard deviation of the generality	$\sigma_G = \sqrt{8/(3 + 6C) - 1}$
Prey-predators correlation coefficient	$r_{\text{cor}} = -\frac{1}{\sqrt{5(1+3/z)}}$
Assortativity	$A = -\frac{(2z/3 - 1/2)^2}{26z^2/9 + z/3 + 1/4}$

Table 1

Summary of the analytical expressions obtained for the niche model in the limits $S \gg 1$ and $C \ll 1$.

- (1) The distribution of number of prey depends on the functional form of $p(x)$, but Eq. (9) will still be obtained for any $p(x)$ decaying exponentially as C (or \bar{x}) tends to zero.
- (2) The distribution of number of predators does not depend on the specific form of $p(x)$. The only requirement is that the connectance $C = \bar{x}/2$ tends to zero under some limit, so that $2z = SC$ remains finite when $S \rightarrow \infty$.

Thus, it is safe to conclude that our findings are robust under quite general conditions, a result that is not possible to obtain without an analytic treatment of the problem. Moreover, as we will show in the next section, the quantitative patterns uncovered for the niche model (Table 1) are important guides for the study of empirical food webs.

3 Patterns in food web structure

In this section, we compare the predictions of our theoretical solution of the niche model with data from empirical food webs obtained from a variety of environments. Remarkably, we find that the quantitative patterns uncovered for the niche model also describe the properties of food webs pertaining to very diverse habitats, such as freshwater, marine-freshwater interfaces, and terrestrial ecosystems.

3.1 Cumulative distributions

The reported empirical food webs generally contain a small number of trophic species. This fact implies that the empirical distributions of the number of prey and number of predators will be quite noisy. For this reason, we consider here the cumulative distributions instead of probability density functions.

We first derive an analytical expression for the cumulative distribution of number of prey, $P_{\text{prey}}(> k) = \sum_{k' \geq k} p_{\text{prey}}(k')$, in the limit of large sizes and small connectances. Since $1/2z$ is generally quite smaller than 1, one can substitute the sum by an integral with negligible error, yielding

$$\begin{aligned} P_{\text{prey}}(> k) &= E_2 \left(\frac{k}{2z} \right) \\ &= \exp \left(-\frac{k}{2z} \right) - \frac{k}{2z} E_1 \left(\frac{k}{2z} \right), \end{aligned} \quad (48)$$

which, similarly to Eq. (9), predicts an exponential decay for large k . In terms of the scaled variable $\tilde{k} = k/2z$, we obtain

$$P_{\text{prey}}(> \tilde{k}) = \exp \left(-\tilde{k} \right) - \tilde{k} E_1 \left(\tilde{k} \right), \quad (49)$$

which, as Eq. (9), contains *no free parameters*.

Next, we derive an analytical expression for the cumulative distribution of number of predators.

In the limits $S \gg 1$ and $C \ll 1$, one can use Eq. (14) to obtain

$$P_{\text{pred}}(> m) = \frac{1}{2z} \sum_{m'=m}^{\infty} \gamma(m' + 1, 2z), \quad (50)$$

As we have already noted, for $m < 2z$ the gamma function can be approximated as $\gamma(m + 1, 2z) \simeq 1$. We can then rewrite Eq. (50) as

$$P_{\text{pred}}(> m) = 1 - \frac{1}{2z} \sum_{m'=0}^{m-1} \gamma(m' + 1, 2z) \simeq 1 - \frac{m}{2z}, \quad m < 2z. \quad (51)$$

For $m \geq 2z$, Eq. (51) decays as the error function (Gradshteyn and Ryzhik, 2000).

Next, we analyze the empirical data for fifteen food webs with 25 to 155 trophic species. These webs have linkage densities $1.6 < z < 17.7$, and connectances in the interval 0.026–0.315 (Williams and Martinez, 2000; Dunne et al., 2002). We first investigate the distributions of number of prey and number of predators. In Figs. 10–12 we compare the cumulative distributions of the number of prey, number of predators, and number of trophic links for species in the fifteen food web investigated with the numerical predictions of the model for their specific values of S and z .

To quantify the agreement between the empirical data and the model, we apply the Kolmogorov-Smirnov (KS) test to the empirical distributions—prey, predator, and number of trophic links—and 1000 realizations of the model. The average output of the KS test is plotted in Fig. 13. We regard $\overline{P}_{\text{KS}} \leq 0.01$ —shown in black in Fig. 13—as strong evidence for the rejection of the null hypothesis. Our results suggest that eleven of the fifteen food webs studied are well approximated by the niche model: Bridge Brook, Skipwith, Coachella, Caribbean Reef, Benguela, St. Martin, Shelf, Chesapeake, St. Marks, Little Rock, and Grassland. The remaining four—El Verde, Canton, Ythan, and Stony Stream—exhibit rather different behavior, which is visually apparent in Figs. 10–12 and confirmed by the results in Fig. 13.

Equations (48)–(51) and the results of Figs. 10–13 suggest the possibility that, to first approximation, P_{prey} and P_{pred} obey universal functional forms that depend only on z . In order to further investigate whether these patterns of similarity are true, as well as if these distributions are universal, we perform the Kolmogorov-Smirnov test between all pairs of food webs. To do this we take advantage of the scaling determined earlier and use the scaled variables as the basis for comparison. Equation (49) predicts that $P_{\text{prey}}(> k)$ depends only on $k/2z$. The scaling of $P_{\text{pred}}(> m)$ is not as straightforward. As discussed before, “true” scaling holds only for $m/2z < 1/2$, while for larger values of $m/2z$ there is a Gaussian decay of the probability function with an explicit dependence on z . However, the decay for $m > z$ is quite fast and, to first approximation, not very relevant.

We show the results of these tests in Fig. 14. We use the same rejection criterion as before. The results of Fig. 14 also support the hypothesis that there are eleven food webs that conform to the theoretical predictions.

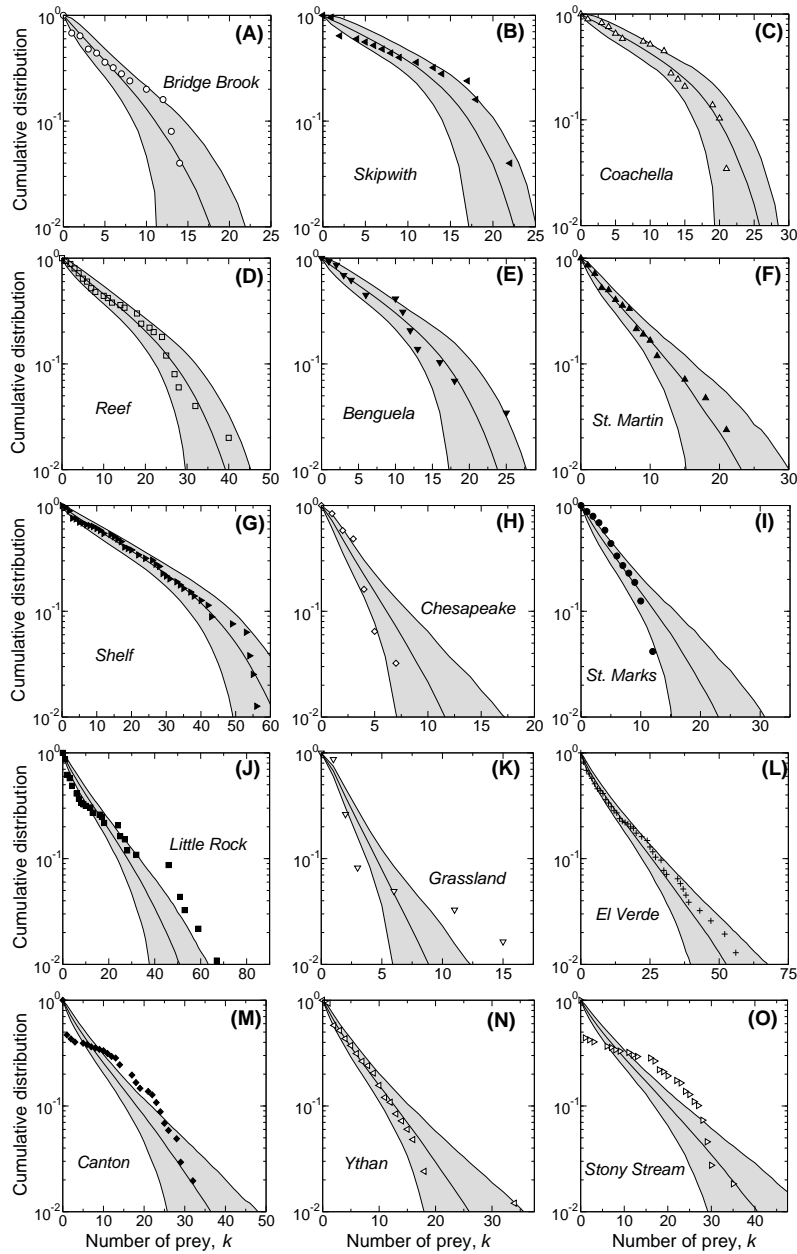


Fig. 10. Cumulative distribution P_{prey} of number k of prey for the fifteen food webs studied: Bridge Brook Lake (Havens, 1992); Skipwith Pond (Warren, 1989); Coachella Valley (Polis, 1991); Caribbean Reef (Opitz, 1996); Benguela (Yodzis, 1998); St. Martin Island (Goldwasser and Roughgarden, 1993); Northeast US Shelf (Link, 2002); Chesapeake Bay (Baird and Ulanowicz, 1989); St. Marks Seagrass (Christian and Luczkovich, 1999); Little Rock Lake (Martinez, 1991); Grassland (Martinez et al., 1999); El Verde Rainforest (Waide and Reagan, 1996); Canton Creek (Townsend et al., 1998); Ythan Estuary (Hall and Raffaelli, 1991); and Stony Stream (Townsend et al., 1998). The solid black line represents the average value from 1000 simulations of the niche model and the grey region represents two standard deviations above and below the model's predictions.

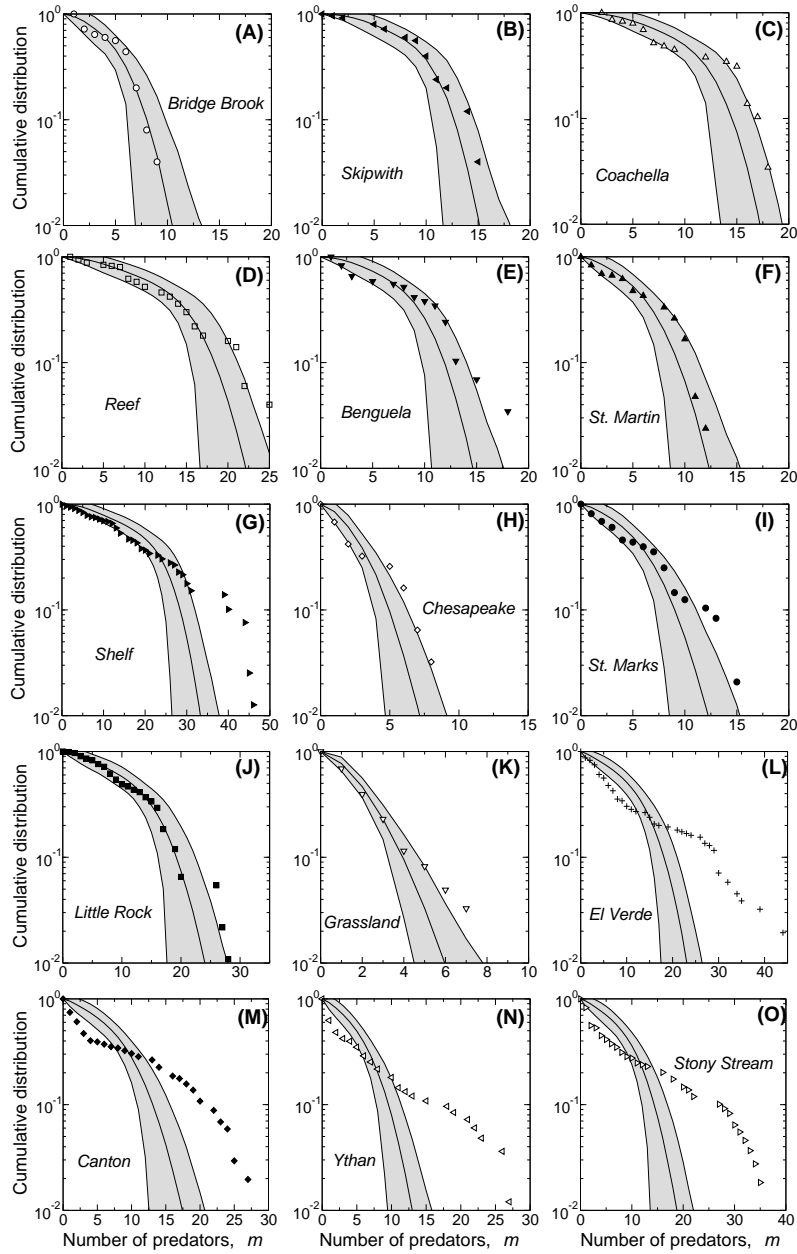


Fig. 11. Cumulative distribution P_{pred} of number m of predators for the fifteen food webs studied (see Fig. 10). The solid black line represents the average value from 1000 simulations of the niche model and the grey region represents two standard deviations above and below the model’s predictions.

For this reason, from this point on we will focus our attention upon these eleven food webs—Bridge Brook, Skipwith, Coachella, Reef, Benguela, St. Martin, Shelf, Chesapeake, St. Marks, Little Rock, and Grassland.

We plot in Figs. 15(A) and (B) the cumulative distributions $P_{\text{prey}}(> k)$ versus the scaled variable $k/2z$ for the eleven similar food webs and find that the data collapse onto a single curve, again supporting the possibility that P_{prey} obeys a universal functional form. Similarly we plot $P_{\text{pred}}(> m)$ versus the scaled variable $m/2z$ for the eleven similar food webs in Figs. 15(B) and (C), again finding a collapse of the data onto a single curve for $m/2z < 0.5$.

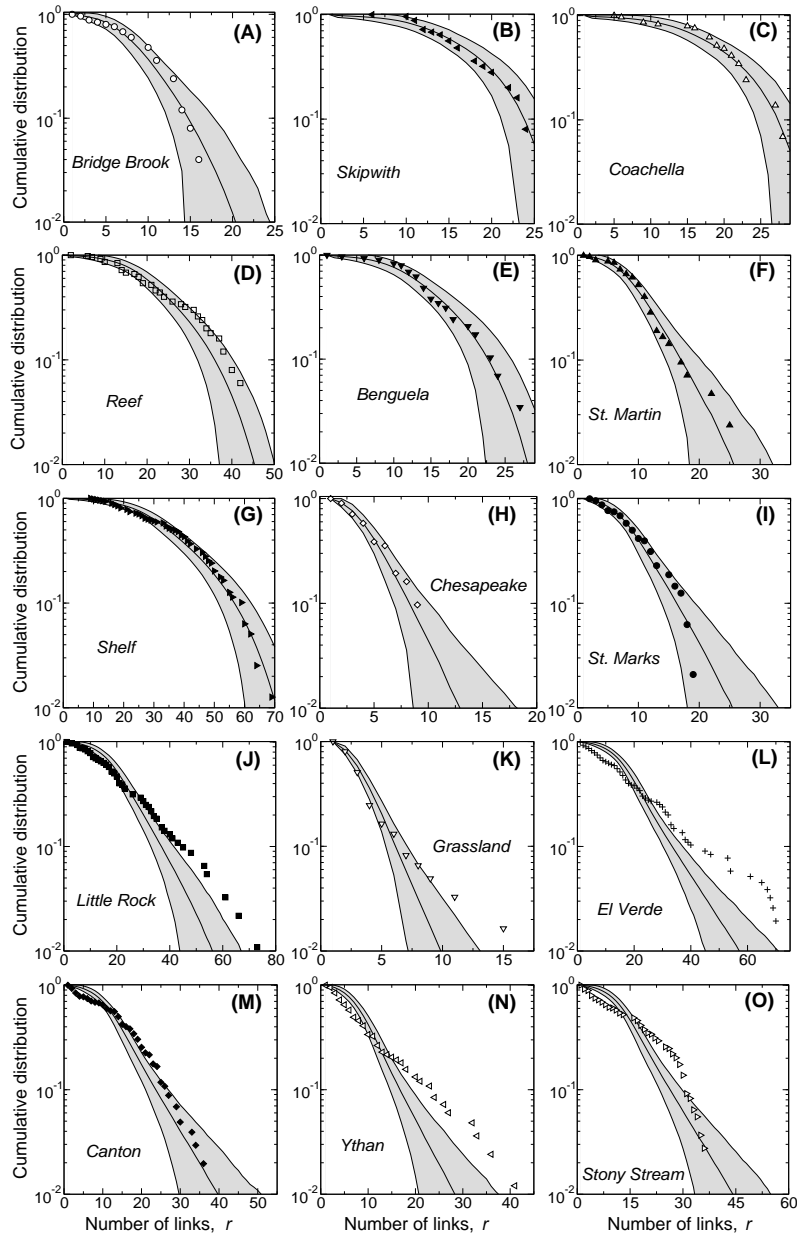


Fig. 12. Cumulative distribution P_{link} of the number r of trophic links for the fifteen food webs studied (see Fig. 10). The solid black line represents the average value from 1000 simulations of the niche model and the grey region represents two standard deviations above and below the model's predictions.

Figure 15 again supports the strong new hypothesis that the distributions of number of prey and number of predators follow *universal functional forms*. To improve statistics and better investigate the specific functional form of these distributions, one may pool the scaled variables, $k/2z$ and $m/2z$, from the different webs into single distributions, P_{prey} and P_{pred} , respectively. Figures 16(A) and (B) show the cumulative distributions of scaled number of prey and scaled number of predators for the pooled webs. Note that the distributions are well approximated by Eqs. (49) and (50) *even though there are no free parameters to fit in the analytical curves*. These results are analogous to the finding of scaling and universality in physical, chemical, and social systems (Stanley, 1971, 1999; Bunde and Havlin, 1994).

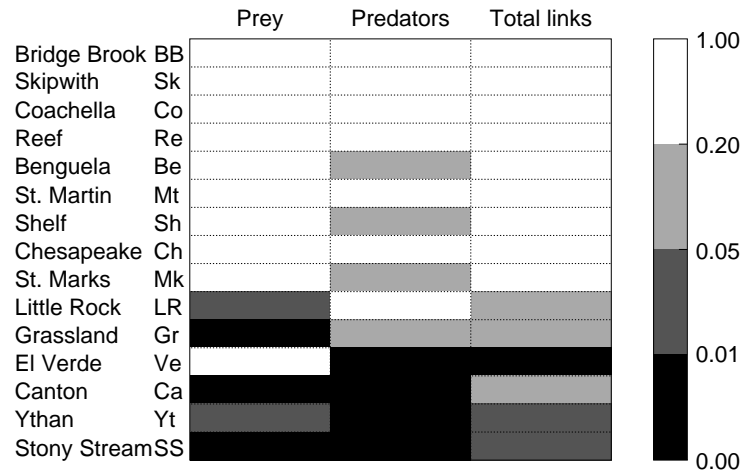


Fig. 13. Comparison of the distributions of prey, predators, and total number of links of the fifteen food webs to the respective distributions obtained from 1000 webs generated by the niche model. We use the KS test for the comparison. Column one is the result for distribution of prey, column two the result for distribution of predators, and column three the result for distribution of total number of links. We regard $\overline{P}_{KS} \leq 0.01$ —shown in black—as strong evidence for the rejection of the null hypothesis.

In Fig. 16 we plot the probability densities for the distribution of number of prey and number of predators for the pooled webs. It is visually apparent that both distributions are different. This is confirmed by the Kolmogorov-Smirnov test which rejects the null hypothesis at the 0.2% level. The distribution of number of prey decays exponentially while the distribution of number of predators is essentially a step function.

One can perform a similar analysis for the distribution p_{link} of the number of trophic links r . As for number of prey or number of predators, the data from the different webs, upon the scaling $r/2z$, collapse onto a single curve, further supporting the hypothesis that scaling holds for food web structure. To better determine the specific functional form of $p_{\text{link}}(r)$, in Fig. 16(D) we pool the scaled variables, $r/2z$, from the eleven food webs into a single distribution. We find that $p_{\text{link}}(r)$ has an exponential decay for $r/2z \gg 1$, in agreement with our theoretical calculations. Therefore, there is a characteristic scale for the linkage density, i.e. food webs do *not* have a scale-free structure, in contrast to erroneous reports in recent studies of food-web fragility (Solé and Montoya, 2001; Montoya and Solé, 2002).

The analytical expression for the distribution of the total number of links, Eq. (16), was derived assuming that the number of prey and number of predators are uncorrelated. In Fig. 17 we compare the empirical r_{cor} with data generated from simulations of the niche model. It can be seen that, just as in the niche model, food webs show, on average, slightly negative correlation between number of prey and number of predators. It is also important to note that for both the empirical data and the niche model, there is a wide range of possible correlation coefficients which may occur.

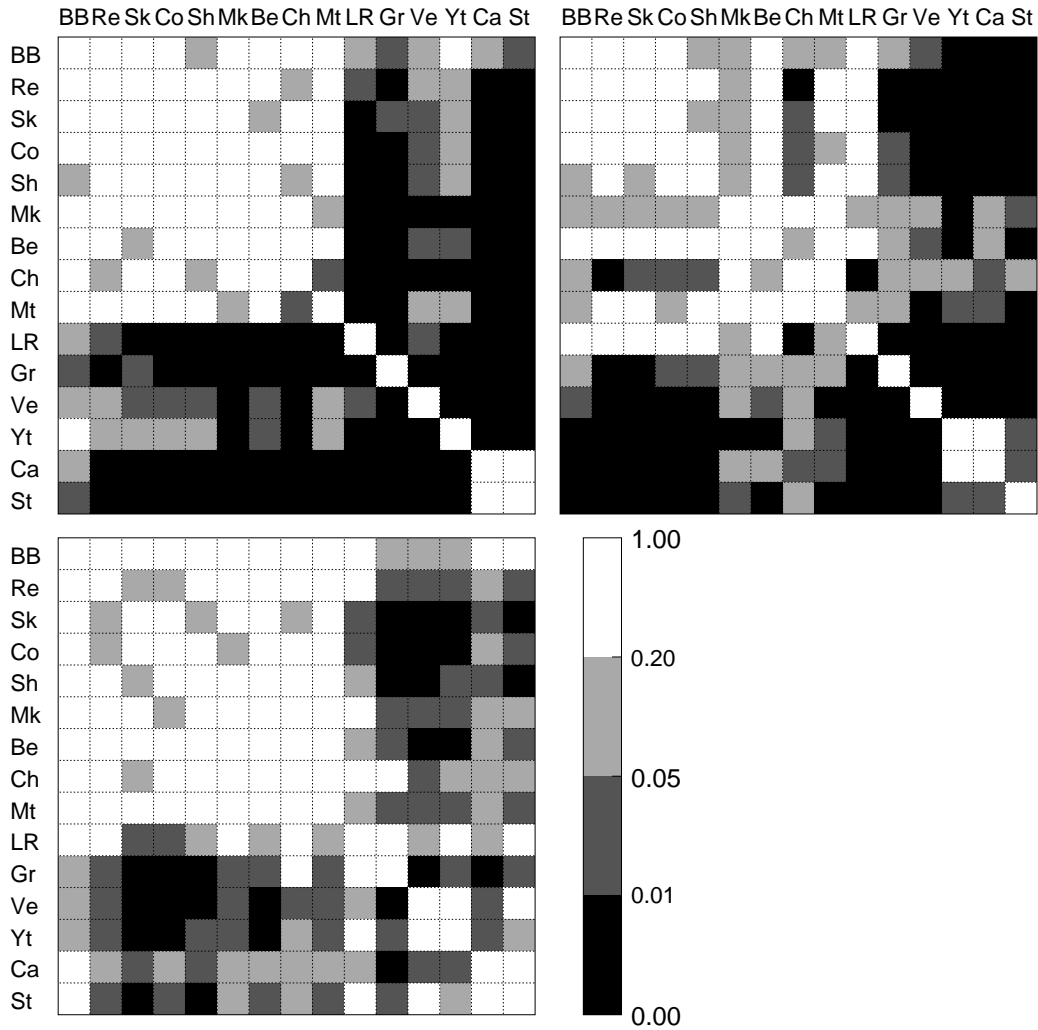


Fig. 14. Comparison of the distributions of prey, predators, and total number of links of the fifteen empirical food webs. We use the KS test for the comparison. The resulting matrices may be interpreted as similarity matrices with values $0 \leq P_{KS} \leq 1$, the KS probability. We show the matrix for distribution of prey k in the top left, the matrix for distribution of predators m in the top right, and the matrix for distribution of total number of links r in the bottom left. We regard $\overline{P}_{KS} \leq 0.01$ —shown in black—as strong evidence for the rejection of the null hypothesis.

3.2 Assortativity

It has been reported that technological and biological networks are most commonly disassortive, in contrast to social networks which are most commonly assortative in nature (Newman, 2002). We analyze the assortativities of the fifteen empirical food webs investigated and compare this data to numerical simulations of the model. We confirm that food webs are mostly disassortative. However, the assortativities cover a wide range of values from moderately disassortative to slightly assortative. The numerical results from simulations of the niche model also show a similarly wide range of possible values, as shown in Fig. 18.

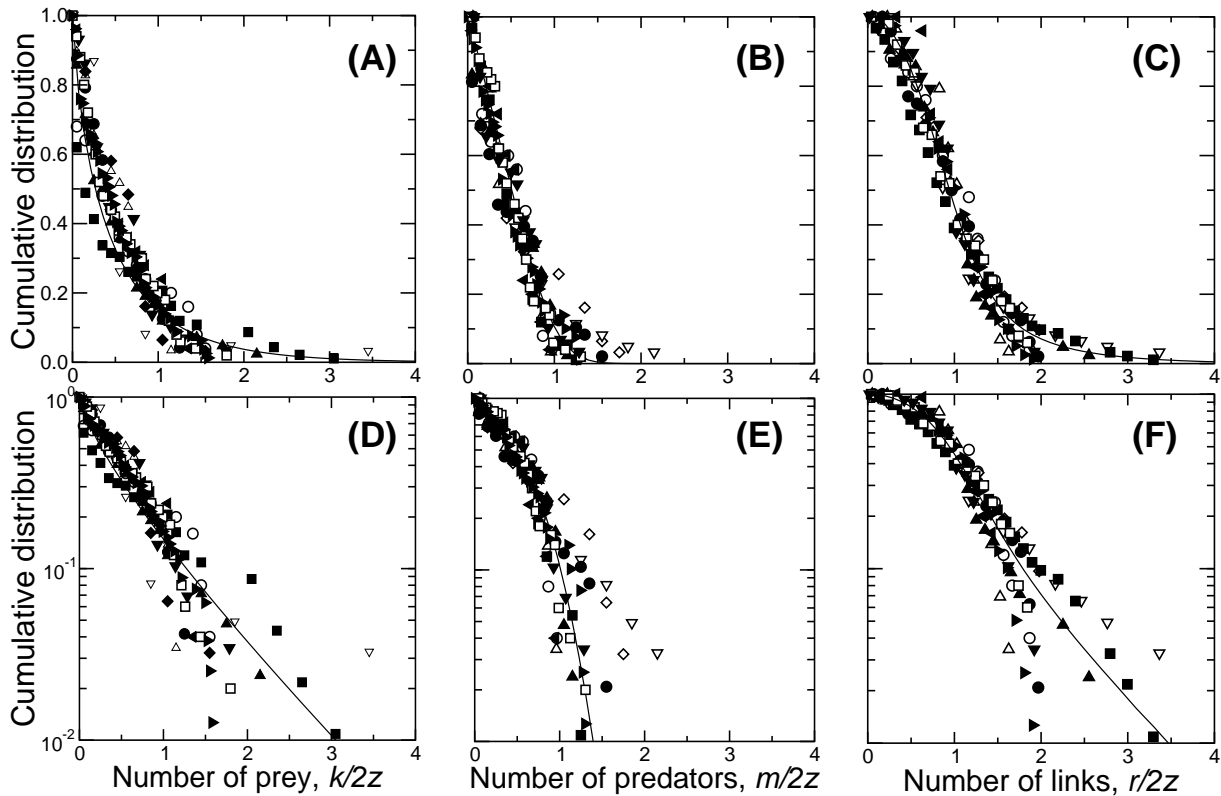


Fig. 15. Visual test of the “scaling hypothesis” that the distributions of number of prey, predators, and trophic links have the same functional form for different food webs. (A) Cumulative distribution P_{prey} of the scaled number of prey $k/2z$ for the eleven food webs. The solid line is the prediction of Eq. (48). The data “collapse” onto a single curve that agrees well with the analytical results. (B) Cumulative distribution P_{pred} of the scaled number of predators $m/2z$ for the eleven webs. The solid line is the analytical prediction of Eq. (50) for the average value of z in the empirical data, $z = 8.44$ (C) Cumulative distribution P_{links} of the scaled number of predators $r/2z$ for the eleven webs. The solid line is the analytical prediction. Semi-logarithmic plot of the scaled distributions of (D) number of prey, (E) number of predators, and (F) total number of links. The symbols are those introduced in Fig. 10.

3.3 Cannibalism

We determine the fraction of cannibal species as a function of the connectance C for the empirical webs considered, and compare it with the predictions of the niche and the random model analyzed in Appendix B. We find that the analytical expression for cannibalism supplied by the niche model captures the behavior of all fifteen food webs we study. In contrast, the predictions of the random model, shown in Fig. 19(B), cannot capture the behavior of Coachella or the Caribbean Reef webs.

It is observed that the empirical values are generally smaller than the predictions. Indeed, some food webs have no cannibals at all. Such low ratios of cannibalism is seen by some authors as unrealistic, and possibly due to the large effort needed to build a comprehensive food web (Polis, 1991).

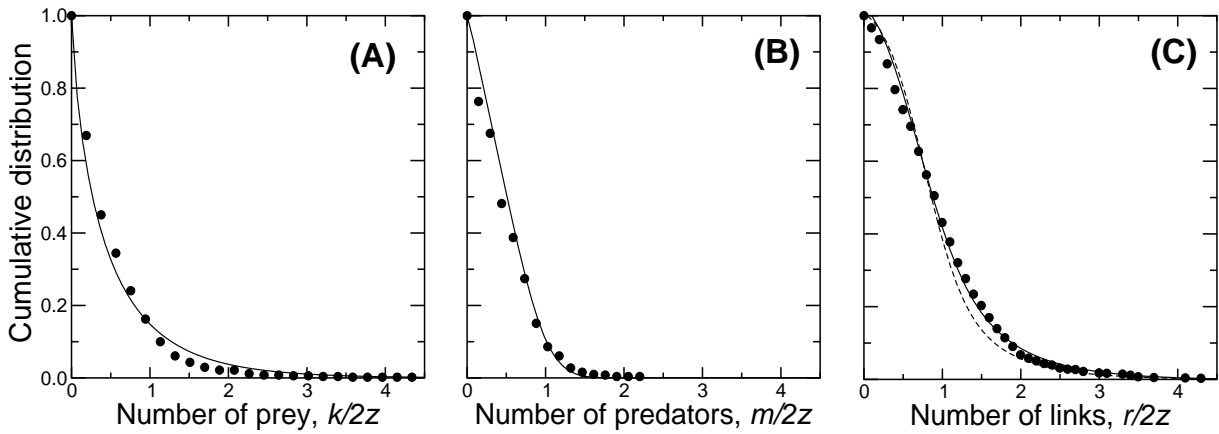


Fig. 16. Cumulative distributions (A) P_{prey} of the scaled number of prey, $k/2z$, and (B) P_{pred} of the scaled number of predators, $m/2z$, for the eleven pooled webs. The solid lines are, respectively, the analytical predictions Eq. (49) and (50), the latter with the average value $z = 8.44$. (C) Cumulative distribution of the number of trophic interactions per species $r = k + m$ for the eleven pooled webs. The solid line is obtained by numerically convolving the distributions Eq. (49) and (50) while the dashed line is obtained by numerical simulations of the niche model for $S = 511$ and $z = 8.44$, the parameter values of the pooled distributions. The tail of the distribution decays exponentially, indicating that food webs do *not* have a scale-free structure.

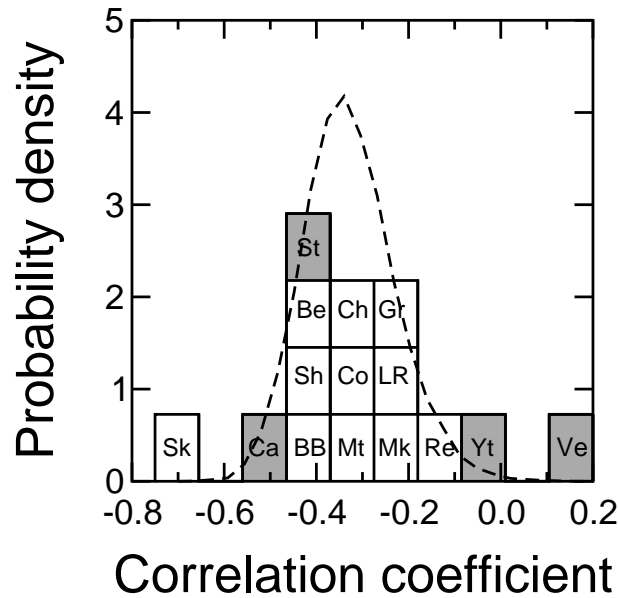


Fig. 17. Correlation coefficient between number of prey and number of predators for the fifteen food webs investigated. Probability density plot for the correlation coefficients obtained empirically compared to the probability density of r_{cor} from 5000 realizations of the niche model at the average properties of the empirical webs studied, $S = 64$ and $z = 8.06$. Note the good agreement between data and model predictions. The white boxes represent the eleven food webs which are well described by our analytical expressions and the niche model while the grey boxes represent the four food webs which are not.

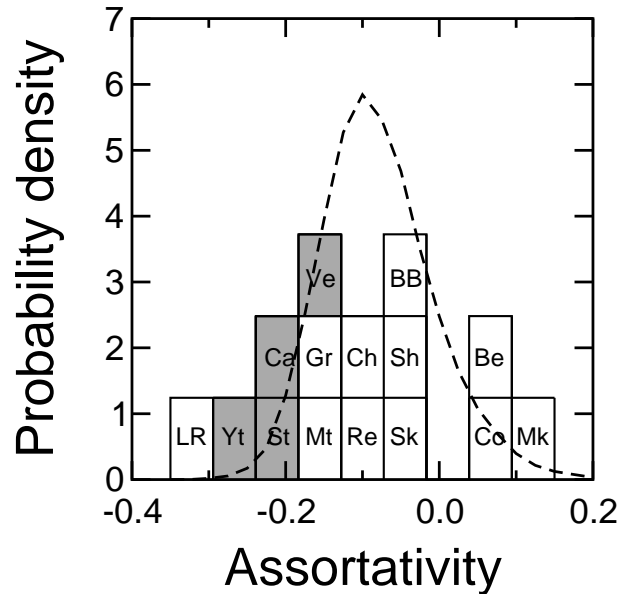


Fig. 18. Assortativity of the fifteen food webs investigated. Probability density plot for the assortativities obtained empirically compared to the probability density of assortativity from 5000 realizations of the niche model at the average properties of the empirical webs studied, namely $S = 64$ and $z = 8.06$. The food webs are, on average, slightly disassortative; however, they cover a wide range of values from moderately disassortative to slightly assortative and all of these behaviors are captured in the predictions of the model. Assortativities were calculated as outlined in (Newman, 2002). As in Fig. 17, the white boxes represent the eleven food webs which are well described by our analytical expressions and the niche model while the grey boxes represent the four food webs which are not.

3.4 Network theory measures

Next, we investigate if the scaling hypothesis suggested by the analysis of distribution of trophic links also applies to other quantities characterizing food web structure. We consider two quantities with ecologic implications: (i) the average trophic distance d between species (Watts and Strogatz, 1998), which is defined as the typical number of species needed to trophically connect two given species, and (ii) the clustering coefficient C (Watts and Strogatz, 1998) which quantifies the fraction of species triplets that form fully-connected triangles.

In Fig. 20(A) we compare our numerical results for the average trophic distance d for the niche model (Williams and Martinez, 2000) with the values calculated for the food webs analyzed. We observe that there is agreement between the niche model and the empirical data. Remarkably, the behavior predicted by the model also holds for randomization of the empirical data, where the randomization of links between species is such that the distributions of number of predators and number of prey remain the same. We also find that d increases with web size as $\log S$ both for the model and for the data.

The results of Fig. 20(A) also support the scaling hypothesis and suggest that the average distance in a food web may also follow a unique functional form for different food webs. This feature is predicted remarkably well by the niche model, even for the four webs for which the

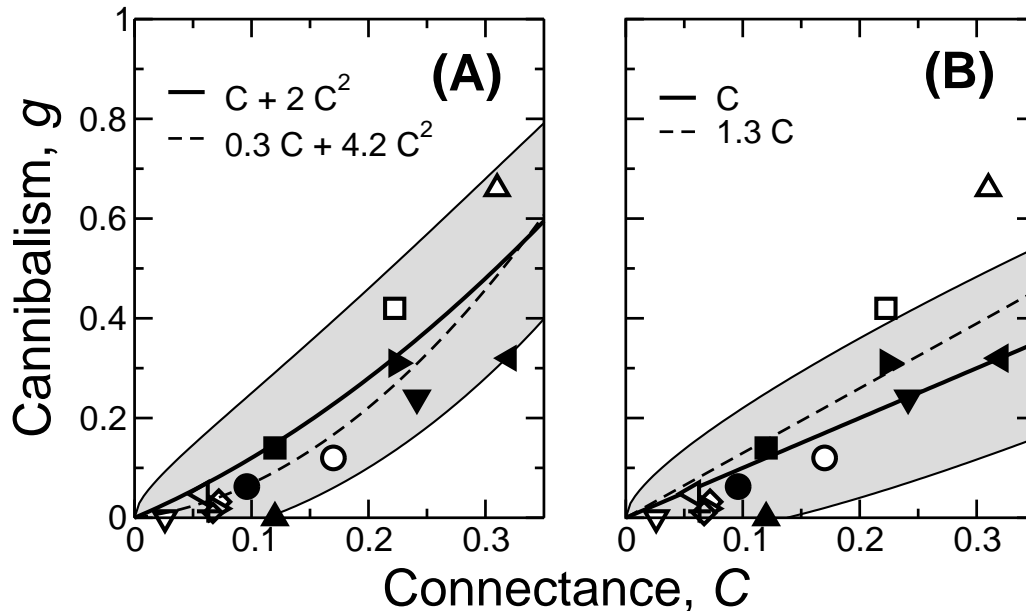


Fig. 19. Fraction of cannibal species as a function of the connectance C for empirical webs relative to (A) the niche model and (B) the random model. Thick solid lines correspond to model predictions Eqs. (24) and (B.1), respectively. The shaded area is within two standard deviations from the model average. One observes that the empirical results are in good agreement with the niche model. The dashed lines represent the best fit quadratic (A) and linear (B) expressions for the empirical data. The symbols are those introduced in Fig. 10 except that the four poorly approximated food webs are filled in grey.

model fails to describe other topological properties.

Figure 20(B) shows our results for the clustering coefficient \mathcal{C} of the food webs studied and for the niche model. We find that the data is well approximated by the model predictions which show that \mathcal{C} decreases to zero as $1/S$ as web size S increases. This result does not hold for Grassland—which has a much higher value than predicted—nor for Canton or Stony Stream—which feature much lower values than predicted. As for the average distance d , the niche model accurately predicts the values of the clustering coefficient even after the randomization of the empirical food webs. Again, this suggests that there are robust features inherent in the simple framework of the niche model which allows it to describe well the most complete empirical data available.

4 Concluding remarks

The major finding of this study is the uncovering of unifying quantitative patterns characterizing the structure of food webs from diverse environments. Specifically, we find that, for the majority of the most complete empirical food webs, the distributions of the number of prey, number of predators, and number of trophic links obey universal scaling functions, where the scaling quantity is the linkage density. Remarkably, these scaling functions are consistent with analytical predictions we derived for the niche model. Therefore, our results suggest that these

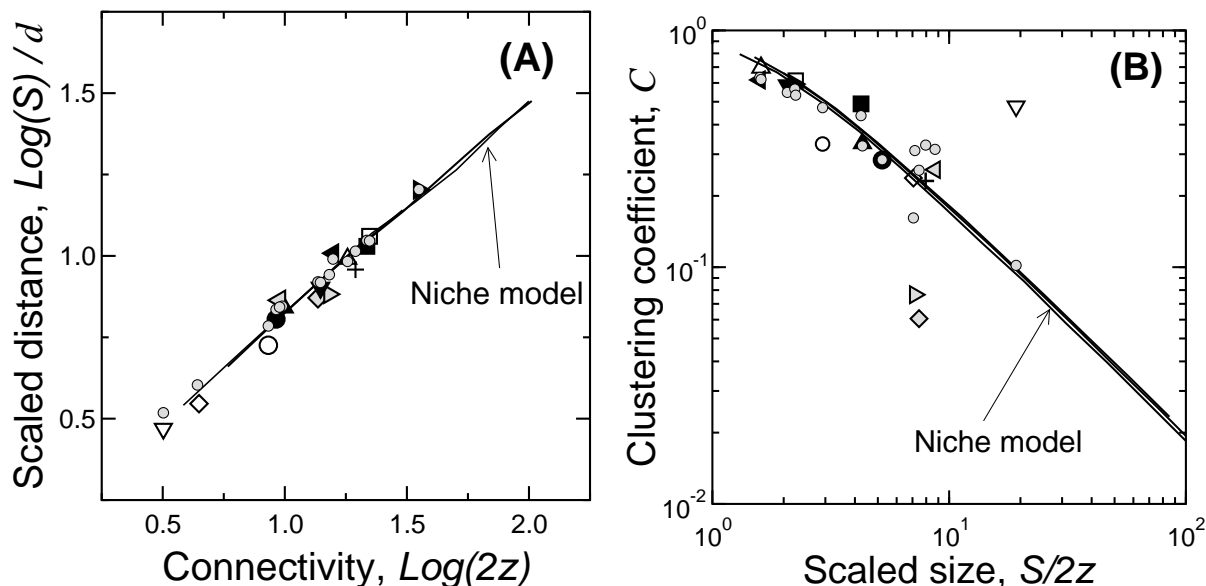


Fig. 20. (A) Scaled average trophic distance d between species versus linkage density z . We compare the data with the numerical simulations of the niche model (Williams and Martinez, 2000) for web sizes $S = 100, 500, 1000$ (thin solid lines). We find a logarithmic increase of the average distance with web size S for the empirical food webs, in good agreement with the model predictions. (B) Double-logarithmic plot of the clustering coefficient C versus the scaled web size $S/2z$. We compare the data with numerical results for the niche model (Williams and Martinez, 2000) for three values of the linkage density in the empirically-relevant range ($z = 2.5, 5, 10$). We find that the clustering coefficient of the food webs is inversely proportional to the web size S , in good agreement with the model predictions and with the asymptotic behavior predicted for a random graph (Watts and Strogatz, 1998). For both (A) and (B) the grey circles represent the average values calculated from 1000 randomizations of the links of the empirical food webs keeping the same distributions of number of prey and number of predators. Note that the behavior of these randomized webs is still captured quite well by the niche model implying that the underlying distributions themselves are responsible for this behavior. The symbols are those introduced in Fig. 10 except that the four poorly approximated food webs are filled in grey.

distributions can be theoretically predicted merely by knowing the food web's linkage density, a parameter readily accessible empirically.

Our results are of interest for a number of reasons. First, the results of Fig. 20—which also support the scaling hypothesis—indicate that there is very little, if any, compartmentalization in ecosystems (Pimm and Lawton, 1980), suggesting the possibility that ecosystems are highly interconnected and that the removal of any species may induce large disturbances. Second, regularities such as these are interesting as descriptors of trophic interactions inside communities because they may enable us to make predictions in the absence of high-quality data, and provide insight into how ecologic communities function and are assembled. Third, the structure of food webs is different from that of many other biological networks in two important aspects: the links are uni-directional and the in- and out-degree distribution are different. These two facts are a result of the *directed* character of the trophic interactions and of the asymmetry it creates. Interestingly, the niche model captures this asymmetry in its rules, which may explain its success in explaining the empirical results. Fourth, food webs do not have a scale free distribution of number of links (total, prey, or predators). This may be viewed as surprising

since one could expect most species to try to prey on the most abundant species in the ecosystem (an “abundant-get-eaten” kind of mechanism). Such a preferential attachment would lead to a scale-free distribution of links; instead, we find a single-scale distribution, suggesting that species specialize and prey on a small set of other species.

We do not intend to provide here a detailed study of why four of the fifteen food webs we consider do not accommodate to the patterns we report. We will, nonetheless, offer a few remarks on this matter. The four webs that do not conform to our patterns are Ythan Estuary, Canton Creek, Stony Stream, and El Verde Rainforest. Let us first consider the Ythan Estuary web. It has been noted already that this web displays an over-representation of top bird species (Williams and Martinez, 2000) which could account for the differences. For Canton Creek and Stony Stream two aspects distinguish them from the other food webs. First, they are quantitatively quite similar. This fact is illustrated by noting that the Kolmogorov-Smirnov test provides P_{KS} of 0.565, 0.045, and 0.794 under direct comparison of their respective distributions of number of prey, number of predators, and total number of links. Upon further investigation this is easily explained as the original authors stated plainly that it was their intention to compile food webs from habitats that were as similar as possible to each other (Townsend et al., 1998). Second, the Canton Creek and Stony Stream webs are also time specific (Townsend et al., 1998)—meaning their data was collected solely on one occasion—as opposed to cumulative—which are based upon data collected on multiple occasions until reaching some state of “completeness.” This time specific nature not only distinguishes them from all of the other food webs which we have studied, but also implies that they contain rather different information and are not directly comparable to cumulative food webs in our framework. We lastly address the El Verde Rainforest food web by pointing out that over one third of the links were not observed in the field but rather are based upon interactions involving closely related species in the forest or observations or published accounts of the interaction outside of the forest (Waide and Reagan, 1996).

To conclude, we want to stress that our findings are remarkable for two main reasons: (i) they hold for eleven out of fifteen of the most complete food webs studied, in contrast to previously reported patterns, and (ii) they support the conclusion that fundamental concepts of modern statistical physics such as scaling and universality—which were developed for the study of inanimate systems—may be successfully applied in the study of food webs—which comprise animate beings (Camacho et al., 2002*a,b*; Garlaschelli et al., 2003). Indeed, our results are consistent with the underlying hypothesis of scaling theory, i.e., food webs display “universal” patterns in the way trophic relations are established despite apparently “fundamental” differences in factors such as the environment (e.g. marine versus terrestrial), ecosystem assembly, and past history.

Acknowledgments

We thank J. Bafaluy, M. C. Barbosa, M. Barthélémy, J. Faraudo, G. Franzese, S. Mossa, R. V. Solé, H. E. Stanley, and especially N. D. Martinez and R. J. Williams for stimulating discussions

and helpful suggestions. We also thank J. E. Cohen, N. D. Martinez, R. J. Williams, and J. Dunne for making their electronic databases of food webs available to us. JC thanks the Spanish CICYT (BFM2003-06033) for support. RG thanks the Fulbright Program and the Spanish Ministry of Education, Culture & Sports. DBS thanks the NU ChBE Murphy Fellowship and the NSF-IGERT program "Dynamics of Complex Systems in Science and Engineering" (DGE-9987577).

Literature cited

- Albert, R., A.-L. Barabási, and H. Jeong. 2000. Error and attack tolerance in complex networks. *Nature*, **406**:378.
- Baird, D. and R. E. Ulanowicz. 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.*, **59**:329–364.
- Berlow, E. L. 1999. Strong effects of weak interactions in ecological communities. *Nature*, **398**:330–334.
- Briand, F. and J. E. Cohen. 1984. Community food webs have scale invariant structure. *Nature*, **307**:264–266.
- Briand, F. and J. E. Cohen. 1987. Environmental correlates of food chain length. *Science*, **238**:956–960.
- Bunde, A. and S. Havlin, editors. 1994. *Fractals in Science*. World Scientific, Singapore.
- Camacho, J., R. Guimerà, and L. A. N. Amaral. 2002a. Analytical solution of a model for complex food webs. *Phys. Rev. E*, **65**:030901(R).
- Camacho, J., R. Guimerà, and L. A. N. Amaral. 2002b. Robust patterns in food web structure. *Phys. Rev. Lett.*, **88**:228202.
- Chapin, F. S., E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Díaz. 2000. Consequences of changing biodiversity. *Nature*, **405**:234.
- Christian, R. R. and J. J. Luczkovich. 1999. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecological Modelling*, **117**:99–124.
- Cohen, J. E., F. Briand, and C. M. Newman. 1990. *Community Food Webs: Data and Theory*, volume 20 of *Biomathematics*. Springer, Berlin.
- Covington, W. W. 2000. Helping western forests heal. *Nature*, **408**:135.
- Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Food-web structure and network theory: The role of connectance and size. *Proc. Nat. Ac. Sci. USA*, **99**:12917–12922.
- Garlaschelli, D., G. Caldarelli, and L. Pietronero. 2003. Universal scaling relations in food webs. *Nature*, **423**:165–168.
- Goldwasser, L. and J. Roughgarden. 1993. Construction of a large Caribbean food web. *Ecology*, **74**:1216–1233.
- Gradshteyn, I. and I. M. Ryzhik. 2000. *Table of Integrals, Series, and Products*. Academic Press, New York, 6th edition.
- Hall, S. J. and D. Raffaelli. 1991. Food-web patterns: lessons from a species-rich web. *J. Anim. Ecol.*, **60**:823–842.
- Hall, S. J. and D. Raffaelli. 1993. Food webs: theory and reality. *Adv. Ecol. Res.*, **24**:187–239.
- Havens, K. 1992. Scale and structure in natural food webs. *Science*, **257**:1107–1109.

- Hutchings, J. A. 2000. Collapse and recovery of marine fishes. *Nature*, **406**:882.
- Lawton, J. H. and P. H. Warren. 1988. Static and dynamic explanations for patterns in food webs. *Trends Ecol. Evol.*, **3**:242–245.
- Link, J. 2002. Does food web theory work for marine ecosystems? *Marine Ecology Progress Series*, **230**:1–9.
- Martinez, N. 1993. Effects of scale on food web structure. *Science*, **260**:242–243.
- Martinez, N. D. 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monog.*, **61**:367–392.
- Martinez, N. D., B. A. Hawkins, H. A. Dawah, and B. P. Feifarek. 1999. Effects of sampling effort on characterization of food-web structure. *Ecology*, **80**:1044–1055.
- McKann, K. S. 2000. The diversity–stability debate. *Nature*, **405**:228–233.
- Montoya, J. M. and R. V. Solé. 2002. Small world patterns in food webs. *Journal of Theoretical Biology*, **214**:405.
- Newman, M. E. J. 2002. Assortative mixing in networks. *Phys. Rev. Lett.*, **89**:208701.
- Opitz, S. 1996. Trophic interactions in caribbean coral reefs. ICLARM Technical Reports, **43**:341.
- Pathria, R. K. 1972. *Statistical Mechanics*. Pergamon Press, New York.
- Pimm, S. L. and J. H. Lawton. 1980. Are food webs divided into compartments? *J. Animal Ecol.*, **49**:879–898.
- Pimm, S. L., J. H. Lawton, and J. E. Cohen. 1991. Food web patterns and their consequences. *Nature*, **350**:669–674.
- Pimm, S. L., G. J. Russell, and J. L. Gittleman. 1995. The future of biodiversity. *Science*, **269**:347–350.
- Polis, G. A. 1991. Complex trophic interactions in deserts: An empirical critique of food-web theory. *American Naturalist*, **138**:123–155.
- Rejmanek, M. and P. Stary. 1979. Connectance in real biotic communities and critical values for stability of model ecosystems. *Nature*, **280**:311.
- Solé, R. V., D. Alonso, and A. McKane. 2002. Self-organized instability in complex systems. *Phil. Trans. Royal Soc.*, **357**:667–681.
- Solé, R. V. and J. M. Montoya. 2001. Complexity and fragility in ecological networks. *Proc. R. Soc. B*, **268**:2039–2045.
- Stanley, H. E. 1971. *Introduction to Phase Transitions and Critical Phenomena*. Oxford University Press, Oxford.
- Stanley, H. E. 1999. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Rev. Mod. Phys.*, **71**:S358–S366.
- Stone, R. 1999. A plan to save Hawaii's threatened biodiversity. *Science*, **285**:817.
- Townsend, C. R., R. M. Thompson, A. R. McIntosh, C. Kilroy, E. Edwards, and M. R. Scarborough. 1998. Disturbance, resource supply, and food-web architecture in streams. *Ecology Letters*, **1**:200–209.
- Waide, R. B. and W. B. Reagan, editors. 1996. *The Food Web of a Tropical Rainforest*. University of Chicago Press, Chicago, IL.
- Warren, P. H. 1989. Spatial and temporal variation in a freshwater food web. *Oikos*, **55**:299–311.
- Watts, D. J. and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature*, **393**:440.

- Williams, R. J. and N. D. Martinez. 2000. Simple rules yield complex food webs. *Nature*, **404**:180–183.
- Yodzis, P. 1998. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *Journal of Animal Ecology*, **67**:635–658.

A Variables

Model variable	Variable definition
S	Number of trophic species
L	Total number of trophic links
C	Directed connectance, $C = L/S^2$
z	Linkage density, $z = L/S$
n	Niche number in niche model, $n \in [0, 1]$
a	Range of predation in niche model
x	Beta-distributed random variable
b	Characteristic parameter for beta-distribution of x

Ecological variable	Variable definition
k	Number of prey
\tilde{k}	Scaled number of prey, $\tilde{k} \equiv k/2z$
m	Number of predators
\tilde{m}	Scaled number of predators, $\tilde{m} \equiv m/2z$
r	Total number of trophic links, $r \equiv k + m$
\tilde{r}	Scaled total number of trophic links, $\tilde{r} \equiv r/2z$
T	Fraction of top species
B	Fraction of basal species
I	Fraction of intermediate species
g	Fraction of cannibals
σ_V	Standard deviation of vulnerability
σ_G	Standard deviation of generality
r_{cor}	Correlation coefficient between species number of prey and number of predators
A	Assortativity

B Fraction of cannibal species for a random linking model

In this Appendix we calculate the fraction of cannibal species for a random model, i.e. a model where the links between species are placed at random. For a system with S species and z average number of links per species, the probability that a species feeds on itself will be

$$g_{\text{rand}} = \frac{z}{S} = C. \quad (\text{B.1})$$

The standard deviation of g_{rand} is obtained from Eq. (26) by using $g = g_{\text{rand}}$, yielding

$$\sigma_{g_{\text{rand}}} = \sqrt{\frac{g_{\text{rand}}(1 - g_{\text{rand}})}{S}}. \quad (\text{B.2})$$

In the limit of small C , the predictions of both models are similar. Note that this occurs regardless of the particular form of $p_x(x)$. In fact, $C \rightarrow 0$ is equivalent to $\bar{x} \rightarrow 0$, so that $p_x(x)$ takes non-vanishing values only for very small x . One can then neglect $x/2$ versus 1 in the

denominator of Eq. (23) and get

$$g = \int_0^1 p_x(x) \frac{x/2}{1-x/2} dx \simeq \frac{\bar{x}}{2} = C = g_{\text{rand}}. \quad (\text{B.3})$$

Therefore, only the second term in the right hand side of expression Eq. (24) depends on the specific form of $p_x(x)$.

C Calculation of the moments of the distributions of number of prey and number of predators

In this Appendix we calculate averages of the type $\overline{k^i m^j}$, with i and j integers, in the limit of large web sizes and small connectances. These averages are used in the calculation of the assortativity, in Section 2.8.

In Section 2.7, we obtained Eq. (35) for the join probability $p(k, m)$, namely

$$p(k, m) = \int dn p(k, m, n) = \int dn p(k|n)p(m|n). \quad (\text{C.1})$$

With this, the average we want to calculate writes

$$\overline{k^i m^j} = \int_0^1 dn \left(\sum_{k=0}^S k^i p(k|n) \right) \left(\sum_{m=0}^S m^j p(m|n) \right) \equiv \int_0^1 dn \overline{k_n^i} \overline{m_n^j}. \quad (\text{C.2})$$

The probability $p(k|n)$ of a species with niche value n to have k prey can be calculated as follows. According to the rules of the niche model, in order to have k prey, the parameter x must take the value $x = k/Sn$. Then, by performing a change of variables, one obtains

$$p(k|n) = \frac{1}{nS} p_x\left(\frac{k}{nS}\right), \quad (\text{C.3})$$

where p_x is given by Eq. (1). In the limit of small C , p_x can be approximated by

$$p_x(x) \simeq b \exp(-bx) \simeq 2C \exp(-x/2C). \quad (\text{C.4})$$

Then, $p(k|n)$ is given by

$$p(k|n) = \frac{1}{2nz} \exp\left(-\frac{k}{2nz}\right). \quad (\text{C.5})$$

The average values $\overline{k_n^i}$ can be evaluated, in the limit of small C , as

$$\overline{k_n^i} = \int_0^{\infty} k^i \frac{1}{2nz} \exp\left(-\frac{k}{2nz}\right) = (2nz)^i i! \quad (\text{C.6})$$

In order to evaluate $\overline{m_n^j}$ we need $p(m|n)$. In the limit $S \gg 1$ and $C \ll 1$, the number of potential predators of a species of niche value n (that is, the number of species with a niche number larger than n) is $S(1-n)$. As seen in Section 2.3, each of them has a probability \bar{x} to prey on the species. Thus, the probability that a species with niche parameter n has m predators is simply the binomial distribution,

$$p(m|n) = \binom{S(1-n)}{m} \bar{x}^m (1-\bar{x})^{S(1-n)-m}, \quad (\text{C.7})$$

that in our limit yields the Poisson distribution, namely

$$p(m|n) = \frac{\lambda^m \exp(-\lambda)}{\lambda!}, \quad (\text{C.8})$$

with

$$\lambda = S(1-n)\bar{x} = 2z(1-n). \quad (\text{C.9})$$

One can now obtain the moments

$$\overline{m_n} = \lambda \quad (\text{C.10})$$

$$\overline{m_n^2} = \lambda^2 + \lambda \quad (\text{C.11})$$

$$\overline{m_n^3} = \lambda^3 + 3\lambda^2 + \lambda. \quad (\text{C.12})$$

Finally, using Eq. (C.6), (C.9) and (C.10)–(C.12) into Eq. (C.1), yields

$$\overline{k} = \overline{m} = z, \quad (\text{C.13})$$

$$\overline{k^2} = 8z^2/3, \quad (\text{C.14})$$

$$\overline{k^3} = 12z^3, \quad (\text{C.15})$$

$$\overline{m^2} = z + 4z^2/3, \quad (\text{C.16})$$

$$\overline{m^3} = z + 4z^2 + 2z^3, \quad (\text{C.17})$$

$$\overline{k m} = 2z^2/3, \quad (\text{C.18})$$

$$\overline{k m^2} = 2z^2(1+z)/3, \quad (\text{C.19})$$

and

$$\overline{k^2 m} = 4z^2/3. \quad (\text{C.20})$$