

## QUANTITATIVE PATTERNS IN THE STRUCTURE OF MODEL AND EMPIRICAL FOOD WEBS

D. B. STOFFER,<sup>1,3</sup> J. CAMACHO,<sup>1,2</sup> R. GUIMERÀ,<sup>1</sup> C. A. NG,<sup>1</sup> AND L. A. NUNES AMARAL<sup>1</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208 USA

<sup>2</sup>Departament de Física (Física Estadística), Universitat Autònoma de Barcelona, E-08193 Bellaterra, Catalonia, Spain

**Abstract.** Understanding the structure of food webs, and the mechanisms that give rise to this structure, is one of the most challenging problems in ecology. We analyze from a statistical physics perspective the network structure of model food webs and of 15 community food webs from a variety of environments, including freshwater, marine, estuarine, and terrestrial environments. We perform a theoretical analysis of two recently proposed models for food webs, the niche model of R. J. Williams and N. D. Martinez and the nested-hierarchy model of M.-F. Cattin et al. We find that the two models generate distributions of numbers of prey, predators, and links that are described by the *same* analytical expressions. Our analytical treatment reveals that a model's capacity to reproduce empirical data is principally determined by its ability to satisfy two conditions: (1) the species' niche values form a *totally ordered set* and (2) each species has a specific exponentially decaying probability of preying on a given fraction of the *species with lower niche values*. To test this hypothesis, we generalize the cascade model of J. E. Cohen and C. M. Newman so that it satisfies condition 2 and find that the new model is able to reproduce the properties of empirical food webs, validating our hypothesis. We use our analytical predictions as a guide to the analysis of 15 of the most complete empirical food webs available. We demonstrate that the quantitative unifying patterns that describe the properties of the food-web models considered earlier also describe the majority of the empirical webs considered. We find good agreement between the empirical data and the models for the average distance between species and the average clustering coefficient. Our results strongly support two hypotheses: first, that any model satisfying the two conditions we identify will accurately reproduce a number of the statistical properties of empirical food webs, and second, that the empirical distributions of number of prey and number of predators follow *universal functional forms* that, without free parameters, match our analytical predictions.

**Key words:** complex networks; food-web patterns; food-web structure; food webs, model vs. empirical; network structure; scaling; universality.

### INTRODUCTION

Species in an ecosystem are connected through trophic relationships to form highly complex networks, termed “food webs” (Briand and Cohen 1984, Cohen et al. 1990). Understanding the structure of food webs is of great importance because it provides insights into, for example, how ecosystems behave under perturbations (Berlow 1999, Chapin et al. 2000, McCann 2000). Recent increases in the availability of computational tools have facilitated the development of numerical simulations to explore these issues. In particular, several models have been proposed in the last two decades that aim to describe the structure of food webs (Cohen and Newman 1985, Williams and Martinez 2000, Cattin et al. 2004). Unfortunately, such simulation models alone are rarely able to bridge the gap between stochastic simulations and underlying mechanisms in a way that is useful or relevant to ecologists. Because

simulations are for the most part analytically inaccessible, it is difficult to understand exactly *why* a model behaves as it does, and how its behavior may be sensitive to the parameters chosen.

In an effort to shed new light on the factors behind the success, or failure, of food-web models, we present here a detailed study of three food-web models: the cascade model (Cohen and Newman 1985), the niche model (Williams and Martinez 2000), and the nested-hierarchy model (Cattin et al. 2004). We investigate the cascade model because it is perhaps the simplest numerical model that seeks to capture food-web structure using simple rules to define predator–prey interactions (Cohen and Newman 1985). We also select it because it fails to capture the structure of food webs with many species. By studying the differences between this model and more successful models we can elucidate the important mechanisms it lacks. We select the niche model because it is capable of accurately describing several statistical properties of empirical food webs from diverse environments, while maintaining computational simplicity (Williams and Martinez 2000). Finally, we consider the

Manuscript received 11 June 2004; revised 13 October 2004; accepted 20 October 2004. Corresponding Editor: M. Holyoak.

<sup>3</sup> E-mail: d-stouffer@northwestern.edu

TABLE 1. Definitions of the model and ecological variables used.

Variable	Definition
Model variables	
$S$	number of trophic species
$L$	total number of trophic links
$C$	directed connectance, $C \equiv L/S^2$
$z$	linkage density, $z \equiv L/S$
$n$	niche value, $n \in [0, 1]$
$a$	range of predation in niche model
$x$	probability of preying on other species, $x \in [0, 1]$
$\beta$	characteristic parameter for beta-distribution of $x$
Ecological variables	
$k$	number of prey
$\tilde{k}$	scaled number of prey, $\tilde{k} \equiv k/2z$
$m$	number of predators
$\tilde{m}$	scaled number of predators, $\tilde{m} \equiv m/2z$
$r$	number of links, $r \equiv k + m$
$\tilde{r}$	scaled number of links, $\tilde{r} \equiv r/2z$
$T$	fraction of top species
$B$	fraction of basal species
$\sigma_v$	standard deviation of vulnerability
$\sigma_G$	standard deviation of generality
$d$	average trophic distance between species
$C$	clustering coefficient

nested-hierarchy model because it claims to improve on the niche model by considering phylogenetic constraints and adaptation, thereby being more “ecologically sound” (Cattin et al. 2004). We employ an analytical treatment of these models, which allows us to go beyond their descriptions and identify the underlying mechanisms that are being implemented.

We demonstrate two simple conditions that are sufficient for a model to accurately predict a number of statistical properties of empirical food webs. Our results are consistent with the underlying hypothesis of scaling theory, which states: in the vicinity of the critical point, there exists a scaling power and a characteristic scaling variable  $\lambda$  such that a property  $f(x)$  of the system will obey the functional equation  $f(\lambda^\alpha x) = \lambda f(x)$  (Stanley 1999). When scaled, properties from diverse examples of such systems should collapse onto a single curve. Indeed, we find that food webs display universal patterns in the way trophic relations are established with  $\alpha = 1$  and  $\lambda = z$ , where  $z$  is the linkage density (see, e.g., Eqs. 4 and 6), despite what seem to be fundamental differences in ecosystem type (e.g., aquatic vs. terrestrial), ecosystem assembly, or past history.

The organization of the paper is as follows. In the first section, we expand on recently published analytical and numerical studies of the niche model (Camacho et al. 2002a, b) by comparing the niche model to the nested-hierarchy model and a generalized cascade model. In the second section, we analyze empirical food webs and show the existence of robust quantitative patterns that are predicted by the models.

#### ANALYTICAL TREATMENT

##### The niche model

We consider here ecosystems with  $S$  species and  $L$  trophic interactions between these species (Table 1;

also see Appendix A). In the niche model (Williams and Martinez 2000), one randomly assigns the  $S$  species to “trophic niches” with niche values  $n_i$  mapped uniformly onto the interval  $[0, 1]$ . Species can be ordered according to their niche value,  $n_1 < n_2 < \dots < n_S$ , that is, the niche values form a totally ordered set.

A species  $i$  is characterized by its niche value  $n_i$  and by its list of prey. Prey are chosen for all species according to the following rule (Appendix B, Fig. B1): A species  $i$  preys on all species  $j$  with niche value  $n_j$  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

$$p(x) = \beta(1-x)^{\beta-1}. \quad (1)$$

Williams and Martinez appear to have chosen this functional form for convenience, but, as we will show later, the predictions of the model are robust to changes in the specific form of  $p(x)$ . The values of the parameters  $\beta$  and  $S$  determine the linkage density  $z \equiv L/S$  of the food web and its directed connectance  $C \equiv L/S^2$ , as  $\beta = (1/2C) - 1$ .

Camacho et al. (2002a) derived analytical expressions for the distributions of numbers of prey, predators, and links for the niche model in the limits  $S \gg 1$  and  $C \ll 1$ . In particular, the analytical expression for the distribution of number  $k$  of prey is

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

where  $E_1(x)$  is the exponential-integral function (Gradstheyn and Ryzhik 2000).

Camacho et al. (2002a) showed further that the distribution of number of prey does not depend on the exact functional form of  $p(x)$  (which represents the probability of preying on a fraction  $x$  of species with lower niche values) as long as  $p(x)$  depends exponentially on  $x$ . This arises because the limit  $C \ll 1$  corresponds to  $\beta \rightarrow \infty$ . Thus,  $p(x)$  is negligible except when  $x \ll 1$ , where  $e^{-x} \simeq (1-x)$ ; in other words,

$$\lim_{\beta \rightarrow \infty} p(x) = \beta e^{-\beta x}. \quad (3)$$

Fig. 1 compares Eq. 1 and Eq. 3 for various values of  $C$ , confirming that the former tends to the exponential function of Eq. 3 for small  $C$ . Therefore, the distribution of number of prey in this limit (Eq. 2) does not depend on the particular form of  $p(x)$  as long as  $p(x)$  depends exponentially on  $x$ .

Furthermore, for any value of  $z$ , the scaled variable  $\tilde{k} = k/2z$  is distributed according to the same probability density function:

$$p_{\text{prey}}(\tilde{k}) = E_1(\tilde{k}). \quad (4)$$

Provided that  $C$  is small, this probability density function is therefore universal, that is, it is identical for any values of  $S$  and  $z$  (Camacho et al. 2002b).

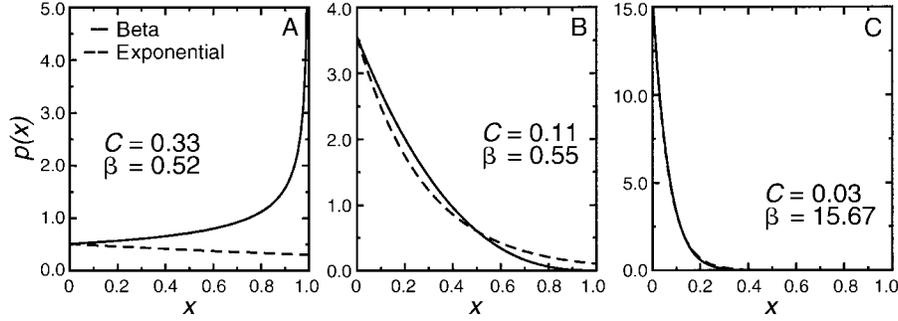


FIG. 1. Comparison between the beta distribution (Eq. 1) and the exponential distribution (Eq. 3), across the range of empirically observed values of  $C$ . For  $C \leq 0.11$ , the exponential distribution is a reasonable approximation of the beta distribution.

Camacho et al. (2002a) also derived an expression for the distribution of number  $m$  of predators, in the limits  $S \gg 1$  and  $C \ll 1$ :

$$p_{\text{pred}}(m) = (1/2z)\gamma(m + 1, 2z) \quad (5)$$

where  $\gamma$  is the incomplete gamma function (Gradshteyn and Ryzhik 2000). For  $m < 2z$ ,  $p_{\text{pred}}$  is approximately constant because  $\gamma(m + 1, 2z) \approx 1$ ; for  $m > 2z$ ,  $p_{\text{pred}}$  decays to zero with a Gaussian tail (Camacho et al. 2002a). Unlike the distribution of number of prey, the distribution of number of predators is completely independent of the form of  $p(x)$  (Camacho et al. 2002a).

Unlike the scaling seen for the distribution of number of prey, Eq. 5 is not simply a function of the scaled variable  $m/2z$ . However, for small values of  $m/2z$ ,  $\gamma$  is a constant and thus it does not depend on  $m$  or  $z$ . The probability density for the scaled variable  $\tilde{m} = m/2z$  is thus

$$p_{\text{pred}}(\tilde{m}) = \gamma(2z\tilde{m} + 1, 2z) \approx 1 \quad \tilde{m} < 1 \quad (6)$$

for any  $z$ . For  $\tilde{m} > 1$ ,  $p_{\text{pred}}(\tilde{m})$  decays quite rapidly (it is a Gaussian), so its exact functional form is not important.

If one neglects correlations between the number of prey and number of predators of a species, one can compute the probability density for the number of links using Eqs. 2 and 5. Under this approximation, the probability density for the number of links is simply given by the convolution of the two distributions. In the limits  $S \gg 1$  and  $C \ll 1$ , one has

$$p_{\text{links}}(r) = \int_0^r p_{\text{prey}}(t)p_{\text{pred}}(r - t) dt = \frac{1}{(2z)^2} \int_0^r E_1\left(\frac{t}{2z}\right)\gamma(r - t + 1, 2z) dt \quad (7)$$

which cannot be integrated analytically in terms of recognizable functions but can be calculated numerically.

Fig. 2 presents a comparison of the analytical expressions for the distributions of numbers of prey, predators, and links to simulations of the niche model. It is important to emphasize that the distributions of numbers of prey, predators, and links characterizing these diverse food webs obey *universal functional forms* that depend on a single parameter—the linkage density  $z$  of the food web (Camacho et al. 2002b). It is therefore

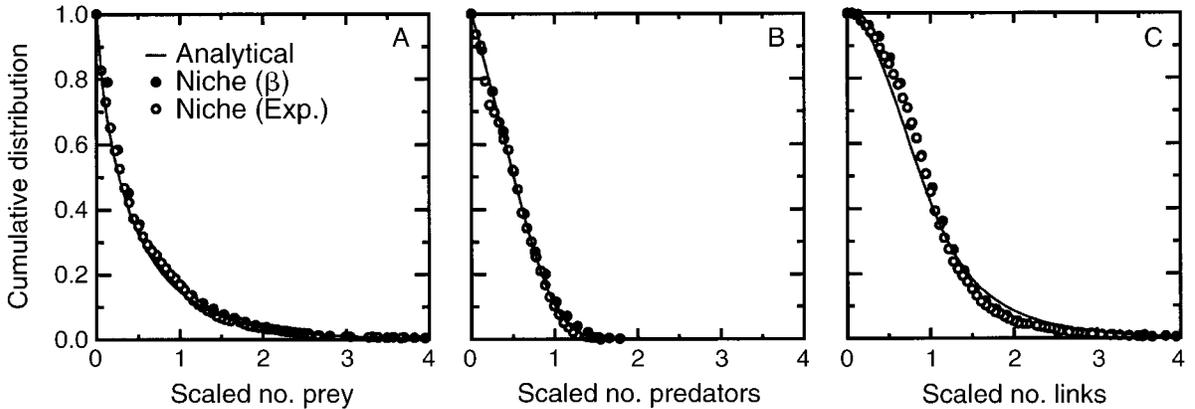


FIG. 2. Comparison between the analytical expressions and numerical simulations of the niche model for the cumulative distributions of scaled (A) number of prey,  $k/2z$ , (B) number of predators,  $m/2z$ , and (C) number of links,  $r/2z$ . The distributions are for a single realization of the niche model with  $S = 1000$  species and the linkage density  $z = 5$ . The model was implemented with  $p(x) = \beta(1 - x)^{\beta-1}$  (solid symbols) and  $p(x) = \beta e^{-\beta x}$  (open symbols).

TABLE 2. Summary of the analytical expressions obtained for the niche model in the limits  $S \gg 1$  and  $C \ll 1$ . These properties also hold for the nested-hierarchy and generalized cascade models in the same limits.

Property	Expression
Distribution of number of prey	$p_{\text{prey}}(k) = (1/2z)E_1(k/2z)$
Distribution of number of predators	$p_{\text{pred}}(m) = (1/2z)\gamma(m + 1, 2z)$
Distribution of number of links	$p_{\text{links}}(r) = \int_0^r p_{\text{prey}}(t)p_{\text{pred}}(r - t)dt$ $= 1/(2z)^2 \int_0^r E_1(t/2z)\gamma(r - t + 1, 2z)dt$
Fraction of top species	$T = (1 - e^{-2z})/2z$
Fraction of basal species	$B = \ln(1 + 2z)/2z$
Standard deviation of the vulnerability	$\sigma_v = \sqrt{1/3 + 1/z}$
Standard deviation of the generality	$\sigma_G = \sqrt{8/(3 + 6C) - 1}$

possible, given the linkage density of a web, to describe these distributions without any additional information, provided  $C \ll 1$ .

The importance of these distributions lies in the fact that the values of a number of ecologically relevant quantities may be estimated from them (Table 2). Indeed, our estimates compare quite well to simulations of the niche model (Camacho et al. 2002a). Moreover, they represent a significant fraction of the properties considered in the original validation of the cascade (Cohen and Newman 1985), niche (Williams and Martinez 2000), and nested-hierarchy (Cattin et al. 2004) models. Note that other quantities may also be derived in a straightforward manner for these models (Table C1 in Appendix C).

#### The nested-hierarchy model

The nested-hierarchy model claims to improve on the niche model by incorporating “phylogenetic constraints and adaptation” into the prey-selection process, rather than relying solely on niche theory (Cattin et al. 2004). This approach is said to better reflect the complexity and statistical properties of real food webs. We have examined the nested-hierarchy model in detail and demonstrate here that, though it appears to be quite different in its description, it nevertheless generates

webs characterized by the same *universal* distributions of numbers of prey, predators, and links described in the previous subsection.

In the nested-hierarchy model, the number of prey  $k_i$  of a species  $i$  is obtained by multiplying the predator’s niche value  $n_i$  by a value randomly drawn from the interval  $[0, 1]$  according to the beta distribution, *exactly* as in the niche model. The nested-hierarchy model thus generates the same distribution of number of prey as the niche model (Fig. 3A).

The selection of prey, and thus the distribution of number of predators, in this model is determined using a two-stage, multi-step process (Fig. B2 in Appendix B). In stage one, the first prey of species  $i$  is selected at random from among species with lower niche numbers than  $i$ . Let  $j$  be the first prey of  $i$ . If  $j$  is also a prey of another species, then the next prey of  $i$  is chosen from the pool of species eaten by “the group of  $j$ ’s consumers.” This group includes all consumers sharing at least one prey, with at least one of these consumers feeding on  $j$ . If the required number of links for species  $i$  cannot be satisfied by this pool, the remaining prey are chosen randomly from among the species with no predators that have niche values lower than  $i$ . The species enters stage two if all possible species with niche values lower than  $i$  have been selected and  $i$  requires additional

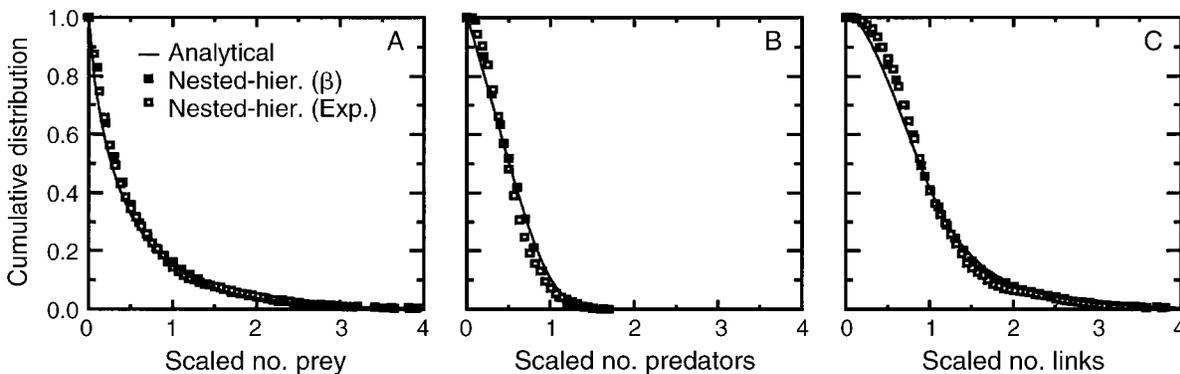


FIG. 3. Comparison between the analytical expressions and numerical simulations of the nested-hierarchy model for the cumulative distributions of scaled (A) number of prey,  $k/2z$ , (B) number of predators,  $m/2z$ , and (C) number of links,  $r/2z$ . The distributions are for a single realization of the nested-hierarchy model with  $S = 1000$  species and  $z = 5$ . The model was implemented with  $p(x) = \beta(1 - x)^{\beta-1}$  (solid symbols) and  $p(x) = \beta e^{-\beta x}$  (open symbols).

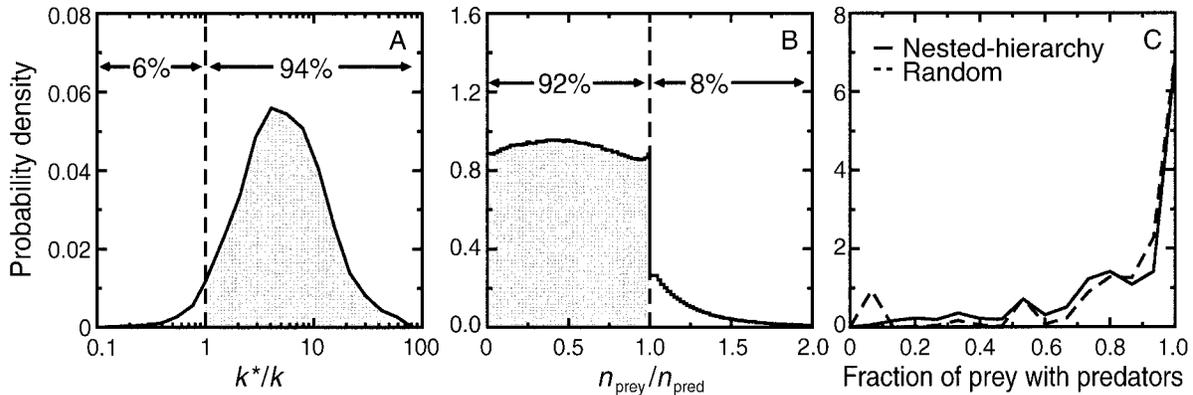


FIG. 4. Prey selection in the nested-hierarchy model for webs with  $S = 100$  species and  $z = 8$ . (A) Probability density function of number  $k^*$  of prey available, normalized by number  $k$  of prey desired. The prey available to species  $i$  is defined as: (1) species with niche values  $< n_i$  that have no predators, and (2) if species  $i$  has selected a prey with other predators, all members of the pool defined in Appendix B (Fig. B2) that have a lower niche value than species  $i$ . Our results demonstrate that a species  $i$  will consume species with niche values greater than or equal to its own only 6% of the time, explaining the very low rate of cannibalism seen in the nested-hierarchy model. (B) Probability density function of  $n_{\text{prey}}/n_{\text{pred}}$ , where  $n$  is the niche value, for links created in the nested-hierarchy model. Of all links, 92% have prey with niche values lower than the predator, with an *approximately uniform distribution* for niche values lower than that of the predator, indicating that species in this region are *effectively chosen at random*. Consuming species with larger niche values represent only 8% of all links. (C) Probability density function of the fraction of selected prey that already had at least one predator for the nested-hierarchy model and a random-selection model. The resulting distributions are almost indistinguishable, explaining why, despite the intended “phylogenetic constraints,” the nested-hierarchy model is, in fact, a model implementing a random selection of species with lower niche values, which shows no bias toward species having either many or few predators.

prey. In this stage prey are chosen from among species with niche values greater than or equal to the niche value of  $i$ .

In the nested-hierarchy model, preferentially consuming a species to which a species  $i$  is linked via a consumer group is intended to mimic phylogenetic constraints. Unfortunately, as demonstrated in Fig. 4, because the prey pool is built by sampling species at up to *four degrees of separation* from the original predator—that is, potential prey are chosen from among species that are separated from the predator by up to four trophic links—it turns out that species undergoing this process will be effectively picking their prey randomly from among *all species with lower niche values*. This explains why, despite apparent differences in the prey selection rules for the nested-hierarchy model, the distribution of number of predators and subsequently number of links are described by the same *universal forms* derived for the niche model (Fig. 3B, C).

#### The generalized cascade model

The results in the two previous subsections are congruent with the existence of two conditions that are sufficient for a model to reproduce the properties of empirical food webs:

*Condition 1:* The niche values to which species are assigned form a totally ordered set.

*Condition 2:* Each species has a specific probability  $x$  of preying on species with lower niche values, where  $x$  is drawn from an approximately exponential distribution.

In the following, we test the hypothesis that any model satisfying these two conditions will generate the

same distributions of trophic connections as the niche and nested-hierarchy models. To this end, we generalize the cascade model of Cohen and Newman (1985) so that it satisfies Condition 2.

In the cascade model of Cohen and Newman (1985), a species  $j$  with  $n_j < n_i$  becomes a prey of  $i$  with fixed probability  $x_0 = 2CS/(S - 1)$ . Williams and Martinez (2000) demonstrated that this model is not able to reproduce the properties of real food webs. Our analytical results for the niche and nested-hierarchy models suggest that the reason why the cascade model does not reproduce the data is that it does not satisfy Condition 2. To solve this problem, we generalize the cascade model in the following manner. A species  $i$  preys on species  $j$  with  $n_j \leq n_i$ , with a species-specific probability  $x$  drawn—from the beta distribution or an exponential distribution—from the interval  $[0, 1]$  (Appendix B, Fig. B3). Note that the idea of a predator-specific—or predator-dominant— $x$  was already discussed by Cohen (1990), though not in the manner proposed here.

Our generalization of the cascade model is similar to the implementation of the niche model but without the constraint that predation occurs within a continuous niche range. Significantly, we find that the generalized cascade model generates the same distributions of numbers of prey, predators, and links as those generated by the niche and nested-hierarchy models (Fig. 5).

#### The importance of the distributions

As noted previously, the original testing of the models by Cohen and Newman (1985), Williams and Martinez (2000), and Cattin et al. (2004) relied on the com-

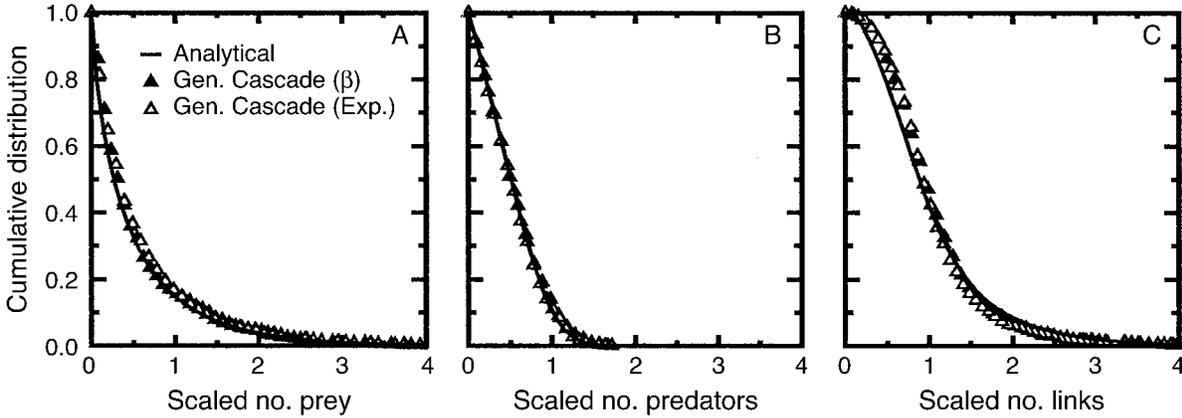


FIG. 5. Comparison between the analytical expressions and numerical simulations of the generalized cascade model for the cumulative distributions of scaled (A) number of prey,  $k/2z$ , (B) number of predators,  $m/2z$ , and (C) number of links,  $r/2z$ . The distributions are for a single realization of the generalized cascade model with  $S = 1000$  species and  $z = 5$ . The model was implemented with  $p(x) = \beta(1 - x)^{\beta-1}$  (solid symbols) and  $p(x) = \beta e^{-\beta x}$  (open symbols).

parison of the values of quantities for model-generated and empirical webs. In Table 2 we list a number of the quantities compared, which may be derived explicitly from the distributions of number of prey and number of predators.

It is important to stress, however, that this approach is far from ideal. The authors of these studies rely principally on comparisons between empirical data and properties derived from the distributions of numbers of prey and predators. We believe that it is more direct and elucidating to compare the distributions themselves. By following this approach and by solving the models analytically we are able to demonstrate *how and why* three seemingly distinct models accurately predict the statistical properties of empirical food webs.

#### PATTERNS IN FOOD-WEB STRUCTURE

In this section we compare data from empirical food webs with the predictions of the analytical solutions for the distributions of numbers of prey, predators, and links. Remarkably, we find that the quantitative analytical patterns describe the properties of empirical food webs pertaining to very diverse habitats, including freshwater, marine, estuarine, and terrestrial ecosystems.

##### *Distributions of numbers of prey, predators, and links*

Empirical food webs found in the literature generally contain a small number of trophic species (typically 25–155 species). This fact implies that the empirical distributions of the numbers of prey, predators, and links will be quite noisy. For this reason, we consider here the cumulative distributions, defined as

$$P(>y) = \sum_{y'=y}^{\infty} p(y') \quad (8)$$

instead of the probability density functions considered in the previous section.

Eq. 2 implies that the cumulative distribution of number of prey is given by the following:

$$P_{\text{prey}}(>k) = e^{-k/2z} - \frac{k}{2z} E_1\left(\frac{k}{2z}\right). \quad (9)$$

In terms of the scaled variable  $\tilde{k} = k/2z$ , we obtain

$$P_{\text{prey}}(>\tilde{k}) = e^{-\tilde{k}} - \tilde{k} E_1(\tilde{k}). \quad (10)$$

Like Eq. 4, this expression contains no free parameters and is *universal*. That is, it is independent of model details and parameter values.

Eq. 5 implies that the cumulative distribution of number of predators follows the form

$$P_{\text{pred}}(>m) = \frac{1}{2z} \sum_{m'=m}^{\infty} \gamma(m' + 1, 2z). \quad (11)$$

As previously noted, for  $m < 2z$  the incomplete gamma function  $\gamma$  can be approximated as  $\gamma(m + 1, 2z) \approx 1$ . One can therefore rewrite Eq. 11 as

$$\begin{aligned} P_{\text{pred}}(>m) &= 1 - \frac{1}{2z} \sum_{m'=0}^{m-1} \gamma(m' + 1, 2z) \\ &\approx 1 - \frac{m}{2z}, \quad m < 2z. \end{aligned} \quad (12)$$

In terms of the scaled variable  $\tilde{m} = m/2z$ , we obtain

$$P_{\text{pred}}(>\tilde{m}) = 1 - \tilde{m}, \quad \tilde{m} < 1. \quad (13)$$

For  $\tilde{m} \geq 1$ ,  $P_{\text{pred}}(>\tilde{m})$  decays to zero as the error function (Gradshteyn and Ryzhik 2000).

As in Eq. 8, the cumulative distribution of number of links is given by

$$P_{\text{links}}(>r) = \sum_{r'=r}^{\infty} p_{\text{links}}(r') \quad (14)$$

which cannot be determined analytically but may be computed numerically.

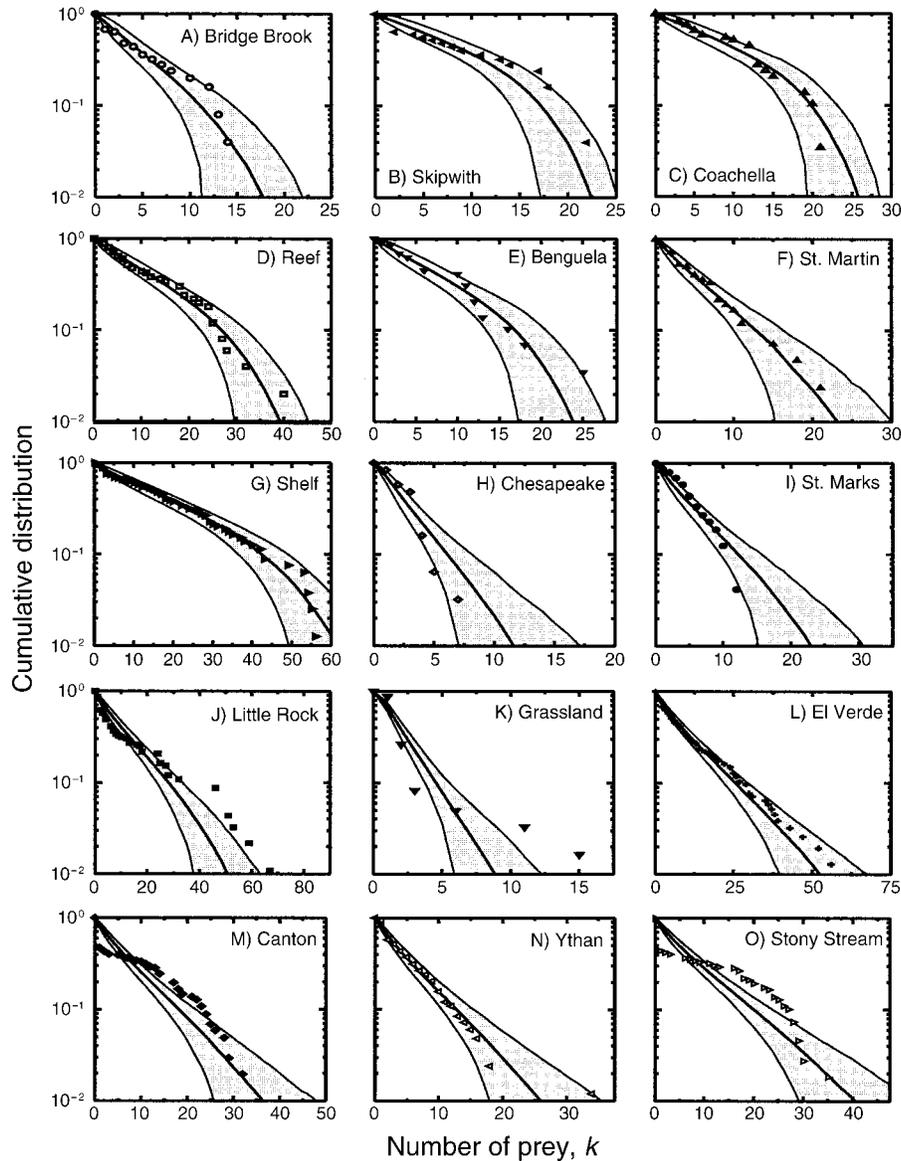


FIG. 6. Cumulative distribution  $P_{\text{prey}}$  of number of prey  $k$  for the 15 webs studied: Bridge Brook Lake (Havens 1992); Skipwith Pond (Warren 1989); Coachella Valley (Polis 1991); Caribbean Reef (Opitz 1996); Benguela (Yodzis 1998); St. Martin Island (Goldwasser and Roughgarden 1993); Northeast United States 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 gray region represents two standard deviations above and below the model's predictions.

We analyze the empirical distributions of numbers of prey, predators, and links (Fig. 6; see also Figs. D1 and D2 in Appendix D) for 15 food webs with 25 to 155 trophic species using several techniques. These webs have linkage densities  $1.6 < z < 17.7$ , and connectances in the interval 0.026–0.315.

In our first analysis, we quantify the agreement between the distribution of numbers of prey, predators, and links using the Kolmogorov-Smirnov (KS) test (Fig. 7). We use the niche model as a representative

model, but could have used any of the models described earlier and reached similar conclusions.

Our results suggest that 11 of the 15 food webs studied are well described by the models: 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 Fig. 6 (see also Figs. D1 and D2 in Appendix D) and confirmed by the

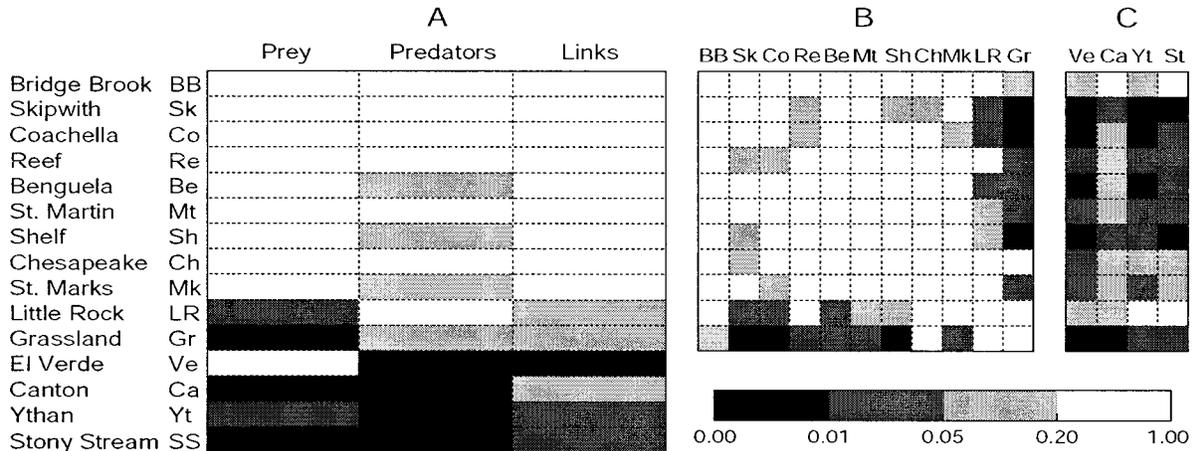


FIG. 7. (A) Comparison of the distributions of numbers of prey, predators, and links of the 15 food webs to the respective distributions obtained from 1000 webs generated by the niche model. (B) Comparison of the distributions of number of links for the 11 empirical food webs that we hypothesize obey universal properties. (C) Comparison of these 11 food webs with four food webs that we hypothesize do not obey universal properties. We use the Kolmogorov-Smirnov (KS) test for all comparisons. The results in panels (B) and (C) may be interpreted as similarity matrices with values  $0 \leq \bar{P}_{KS} \leq 1$ , the KS probability. We regard  $\bar{P}_{KS} \leq 0.01$ , shown in black, as strong evidence for the rejection of the null hypothesis.

results in Fig. 7A. The results of Figs. 6, D1, D2, and 7A suggest the possibility that models that satisfy our two conditions do indeed reproduce the properties of empirical food webs.

To test the hypothesis that the empirical distributions of numbers of prey, predators, and links follow universal functional forms, we quantify the agreement between all pairs of food webs using the KS test (Fig. 7B, C; see also Figs. D3 and D4 in Appendix D). We apply the  $\chi^2$  tests to the values from Fig. 7B and 7C to test their statistical significance. Remarkably, we find that we can accept the null hypothesis in the case of Fig. 7B at the 17% significance level while the null hypothesis is rejected for the case of Fig. 7C ( $p \approx 3.5 \times 10^{-35}$ ). Eqs. 9–14 and the results of Fig. 7B and 7C (and Figs. D3 and D4 in Appendix D) validate the

assertion that  $P_{\text{prey}}$ ,  $P_{\text{pred}}$ , and  $P_{\text{links}}$  obey universal functional forms (Camacho et al. 2002b).

For this reason, from this point on we will focus our attention on these 11 food webs: Bridge Brook, Skipwith, Coachella, Reef, Benguela, St. Martin, Shelf, Chesapeake, St. Marks, Little Rock, and Grassland. We plot in Fig. 8A the cumulative distributions  $P_{\text{prey}}(>k)$  vs. the scaled variable  $\tilde{k} = k/2z$  for the 11 similar food webs and find that the data collapse onto a single curve, supporting the possibility that  $P_{\text{prey}}$  obeys a universal functional form (Camacho et al. 2002b). We also plot  $P_{\text{pred}}(>\tilde{m})$  vs. the scaled variable  $\tilde{m} = m/2z$  for the 11 similar food webs in Fig. 8B finding a similar collapse of the data onto a single curve for  $m/2z < 1$  (Camacho et al. 2002b). Similarly, we plot in Fig. 8C  $P_{\text{links}}(>\tilde{r})$  vs. the scaled variable  $\tilde{r} = r/2z$ , again noting a collapse

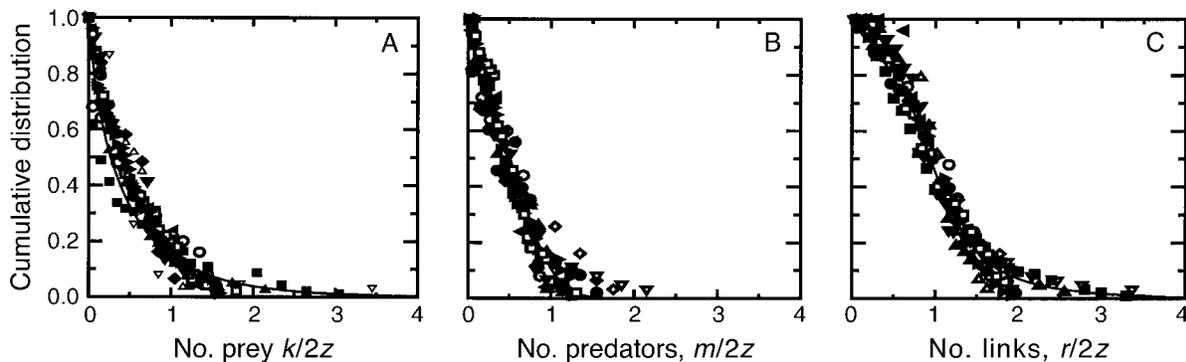


FIG. 8. Visual test of the “scaling hypothesis” that the distributions of numbers of prey, predators, and links have the same functional form for different food webs. (A) Cumulative distribution  $P_{\text{prey}}$  of the scaled number of prey for the 11 food webs. The solid line is the analytical prediction of Eq. 9. The data collapse onto a single curve consistent with our analytical results. (B) Cumulative distribution  $P_{\text{pred}}$  of the scaled number of predators for the 11 webs. The solid line is the analytical prediction of Eq. 11 for the average value of  $z$  in the empirical data,  $z = 8.44$ . (C) Cumulative distribution  $P_{\text{links}}$  of the scaled number of links for the 11 webs. The solid line is the prediction of Eq. 14. See Fig. 6 for source of the food-web symbols.

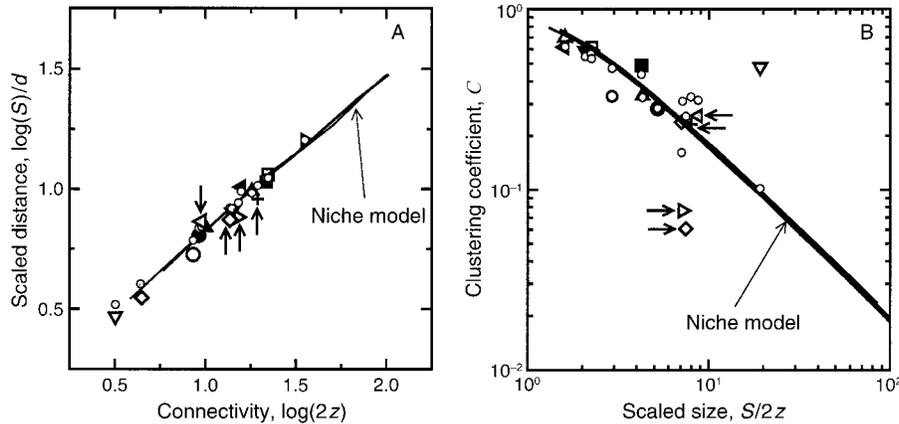


FIG. 9. (A) Scaled average trophic distance  $d$  between species vs. linkage density  $z$ . We compare the data with the numerical simulations of the niche model for web sizes  $S = 100, 500, 1000$  (thin solid lines). We find a logarithmic increase of the average distance with web size  $S$  for the empirical food webs, in good agreement with the model predictions. These results are consistent with the sensitivity of  $d$  to a web's directed connectance  $C$  and size  $S$ , as determined by Williams et al. (2002). (B) Clustering coefficient  $C$  vs. the scaled web size  $S/2z$ . We compare the data with numerical results for the niche model for three values of the linkage density in the empirically relevant range ( $z = 5, 10, 20$ ). 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. For both (A) and (B), the gray circles represent the average values from 1000 randomizations of the empirical food webs, keeping the distributions of number of prey and number of predators unchanged. Note that the behavior of these randomized webs is still captured by the niche model, implying that the underlying distributions themselves are responsible for this behavior. The symbols are those introduced in Fig. 6, except that the four poorly approximated food webs are filled in gray and indicated with arrows.

of the data onto a single curve, further supporting the hypothesis that scaling holds for food-web structure (Camacho et al. 2002b).

To improve statistics, and better investigate the specific functional form of these distributions, one may pool the scaled variables,  $k/2z$ ,  $m/2z$ , and  $r/2z$  from the different webs into single distributions,  $P_{\text{prey}}$ ,  $P_{\text{pred}}$ , and  $P_{\text{links}}$ , respectively. The cumulative distributions of the scaled numbers of prey, predators, and links for the pooled webs are well approximated by Eqs. 9, 11, and 14 even though there are no free parameters to fit in the analytical curves (Appendix D, Fig. D5). We find that  $P_{\text{links}}(r)$  has an exponential decay for  $r/2z \gg 1$ . There is, therefore, a characteristic scale of the linkage density. Therefore, 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).

#### Network theory measures

Next, we investigate whether or not the scaling hypothesis also applies to other properties characterizing food-web structure. We consider two quantities with ecological implications: (1) 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 (2) the clustering coefficient  $C$  (Watts and Strogatz 1998), which quantifies the fraction of species triplets that form fully connected triangles.

In Fig. 9A we compare our numerical results for the average trophic distance  $d$  for the niche model with the

values calculated for the food webs analyzed. We find that  $d$  increases with web size as  $\log(S)$  for both the model and the data. In Fig. 9B, we show that the clustering coefficient  $C$  of the food webs studied decreases to zero as  $1/S$  with increasing web size  $S$ , in good agreement with the asymptotic behavior predicted for a random graph (Watts and Strogatz 1998). Remarkably, the behavior predicted by the model also holds for the randomization of the empirical data, which leaves the distributions of number of prey and number of predators unchanged. This finding suggests that the underlying distributions are responsible for the observed behavior. This also explains why the nested-hierarchy and generalized cascade models, both of which are a sort of "randomized" niche model, provide similar predictive capabilities.

#### Why do model predictions only work for 11 of the 15 food webs?

The Canton Creek, Stony Stream, Ythan Estuary, and El Verde Rainforest food webs do not conform to the patterns we have reported. The lack of fit to our analytical expressions cannot be attributed to the fact that the expressions are derived for low values of the connectance  $C$ , since all 15 webs studied have  $C < 0.33$ .

Concerning the first two, Canton Creek and Stony Stream, we find two aspects that distinguish them from the other food webs studied. First, they are remarkably similar, as evidenced by the Kolmogorov-Smirnov test, which provides  $P_{\text{KS}}$  of 0.565, 0.045, and 0.794 under direct comparison of their respective distributions of numbers of prey, predators, and links, respectively.

This is explained by the fact that the original authors' stated intention was to compile food webs from habitats that were as similar as possible (Townsend et al. 1998). Second, these webs are based on data collected solely on one occasion, in contrast with the other cumulative webs, which are based on data collected or accumulated over time until reaching some state of "completeness." The data therefore contain rather different information and meaning relative to the other 13 webs.

With regard to the Ythan Estuary and El Verde Rainforest, questions have been raised concerning the data's accuracy. It has been pointed out that Ythan Estuary, for example, displays an over-representation of top bird species (Williams and Martinez 2000). In the El Verde Rainforest food web, over one third of the links were not observed in the field but rather are based on interactions involving closely related species in the forest or accounts of their interactions outside of the forest (Waide and Reagan 1996).

#### CONCLUDING REMARKS

We report two major findings:

1) There are two critical conditions that a food web model must satisfy in order to successfully predict the properties of empirical food webs: (1) the niche values to which species are assigned form a *totally ordered set* and (2) each species has a specific, exponentially decaying probability  $p(x)$  of preying on a fraction  $x$  of the species with lower niche values. It is remarkable that even an extremely simple model, such as our generalized cascade model, can achieve this.

2) We uncover unifying quantitative patterns characterizing the structure of empirical food webs from diverse environments. Specifically, we find that, for the majority of the most complete empirical food webs, the distributions of the numbers of prey, predators, and links obey universal scaling functions, where the scaling factor is the linkage density. Remarkably, these scaling functions are consistent with analytical predictions we derived for the niche, nested-hierarchy, and generalized cascade models. Therefore, our results suggest that these distributions can be theoretically predicted merely by knowing the food-web's linkage density, a parameter readily accessible empirically.

Our results are also of interest for a number of additional reasons. First, the results are insensitive to the precise distribution of niche values. If species are ranked according to body size or mass, in accord with prevailing theories, a uniform distribution of niche values is not plausible (Warren and Lawton 1987, Cohen 1989, Cohen et al. 1993, Neubert et al. 2000). However, provided our second condition is satisfied—that there is an exponentially decaying, species-specific probability of preying on a fraction of species with lower niche values—all that matters is whether the niche values can be ordered. Second, the results of Fig. 9—which also support the scaling hypothesis—indicate that there is very little, if any, compartmentalization

in communities (Pimm and Lawton 1980), suggesting the possibility that communities are highly interconnected and that the removal of any species may induce large disturbances (Pimm 1979, 1980, Pimm et al. 1991, Borrvall et al. 2000, Dunne et al. 2002). Third, 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 ecological communities function and are assembled. Fourth, food webs do *not* have a scale-free distribution of numbers of prey, predators, or links.

To conclude, we want to stress the three main reasons that our findings are significant: (1) we demonstrate that only two simple conditions are sufficient for a food-web model to accurately predict the statistical properties of empirical food webs, (2) they hold for 11 out of 15 of the most complete food webs studied, in contrast to previously reported patterns, and (3) they support the conclusion that fundamental concepts of modern statistical physics such as scaling and universality may be successfully applied in the study of food webs (Camacho et al. 2002a, b, Garlaschelli et al. 2003). Indeed, our results are consistent with the underlying hypothesis of scaling theory, that is, food webs display "universal" patterns in the way trophic relations are established despite apparently fundamental differences in factors such as the environment (e.g., aquatic vs. terrestrial), ecosystem assembly, and past history.

#### ACKNOWLEDGMENTS

We thank A. Moreira and M. Sales-Pardo 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. D. B. Stouffer acknowledges the Northwestern University ChBE Murphy Fellowship and NSF-IGERT "Dynamics of Complex Systems in Science and Engineering" (DGE-9987577). J. Camacho thanks the Spanish CICYT (BFM2003-06033) for support. R. Guimerà thanks the Fulbright Program and the Spanish Ministry of Education, Culture & Sports. C. A. Ng thanks the NSF Graduate Research Fellowship program. L. A. Nunes Amaral acknowledges a Searle Leadership Fund Award.

#### LITERATURE CITED

- Baird, D., and R. E. Ulanowicz. 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecological Monographs* **59**:329–364.
- Berlow, E. L. 1999. Strong effects of weak interactions in ecological communities. *Nature* **398**:330–334.
- Borrvall, C., B. Ebenman, and T. Jonsson. 2000. Biodiversity and the risk of cascading extinction in model food webs. *Ecology Letters* **3**:131–136.
- Briand, F., and J. E. Cohen. 1984. Community food webs have scale-invariant structure. *Nature* **307**:264–266.
- Camacho, J., R. Guimerà, and L. A. Nunes Amaral. 2002a. Analytical solution of a model for complex food webs. *Physical Review E* **65**, article number 030901(R).
- Camacho, J., R. Guimerà, and L. A. Nunes Amaral. 2002b. Robust patterns in food-web structure. *Physical Review Letters* **88**, article number 228102.
- Cattin, M.-F., L.-F. Bersier, C. Banašek-Richter, R. Baltensperger, and J.-P. Gabriel. 2004. Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427**:835–839.

- Chapin, F. S., III, E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavoie, 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 food web network through effective trophic levels. *Ecological Modelling* **117**: 99–124.
- Cohen, J. E. 1989. Food webs and community structure. Pages 181–202 in J. Roughgarden, R. M. May, and S. A. Levin, editors. *Perspectives in ecological theory*. Princeton University Press, Princeton, New Jersey, USA.
- Cohen, J. E. 1990. A stochastic theory of community food webs. VI. Heterogeneous alternatives to the cascade model. *Theoretical Population Biology* **37**:55–90.
- Cohen, J. E., F. Briand, and C. M. Newman. 1990. Community food webs: data and theory. Springer-Verlag, Berlin, Germany.
- Cohen, J. E., and C. M. Newman. 1985. A stochastic theory of community food webs. I. Models and aggregated data. *Proceedings of the Royal Society of London Series B* **224**: 421–448.
- Cohen, J. E., S. L. Pimm, P. Yodzis, and J. Saldaña. 1993. Body sizes of animal predators and animal prey in food webs. *Journal of Animal Ecology* **62**:67–78.
- Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters* **5**:559–567.
- 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.
- Gradstheyn, I. S., and I. M. Ryzhik. 2000. *Table of integrals, series, and products*. Sixth edition. Academic Press, New York, New York, USA.
- Hall, S. J., and D. Raffaelli. 1991. Food-web patterns: lessons from a species-rich web. *Journal of Animal Ecology* **60**: 823–842.
- Havens, K. 1992. Scale and structure in natural food webs. *Science* **257**:1107–1109.
- Link, J. 2002. Does food-web theory work for marine ecosystems? *Marine Ecology Progress Series* **230**:1–9.
- Martinez, N. D. 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecological Monographs* **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.
- McCann, 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**:405–412.
- Neubert, M. G., S. C. Blumenshine, D. E. Duplisea, T. Jonsen, and B. Rashleigh. 2000. Body size and food-web structure: testing the equiprobability assumption of the cascade model. *Oecologia* **123**:241–251.
- Opitz, S. 1996. Trophic interactions in caribbean coral reefs. ICLARM [International Center for Living Aquatic Resources Management] Technical Reports **43**:341.
- Pimm, S. L. 1979. Complexity and stability: another look at MacArthur's original hypothesis. *Oikos* **33**:357.
- Pimm, S. L. 1980. Food web design and the effect of species deletion. *Oikos* **35**:139–149.
- Pimm, S. L., and J. H. Lawton. 1980. Are food webs divided into compartments? *Journal of Animal Ecology* **49**:879–898.
- Pimm, S. L., J. H. Lawton, and J. E. Cohen. 1991. Food-web patterns and their consequences. *Nature* **350**:669–674.
- Polis, G. A. 1991. Complex trophic interactions in deserts: an empirical critique of food-web theory. *American Naturalist* **138**:123–155.
- Solé, R. V., and J. M. Montoya. 2001. Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London Series B* **268**:2039–2045.
- Stanley, H. E. 1999. Scaling, universality, and renormalization: three pillars of modern critical phenomena. *Reviews of Modern Physics* **71**:S358–S366.
- Townsend, C. R., R. M. Thompson, A. R. McIntosh, C. Kilroy, E. Edwards, and M. R. Scarsbrook. 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, Illinois, USA.
- Warren, P. H. 1989. Spatial and temporal variation in a freshwater food web. *Oikos* **55**:299–311.
- Warren, P. H., and J. H. Lawton. 1987. Invertebrate predator–prey body size relationships: an explanation for upper triangular food webs and patterns in food-web structure? *Oecologia* **74**:231–235.
- Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* **393**:440–442.
- Williams, R. J., E. L. Berlow, J. A. Dunne, A.-L. Barabási, and N. D. Martinez. 2002. Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences (USA)* **99**:12913–12916.
- 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.

#### APPENDIX A

A table of definitions of model and ecological variables is available in ESA's Electronic Data Archive, *Ecological Archives* E086-072-A1.

#### APPENDIX B

An overview of the rules for the three food-web models (niche model, nested-hierarchy model, and generalized cascade model) is available in ESA's Electronic Data Archive, *Ecological Archives* E086-072-A2.

#### APPENDIX C

An analytical solution of the niche model is available in ESA's Electronic Data Archive, *Ecological Archives* E086-072-A3.

#### APPENDIX D

A comparison of the cumulative distributions of the number of predators and number of links for species in the 15 community food webs investigated (empirical data) with the numerical predictions of the niche model is available in ESA's Electronic Data Archive, *Ecological Archives* E086-072-A4.