NORTHWESTERN UNIVERSITY

A complex systems approach to the modeling and characterization of natural ecosystems

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical Engineering

By

Daniel B. Stouffer

EVANSTON, ILLINOIS

December 2007

Abstract

A complex systems approach to the modeling and characterization of natural ecosystems

Daniel B. Stouffer

The linear food chain of high school biology textbooks has been replaced by the food web, the increasingly complex network of trophic interactions in an ecosystem. This complexity, however, masks a number of robust statistical properties. Yet much research to date has concentrated upon predicting and testing the existence of such patterns rather than on the factors giving rise to them. In order to study such a complex problem, I incorporate concepts and methods from fields not traditionally aligned with ecology—such as chemical engineering and statistical physics—to develop an integrated approach.

By following this approach, I have demonstrated that there are key universal features common to all ecosystems, independent of variables such as the population and type of species present, assembly history, or particular environment. These universal features include a number of important descriptors of the structure of food webs, as well as the empirically observed mechanism of prey selection. My research provides insight into, for example, how best to develop dynamic ecosystem models to probe the effect of exotic species.

Acknowledgments

First and foremost, I am indebted to my advisor and mentor Luís A. N. Amaral. I would also like to thank all of my collaborators in the work that is presented in this thesis, Juan Camacho, Roger Guimerà, Wenxin Jiang, and Carla A. Ng.

I am compelled to thank the members of my proposal and thesis committees, Kimberly A. Gray, Richard M. Lueptow, Julio M. Ottino, and Bruce E. Rittman, for the donation of their time and expertise in the evaluation of my work.

I would also like to thank everyone who has contributed to this thesis through stimulating discussions, helpful suggestions, or worse. This includes Jordi Bascompte, Joel E. Cohen, Iain D. Couzin, Jennifer Dunne, Jeremy W. Fox, Einat Fuchs, Joanna González-Lergier, Lev Guzmán-Vargas, Simon A. Levin, R. Dean Malmgren, Neo D. Martinez, Robert May, Patrick D. McMullen, André A. Moreira, Marta Sales-Pardo, Erin N. Sawardecker, Sam M. D. Seaver, Michael J. Stringer, Joshua S. Weitz, numerous anonymous referees and almost as many thoughtful editors. I also acknowledge the NU ChBE Murphy Fellowship, NU Teaching Apprenticeship Program, and NSF-IGERT "Dynamics of Complex Systems in Science and Engineering" (DGE-9987577) for their generous financial support.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The global biodiversity challenge

There are between seven and fifty million different species of plants and animals on Earth[1]. About two-thirds of these species live in the tropics, largely in the tropical forests[1]. In fact, about 30–50% of plant, amphibian, reptile, mammal, and bird species occur in 25 hotspots that occupy no more than 2% of the terrestrial land mass[2]. It is believed that fish and other marine organisms are similarly concentrated[3].

The concentration of natural species demands that hotspots be managed with particular attention and caution. Unfortunately, the tropical rain forests receive the complete opposite as only about one half of the original 16 million square kilometers remains[4] and clearing now eliminates about 0.2 million square kilometers every year[5,6]. This factor and others such as overpopulation and global warming place us in the midst of the sixth major extinction event in natural history[7].

Extinctions of species represent one of the most dramatic ecosystem perturbations and take place on far quicker time scales than evolution and introduction of new species into a habitat[7].

Extinctions can greatly alter an ecosystem's biodiversity; they can affect ecosystem stability, its resilience to environmental change, or its resistance to invasion of exotic species[8].

On the shore of Lake Michigan, the situation is no better. High levels of PCBs and heavy metal contamination have led to fish consumption advisories. Pollution, however, is not the only danger. On October 13, 2004, a federal task force announced funding for a $9.1 million permanent electrified barrier on a waterway near Chicago to prevent the invasive Asian carp from migrating from the Mississippi River into Lake Michigan. At about the same time, a Northern snakehead, a fish native to China, Korea, and Russia, was caught in Lake Michigan. This was the first appearance of the fish in the Lake, despite both Illinois and Chicago imposing snakehead bans two years ago. Both species are characterized by voracious appetites, and could potentially wreak havoc upon the Lake's ecosystem. Changes to ecosystems also carry high social and economic costs. Estimates of the overall cost of invasions by exotic species in the United States alone range up to $137 billion annually[8].

1.2 The complexity of ecological problems

The management of ecosystems must be firmly based upon environmental and ecological knowledge. However, this "principle" is hard to implement for two types of reasons: First, its consequences extend far into the realms of social and economic policy. Second, in order to optimize the policies for management of wildlife habitats and fisheries, sustainability of natural resources, and allocation of areas for habitat preservation, we need a better understanding of ecosystems.

Unfortunately, there remain a number of open questions regarding ecosystem structure, stability, and dynamics. Studies demonstrate that a linear conception of ecosystems, represented by the simplistic, but widely used, "food chain" model, is not accurate; rather, trophic interactions are

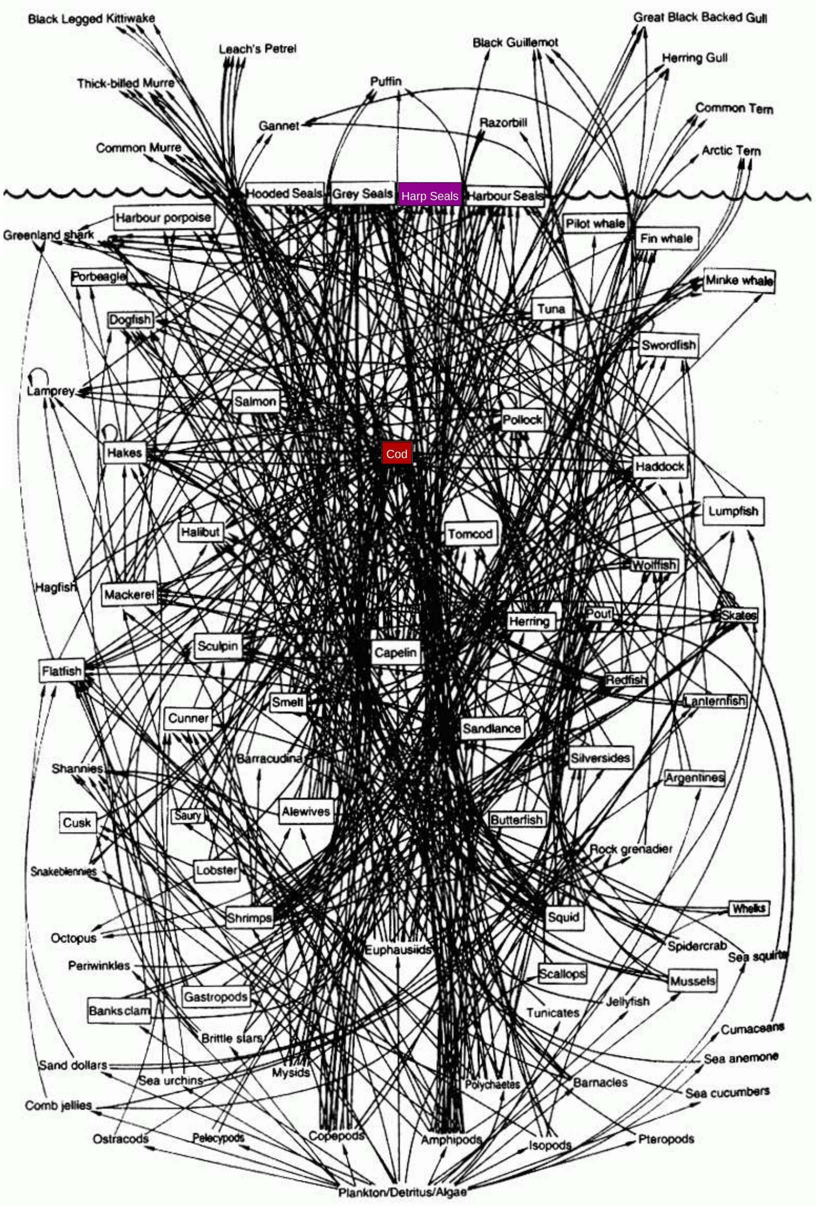

Figure 1.1: Graphical representation of the food web for the Northwestern Atlantic fishery. An arrow connects species in predator-prey relationships, with the arrow pointing from the prey to the predator. Some experts blamed harp seals (purple box) for the collapse of the cod (red box) population. (Note that cod is also a prey of other species such as hooded seals, grey seals, and harbor porpoises. Adapted from Buchanan[9].

known to give rise to complex webs that are still far from being understood[10,11,12] (Fig. 1.1). Much of this lack of understanding stems from insufficient progress in the characterization of empirical food webs due to the difficulties involved in collecting high quality empirical data[13]. These difficulties including the long hours required for direct observation and data collection for stomach content analysis or scatology. Additionally, it is still an issue to know when a scientist may faithfully declare any data as “complete.” Because of these and other hurdles, there exist less than twenty empirical data sets which are widely regarded as “sufficiently comprehensive.”

Though ecologists have been studying the problem for decades, there has been little progress in taking ecology from an observational science to a *predictive* one. This is due partially to the fact that, until recently, ecology has also stood as more of a qualitative than quantitative science. More importantly, however, is the typical scale of ecological problems and what is required in order to solve them. Consider, for example, the issue of invasive species in Lake Michigan. There is much speculation regarding the impact of introduction of the Asian carp or Northern snakehead, yet there exists no legitimate means to predict what the effects will truly be. As experimental reproduction of such a scenario is impossible, not to mention unwise environmentally, the development of theoretical and computational means to model such processes provides the strongest possibility of transforming the predictive capabilities of the science of ecology.

1.3 The “philosopher’s stone” of ecological research

The tremendous diversity of ecosystems around the globe is apparent to even the most casual of observers. These ecosystems can differ in the population sizes and type of species present, the type of environment, the assembly history, and the rate of change. This diversity poses a very real challenge to the development of a *general* understanding of ecosystem behavior and stability. However, recalling the tale of the Northwest Atlantic fishery, how can we hope to begin to understand the consequences of our actions on a general ecosystem if experts cannot reach a consensus regarding a *single* ecosystem?

During the 80's, ecologists have attempted to draw generalizations about the structure of food webs and three food web "laws" were proposed[10,14,15]. The first, known as the "species scaling law," states that the fractions of top, intermediate, and basal species[*] do not vary with the number of species S in the food web. The second, the "link scaling law," states that the fractions of top-intermediate, top-basal, intermediate-intermediate, and intermediate-basal links also do not vary with S. The last, referred to as the "link-species scaling law," states that the number of links L is proportional to S and thus that the linkage density z does not change with S.

These "laws" were "uncovered" for a number of poorly resolved food webs, *some with as few as five species*[15,16,17]. For example, data analyzed to validate all three of these laws in Havens[18] was reanalyzed by Martinez[17] whose study discredited this validation. Based on his analysis, Martinez proposed that the directed connectance C, not z, is invariant with S[17]. Unfortunately, even the scale invariance of direct connectance seems inaccurate.

Due to the complexity of the new highly-resolved webs it has become quite challenging to obtain quantitative patterns with which to "substitute" the old ones[12]. However, the recent application of more tools from statistical physics has helped uncover new patterns which are robust and universal among the empirical food webs.

---

[*] Top species are species who have no predators. Intermediate species are species who have both predators and prey. Basal species are species who have no prey.

1.4 Universality in food web structure

The central hypothesis of my research continues the idea that, in spite of all the aspects unique to each ecosystem, there are a number of universal features that hold for a large number of ecosystems. This hypothesis is based on the principle that there are emergent properties in complex systems which arise from constraints acting upon the system. This is the origin of the “universality” found in food webs. In the case of ecosystems, bioenergetic constraints could be considered a major factor controlling ecosystem structure, though as we will see may not be the strongest driving force.

While much research to date has concentrated upon predicting and testing the existence of such patterns rather than on the factors giving rise to them. In order to study such a complex problem, I incorporate concepts and methods from fields not traditionally aligned with ecology—such as chemical engineering and statistical physics—to develop an integrated approach.

By following this approach, I have demonstrated that there are key universal features common to all ecosystems, independent of variables such as the population and type of species present, assembly history, or particular environment. These universal features include a number of important descriptors of the structure of food webs, as well as the empirically observed mechanism of prey selection. In addition, I have been able to demonstrate that some of these universal features provide an empirical validation of long standing, but heretofore unproven, ecological theories. My research provides insight into, for example, how best to develop dynamic ecosystem models to probe the effect of exotic species.

Chapter 2

Static food web models

A number of models were proposed the past twenty years in an attempt to explain the source of patterns or regularities in empirical food webs. The concept of “niche theory” or “niche space” is a fundamental concept in these models. According to Cohen[19], niche theory has origins dating back to the beginning of the twentieth century. Niche space is defined as an “n-dimensional hyperspace” with n corresponding to innumerable ecological and environmental characteristics[20]. Therefore, each species’ niche is the “result” of all n factors acting upon it and the niche represents the functional role and position of the organism in its community.

Upon reviewing an increasing number of ecological factors, it is clear that each niche contains a single species because no pair of species is exactly the same. Because of this fact, the concept of an ecological niche has been inextricably linked to inter-species competition and resource utilization. In this context, a niche is not a “point” but rather a niche “range” corresponding to the resource spectrum a species exploits[21]. The niche overlap, or the region of niche space shared by two species, represents the amount of shared resources and has been used as a gauge of the level of inter-species competition. It is believed that each species exists in its own, unique, niche in order to avoid substantial niche overlap, and hence competition. Likewise, evolutionary pressure leads to the filling of vacant regions of niche space by species adaptation or even the evolution of new species[20].

However, substantial niche overlap does not translate directly to exclusionary competition; in fact, there may be an *inverse* relationship between overlap and competition[22]. If resources are plentiful, two species may have significant overlap without detriment to each other. In this case, extensive overlap could be correlated with *reduced* competition. A lack of niche overlap on an ecosystem-wide scale could represent a resource-scarce environment, where the disadvantage of sharing resources is far more severe. This concept, Pianka's "niche overlap hypothesis," states that the maximal tolerable niche overlap should be lower in intensely competitive situations than in environments with plenty of resources[22].

A more recent interpretation of niche theory, however, relates to the niche providing species an ordering or hierarchy[23,24]. The most common hypothesis is that species may be ordered in *one-dimension* based solely upon the mass of an individual. This formulation of niche theory provides a much simpler criterion than Hutchinson's "n-dimensional hyperspace" and in fact, with few exceptions, species' diets have been found to agree quite well empirically[25,26]. This notion of the niche again representing a point, but here in a trophic hierarchy, has provided the foundation for recent advances in the understanding of how food webs are "assembled".

Several models have been proposed in the last two decades that aim to describe the structure of food webs[10,27,28]. 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 field ecologists. Because the models may be analytically intractable, 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, I have performed a detailed study of three food web models: the cascade model[10], the niche model[27], and the nested-hierarchy model[28]. 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[10]. 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[27]. Finally, we consider the nested-hierarchy model because it claims to improve on the niche model by considering phylogenetic constraints and adaptation, thereby being more "ecologically sound"[28].

2.1 The cascade model

In the cascade model, one first randomly assigns S species to "trophic niches" with niche values n_i mapped uniformly onto the interval $[0, 1]$. Each species j with $n_j < n_i$ can become a prey of i with probability $x_0 = 2CS/(S-1)$, where $C = L/S^2$ is the food web's directed connectance (Fig. 2.1). This results in a food web with S species and L trophic interactions between those species.

2.2 The niche model

In the niche model[27], one first randomly assigns S species to "trophic niches" with niche values n_i mapped uniformly onto the interval $[0, 1]$, just as in the cascade model.

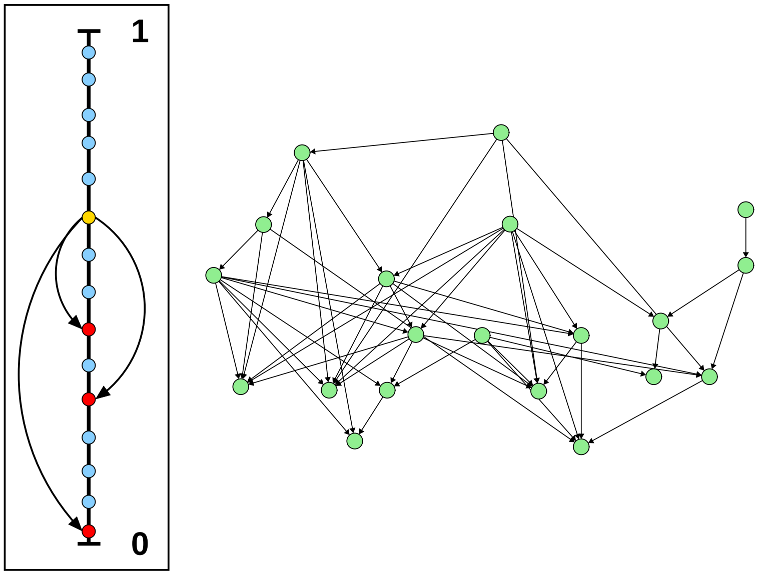

Figure 2.1: We show the prey selection in the cascade model in the figure on the left. The yellow predator can randomly select prey from the set of all species with lower niche values, with equal probability x_0. Here $x_0 = 0.15$ and the predator selects the three species marked in red as prey. On the right, we show a graphical representation of a cascade model-generated food web with $S = 20$ species and directed connectance $C = 0.15$.

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 (Fig. 2.2). 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 $x \in [0, 1]$ is a random variable with probability density function

$$p(x) = \beta\,(1-x)^{(\beta-1)}\ . \tag{2.1}$$

Williams & Martinez[27] 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 $p(x)$. Note that this *species-specific* probability x is an important distinction between the cascade and niche

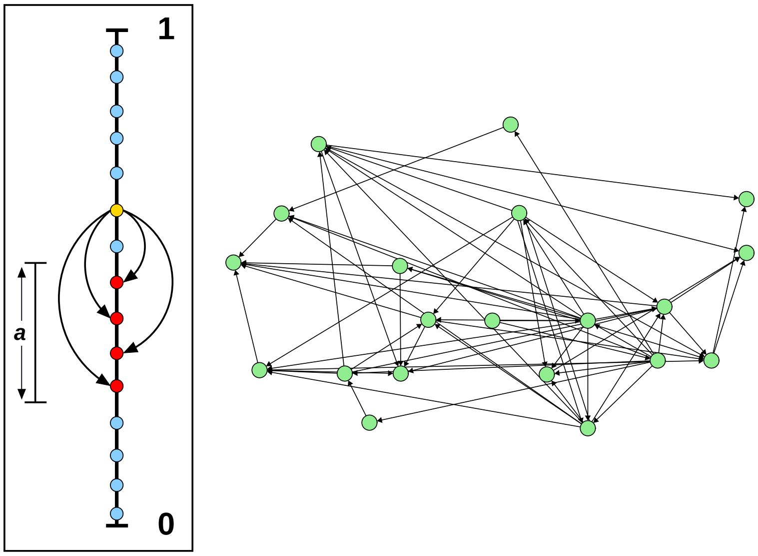

Figure 2.2: We show the prey selection mechanism in the niche model in the figure on the left. In the niche model, each predator consumes all species lying within a segment of the niche axis with length a_i. The center of this segment, c_{a_i}, is uniformly drawn from the range $[a_i/2, n_i]$. Note that this will allow up to half of the potential prey to have niche values higher than the predator. In this example, the yellow predator consumes the four red prey all with niche values less than its own. On the right, we show a graphical representation of a niche model-generated food web with $S = 20$ species and directed connectance $C = 0.15$. For the sake of comparison, the species are arranged spatially just as in Fig. 2.1.

models. The values of the parameters β and S determine the linkage density $z = L/S$ of the food web and its directed connectance $C = L/S^2$, as $\beta = \frac{1}{2C} - 1$.

2.3 The nested-hierarchy model

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 x_i drawn from the interval $[0, 1]$ according to Eq. (2.1).

Prey selection in this model obeys a two-stage, multi-step process (Fig. 2.3). In stage one, the first prey of species i is selected at random from among species with lower niche values than i.

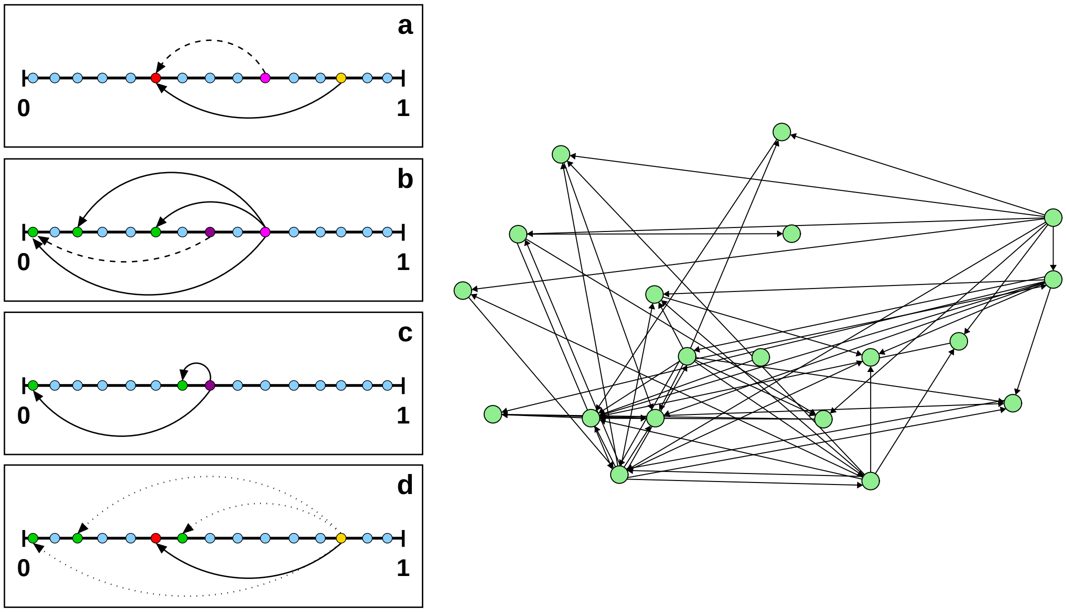

Figure 2.3: We diagram the implementation of prey selection in the nested-hierarchy model in the figures on the left. **a**, The yellow species randomly selects a prey from the set of all species with lower niche values. The chosen species, indicated in red, is also a prey of the magenta species. **b** and **c**, The next prey of the yellow species is then randomly chosen from among the set of prey of the group of the red species' consumers. The group of the red species' consumers are the consumers sharing at least one prey, at least one of which must be the red species; this group is thus the magenta species, the purple species, and now the yellow species. **d**, The pool of potential prey for the yellow species is then each of the green prey species. Note that it contains all prey from the group of the red species' consumers. On the right, we show a graphical representation of a nested-hierarchy model-generated food web with $S = 20$ species and directed connectance $C = 0.15$. For the sake of comparison, the species are arranged spatially just as in Figs. 2.1 and 2.2.

Let j be the first prey of i. If j is 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. The group of j's consumers is defined as 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.

As a last resort, if all possible species with niche values lower than i have been selected and i requires additional prey, selection enters stage two, and prey are chosen at random from among species with niche values greater than or equal to the niche value of i.

2.4 Validation of food web models

The original testing of the models by Cohen & Newman[10], Williams & Martinez[27], and Cattin *et al.*[28] relied on the comparison of different properties between model-generated and empirical webs. The properties compared include the number of top, intermediate, and basal species, the standard deviations of generality and vulnerability, the mean maximum similarity between species, the mean and standard deviation of food chain length, the number of food chains, the number of trophic loops, the fraction of cannibal species, and the fraction of omnivores[10,27,28]. For each of these properties, the authors compared the empirical values to averages obtained from multiple model realizations to determine whether or not they fell within the confidence intervals.

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 most of which can be derived from the distributions of numbers of prey and predators (Table 2.1). We believe that it is more direct and elucidating to compare the distributions themselves. In the next section, we will solve analytically and numerically for these distributions for each of the models discussed with that goal of quantifying both *how and why* three seemingly distinct models accurately predict the statistical properties of empirical food webs.

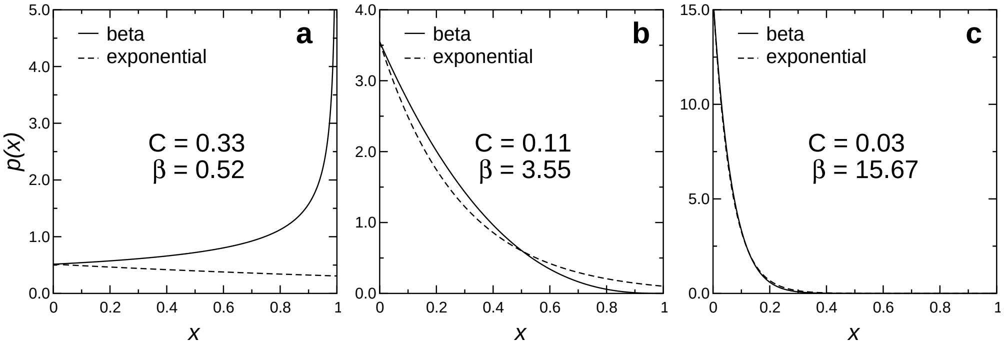

Figure 2.4: Comparison between the beta-distribution, Eq. (2.1), and the exponential distribution, Eq. (2.3), across the range of empirically observed values of C. For $C \leq 0.11$, the exponential distribution provides a good approximation of the beta-distribution.

2.5 Distribution of number of prey

Camacho *et al.*[29] 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

$$p_{\text{prey}}(k) = (1/2z)\, E_1(k/2z)\,, \tag{2.2}$$

where $E_1(x)$ is the exponential-integral function[30].

Camacho *et al.*[29] 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 $\beta \to \infty$. Thus, $p(x)$ is negligible except when $x \ll 1$, where $e^{-x} \simeq (1-x)$, in other words

$$\lim_{\beta\to\infty} p(x) = \beta e^{-\beta x}\,. \tag{2.3}$$

Figure 2.4 compares Eq. (2.3) and Eq. (2.1) for various values of C, confirming that the former tends to the exponential function of Eq. (2.1) for small C. Therefore, the distribution of number of prey in this limit—Eq. (2.2)—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,

$$p_{\text{prey}}(\tilde{k}) = E_1(\tilde{k})\,. \tag{2.4}$$

Provided that C is small, this probability density function is therefore universal, that is, it is identical for any values of S and z[31].

2.6 Distribution of number of predators

Camacho *et al.*[29] also derived an expression for the distribution of number m of predators, in the limits $S \gg 1$ and $C \ll 1$,

$$p_{\text{pred}}(m) = (1/2z)\,\gamma(m+1, 2z)\,, \tag{2.5}$$

where $\gamma(x)$ is the incomplete gamma function[30]. For $m < 2z$, p_{pred} is approximately constant because $\gamma(m+1, 2z) \approx 1$; for $m > 2z$, p_{pred} decays to zero with a Gaussian tail[29]. Unlike the distribution of number of prey, the distribution of number of predators is completely independent of the form of $p(x)$[29].

Unlike the scaling seen for the distribution of number of prey, Eq. (2.5) 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. The probability density for the scaled variable $\tilde{m} = m/2z$ is thus

$$p_{\text{pred}}(\tilde{m}) = \gamma(2z\,\tilde{m} + 1, 2z) \approx 1 \qquad \tilde{m} < 1 \tag{2.6}$$

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.

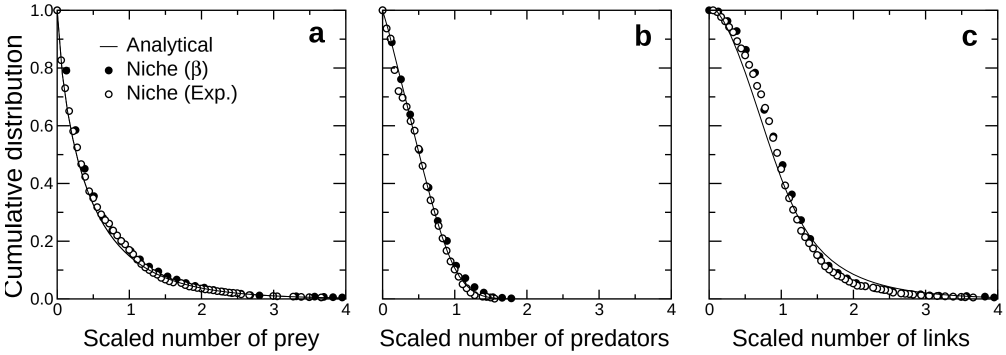

Figure 2.5: Comparison between the analytical expressions and numerical simulations of the niche model for the cumulative distributions of scaled **a** number of prey, $k/2z$, **b** number of predators, $m/2z$, and **c** number of links, $r/2z$. The distributions are for a single realization of the niche model with $S = 1000$ and $z = 5$. The model was implemented with $p(x) = \beta(1-x)^{(\beta-1)}$ (filled symbols) and $p(x) = \beta e^{-\beta x}$ (open symbols).

2.7 Distribution of number of links

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. (2.2) and (2.5). 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

$$\begin{aligned} p_{\text{links}}(r) &= \int_0^r p_{\text{prey}}(t) p_{\text{pred}}(r-t)\, dt \\ &= \frac{1}{(2z)^2} \int_0^r E_1(t/2z)\gamma(r-t+1, 2z)\, dt\,, \end{aligned} \tag{2.7}$$

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

Figure 2.5 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

Table 2.1: 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	$p_{\text{links}}(r) = \int_0^r p_{\text{prey}}(t) p_{\text{pred}}(r-t)\, dt$ $= \frac{1}{(2z)^2} \int_0^r \; E_1(t/2z)\gamma(r-t+1, 2z)\, dt$
Fraction of top species	$T = \frac{1-e^{-2z}}{2z}$
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	$\sigma_G = \sqrt{8/(3+6C) - 1}$
Predator-prey correlation coefficient	$r_{\text{cor}} = -\frac{1}{\sqrt{5(1+3/z)}}$
Assortativity	$A = -\frac{(2z/3-1/2)^2}{26z^2/9+z/3+1/4}$

universal functional forms that depend on a single parameter—the linkage density z of the food web[31]. It is therefore possible, given the linkage density of a web, to describe these distributions without any additional information, provided $C \ll 1$.

As discussed earlier, 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 2.1). Indeed, our estimates compare quite well to simulations of the niche model[29].

2.8 Numerical analysis of 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[28]. This approach is said to better reflect the complexity and statistical properties of real food

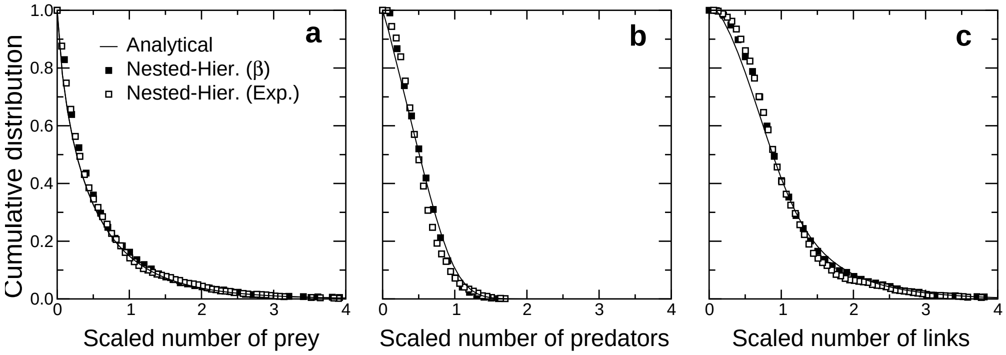

Figure 2.6: Comparison between the analytical expressions and numerical simulations of the nested-hierarchy model for the cumulative distributions of scaled **a** number of prey, $k/2z$, **b** number of predators, $m/2z$, and **c** number of links, $r/2z$. The distributions are for a single realization of the nested-hierarchy model with $S = 1000$ and $z = 5$. The model was implemented with $p(x) = \beta(1-x)^{(\beta-1)}$ (filled symbols) and $p(x) = \beta e^{-\beta x}$ (open symbols).

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[32].

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[32] (Fig. 2.6).

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. 2.7, 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[32]. Our numerical results demonstrate that a species will consume species with niche values greater than or equal to its own only 6% of the time

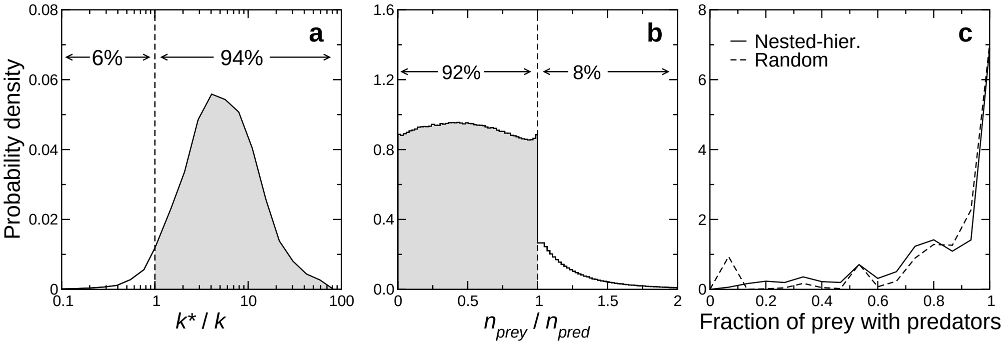

Figure 2.7: Prey selection in the nested-hierarchy model for food webs with $S = 100$ and $z = 8$. **a** Probability density function of number k^* of prey available, normalized by number k of prey desired. The prey available to species i is defined as: (i) the species with niche values smaller than n_i which have no predators and, (ii) if species i has selected a prey with other predators, all members of the pool defined in Fig. 2.3 which have a lower niche value than species i. A species i will consume species with niche values greater than or equal to its own only 6% of the time. **b** Probability density function of $n_{\rm prey}/n_{\rm pred}$—where n is the niche value—for links created in the nested-hierarchy model. 92% of all links have prey with niche values lower than the predator, with an *approximately uniform distribution* for niche values lower than that of the predator. **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, implying that the nested-hierarchy model is very closely implementing random selection of species with lower niche values, which shows no bias toward species having either many or few predators.

(Fig. 2.7**a**), explaining the very low rate of cannibalism seen in the nested-hierarchy model[32]. In addition, 92% of all links have prey with niche values lower than the predator (Fig. 2.7**b**), 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*[32] (Fig. 2.7**c**). Consuming species with larger niche values represents only 8% of all links[32]. 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[32] (Figs. 2.6).

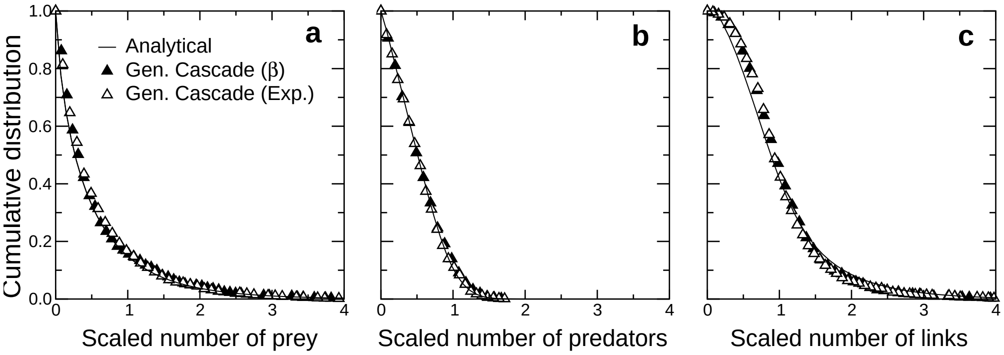

Figure 2.8: Comparison between the analytical expressions and numerical simulations of the generalized cascade model for the cumulative distributions of scaled **a** number of prey, $k/2z$, **b** number of predators, $m/2z$, and **c** number of links, $r/2z$. The distributions are for a single realization of the generalized cascade model with $S = 1000$ and $z = 5$. The model was implemented with $p(x) = \beta(1-x)^{(\beta-1)}$ (filled symbols) and $p(x) = \beta e^{-\beta x}$ (open symbols).

2.9 The generalized cascade model

The results for the niche and nested-hierarchy models are congruent with the existence of two conditions that are sufficient for a model to reproduce the properties of empirical food webs[32]:

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

Condition II: 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 & Newman[10] so that it satisfies Condition II.

In the cascade model of Cohen & Newman[10], a species j with $n_j < n_i$ becomes a prey of i with fixed probability $x_0 = 2CS/(S-1)$. Williams & Martinez[27] 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 II. To solve this problem, we generalize the cascade model in the following manner[32]. 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 $[0, 1]$. Note that the idea of a predator specific—or predator dominant—x was already discussed by Cohen[33], 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[32] (Fig. 2.8).

2.10 Concluding remarks

In this chapter, we demonstrated that there are two critical conditions that a food web model must satisfy in order to successfully predict the properties of empirical food webs[32]:

1. The niche values to which species are assigned form a *totally ordered set*
2. Each species has a specific, exponentially-decaying probability $p(x)$ of preying on a fraction x of the species with lower niche values.

As noted previously, the original model validation by Cohen & Newman[10], Williams & Martinez[27], and Cattin *et al.*[28] relied on the comparison of different properties between model-generated and empirical webs. In Table 2.1, we list a number of the properties compared which may be derived explicitly from the distributions of number of prey and number of predators.

We argued, however, that it is important to stress that this approach is far from ideal and that it is more direct and elucidating to compare the distributions themselves. In the next chapter, we will follow this approach and use our analytical solution of the models to quantify both *how* and *why* three seemingly distinct models accurately predict the statistical properties of empirical food webs.

Chapter 3

Two conditions explain many global aspects of food web structure

In this chapter, 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[32].

3.1 Empirical food webs

Throughout this thesis, we will examine various properties of empirical food webs. We consider seventeen food webs from a variety of environments: four estuarine—Chesapeake Bay[34], St. Marks[35], Ythan (1)[36], and Ythan (2)[37]; five freshwater—Bridge Brook Lake[18], Canton Creek[38], Little Rock Lake[39], Skipwith Pond[40], and Stony Stream[38]; three marine—Benguela[41], Caribbean

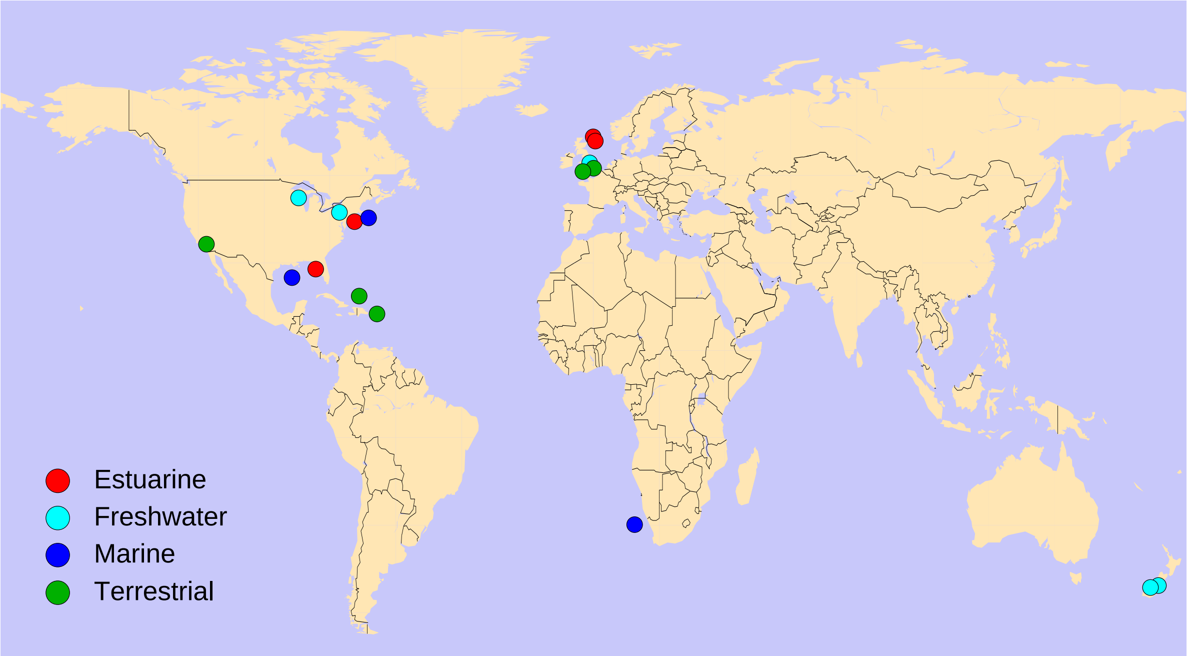

Figure 3.1: Map of empirical food webs studied. Red circles indicate the four estuarine food webs—Chesapeake Bay, St. Marks, Ythan (1), and Ythan (2); cyan circles indicate the five freshwater food webs—Bridge Brook Lake, Canton Creek, Little Rock Lake, Skipwith Pond, and Stony Stream; blue circles indicate the three marine food webs—Benguela, Caribbean Reef, and Northeast US Shelf; and green circles indicate the five terrestrial food webs—Coachella Valley, El Verde Rainforest, Grassland, Scotch Broom, and St. Martin.

Reef[42], and Northeast US Shelf[43]; and five terrestrial—Coachella Valley[44], El Verde Rainforest[45], Grassland[11], Scotch Broom[46], and St. Martin[47] (Table 3.1 and Fig. 3.1).

Aside from the specifics of their environment, there are additional factors which distinguish some food webs from the others and should be noted. First, three sets of food webs shared investigators in the compilation of the data: (i) Grassland, Little Rock Lake, and Scotch Broom[11,39,46], (ii) Canton Creek and Stony Stream[38], and (iii) Ythan Estuary (1) and Ythan Estuary (2)[36,37]. Second, Scotch Broom is a source web[46]—a food web generated by following all trophic links upward from a single basal species—while the remainder are community food webs. Third, Canton Creek

Table 3.1: Empirical food webs studied. S is the number of trophic species in the food web, L is the number of trophic predator-prey interactions, z is the linkage density of the food web, and C is the food web's direct connectance.

Food web	S	L	$z = L/S$	$C = L/S^2$	Reference
Benguela	29	203	7.00	0.24	Yodzis[41]
Bridge Brook Lake	25	106	4.42	0.17	Havens[18]
Canton Creek	102	697	6.83	0.067	Townsend *et al.*[38]
Caribbean Reef	50	556	11.12	0.22	Opitz[42]
Chesapeake Bay	31	68	2.19	0.071	Baird & Ulanowicz[34]
Coachella Valley	29	262	9.034	0.31	Polis[44]
El Verde Rainforest	155	1509	9.74	0.063	Waide & Reagan[45]
Grassland	61	97	1.59	0.026	Martinez *et al.*[11]
Little Rock Lake	92	997	10.84	0.18	Martinez[39]
Northeast US Shelf	79	1400	17.72	0.22	Link[43]
Scotch Broom	85	223	2.62	0.031	Hawkins *et al.*[46]
Skipwith Pond	25	197	7.88	0.32	Warren[40]
St. Marks Seagrass	48	221	4.61	0.096	Christian & Luczkovich[35]
St. Martin Island	42	205	4.88	0.12	Goldwasser & Roughgarden[47]
Stony Stream	109	829	7.61	0.070	Townsend *et al.*[38]
Ythan Estuary (1)	83	395	4.76	0.057	Hall & Raffaelli[36]
Ythan Estuary (2)	174	579	3.33	0.019	Hall & Raffaelli[37]

and Stony Stream[38] are time-specific food webs, that is the data was collected one a single day, whereas the other foods webs are cumulative food webs composed of data collected multiple times and spanning multiple seasons. Lastly, Ythan Estuary (2) is the same food web as Ythan Estuary (1) but with parasitism included.

3.2 Empirical distributions of numbers of prey, predators, and links

Empirical food webs found in the literature generally contain a small number of trophic species (Table 3.1). 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

$$P(> y) = \sum_{y'=y}^{\infty} p(y')\,, \tag{3.1}$$

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

Equation (2.2) implies that the cumulative distribution of number of prey is given by

$$P_{\text{prey}}(> k) \;\; = \;\; \exp\left(-\frac{k}{2z}\right) - \frac{k}{2z}\, E_1\left(\frac{k}{2z}\right)\;. \tag{3.2}$$

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

$$P_{\text{prey}}(> \tilde{k}) = \exp\left(-\tilde{k}\right) - \tilde{k}\, E_1\left(\tilde{k}\right)\;. \tag{3.3}$$

Like Eq. (2.4), this expression contains *no free parameters* and is *universal*. That is, it is independent of model details and parameter values.

Equation (2.5) implies that the cumulative distribution of number of predators follows the form

$$P_{\text{pred}}(> m) = \frac{1}{2z} \sum_{m'=m}^{\infty} \gamma(m'+1, 2z)\,. \tag{3.4}$$

As previously noted, for $m < 2z$ the incomplete gamma function γ can be approximated as $\gamma(m+1, 2z) \simeq 1$. One can therefore rewrite Eq. (3.4) as

$$P_{\text{pred}}(> m) = 1 - \frac{1}{2z} \sum_{m'=0}^{m-1} \gamma(m'+1, 2z) \simeq 1 - \frac{m}{2z}\,, \quad m < 2z\,. \tag{3.5}$$

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

$$P_{\text{pred}}(> \tilde{m}) = 1 - \tilde{m}\,, \quad \tilde{m} < 1\,. \tag{3.6}$$

For $\tilde{m} \geq 1$, $P_{\text{pred}}(> \tilde{m})$ decays to zero as the error function[30].

As in Eq. (3.1), the cumulative distribution of number of links is given by

$$P_{\text{links}}(> r) = \sum_{r'=r}^{\infty} p_{\text{links}}(r')\,, \tag{3.7}$$

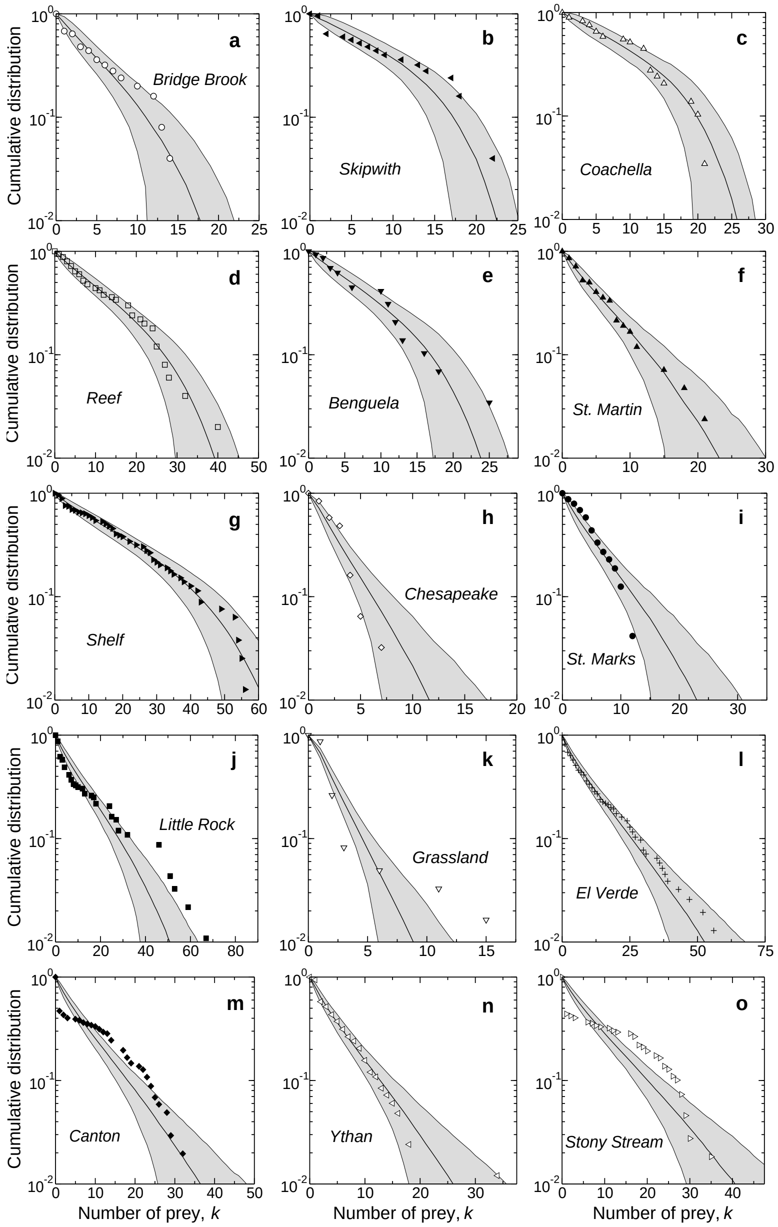

Figure 3.2: Cumulative distribution P_{prey} of number of prey k for the 15 food webs studied: Bridge Brook Lake[18]; Skipwith Pond[40]; Coachella Valley[44]; Caribbean Reef[42]; Benguela[41]; St. Martin Island[47]; Northeast US Shelf[43]; Chesapeake Bay[34]; St. Marks Seagrass[35]; Little Rock Lake[39]; Grassland[11]; El Verde Rainforest[45]; Canton Creek[38]; Ythan Estuary[36]; and Stony Stream[38]. 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.

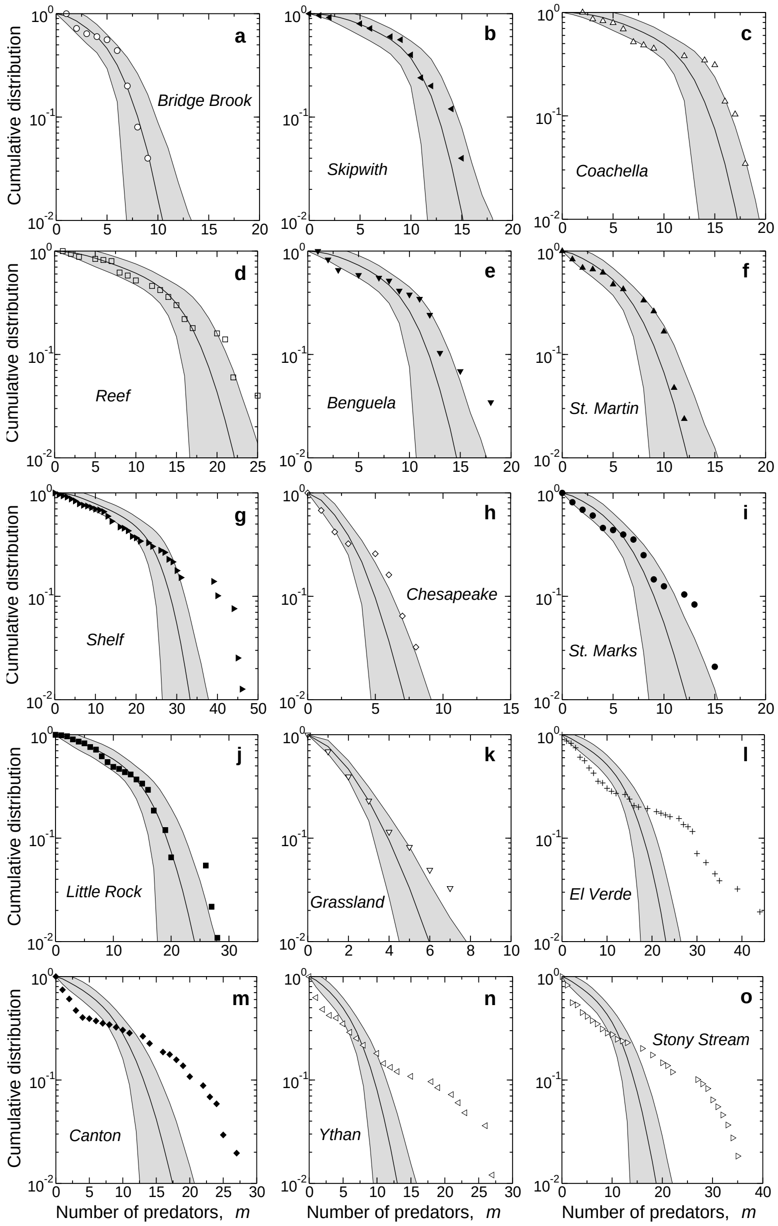

Figure 3.3: Cumulative distribution P_{pred} of number of predators m for the 15 food webs studied (see Fig. 3.2). The solid black line represents the average value from 1000 simulations of the niche model and the gray region represents two standard deviations above and below the model's predictions.

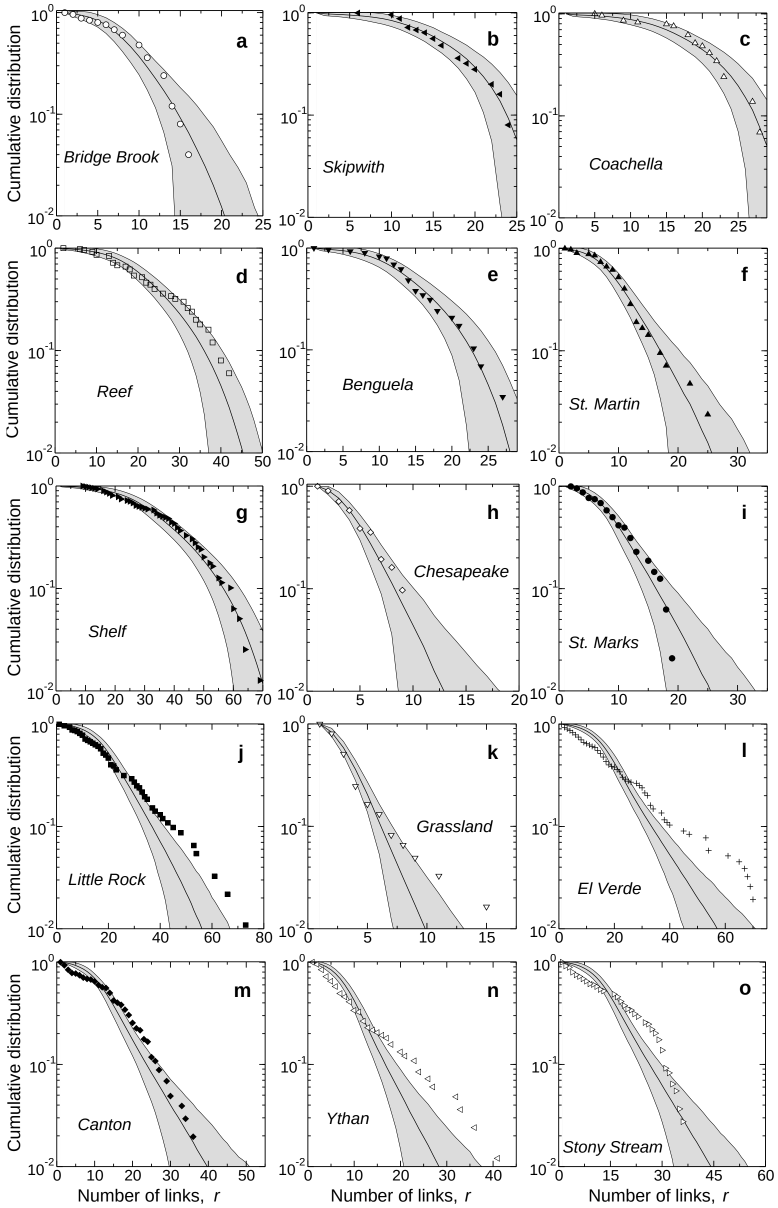

Figure 3.4: Cumulative distribution P_{link} of the number of links r for the 15 food webs studied (see Fig. 3.2). The solid black line represents the average value from 1000 simulations of the niche model and the gray region represents two standard deviations above and below the model's predictions.

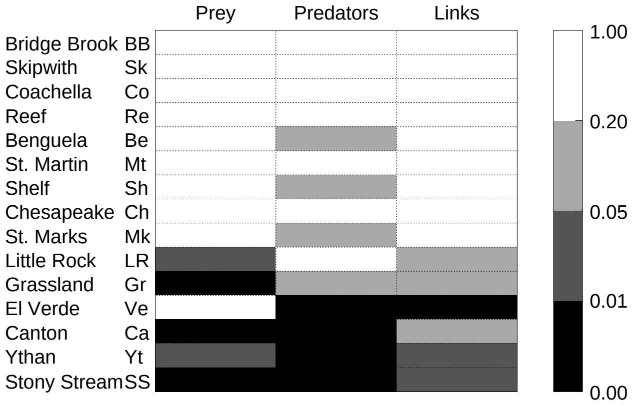

Figure 3.5: Comparison of the distributions of numbers of prey, predators, and links of the 15 food webs to the respective distributions obtained from 1000 webs generated by the niche model. We regard $\overline{P}_{\mathrm{KS}} \leq 0.01$—shown in black—as strong evidence for the rejection of the null hypothesis.

which cannot be determined analytically but may be computed numerically.

In Figures 3.2, 3.3, and 3.4, we analyze the empirical distributions of numbers of prey, predators, and links 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 (Table 3.1).

3.3 Quantifying the agreement between the models and the empirical data

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. 3.5). We use the niche model as a representative model, but identical results are obtained for any of the models described earlier.

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.

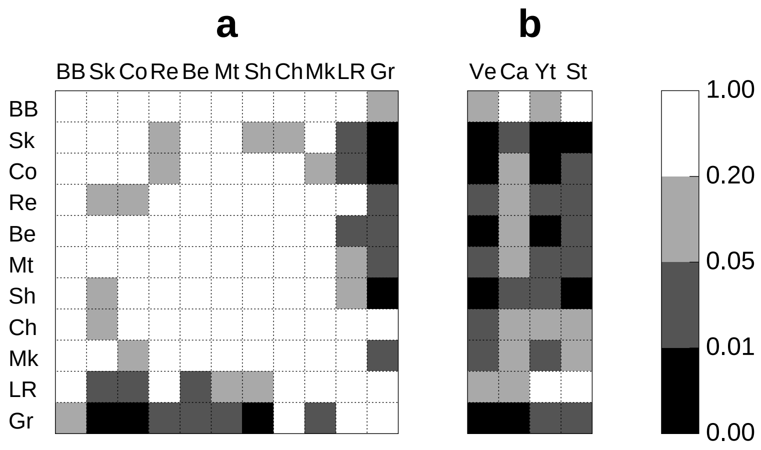

Figure 3.6: **a** Comparison of the distributions of number of links for the 11 empirical food webs we hypothesize obey universal properties. **b** Comparison of these 11 food webs with four food webs that we hypothesize do not obey universal properties. We use the KS test for all comparisons. The results in **a** and **b** may be interpreted as similarity matrices with values $0 \leq P_{\rm KS} \leq 1$, the KS probability. We regard $\overline{P}_{\rm KS} \leq 0.01$—shown in black—as strong evidence for the rejection of the null hypothesis.

Marks, Little Rock, and Grassland[32]. The remaining four—El Verde, Canton, Ythan, and Stony Stream—exhibit rather different behavior, which is visually apparent in Figs. 3.2, 3.3, and 3.4 and confirmed by the results in Fig. 3.5[32]. The results of Figs. 3.2, 3.3, 3.4, and 3.5 suggest the possibility that models which satisfy our two conditions do indeed reproduce the properties of many empirical food webs[32].

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 (Figs. 3.6, 3.7, and 3.8). We apply the χ^2 tests to the values from these figures to test their statistical significance. Remarkably, we find that we can accept the null hypothesis in the case of Fig. 3.6**a** at the 17% significance level while the null hypothesis is rejected for the case of Fig. 3.6**b** ($p \approx 3.5 \times 10^{-35}$)[32].

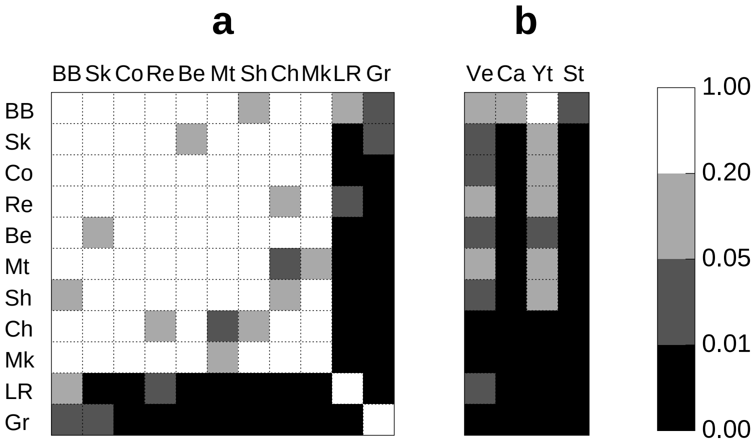

Figure 3.7: **a** Comparison of the distributions of number of prey for the 11 empirical food webs we hypothesize obey universal properties. **b** Comparison of these 11 food webs with four food webs that we hypothesize do not obey universal properties. We use the KS test for all comparisons. The results in **a** and **b** may be interpreted as similarity matrices with values $0 \leq P_{\mathrm{KS}} \leq 1$, the KS probability. We regard $\overline{P}_{\mathrm{KS}} \leq 0.01$—shown in black—as strong evidence for the rejection of the null hypothesis.

We next apply the χ^2 tests to the values from Figs. 3.7 and 3.8 to test their statistical significance. The results for these figures are not as straightforward as for Fig. 3.6. For the distributions of number of prey, we find that the analysis is sensitive to the inclusion of the Little Rock Lake and Grassland food webs. Considering the complete set of 11 food webs, the null hypothesis is rejected in the case of Fig. 3.7**a** ($p \approx 5.1 \times 10^{-8}$). If we eliminate Grassland from this group, the null hypothesis is again rejected ($p \approx 0.027$). Upon eliminating both Little Rock and Grassland from this group, the null hypothesis is accepted ($p \approx 0.11$). In the case of Fig. 3.7**b**, the null hypothesis is rejected at the onset with $p \approx 5.8 \times 10^{-64}$[32].

For the distributions of number of predators, we also find that the significance of the matrices is sensitive to the inclusion of the Little Rock Lake and Grassland food webs. Considering the complete set of 11 food webs, the null hypothesis is rejected in the case of Fig. 3.8**a** ($p \approx 1.2 \times 10^{-6}$). If we eliminate Grassland from this group, the null hypothesis is again rejected ($p \approx$

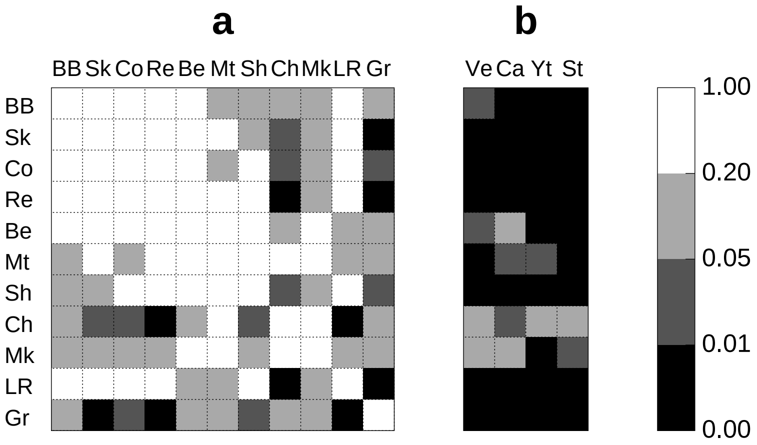

Figure 3.8: **a** Comparison of the distributions of number of predators for the 11 empirical food webs we hypothesize obey universal properties. **b** Comparison of these 11 food webs with four food webs that we hypothesize do not obey universal properties. We use the KS test for all comparisons. The results in **a** and **b** may be interpreted as similarity matrices with values $0 \leq P_{\text{KS}} \leq 1$, the KS probability. We regard $\overline{P}_{\text{KS}} \leq 0.01$—shown in black—as strong evidence for the rejection of the null hypothesis.

0.0079). Eliminating both Little Rock and Grassland from this group, the null hypothesis is still rejected ($p \approx 0.014$). In the case of Fig. 3.8**b**, the null hypothesis is rejected at the onset with $p \approx 8.0 \times 10^{-68}$[32].

Considering the results of Figs. 3.7 and 3.8 alone, the validity of including the Little Rock and Grassland food webs may be questioned. However, when additionally taking into account the results of Figs. 3.5 and 3.6 as well as additional ecological characteristics which will be discussed later, we conclude that our separation of the food webs into 11 which obey universal properties and four which do not is a valid first approximation. In addition, we conclude that the results of Figs. 3.5, 3.6, 3.7, and 3.8 validate the hypothesis that P_{prey}, P_{pred}, and P_{links} obey universal functional forms[31,32].

For this reason, for the remainder of this chapter we will focus our attention on these 11 food webs: Bridge Brook, Skipwith, Coachella, Reef, Benguela, St. Martin, Shelf, Chesapeake, St.

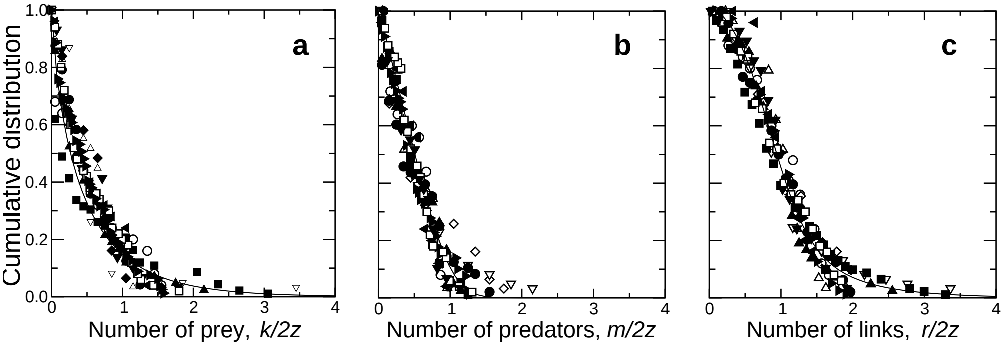

Figure 3.9: Visual test of the "scaling hypothesis" that the distributions of numbers of prey, predators, and links have the same functional form for different food webs. **a** Cumulative distribution P_{prey} of the scaled number of prey for the 11 food webs. The solid line is the analytical prediction of Eq. (3.2). The data collapse onto a single curve consistent with our analytical results. **b** Cumulative distribution P_{pred} of the scaled number of predators for the 11 webs. The solid line is the analytical prediction of Eq. (3.4) for the average value of z in the empirical data, $z = 8.44$. **c** Cumulative distribution P_{links} of the scaled number of links for the 11 webs. The solid line is the prediction of Eq. (3.7). The symbols are those introduced in Fig. 3.2.

Marks, Little Rock, and Grassland. We plot in Fig. 3.9**a** the cumulative distributions $P_{\text{prey}}(>\tilde{k})$ versus 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_{prey} obeys a universal functional form[31,32]. We also plot $P_{\text{pred}}(>\tilde{m})$ versus the scaled variable $\tilde{m} = m/2z$ for the 11 similar food webs in Fig. 3.9**b** finding a similar collapse of the data onto a single curve for $m/2z < 1$[31]. Similarly, we plot in Fig. 3.9**c** $P_{\text{links}}(>\tilde{r})$ versus the scaled variable $\tilde{r} = r/2z$, again noting a collapse of the data onto a single curve further supporting the hypothesis that scaling holds for food web structure[31,32].

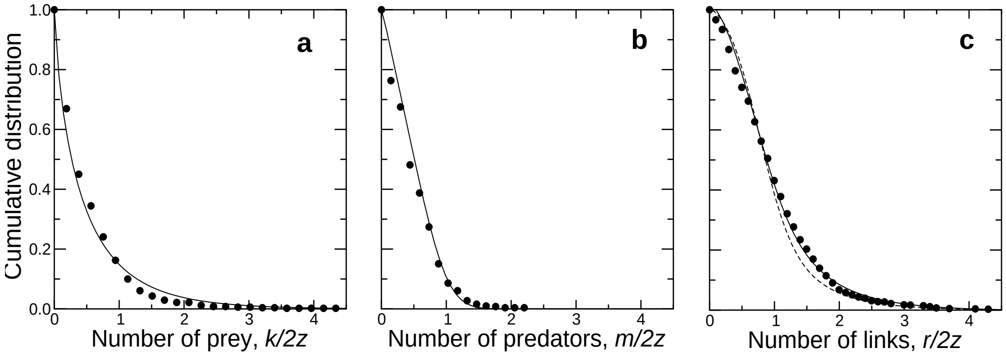

Figure 3.10: Cumulative distributions **a** P_{prey} of the scaled number of prey and **b** P_{pred} of the scaled number of predators for the 11 pooled webs. The solid lines are, respectively, the analytical predictions Eq. (3.2) and (3.4), the latter with the average value $z = 8.44$. **c** Cumulative distribution of scaled number of links $r = k + m$ for the 11 pooled webs. The solid line is obtained by numerically convolving the distributions Eq. (3.2) and (3.4) 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.

3.4 Examining the pooled distribution from all empirical food webs

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_{prey}, P_{pred}, and P_{links} respectively. The cumulative distributions of the *scaled* numbers of prey, predators, and links for the pooled webs are well approximated by Eqs. (3.2), (3.4), and (3.7) *even though there are no free parameters to fit in the analytical curves*[32] (Fig. 3.10). We find that $P_{\mathrm{links}}(r)$ 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 other studies of food-web structure[48,49].

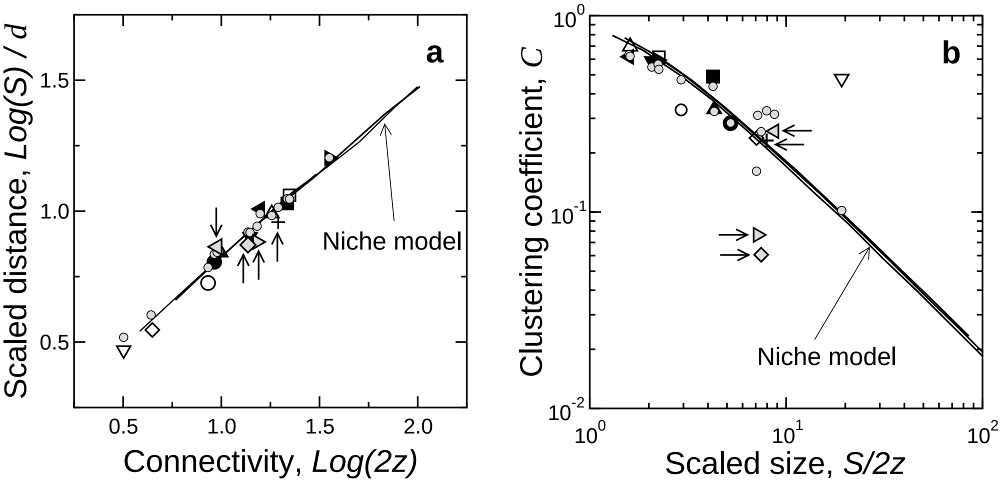

Figure 3.11: **a** Scaled average trophic distance d between species versus linkage density z. We compare the data with the numerical simulations of the niche model for web sizes $S = 100, 500, 1000$ (thin solid lines). We find a logarithmic increase of the average distance with web size S for the empirical 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.*[50]. **b** Clustering coefficient $\mathcal{C}$ versus 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. 3.2 except that the four poorly approximated food webs are filled in gray.

3.5 Other measures

Next, we investigate whether or not the scaling hypothesis also applies to other properties characterizing food web structure[32]. We consider two quantities with ecological implications: (i) the average trophic distance d between species[51], which is defined as the typical number of species needed to trophically connect two given species; and (ii) the clustering coefficient $\mathcal{C}$[51], which quantifies the fraction of species triplets that form fully-connected triangles.

In Fig. 3.11**a** 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 logarithmically with web size for both the model and the data[32]. In Figure 3.11**b**, we show that the clustering coefficient $\mathcal{C}$ 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[51]. 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[32]. 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[32].

3.6 Concluding remarks

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 which distinguish them from the other food webs studied. First, they are remarkably similar, as evidenced by the Kolmogorov-Smirnov test, which provides P_{KS} of $0.565, 0.045,$ and 0.794 under direct comparison of their respective distributions of numbers of prey, predators, and links. 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[38]. 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[27]. 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[45].

In this chapter, 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[32]. 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[32].

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[23,25,26,52]. 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. 3.11—which also support the scaling hypothesis—indicate that there is very little, if any, compartmentalization in communities[53], suggesting the possibility that communities are highly interconnected and that the removal of any species may induce large disturbances[54,55,56,57,58]. 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 that the results of this chapter indicated that fundamental concepts of modern statistical physics such as scaling and universality may be successfully applied in the study of food webs[29,31,32]. 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[32].

Chapter 4

Local structure and prey selection in model food webs

In the previous chapters, we demonstrated analytically that food web model yield the same distributions for the number of prey and number of predators[32], which imply, for example, the same fractions of top and basal species or the standard deviations of generality and vulnerability, just as observed numerically[28]. Remarkably, these distributions are in good agreement with most of the highest quality empirical food webs in the literature, providing a general pattern of food web topology[29,32].

Our study of these models, however, mainly characterized the global structure of food web topology. In this chapter, in contrast, we focus on the analysis of the *local* structure of food webs through the study of the so-called food web subgraphs or motifs[59,60] (Fig. 4.1). This methodology has been applied successfully to a number of empirical networks, including biological, technological and sociological systems, to uncover the underlying structure at a scale in between the entire community and single or pair-wise population dynamics[61,62].

Our approach follows on a rich literature in ecology that attempts to gain insight into the dynamics and stability of natural ecosystems in terms of small sub-webs containing species strongly connected, the so-called ”community modules”[63,64]. These models attempt to bridge the gap between “the baroque complexity of entire communities and the bare bones of single and pair-wise population dynamics”[63]. Community modules provide a set of ecologically relevant subgraphs upon which to perform dynamic studies.

Our perspective is complementary: whereas the latter approach is dynamical and considers only strong links, we focus on structural properties of the food webs; to do so we consider all links, and not only the strong ones. Our perspective is thus similar to that followed by Bascompte & Melián[65], though it differs in several aspects, such as the theoretical approach and our systematic analysis of all three-node motifs[59,60].

Motivated by this fact, we investigate here whether there are any robust patterns in the structural counter-parts to community modules. Specifically, we consider n-species “food-web” motifs[59,60], which consist of the complete set of unique connected subgraphs containing n species[61]. We focus on the case $n = 3$, for which there are 13 distinct motifs (Fig. 4.1). For comparison, there are 199 and 9364 distinct motifs for $n = 4$ and $n = 5$, respectively. Notably our investigation includes subgraphs less explored previously in dynamic studies. As a consequence our results could provide new templates upon which to conduct future dynamic microcosm, mesocosm, and field studies.

Past studies of network motifs have uncovered relevant subgraphs within a variety of empirical networks[61,62], be they biological, technological, and social, or model networks[66,67,68]. Studies have also investigated a subset of motifs in food webs[61,65,69]. However, none of these investigations has been able to mechanistically predict a complete pattern of motif over- and under-representation, analytically or otherwise. Thus, the approach we follow here is a clear departure from what has been done before.

4.1 Food-web motifs

When neglecting cannibalism, i.e., self links, there are two possible motifs comprising a pair of species: (i) single links, $A \to B$, and (ii) double links, $A \leftrightarrow B$. For triplets of species, there are 13 possible distinct motifs when neglecting cannibalism (Fig. 4.1). We label these three-species motifs S1–S5 and D1–D8, where the “S” denotes that the motif includes only single links and the “D” denotes that the motif includes at least one double link.

We study the statistics of motifs in model-generated food webs in order to check if the two basic ingredients for food web construction specified above can satisfactorily describe not only the global properties of empirical food webs, such as the distributions of number of prey, but also its local structure[59,60]. Because it allows for analytical treatment, we will focus first on the generalized cascade model, the simplest model obeying those ingredients[60]. Specifically, we study analytically and numerically the motif probabilities for the generalized cascade model and find agreement between the analytical expressions and the empirical results. We conclude that the model is able to capture the basic properties of the local structure of food webs. Therefore, simple static models as the generalized cascade provide a good unifying description of food web structure both at the global and local levels.

4.2 Testing the significance of a particular motif

To test the significance of the appearance of a particular motif in a network, the number of appearances N_{real} of a motif in the real network is compared to its appearances in an ensemble of randomized networks[61]. This comparison yields the z-score

$$z_i = \frac{N_{\text{real}} - \langle N_{\text{rand}} \rangle}{\sigma_{N_{\text{rand}}}}\,, \tag{4.1}$$

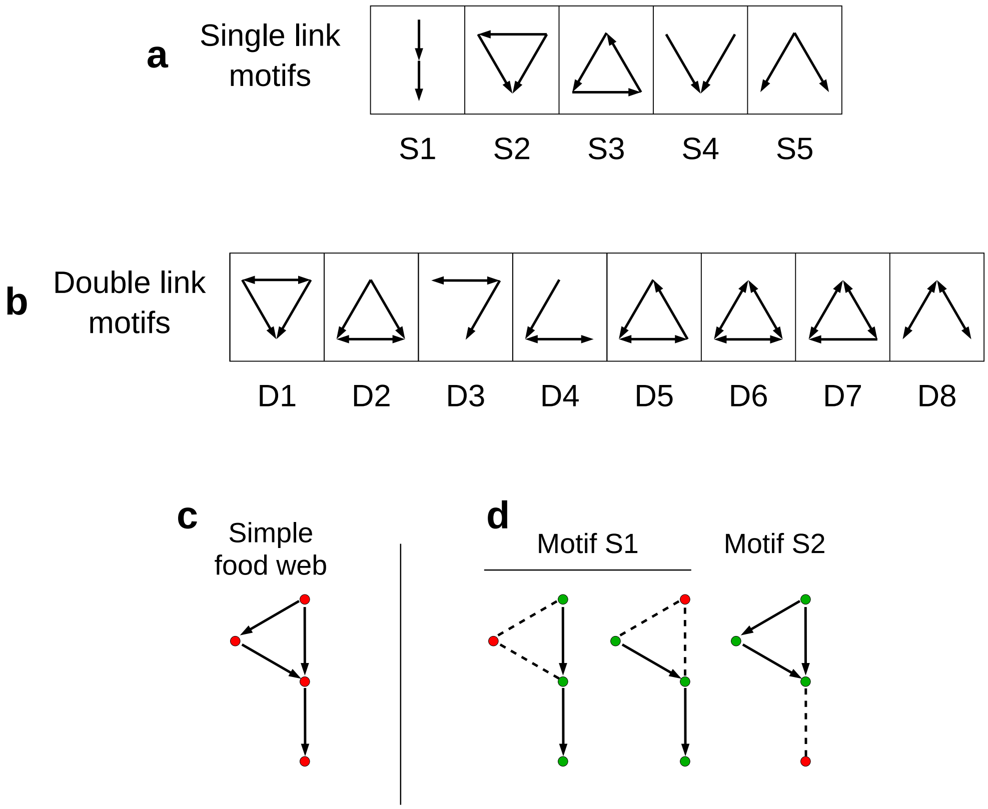

Figure 4.1: Food-web motifs. When neglecting cannibalism, there are 13 unique food-web motifs composed of three species[61]. To simplify our analysis and presentation, we separate the 13 motifs into two groups: **a**, motifs S1–S5 which only include single links and **b**, motifs D1–D8 which include double links (mutual predation). **c**, A simple food web. **d**, If we search the food web in **c** for food-web motifs, we find two instances of motif S1 and one instance of motif S2. Note that enumeration of food-web motifs counts separately all connected species triplets and that many of these subgraphs have a clear ecological relevance, as related to some community modules. For example, motif S1 describes the simple food chain, S2 simple omnivorism, S3 a trophic loop involving three species, S4 isolated exploitative competition, and S5 isolated generalist predation.

where $\langle N_{\text{rand}} \rangle$ and $\sigma_{N_{\text{rand}}}$ are the average and standard deviation of the random ensemble respectively. The z-score of motif i thus measures the significance of deviations of the network from the null hypothesis. We represent the set of z-scores for three-species motifs as a vector $\mathcal{Z} = \{z_i\}$, which has one component each for the motifs S1–S5 and D1–D8. Williams & Martinez[27] used the

z-score as an important means to validate the niche model, as did Cattin *et al.*[28] in their validation of the nested-hierarchy model. It is important to note that use of the z-score implicitly assumes the underlying values are normally distributed. We have verified that this assumption is valid for the distribution of N_{rand}.

The profile of z-scores indicates whether particular motifs are over- or under-represented relative to the null hypothesis provided by randomized networks. A key point to emphasize is that these are relative descriptors for the appearances of a given motif. Therefore, over-representation implies that a motif appears more frequently than in the randomized network, while under-representation implies less frequent occurrence. Under-representation rarely implies absence, particularly in larger networks.

4.3 Generating an ensemble of random networks

For the purpose of generating the ensemble of random networks, one must specify the constraints to be fulfilled while randomizing the network[66,70]. These constraints correspond in practice to specifying the correct null hypothesis. In our analysis of the motifs, we conserve the following attributes for every species during the randomizations: (i) number of prey, (ii) number of predators, (iii) number of single links, $A \to B$, (iv) number of double links, $A \leftrightarrow B$, and (v) whether or not a species is a cannibal. This preserves the overall distributions of each of these properties, numbers of prey and predators, single links, double links, and cannibal links. This process guarantees that all subsequent results are due to other network properties[61,66,70]. It is clear that this randomization scheme maintains that the same species as cannibals before and after. Because of this fact, and the fact that the appearance of cannibalism in food webs is consistent with the random hypothesis[27,32], we neglect cannibalism in our investigation of prey selection mechanisms.

It is important for the randomization process to preserve the distributions of lower-order motifs when considering motifs of a specified size to maintain consistency and validity between results at different levels[61]. Therefore, when examining motifs of size n one must preserve motifs of size $n-1$. This is why we preserve the distributions of single and double links in the examination of three-species motifs.

To randomize a network while conserving all of these properties, we use the Markov-chain Monte Carlo switching algorithm[70] and treat single, double, and cannibal links separately. For example, two single links $A \rightarrow B$ and $C \rightarrow D$ can become $A \rightarrow D$ and $C \rightarrow B$, provided both $A \rightarrow D$ and $C \rightarrow B$ do not already exist in the network and they do not form new double links. Similarly, two double links $A \leftrightarrow B$ and $C \leftrightarrow D$ can become $A \leftrightarrow D$ and $C \leftrightarrow B$, provided both A, D and C, B are unconnected by a link in any direction.

4.4 Prey selection in food web models

Let us examine further the implications of our two conditions which explain global food web structure. The first condition was first proposed in the cascade model of Cohen and Newman[10]. They hypothesized that intermediate species and top predators in an ecosystem can be ranked, i.e., assigned an ordered set of "niche values". Remarkably, a single factor—species' mass—is thought to provide an excellent proxy for niche value[23,25,52,71,72].[*]

The latter condition, again, is shared by the three models; however, the manner in which prey are selected is distinctly different in each. In the generalized cascade model[32]—and, approximately, in the nested-hierarchy model[28]—predators select their prey at random from species with

---

[*] This idea can be generalized to include basal species by setting their niche value to zero thereby placing them at the bottom of the trophic hierarchy.

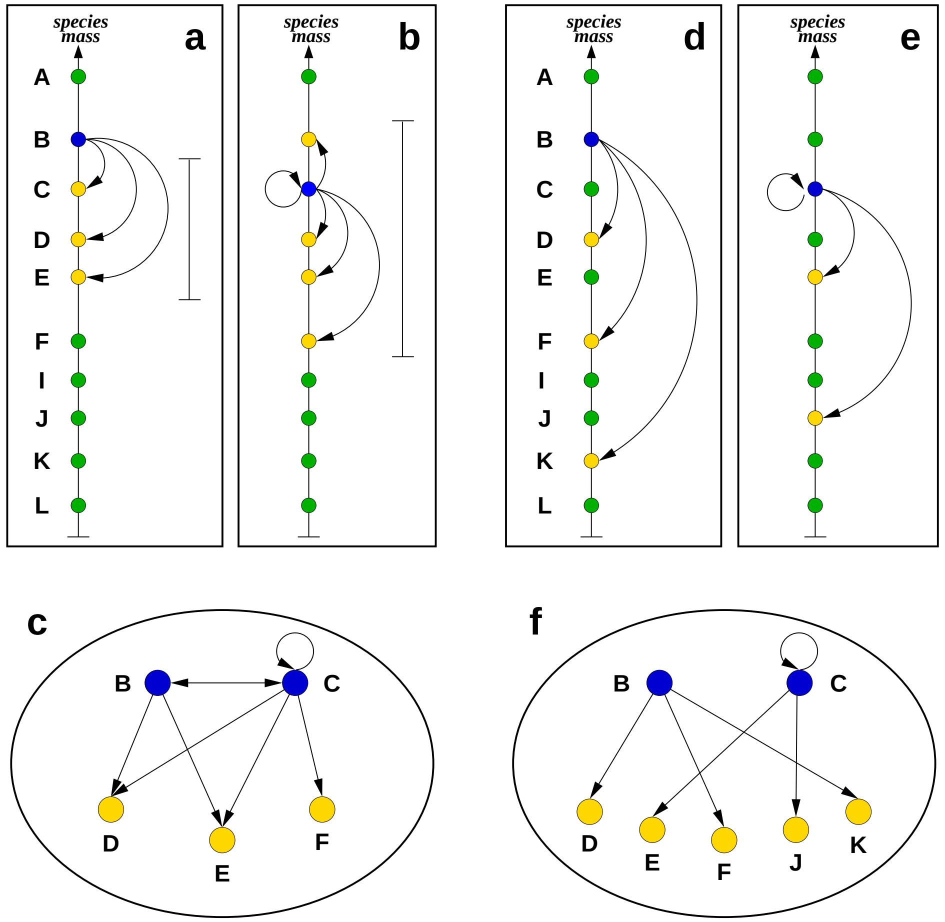

Figure 4.2: Two possible mechanisms for prey selection. Top and intermediate species can be assigned to trophic niches according to some "niche value". This value is typically thought to be species' mass as predators predominantly prey on other species with mass less-than or equal-to their own. We represent predators as blue circles and prey as yellow circles. **a** and **b**, The predator preys upon species within a contiguous range, represented by the solid bar, as in the niche model. **c**, Network representation of the corresponding food web. Under this prey selection mechanism, if two species appear in a double link and consume each other, they must be close together on the niche axis. In addition, their contiguous range of prey will overlap, making it likely that they have common prey. Similarly, two species which are close together on the niche axis are very likely to share predators because of the contiguous range. **d** and **e**, The predator consumes prey with lower or equal niche value at random, as in the generalized cascade model. **f**, Network representation of the corresponding food web. Because of the random prey selection mechanism, there is a substantially weaker relationship between the location of the two predators on the niche and the likelihood that they share predators and prey.

niche values less than or equal to their own, i.e., species of smaller or equivalent mass (Fig. 4.2). In the niche model[27], in contrast, predators consume prey falling within a contiguous range of niche values (Fig. 4.2). Using mass as a proxy, the generalized cascade and nested-hierarchy models implicitly assume that predators are indifferent to the size of their prey, provided only that they are of smaller or equivalent mass; the niche model implicitly assumes that selection pressures force predators to most efficiently prey upon species whose mass is close to a “characteristic value”.

4.5 Motifs related to niche ordering

Before we consider how these prey selection mechanisms would translate into food-web motifs, let us attempt to predict the motif expression using only the ranking induced by species’ mass. We would find it unsurprising to observe over-representation of motif S1, as this motif is the simple food chain[73] (Fig. 4.1). Likewise, we would expect to find over-representation of motif S2 since it represents omnivory[44]—a predator consuming species from two different, but lower, trophic levels—and thus continues to uphold the mass ordering within the food web (Fig. 4.1). In contrast, we would expect under-representation of motif S3, again on a basis of species having ordered niche values (Fig. 4.1). Motif S3 is a clear example of species “violating” the presumed ordering by consuming a species of greater mass.

4.6 Motifs in the generalized cascade model

Because of its simplicity, the generalized cascade model can be treated analytically to determine the expected over- and under-representation of food-web motifs[60]. Since predators in the generalized cascade model cannot feed on species having a larger niche value, no trophic loops of any size

Table 4.1: We show the number of single, $A \to B$, double, $A \leftrightarrow B$, and cannibal $A \to A$ links in each empirical food web. Here an arrow goes from predator to prey.

Food web	S	L	Single links	Double links	Cannibal links	Reference
Benguela	29	203	186	5	7	Yodzis[41]
Bridge Brook Lake	25	106	102	1	3	Havens[18]
Canton Creek	102	697	696	0	1	Townsend *et al.*[38]
Caribbean Reef	50	556	471	32	21	Opitz[42]
Chesapeake Bay	31	68	67	0	1	Baird & Ulanowicz[34]
Coachella Valley	29	262	199	22	19	Polis[44]
El Verde Rainforest	155	1509	1369	69	2	Waide & Reagan[45]
Grassland	61	97	97	0	0	Martinez *et al.*[11]
Little Rock Lake	92	997	936	24	13	Martinez[39]
Northeast US Shelf	79	1400	1363	7	25	Link[43]
Scotch Broom	85	223	219	0	4	Hawkins *et al.*[46]
Skipwith Pond	25	197	181	4	8	Warren[40]
St. Marks Seagrass	48	221	218	0	3	Christian & Luczkovich[35]
St. Martin Island	42	205	205	0	0	Goldwasser & Roughgarden[47]
Stony Stream	109	829	827	0	2	Townsend *et al.*[38]
Ythan Estuary (1)	83	395	389	1	4	Hall & Raffaelli[36]
Ythan Estuary (2)	174	579	573	1	4	Hall & Raffaelli[37]

can exist. As a consequence there is no mutual predation and none of the eight unique motifs which contain double links will be observed in the generalized cascade model. Therefore our analytical derivations focuses on the five motifs S1–S5 (Fig. 4.1), while the analysis of the motifs containing double links will be dealt with later[60]. The present analysis is nonetheless meaningful since single connections account for the vast majority of the links in empirical food webs (Table 4.1). Notice that these motifs have a clear ecological relevance and are additionally related to some proposed community modules; in particular, motif S1 describes the simple food chain, motif S2 simple omnivorism, motif S3 a trophic loop involving three species, motif S4 isolated exploitative competition, and motif S5 isolated generalist predation[60].

The probability p^i of observing motif i is related to the number of appearances of the motif N^i by

$$p^i = \frac{N^i}{S\left(S-1\right)\left(S-2\right)/6}\,, \tag{4.2}$$

where the denominator is the total number of possible triplets of species. We choose the probability p, instead of the number of appearances N because, as we will later demonstrate, the probability is not a function of S, and instead depends on a single variable, the directed connectance C. This property is a very interesting one because it allows a unified description of food webs of different size.

Recall that no trophic loops are possible within the generalized cascade model. Motif S3 is therefore forbidden and

$$p^{\mathrm{S3}} = 0\,. \tag{4.3}$$

We next derive expressions for the remaining motifs, S1, S2, S4, and S5. The probability for a given motif to appear is equivalent to the probability for three arbitrary species to be connected in the specified fashion. Let us now consider three arbitrary species, A, B, and C, with $n_A > n_B > n_C$. We call x_i the probability of species i consuming each species with lower niche values. It then follows that

$$\begin{aligned} p^{\mathrm{S1}} = p^{\mathrm{S4}} &= \left\langle x_A\, x_B \right\rangle - \left\langle x_A^2\, x_B \right\rangle\,, && (4.4)\\ p^{\mathrm{S2}} &= \left\langle x_A^2\, x_B \right\rangle\,, && (4.5) \end{aligned}$$

and

$$p^{\mathrm{S5}} = \left\langle x_A^2 \right\rangle - \left\langle x_A^2\, x_B \right\rangle\,, \tag{4.6}$$

where $\langle \ldots \rangle$ indicates the average over the probability distribution $p(x)$. In addition, because x_A and x_B are independent random variables, Eqs. (4.4)–(4.6) can be rewritten as

$$p^{\mathrm{S1}} = p^{\mathrm{S4}} = \langle x \rangle^2 - \left\langle x^2 \right\rangle \langle x \rangle \ , \tag{4.7}$$

$$p^{\mathrm{S2}} = \left\langle x^2 \right\rangle \langle x \rangle \ , \tag{4.8}$$

and

$$p^{\mathrm{S5}} = \left\langle x^2 \right\rangle - \left\langle x^2 \right\rangle \langle x \rangle \ . \tag{4.9}$$

These expressions are valid for arbitrary distributions $p(x)$. Substituting Eq. (2.1), the beta-distribution, they become

$$p^{\mathrm{S1}} = p^{\mathrm{S4}} = 4C^2 \frac{1-2C}{1+2C} \, , \tag{4.10}$$

$$p^{\mathrm{S2}} = \frac{16C^3}{1+2C} \, , \tag{4.11}$$

and

$$p^{\mathrm{S5}} = 8C^2 \frac{1-2C}{1+2C} \, . \tag{4.12}$$

In Fig. 4.3, we compare the analytical predictions, Eqs. (4.10)–(4.12), with simulations of the generalized cascade model. It becomes visually apparent that the expressions we derived compare quite well with the model-generated data[60]. Notice that the probabilities p only depend on the connectance C.

4.7 Motifs in randomizations of the generalized cascade model

As a second test of the generalized cascade model, we analyze which motifs are typically over- or under-represented as compared to the corresponding randomized networks[61,62]. The randomized networks are obtained by preserving the number of prey k_i and of predators m_i of each species i as

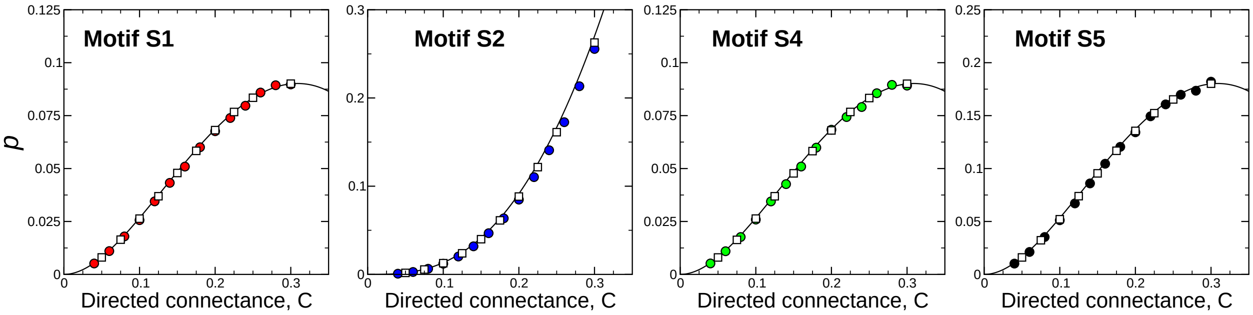

Figure 4.3: Comparison between analytical expressions, Eqs. (4.10)–(4.12), and simulations of the generalized cascade model for motifs S1, S2, S4, and S5. We exclude motif S3 because, by definition, $p^{\mathrm{S3}} = 0$. It is visually apparent that the analytical predictions agree with the model-generated data. Filled circles are for food webs with $S = 50$ and open squares for food webs with $S = 100$. Each data point represents an average over 1000 model-generated food webs.

in the generalized cascade model, but rewiring their trophic links randomly using the Markov-chain Monte Carlo switching algorithm[70,74].

By virtue of the randomization, a species in the randomized network may feed on a species with a higher niche value than itself, a possibility that is excluded in the model. However, because of the formulation of the Markov-chain Monte Carlo switching algorithm, no double links are produced in the randomization and the resulting networks only contain motifs S1–S5[70,74]. Notice however that, by construction, the distributions of number of prey or of number of predators are the same as in the original network. Then, one must not confuse these randomized networks with *completely* random networks, whose distributions are different from the original ones.

Note that, since randomized food webs possess the same degree distributions as the original ones, the occurrence of patterns of over- and under-representation of motifs in empirical food webs would require an explanation. One can think of two principle arguments for their existence: either they are a consequence of the mechanism generating the network[66], or they provide some ecological advantage and have arisen as a result of selection pressure. Here we show that the generalized

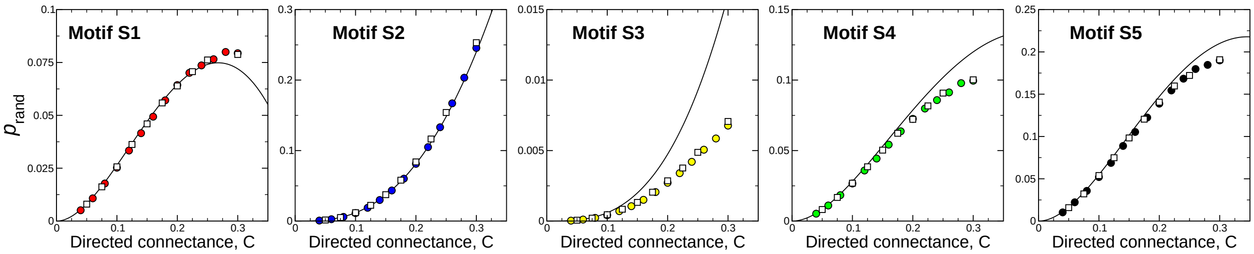

Figure 4.4: Comparison between analytical expressions, Eqs. (4.29)–(4.33), and randomizations of the generalized cascade model for motifs S1–S5. It is visually apparent that the analytical predictions compare well to the model-generated data. Filled circles are for food webs with $S = 50$ and open squares for food webs with $S = 100$. Each data point represents an average over 1000 model-generated food webs.

cascade model yields well-defined patterns of over- and under-representation of motifs[60]. If these predictions compare well with empirical food webs, one might conclude that no need of the second hypothesis is required and all patterns arise as the result of the food web generating mechanisms[60].

Let us now calculate the motif probabilities for the randomizations of the generalized cascade model[60]. Itzkovitz *et al.*[75] derived general expressions for the average number of appearances N^i of motifs in randomized networks[75]. The fraction of motifs p^i_{rand} is obtained dividing N^i by the total number of possible triplets of species, $S_T \equiv S(S-1)(S-2)/6$. For motifs S1–S5, these can be cast as

$$p^{S1}_{rand} = p_1 - p^{S2}_{rand} - 3p^{S3}_{rand}\,, \tag{4.13}$$

$$p^{S4}_{rand} = p_4 - p^{S2}_{rand}\,, \tag{4.14}$$

$$p^{S5}_{rand} = p_5 - p^{S2}_{rand}\,. \tag{4.15}$$

with

$$p_1 = \frac{S}{S_T} < k_i m_i >\,, \tag{4.16}$$

$$p_4 = \frac{S}{2S_T} < m_i(m_i-1) >\,, \tag{4.17}$$

$$\begin{aligned} p_5 &= \frac{S}{2S_T} < k_i(k_i-1) > \,, && (4.18)\\ p^{S2}_{rand} &= \frac{1}{z^3 S_T} < k_i(k_i-1) >< k_i m_i >< m_i(m_i-1) > \,, && (4.19)\\ p^{S3}_{rand} &= \frac{1}{3z^3 S_T} < k_i m_i >^3 \,, && (4.20) \end{aligned}$$

where $z \equiv L/S$ is the average connectivity, k_i and m_i denote the number of prey and number of predators of species i, respectively, and $< \ldots >$ is the average over all species in the randomized network. Since these networks have the same distributions of in- and out-links as the original networks, these averages can be calculated directly from the latter ones.

From Eqs. (4.16)–(4.20), p^{S2}_{rand} and p^{S3}_{rand} can be rewritten as

$$\begin{aligned} p^{S2}_{rand} &= \frac{4S_T^2}{S^3 z^3} p_1 p_4 p_5 \,, && (4.21)\\ p^{S3}_{rand} &= \frac{S_T^2}{3S^3 z^3} p_1^3 \,. && (4.22) \end{aligned}$$

Therefore, p^i_{rand} can be evaluated if one knows p_1, p_4 and p_5. In order to calculate these quantities, let us note that they have a direct interpretation. (i) $N^1 = \sum_{i=1}^S k_i m_i$ is the number of configurations where species A eats B, and B eats C, independently if there is a trophic connection between species A and C (i.e. it is like a generalization of motif S1); therefore, $p_1 = N^1/S_T$ is just the probability for this configuration, namely $< x_A x_B >=< x >^2$, since x_A and x_B are independent random variables. (ii) $N^2 = \sum_{i=1}^S m_i(m_i-1)$ is the number of configurations where species A feeds on C and species B feeds on C, independently if A and B are connected (i.e. like a generalization of motif S4); then, p_4 is the probability $< x_A x_B >=< x >^2$. (iii) Similarly, p_5 is the probability of A eating species B and C, independently of the eventual connection of B and C, namely $< x_A^2 >=< x^2 >$.

By replacing these results in Eqs. (4.13)–(4.22), one finds

$$p^{\rm S1}_{\rm rand} = < x^2 > - \frac{< x >^4 < x^2 >}{9C^3} - \frac{< x >^6}{36C^3} \,, \qquad (4.23)$$

$$p^{\rm S2}_{\rm rand} = \frac{<x>^4 <x^2>}{9C^3}\,, \tag{4.24}$$

$$p^{\rm S3}_{\rm rand} = \frac{<x>^6}{108C^3}\,, \tag{4.25}$$

$$p^{\rm S4}_{\rm rand} = <x>^2 - \frac{16C}{9} <x^2>\,, \tag{4.26}$$

$$p^{\rm S5}_{\rm rand} = <x^2> (1 - \frac{16C}{9})\,. \tag{4.27}$$

Finally, the beta-function, Eq. (2.1) yields

$$\begin{aligned} <x> &= 2C, \\ <x^2> &= \frac{8C^2}{1+2C}. \end{aligned} \tag{4.28}$$

The substitution of these expressions into Eqs. (4.23)–(4.27) gives the calculate the probabilities for the three-species motifs in the randomizations of the generalized cascade model, yielding

$$p^{\rm S1}_{\rm rand} = 4C^2 \frac{1 - 2C - 8C^2/9}{1+2C}\,, \tag{4.29}$$

$$p^{\rm S2}_{\rm rand} = \frac{178}{9} \frac{C^3}{1+2C}\,, \tag{4.30}$$

$$p^{\rm S3}_{\rm rand} = \frac{16}{27} C^3\,, \tag{4.31}$$

$$p^{\rm S4}_{\rm rand} = 4C^2 \frac{1 - 14C/9}{1+2C}\,, \tag{4.32}$$

and

$$p^{\rm S5}_{\rm rand} = 8C^2 \frac{1 - 16C/9}{1+2C}\,. \tag{4.33}$$

Figure 4.4 compares the analytical predictions for the randomizations, Eqs. (4.29)–(4.33), with simulations of the generalized cascade model finding good agreement[60]. The small discrepancies observed have their origin in that some of the expressions derived in [75] are approximate, whereas our expressions for p are exact.

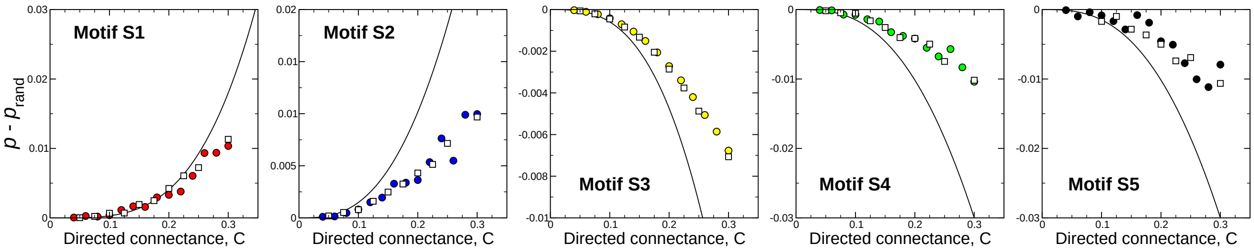

Figure 4.5: Comparison between analytical expressions for $p - p_{\rm rand}$ and simulations of the generalized cascade model for motifs S1–S5. The model predicts over-representation of motifs S1 and S2 and under-representation of motifs S3, S4, and S5. The analytical predictions compare well with the model-generated data, though they generally overestimate the differences due to the approximate character of the expressions used to evaluate p_{ran}. Filled circles are food webs with $S = 50$ and open squares food webs with $S = 100$. Each data point represents an average over 1000 model-generated food webs.

4.8 Over- and under-representation of motifs in the generalized cascade model

Finally, we obtain the differences by subtracting the probability of motifs appearing in the model, Eqs. (4.10)–(4.12), and in their randomizations Eqs. (4.29)–(4.33). Table 4.2 summarizes these results[60]. We show comparisons between the expressions for $p - p_{\rm rand}$ and simulations of the generalized cascade model in Fig. 4.5. Our analytical derivations thus predict that food webs generated by the generalized cascade model have over-expression of motifs S1 (a food chain) and S2 (simple omnivorism) and under-representation of motifs S3 (a trophic loop), S4 (isolated exploitative competition), and S5 (isolated generalist predation)[60]. The percentage of under/over-representation is, however, rather small, generally under the 10%.

Table 4.2: Analytical expressions for over- and under-representation of motifs S1–S5 in the generalized cascade model. The second column gives the difference between the probability of each motif appearing in a generalized cascade model-generated web and in a randomization of a model-generated web. If this value is greater than zero a motif is over-represented; values less than zero imply under-representation.

Three-node motif	$p - p_{\mathrm{rand}}$	Representation
S1	$\frac{32}{9}\frac{C^4}{1+2C}$	Over
S2	$\frac{16}{9}\frac{C^3}{1+2C}$	Over
S3	$-\frac{16}{27}C^3$	Under
S4	$-\frac{16}{9}\frac{C^3}{1+2C}$	Under
S5	$-\frac{16}{9}\frac{C^3}{1+2C}$	Under

4.9 Over- and under-representation of motifs in the niche and nested-hierarchy models

Food webs generated by the generalized cascade model display an over-representation of motifs S1–S2 and an under-representation of motifs S3–S5[60]. Because the niche and nested-hierarchy models are, in essence, specialized implementations of the generalized cascade model [32], we expect them to generate similar profiles for motifs S1–S5. Numerical simulations demonstrate that this is indeed the general case[59] (Fig. 4.6). The nested-hierarchy model, however, exhibits a small tendency to deviate from this pattern for arbitrary combinations of S and L[59].

In Fig. 4.6, we show the results of numerical simulations of the niche and nested-hierarchy models for motifs S1–S5. Our simulations show that food webs generated by the the niche model are characterized by over-representation of motifs S1–S2 and under-representation of motifs S3–

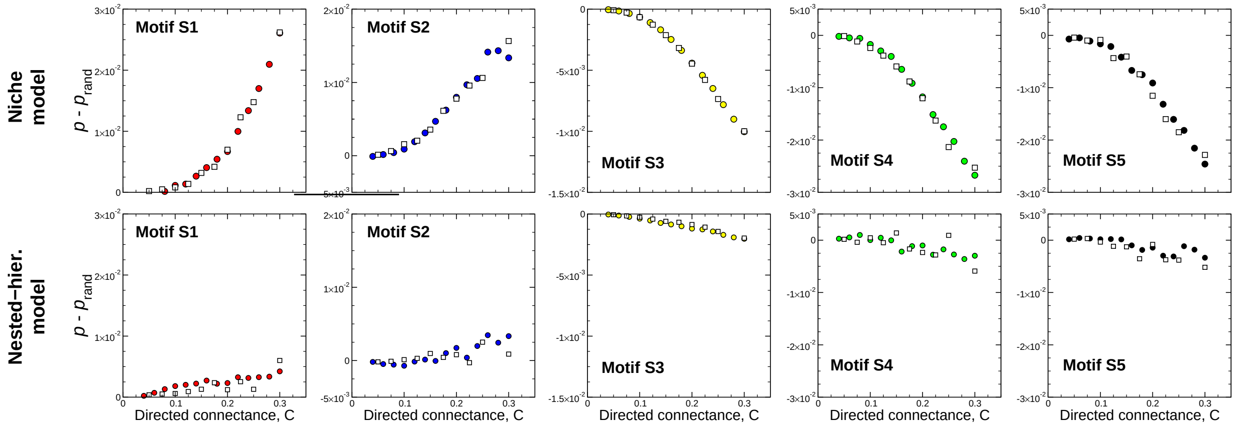

Figure 4.6: We show the over- and under-representation of motifs S1–S5, $p - p_{\rm rand}$, for simulations of the niche and nested-hierarchy models in the first and second rows, respectively. We observe that both models exhibit the same qualitative behavior as each other and as the generalized cascade model. Filled circles are food webs with $S = 50$ and open squares food webs with $S = 100$. Each data point represents an average over 1000 model-generated food webs.

S5, just as we found for the generalized cascade model[59]. Our numerical simulations also demonstrate that food webs generated by the nested-hierarchy model generally feature over-representation of motifs S1–S2 and under-representation of motifs S3–S5, just as in the generalized cascade and niche models[59].

We were unable to obtain analytical results for motifs D1–D8 beyond the trivial case of the generalized cascade model which does not generate double links. Nevertheless, we can still gain insight through heuristic arguments based upon the mechanisms being implemented in the models[59]. Let us first assume that predators consume a contiguous range of species, as in the niche model. Then species with similar niche value are likely to share predators. In addition, for two species to be connected by a double link, they must have similar niche values. This implies that their respective ranges of prey are likely to have substantial overlap (Fig. 4.2). These observations can be trivially translated into predictions of motif representation. Because two species connected by a double link are likely to share prey, we predict over-representation of motif D1 and under-representation of D3. Because two species connected by a double link are likely to share predators, we predict over-representation of motif D2 and under-representation of motif D4. These two cases together further enable us to predict under-representation of motif D5.

Additionally, if $A \leftrightarrow B$ and $C \leftrightarrow B$, it is necessary for all of A, B, and C to have similar niche values. It is then likely that A and C are also connected by a double link and unlikely they have no connection between them. We therefore predict over-representation of motif D6 and under-representation of motif D8.

Let us now consider the case in which prey selection among species of smaller or equivalent mass is random, like in the generalized cascade model, but where some double links are possible. Since prey are chosen at random, species do not need to have similar niche values to be connected by a double link, and further, species connected by a double link are no more (or less) likely to share prey or predators. This is similar to what one expects for the nested-hierarchy model as the ranking imposed by species mass is indeed relaxed in the implementation[32]. For motifs D1–D8, we then expect little difference between food webs generated by the nested-hierarchy model and their randomizations, in contrast to the strongly correlated and systematic pattern exhibited by the niche model.

We show results of simulations of the niche and nested-hierarchy model for motifs D1–D8 in Fig. 4.7. Our numerical simulations of the niche model clearly support our heuristic arguments, which predict over-representation of motifs D1, D2, and D6 and under-representation of motifs D3, D4, D5, and D8[59]. In addition, for motifs D1–D8 in the nested-hierarchy model, we find very different behavior from that predicted by the niche model. We find very little difference between food webs generated by the nested-hierarchy model and their randomizations (Fig. 4.7), as predicted due to the random nature of prey selection in the nested-hierarchy model.

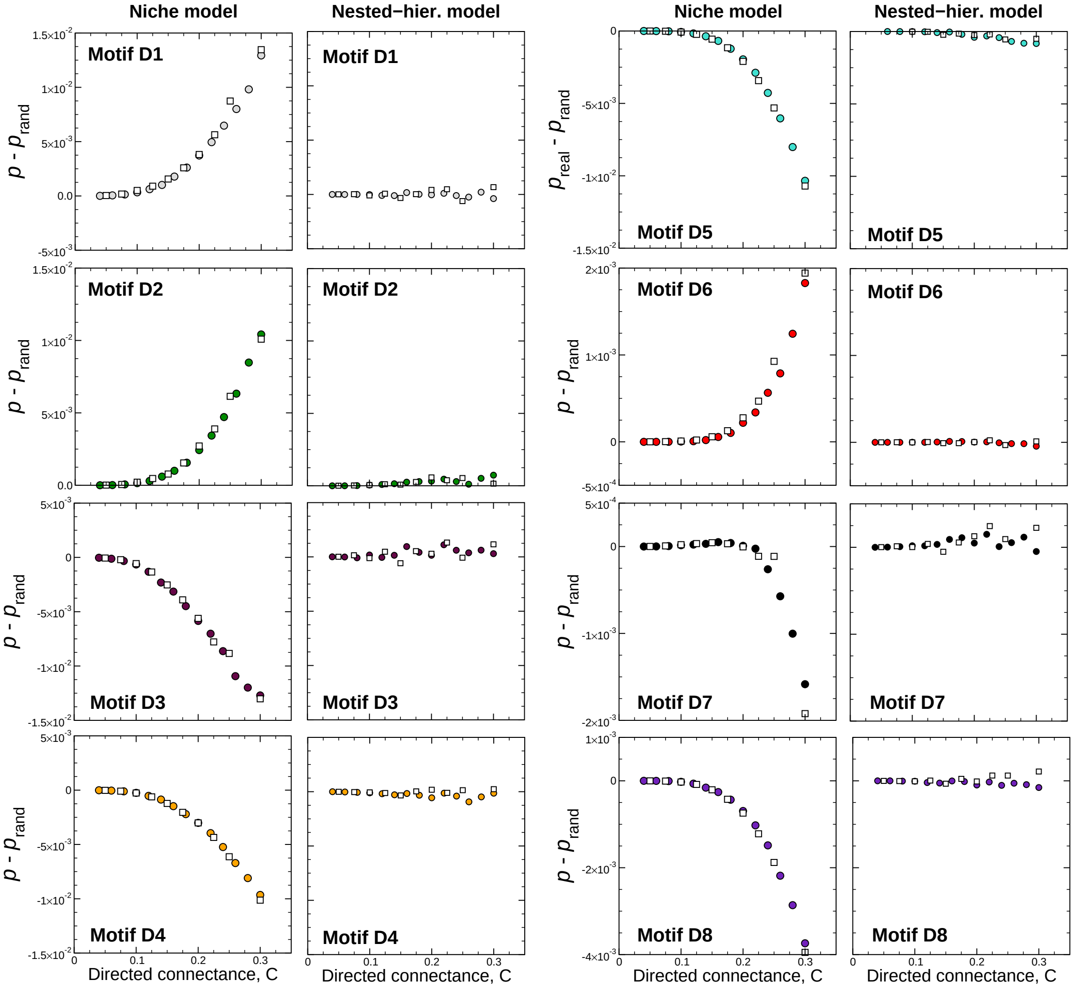

Figure 4.7: We show the over- and under-representation of motifs D1–D8, $p - p_{\mathrm{rand}}$, for simulations of the niche and nested-hierarchy models. The niche model clearly generates food webs with a distinct pattern of over and under-representation. The nested-hierarchy model, however, generates food webs with $p - p_{\mathrm{rand}} \approx 0$, implying that the model-generated and randomized food webs are quantitatively quite similar. Filled circles are food webs with $S = 50$ and open squares are food webs with $S = 100$. Each data point represents an average over 1000 model-generated food webs.

4.10 Concluding remarks

Our analysis of the generalized cascade, niche, and nested-hierarchy models provided insight into how the models' implementations impact the food webs' local structure [59,60]. For motifs S1–S5, the analytical treatment of the generalized cascade model demonstrated over-representation of motifs S1–S2 and an under-representation of motifs S3–S5 [60]. Our numerical simulations of the niche and nested-hierarchy models were also consistent with these patterns [59].

For motifs D1–D8, however, the mechanism of prey selection produced distinct patterns of motif over- and under-representation [59]. We observed that the niche model generates food webs with a consist and robust motif pattern, including over-representation of motifs D1, D2, and D6 and under-representation of motifs D3–D5 and D8. The nested-hierarchy model, in contrast, produces no such conserved pattern. In the next chapter, we will attempt to take advantage of this difference between the two models to determine whether either, if any, model provides a better comparison the the motif pattern observed in empirical food webs.

Chapter 5

Local structure and prey selection in empirical food webs

As in Ch. 3, we now compare the local structure of empirical food webs with our model-based predictions[59]. We observed previously that each of the candidate models—the generalized cascade, niche, and nested-hierarchy models—predict the same general pattern of over- and under-representation of motifs S1–S5[60]. In contrast, the different prey selection mechanisms produce qualitatively and quantitatively distinct patterns of over- and under-representation of motifs D1–D8[59]. We hypothesize examining how the empirical data compares to each of the models will allow us to identify the mechanisms responsible for the observed food-web structure. Interestingly, we find that the over- and under-representation of three-species motifs in empirical food webs can be understood through comparison to a static food-web model, the niche model[59]. Our result conclusively demonstrates that predation upon species with some "characteristic" niche-value is the prey selection mechanism consistent with the local and global structural properties of empirical food webs[59].

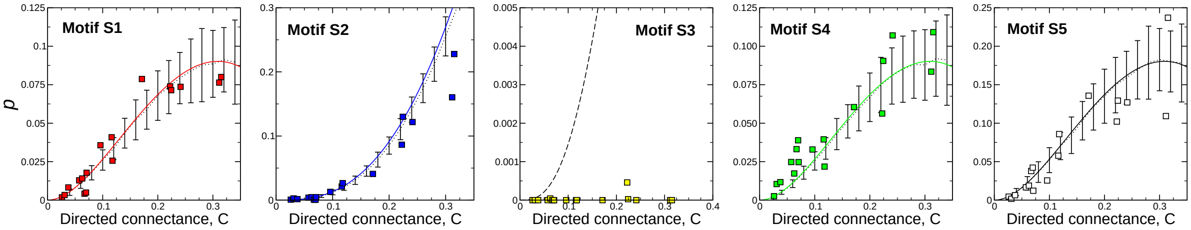

Figure 5.1: Fraction of appearances of subgraphs for empirical food webs (symbols) compared to the analytical predictions for the generalized cascade model (solid lines). Numerical simulations for the generalized cascade model with $S = 50$ are shown by the dotted line where the error bars are two standard deviations. It is visually apparent that the generalized cascade model fits rather well the empirical data for all the motifs. Note that there are no fitting parameters in model estimates.

5.1 Motifs S1–S5 in empirical food webs and their randomizations

One interesting observation from Figs. 4.3–4.6 is that the probabilities generated by model food webs depend on a single variable, the directed connectance C, and very weakly on the size of the food web. This indicates that our representation of the probabilities versus C can be adequate to provide a unified description of empirical data, since it allows us to include in the same plot food webs with different sizes. If empirical food webs behave as model food webs, one expects a common trend for the probabilities as functions of C despite having different S values.

In this section we compute the fraction of appearances for each motif S1–S5 for 17 empirical food webs[60] (Table 3.1). Figures 5.1–5.3 show the results for the empirical food webs, their randomizations and the differences, and also the comparison with the generalized cascade model. One observes that the analytical expressions obtained for the generalized cascade model provide a reasonable agreement with empirical data for p and p_{rand} with no adjustable parameters[60].

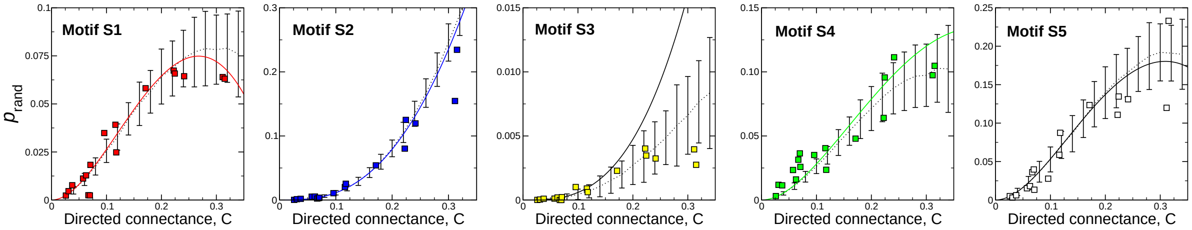

Figure 5.2: Fraction of appearances of motifs for randomizations of the empirical food webs (symbols) compared to the analytical predictions for the generalized cascade model (solid line). Numerical simulations for the generalized cascade model with $S = 50$ are shown by the dotted line where the error bars are two standard deviations. It is visually apparent that the generalized cascade model fits rather well the empirical data for all the motifs. Note that there are no fitting parameters in model estimates.

5.2 Empirical representation pattern for motifs S1–S5

In the plots for the differences, the data generally appear more noisy. This is due to the fact that the empirical values for p and p_{ran} are in general quite similar in magnitude; they commonly differ by less than 10%, in agreement with the model predictions. The general trend however is that motifs S1 and S2 are typically over-represented and motifs S3–S5 are under-represented, in agreement with the qualitative predictions of the model as expressed in Table 4.2. Quantitatively, the theoretical curves generally overestimate the empirical values, while the numerical simulations of the model provide reasonable estimates[60].

There exists, however, more noise in the empirical data than exhibited by the model, in particular for motif S2. To explore this issue further, let us note that it is the same two food webs which seem to deviate from the general trend in the plots of Fig. 5.3: they are Bridge Brook (with $C = 0.17$) and Skipwith Pond ($C = 0.32$). Why exactly those two food webs behave differently from the others is interesting but unclear. We can note that they are the smallest food webs of the ones studied, each with 25 trophic species (see Table 4.1). Although statistical fluctuations grow

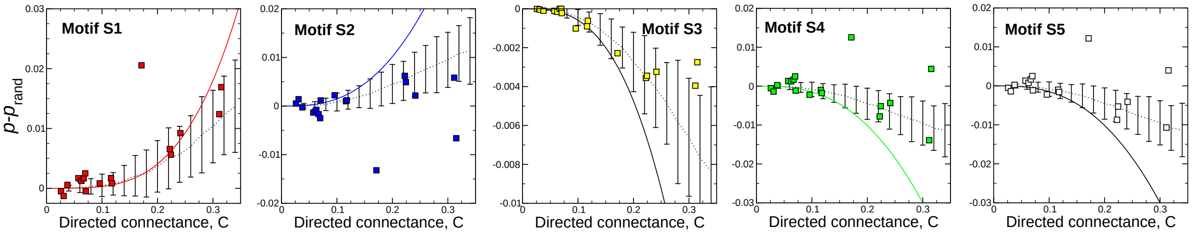

Figure 5.3: Differences between actual appearances of subgraphs and the corresponding randomized food webs for 17 empirical food webs (symbols) as compared to the analytical predictions for the generalized cascade model (solid line). Numerical simulations for the generalized cascade model with $S = 50$ are shown by the dotted line, where the error bars are two standard deviations. Motifs S1 and S2 are typically over-represented and motifs S3–S5 are under-represented, in agreement with the qualitative predictions of the model. The two noticeable deviations correspond to Bridge Brook ($C = 0.17$) and Skipwith Pond ($C = 0.32$). Quantitatively, the analytical curves generally overestimate the differences at larger values of C for both the empirical values and the numerical simulations of the generalized cascade model.

with decreasing size, they do not seem enough to explain this behavior. Note, on the other hand, that they match rather well with the predictions for p and p_{ran} separately (Figs. 5.1–5.2).

In summary, the behavior of the two models indicates that the behavior observed in the empirical data is not a trivial one: not any model would yield a similar behavior for the quantities analyzed. Furthermore, we observe remarkable agreement between the local structure in the generalized cascade model and the empirical data[60].

The predictions of the generalized cascade model for the appearances of motifs S1–S5 provide good comparison to the empirical results[60]. This generalized cascade model was recently shown to fit empirical data for a number of global quantities, including the distributions of the number of prey and predators. Here we show that it also describes the local structure of empirical food webs[60]. This suggests that many features of food web structure could be explained by considering the two principle ingredients inside the model, namely (i) the species' niche values form a totally ordered set, and (ii) each species has a specific exponentially decaying probability of preying on a given fraction of the species with lower niche values. These could then be considered as basic mechanisms actually shaping food webs[60].

Empirical and model food webs predict over-representation of motifs S1 and S2, and under-representation of motifs S3–S5. Empirical data are rather noisy, and the representation of motif S2 predicted by the generalized cascade model is unclear in the empirical data. This is also the result found by Bascompte & Melián[65], who also analyzed the over- and under-representation of a number of ecologically relevant subgraphs, among them, our motifs S1 (food chain) and S2 (simple omnivory). Our methodology, however, differs from theirs in several aspects. On the one hand, we consider trophic species instead of taxonomic ones; on the second hand, we count subgraphs only once (for instance, we do not count the food chains included in motif S2 in the evaluation of S1). Despite the different analysis, we still do not find an unambiguous over-representation of omnivorism, in contrast to what one may expect according to its stabilizing role in trophic interactions[76].

Finally note that, from the quantitative point of view, the percentage of over- and under-representation of motifs is generally small, usually less than a 10%; curiously, this is also the order of magnitude predicted by the generalized cascade model. This high similarity between the number of motifs in empirical food webs and in their randomizations, indicates that there may be no overwhelming evolutionary trend toward over- and under-representation of any motif, and that the small differences observed are more likely a consequence of the mechanisms generating the food web[60].

5.3 Comparing empirical and model-generated motif profiles

Because we do not have analytical predictions for motifs D1–D8, we must follow a different protocol to compare the complete motif profile between model-generated and empirical food webs[59]. To do so, we first generate a model food web with the same number of species S and directed connectance $C = L/S^2$ as the empirical food web, where L is the number of trophic links within the food web. The implementation of the niche model[27], nested-hierarchy model[28], and generalized cascade model[32] is followed as earlier detailed. It is interesting to note that all three models require no free parameters as the only required inputs are the number of species S and number of links L, both of which are obtained directly from the empirical data.

We then compute the z-scores of the model-generated food web. To measure how well the model compares to the empirical data, we calculate two quantities: the uncentered correlation coefficient r and ratio d between the empirical and model-generated data[59]. For each empirical food web this process is repeated at least 250 times per model.

The uncentered correlation coefficient r between the z-score vectors $\mathcal{Z}_a$ and $\mathcal{Z}_b$ of two food webs a and b, respectively, is defined as

$$r = \sum_{j=1}^{m} \left(\frac{z_{a,j}}{\sigma^{(0)}_{\mathcal{Z}_a}} \right) \left(\frac{z_{b,j}}{\sigma^{(0)}_{\mathcal{Z}_b}} \right) , \tag{5.1}$$

where

$$\sigma^{(0)}_{\mathcal{Z}_a} = \sqrt{\sum_{j=1}^{m} \left(z_{a,j} \right)^2 } , \tag{5.2}$$

a specifies the food web, and j is an index over motifs. From the mathematical viewpoint, the uncentered correlation coefficient is the scalar product of the unit vectors in the directions of $\mathcal{Z}_a$ and $\mathcal{Z}_b$ and thus is equal to the cosine of the angle between the two data vectors in an m-dimensional space.

Since the vector $\mathcal{Z}$ describes the shape of over- and under-representation of motifs in a food web (Fig. 5.4), a value of r close to 1 indicates that the two food webs have similar directions, i.e., very similar profiles; a value of r close to 0 indicates little similarity, and close to -1 indicates the two behave oppositely (i.e., the motifs under-represented in one food web are typically over-represented in the other and vice versa). Therefore, two similarly characterized food webs will have r values close to 1 upon comparing their respective z-scores. We would observe the same upon comparing an empirical food web and a model food web which exhibit similar behaviors.

The uncentered correlation coefficient compares the directions of the z-score vectors, but ignores their magnitudes. In order to fully evaluate the similarity of two profiles, we must consider the similarity of the magnitudes as well. To this end, we introduce the ratio d of norms of the z-score vectors $\mathcal{Z}_a$ and $\mathcal{Z}_b$ of two food webs a and b, respectively, defined as

$$d = \frac{|\mathcal{Z}_a|}{|\mathcal{Z}_b|}, \tag{5.3}$$

where

$$|\mathcal{Z}_a| = \sqrt{\sum_{j=1}^{m} (z_{a,j})^2}, \tag{5.4}$$

a specifies the food web, and j is an index over motifs. The ratio d provides a measure of the relative length of the two z-score vectors in an m-dimensional space. This ratio is a measure of how similar two food webs are in the magnitudes of their motif over- and under-representation.

Unlike previous studies which focused upon P-values, e.g. Milo *et al.*[61] and Bascompte & Melián[65], we choose to compare the empirical and model motif patterns using our two metrics r and d for the following reasons. Earlier studies were concerned with individual motifs with respect to some null hypothesis model. By utilizing our two metrics r and d, in contrast, our investigation is able to quantify the representation of the *complete* pattern of motifs simultaneously. This is because a comparison between, for example, a model and empirical food web, can only be considered in good agreement if both r and d give favorable results. P-values, in contrast, could be significant if the measure has a broad distribution and so thus little indication of a lack of strong agreement.

5.4 Empirical representation pattern for all motifs

In order to test our model-based predictions, we study sixteen empirical food webs from a variety of environments[59]: three estuarine—Chesapeake Bay[34], St. Marks[35], and Ythan[36]; five freshwater—Bridge Brook Lake[18], Canton Creek[38], Little Rock Lake[39], Skipwith Pond[40], and Stony Stream[38]; three marine—Benguela[41], Caribbean Reef[42], and Northeast US Shelf[43]; and five terrestrial—Coachella Valley[44], El Verde Rainforest[45], Grassland[11], Scotch Broom[*][46], and St. Martin[47] (Table 3.1). We present the profiles of over- and under-representation for the sixteen empirical food webs studied in Fig. 5.4.

We first consider motifs S1–S5[59]. We find under-representation of motif S3 is common among all sixteen webs and over-representation of motif S1 for all but three webs. These three food webs, Chesapeake Bay, Scotch Broom, and Grassland, have the lowest linkage density L/S of the sixteen webs, where S is the number of trophic species in the food web and L is the number of trophic interactions between these species. It has been shown previously that many properties of the niche model exhibit greater variance at low size, linkage density, and/or directed connectance[32]. Many of these properties represent averages across the food web; therefore, the smaller these three attributes the more sensitive the calculations are to fluctuations within the data.

For the remaining motifs, S2, S4, and S5, we gain additional insight upon comparison to our analytical predictions for the generalized cascade model. Using this basis, we can separate the food webs in two groups[59]: the ten food webs which exhibit over-representation of motif S2 and

---

[*] Note that Scotch Broom is a source web.[46]

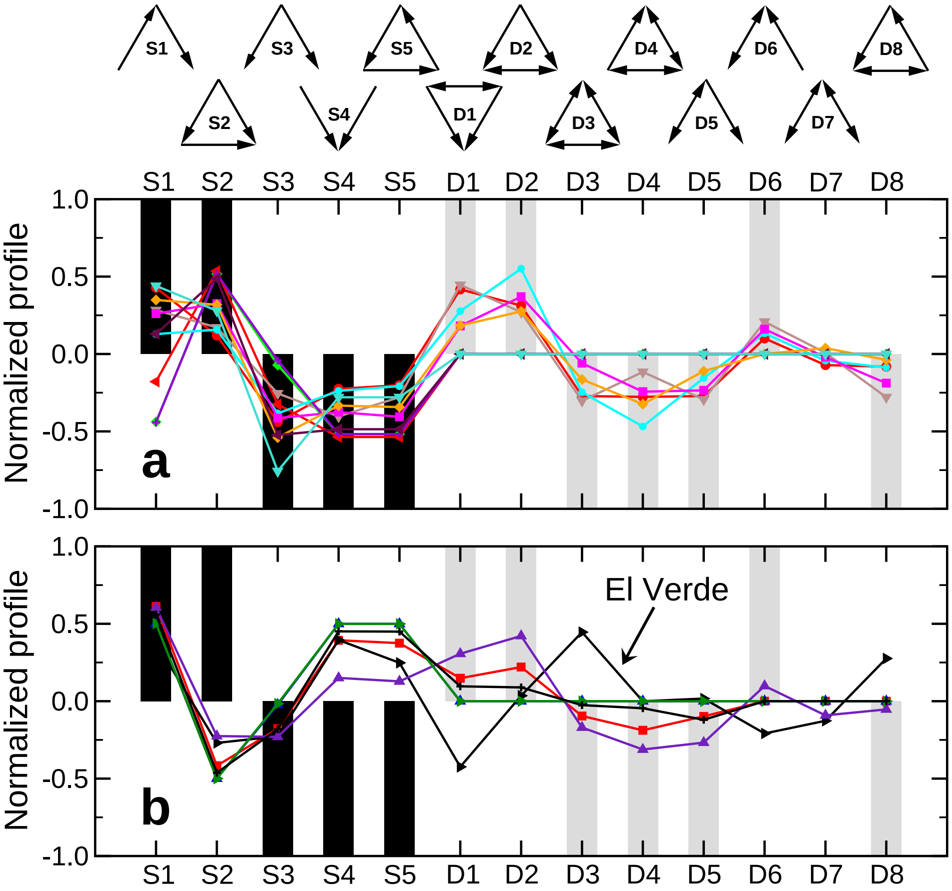

Figure 5.4: Three-species motif over- and under-representation. To facilitate visual comparison between food webs of different sizes and linkage densities, we plot the normalized profile $\mathcal{P}_i = z_i/\sqrt{\sum_j z_j^2}$, which is the vector of z-scores normalized to length 1[62]. The normalization aids in graphical comparison because larger and more densely connected networks tend to exhibit stronger patterns of over- and under-representation; the converse is also true for smaller and less densely connected networks[62]. **a**, Normalized profiles of the 10 empirical food webs characterized by over-representation of motif S2 and under-representation of motifs S4 and S5. **b**, Normalized profiles of the six food webs characterized by under-representation of motif S2 and over-representation of motifs S4 and S5. The bars represent predictions of the generalized cascade and niche models for over- ($\mathcal{P} > 0$) or under-representation ($\mathcal{P} < 0$) of the individual motifs. The black bars represent the predictions of both the generalized cascade and niche models while the grey bars represent the predictions of only the niche model. The empirical webs in **a** are correctly predicted by the niche model, while those in **c** do not match the predictions for motifs S2, S4, and S5. Nevertheless, all food webs with double links, except the El Verde Rainforest, follow the pattern predicted by the niche model for motifs D1–D8.

under-representation of motifs S4–S5, just as in the generalized cascade and niche models, and six remaining food webs which exhibit under-representation of motif S2 and over-representation of motifs S4–S5, opposite the generalized cascade model's predictions. Notably, we find over-representation of S2 and under-representation of S5. Under our null hypothesis, we thus demonstrate that omnivory is much more common relative to other motifs and more than one would expect to find by chance alone, answering definitively a question posed by Pimm[77]. The larger group includes Benguela, Chesapeake Bay, Coachella Valley, Grassland, Little Rock Lake, Caribbean Reef, Northeast US Shelf, Scotch Broom, St. Marks Seagrass, and St. Martin Island.

The differences we find for motifs S2, S4, and S5 in Bridge Brook Lake, Canton Creek, El Verde Rainforest, Skipwith Pond, Stony Stream, and Ythan Estuary are very intriguing. Canton Creek and Stony Stream are, for example, time-specific rather than community food webs[38] and therefore we would expect that they may not be directly comparable to the other food webs considered here. It would be interesting to explore further the ecological or environmental reasons why these two food webs exhibit over-representation of isolated exploitative competition and isolated generalist predation. It is additionally intriguing to note that these are the only food webs for which the nested-hierarchy model has better predictive power than the niche model, due precisely to the deviations for motifs S2, S4, and S5.

Next, we consider motifs D1–D8[59]. Remarkably, we find the same pattern of over- and under-representation for all food webs with double links except El Verde Rainforest (Fig. 5.4). This pattern includes over-representation of motifs D1, D2, and D6 and under-representation of motifs D3–D5 and D8. Significantly, this is the exact pattern we predict when assuming that predators consume a contiguous range of prey, as in the niche model. We emphasize here that a double link is truly mutual predation, for example Hake↔squid or sharks↔seals in Benguela[41]. and that in food webs, particularly aquatic systems, different trophic species are traditionally defined for juveniles and adults of a single taxonomic species when dietary differences are present; see, for example, Warren[40].

5.5 Robustness of the empirical patterns

We can quantify the robustness of these patterns by calculating the uncentered correlation coefficient between the z-score profiles from the 16 empirical food webs[59] (Fig. 5.5). When we consider just motifs S1–S5, as in Fig. 5.5**a**, it is visually apparent that the six food webs exhibit different properties from the rest. Likewise, upon examining Fig. 5.5**b** which considers motifs D1–D8, it is apparent how similar all of the food webs with double links, except El Verde Rainforest, are to each other.

We now examine quantitatively the hypothesis that the contiguous range of prey is responsible for the pattern of over- and under-representation observed in empirical food webs[59]. We also examine the models' ability to reproduce these patterns, beyond the analytical and heuristic arguments[59]. To do this, we compare each empirical food web to simulations of the niche, generalized cascade, and nested-hierarchy models. Each of these models have been shown to accurately reproduce other key properties of empirical food webs[27,28,32]; however, never to this level of structural detail.

To quantify the relationship between the model and empirical food webs, we determine the fraction of the model-generated food webs with uncentered correlation coefficient $0.75 \leq r \leq 1.0$, that is, the fraction of model food webs well aligned with the empirical results by computing the average norm-ratio $\langle d \rangle$, that is, whether the model food webs typically have norm comparable with the empirical food webs.

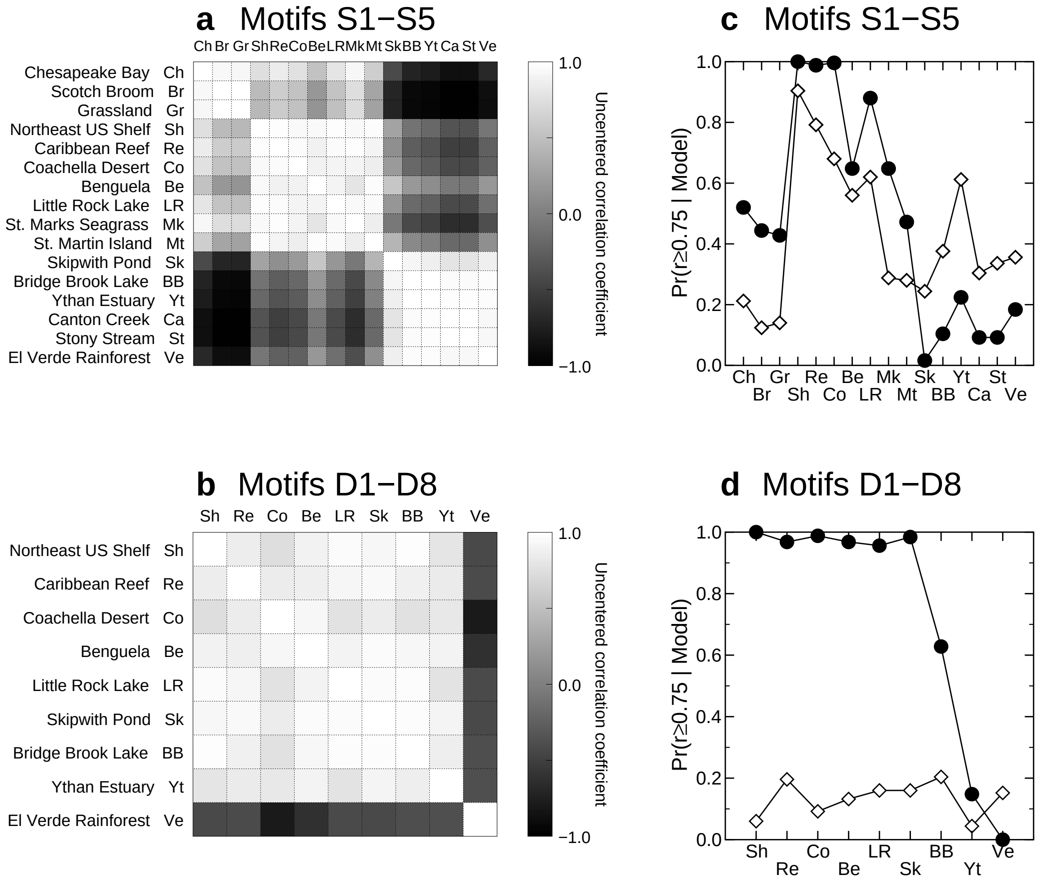

Figure 5.5: Uncentered correlation coefficient between pairs of empirical food webs for **a** motifs S1–S5 and **b** motifs D1–D8. The uncentered correlation coefficient r measures the degree of similarity between two food webs: a value of r close to 1 indicates high similarity, close to 0 indicates little similarity, and close to -1 indicates the two behave oppositely. It is visually apparent in **a** that the last six food webs exhibit different patterns for motifs S1–S5 than the other ten food webs. **c** and **d**, We show here the fraction of model-generated food webs with uncentered correlation coefficient $r \geq 0.75$ when compared to the empirical food webs for the niche (solid circles) and nested-hierarchy (open diamonds) models. When examining the uncentered correlation coefficient, the two models reproduce the empirical profile of over- and under-representation for motifs S1–S5 with similar success (**c**), with the niche model typically performing only slightly better than the nested-hierarchy model. For motifs D1–D8 (**d**), however, only the niche model successfully reproduces the empirical patterns, significantly outperforming the nested-hierarchy model.

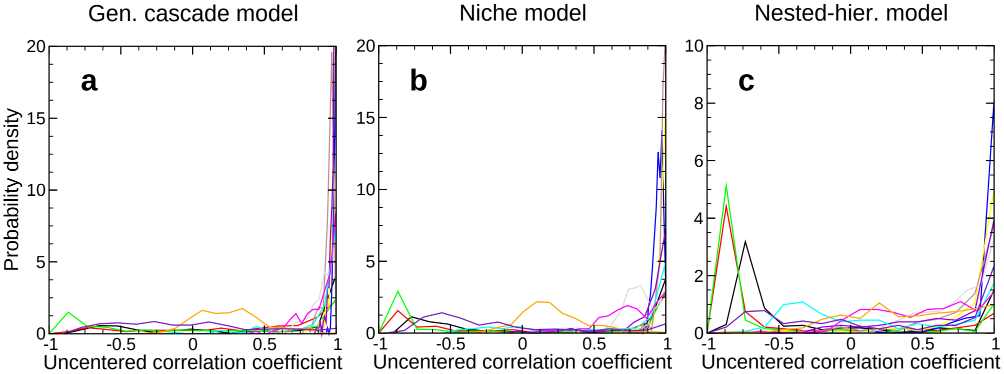

Figure 5.6: Probability density for the uncentered correlation coefficient between each of the empirical food webs and model-generated food webs for motifs S1–S5 for the **a**, generalized cascade, **b**, niche, and **c**, nested-hierarchy. It is visually apparent that all models generally show high positive correlation with the empirical food webs. This result is expected since all models predict the same general over- and under-representation pattern for motifs S1–S5.

We show in Figs. 5.6 and 5.7 the probability densities of uncentered correlation coefficients between z-score profiles of empirical food webs and model-generated food webs with the same S and z. In Figs. 5.5 and 5.6, we observe that the generalized cascade, niche, and nested-hierarchy models exhibit high positive correlations with the empirical data for motifs S1–S5, as expected from earlier results[59]. The differences between the models, however, becomes clear upon examining Figs. 5.5 and 5.7. Here we see that the niche model is *very* strongly correlated with the empirical data[59]. The nested-hierarchy model, instead shows a very broad distribution of correlation values. Even though the nested-hierarchy model still typically generates model food webs which are positively correlated to the empirical data, the performance is much worse than for the niche model[59].

In our comparisons, we find again that the niche model and its underlying mechanism reproduce the empirically observed z-score profiles remarkably well[59]. With regard to motifs S1–S5

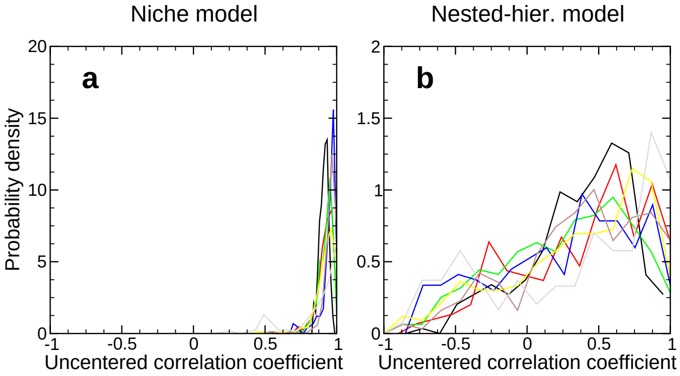

Figure 5.7: Probability density for the uncentered correlation coefficient between the empirical food webs with double links and model-generated food webs for motifs D1–D8 for the **a**, niche and **b**, nested-hierarchy models. It is visually apparent that the niche model exhibits high positive correlation with the empirical food webs, while the nested-hierarchy and random cascade models exhibit very broad distributions of uncentered correlation coefficient values. The niche model, unlike the other models, both qualitatively and quantitatively reproduce the empirically observed pattern of over- and under-representation for motifs D1–D8.

both the niche and generalized cascade models perform almost equally as well in their comparison with the empirical food webs whereas the nested-hierarchy model performs slightly worse[59].

With respect to motifs D1–D8, however, it is clear that the niche model, and its underlying mechanism, reproduce the empirically observed z-score profile with remarkable accuracy[59]. The niche model is highly successful for the food webs which have double links, while the nested-hierarchy model, in contrast, is at best only marginally successful[59].

5.6 Bayes' rule to quantify robustness of empirical mechanism of prey selection

We can use the results of Figs. 5.5 and 5.7 in Bayesian model selection analysis, to further discriminate between the two models of interest, the niche model and the nested-hierarchy model[59].

We take the fraction of model-generated food webs with uncentered correlation coefficient satisfying an empirically observed event $\mathcal{R}$ as a measure of success of a given model M, where $\mathcal{R} = [0.75 \leq r \leq 1.0]$. Mathematically, Bayes' rule states

$$P\left(M|\mathcal{T}\right) = \frac{P\left(\mathcal{T}|M\right) P(M)}{\sum_{M'=1}^{2} P\left(\mathcal{T}|M\right) P(M)}\,, \tag{5.5}$$

where $P(M|\mathcal{T})$ is the posterior probability of selecting a model M given an observed event $\mathcal{T}$, $P(\mathcal{T}|M)$ is the probability of observing the event given a particular model, and $P(M)$ is the prior probability of selecting a model. Assuming no prior knowledge to differentiate between the two models, we select $P(M) = 0.5$. Because we have multiple food webs, we can perform this analysis iteratively to generate a posterior probability of selecting a model given *all* of the data. Using this process we find that, using the uncentered correlation coefficient as a basis, we would *always* select the niche model over the nested-hierarchy model because the posterior probability of the niche model is about 10^6 times larger than that of the nested-hierarchy model[59] (Table 5.6). The ratio of these two posterior probabilities will remain the same even if other candidate models are added to consideration, as long as the two corresponding prior probabilities are equal.

Our results thus demonstrate that network motifs are a robust feature of empirical food webs and that their presence can be explained by predators consuming a contiguous range of prey[59]. There is, of course, the possibility that other models could likewise offer similar predictions with regard to motifs. In fact, while the niche model is able to explain the overall pattern of motif over- and under-representation, it fails to accurately capture the magnitude, as measured by d. This implies that some additional ingredient or understanding is missing.

Table 5.1: Bayesian model selection using the fraction of model-generated food webs with uncentered correlation coefficient satisfying an empirically observed event $\mathcal{R} = [0.75 \leq r \leq 1.0]$ for motifs D1–D8 as the measure of model success. It is clear that with this measure one would essentially *always* select the niche model.

Food web	$P(\text{Niche}\|\mathcal{R})$	$P(\text{Nested}\|\mathcal{R})$
Benguela	0.883	0.117
Bridge Brook Lake	0.747	0.253
Caribbean Reef	0.834	0.166
Coachella Valley	0.918	0.082
El Verde Rainforest	0.026	0.974
Little Rock Lake	0.859	0.141
Northeast US Shelf	0.947	0.053
Skipwith Pond	0.863	0.137
Ythan Estuary	0.771	0.229
All webs	0.999	1×10^{-5}

5.7 Affect of species aggregation on the observed motif patterns

An important question arises with regard to whether or not the observed patterns could be a result of the aggregation of taxonomic species into trophic species commonly observed in empirical food webs[59]. In fact, all of the food webs we consider in our analysis are trophic food webs; furthermore, each of the models we consider were specifically designed to explain observed properties of trophic, not taxonomic, food webs.

To examine the effect of aggregation, we consider two aggregation strategies[59]. First we construct a niche model food web; each of these model webs has $S' = S(1 + f)$ species and linkage density $z = L/S$. Here f is defined as the fraction of species in an observed food web that are the result of aggregating two species; therefore, $f \in [0, 1]$. The limit $f = 0$ corresponds to no aggregated species while $f = 1$ means every species is the result of aggregation. We then lump together fS *pairs* of species, resulting in a food web with S species and linkage density $z \approx L/S$.

In the first aggregation strategy, we lump together the fS pairs of species which are closest together in niche value; this creates a single species with all prey and predators of the original two species. This is synonymous with lumping together species which are likely to share predators. In the second aggregation strategy, we lump together the fS pairs of species which have the greatest overlap in both prey and predators. We quantify the overlap as the number of prey and predators two species share divided by the total number of unique prey and predators of the two species. This is very close to the procedure which occurs in the collection of empirical food web data in lumping similarly behaving species into a single trophic species.

To quantify the effect of aggregation, we examine how changes in the value of $f \in [0, 0.50]$ influence the results of our analysis[59] (Fig. 5.8). For both aggregation strategies, as f increases the agreement between the model webs and empirical data decreases. By the two mechanisms examined here, it is therefore unlikely that the strong agreement we observe between the motifs in the empirical data and niche model could be explained by aggregation of taxonomic species into trophic species[59]. In fact, large amounts of aggregation would have the effect of strongly reducing the strength of the patterns observed empirically[59].

5.8 Concluding remarks

Our comparisons of both the model and empirical food webs indicate that a significant mechanism for prey selection in natural ecosystems is consumption of species whose niche value is close to some characteristic value[59]. The idea that this value is related to species mass is, in fact, well documented. Morphologically, it has been demonstrated that features such as mouth size of birds

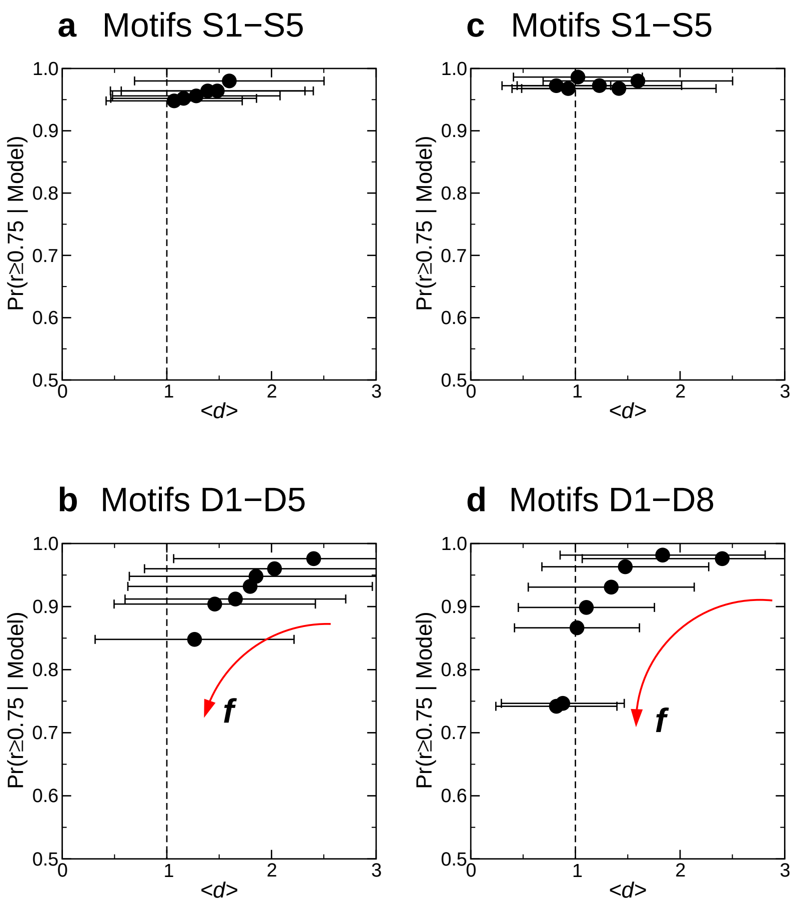

Figure 5.8: Effect of aggregation of taxonomic species into trophic species on the conclusion of diet contiguity in empirical food webs. We show here the fraction of model-generated food webs with uncentered correlation coefficient $r \geq 0.75$ against the average ratio d which compares the magnitudes of model and empirical the z-score vectors. Each model food web is a realization of the niche model with differing fractions of aggregated species f. **a** and **b**, We lump species which are closest together in niche value. This is equivalent to lumping together species which are likely to share predators. **c** and **d**, We lump species which have the greatest overlap in both prey and predators. This is equivalent to the procedure which occurs in the collection of empirical food web data when lumping similarly behaving species into a single trophic species. We see little effect upon motifs S1–S5 but pronounced effects for motifs D1–D8. As f increases (as indicated by the arrows), the agreement between the model webs and empirical data decreases. It is therefore unlikely that the strong agreement we observe between the empirical data and niche model would be due to aggregation of taxonomic species into trophic species, at least by the two mechanisms examined here. All error bars represent two standard deviations of the observed values.

or fish can greatly limit the size of prey that can be consumed[78,79]. Energetically, it has been shown that increasing predator size is associated with increasing mean prey size[80,81,82] and the energy increase from prey consumption must justify that spent during prey capture[82,83,84]. Furthermore, it has been shown that prey size specialization can play a tremendous role in the coexistence of predators and thus maintaining biodiversity[85,86].

We note that our results are also congruent with niche theory and resource-driven sympatric speciation[87,88,89,90]. Our results provide strong statistical evidence for the specialization of predators. This can be seen as supporting evidence for sympatric speciation. Specifically, if a species at some stage in the ecosystem history consumes preys with a broad range of masses, it is likely that over the time it will speciate into two species each specializing on prey covering a smaller range of masses.

Our ecosystem-wide investigation of the mechanism for prey selection provides us with several significant implications. First, our results provide additional evidence for the conclusion that empirical food webs are well approximated by diet contiguity as in the niche model. In the next chapter, we will examine in more detail whether or not the contiguity assumption is well justified.

Second, our results demonstrate that omnivory occurs much more than one would expect to find by chance alone. Moreover, omnivory represents a significant component of a complete food web's structure. This result is congruent with theoretical studies which emphasize the stabilizing impact of omnivory[91]. As our motifs consist of only three species, however, it is not decisive whether they represent true omnivory across trophic levels or simply intraguild predation. This discrepancy helps to explain why Bascompte & Melián[65] concluded that there is an ambiguous role of omnivory in food webs. Therefore, the debate regarding omnivory's true significance remains open[77,92,93,94].

Third, we uncover motifs which are frequently observed within food web structure but to date have not been studied to the depth of others such as apparent competition and exploitative competition[63,64]. These two motifs have been investigated in the laboratory because of their relative simplicity and assumed ecology relevance. It would be intriguing to see if there are dynamic explanations, for example, for the abundance or lack of motifs D1–D8 in addition to the dietary explanation we have uncovered. An intriguing possibility is that future theoretical investigations into motifs with additional species would likewise uncover larger ecologically relevant subgraphs.

Lastly, it is unlikely that there exists systematic bias in food-web data collection or in the strategies for aggregating taxonomic species into trophic species. The existence of any such bias would make it substantially more difficult to obtain the quantitatively robust motif patterns we report.

In conclusion, we would like to note that it is quite remarkable for something as complex as the pattern of representation of motifs in a food web can be understood using only very simple but plausible rules. It is also remarkable that the same patterns emerge independently from the specific details of the food web studied, notably its environment or the particular species present. These fact opens up the possibility that similar solutions may exist for even more daunting ecological problems. We believe that these static food web patterns will enable us to gain insight into, for example, how an invasive species might select its prey upon entering a new environment.

Chapter 6

Food web intervality

Intervality of a food web is related to the number of trophic dimensions characterizing the niches in a community. We introduce here a mathematically robust measure for food web intervality[95]. It has previously been noted that empirical food webs are not strictly interval; however, upon comparison to suitable null hypotheses, we conclude that empirical food webs actually do exhibit a strong bias toward contiguity of prey, that is, toward intervality[95]. Further, our results strongly suggest that empirically observed species and their diets can be mapped onto a single dimension[95]. This finding validates a critical assumption in the recently proposed static niche model and provides guidance for ongoing efforts to develop dynamic models of ecosystems[95].

6.1 Intervality

In earlier chapters, we demonstrated that these three models share two fundamental mechanisms which account for the models' success in reproducing many of the empirical patterns[32]. (i) Species form a totally ordered set in niche space, that is, species can be ordered along a single niche dimension. (ii) Each species has an exponentially-decaying probability of preying on a given fraction of the species with equal or lower niche values. In spite of these similarities, the models exhibit some differences; a crucial distinction concerns how species' prey are organized along the single dimension. In the niche model, species prey upon a contiguous range of prey. In the nested-hierarchy and generalized cascade models, in contrast, the diets are not restricted to a contiguous range.

As we noted, the differences in prey selection lead to very different model motif profiles[59,60]. We further determined that the global and local structure of empirical food webs is consistent with the hypothesis of diet contiguity in empirical food webs. In addition to the local structure, the different prey selection mechanisms lead to drastic differences in the intervality of the food web graph[95] (Fig. 6.1**a** and **b**) [*]. The significance of intervality in complex food webs was first noted by Cohen[96], who reported, as did subsequent studies[19,73,97,98], that the vast majority of empirical food webs in the literature appeared to be interval graphs. Significantly, these studies also suggested that the probability that a food web is interval is strongly dependent upon the number of species represented in the food web, decreasing from approximately one for very small food webs to close to zero for larger webs[73]. The food webs which were analyzed in these studies typically comprised very few species, leaving open the question of whether, or to what degree, larger and more complex

---

[*] Niche model generated food webs are interval by construction[27] while generalized cascade[32] and nested-hierarchy[28] model generated food webs are not. Williams & Martinez[27] found it surprising that a strictly interval model is able to explain strictly non-interval data. They hypothesized that this apparent complication arose from the fact that the "degree of intervality is very high in empirical food webs"[27]. However, the logic behind this hypothesis is incorrect as food webs generated according to the nested-hierarchy and generalized cascade models are not interval but still correctly reproduce many of the same food web properties[32]. In fact, Cattin *et al.*[28] designed the nested-hierarchy model to be explicitly non-interval in an attempt to address an apparent non-empirical basis for contiguous diets assumed by the niche model.

food webs are interval[73]. More recent studies reported persistent non-intervality of highly-resolved empirical food webs[27,28].

Importantly, the degree of intervality of a food web is related to the number of trophic dimensions characterizing the possible niches in a community[19]. More specifically, one may ask what is the minimum number of variables required to describe the factors that influence the trophic organization of the species in a community? Is this number the same or different for different communities?[20,33,99] If a food web is interval, then the species and their diets can be represented along a single dimension. It has been suggested that a single factor—species' mass—provides a suitable proxy for this dimension[10,23,25,52,71,72]. Any departure from intervality has been understood to imply additional complexity in the mechanisms responsible for the structure of the food web.

The number of higher quality food web data sets has steadily increased and these data have enabled researchers to uncover a number of solid empirical regularities[12,27,28,29,31,32,59]. Thus, we believe that a more definitive answer to the question of food web intervality may be at hand[95].

In this chapter we address the question of how "non-interval" empirical food webs truly are[95]. To this end, we define a novel measure of intervality that is more robust than those already in the literature[95]. Notably, our results agree with previous studies which observed that empirical food webs are strictly non-interval; however, we demonstrate that their degree of "intervality" can be understood as a *perturbation* on an underlying interval structure[95]. Our results provide support to the conjecture that species and their diets—that is, ecosystem niches—may be mapped onto a single dimension[95].

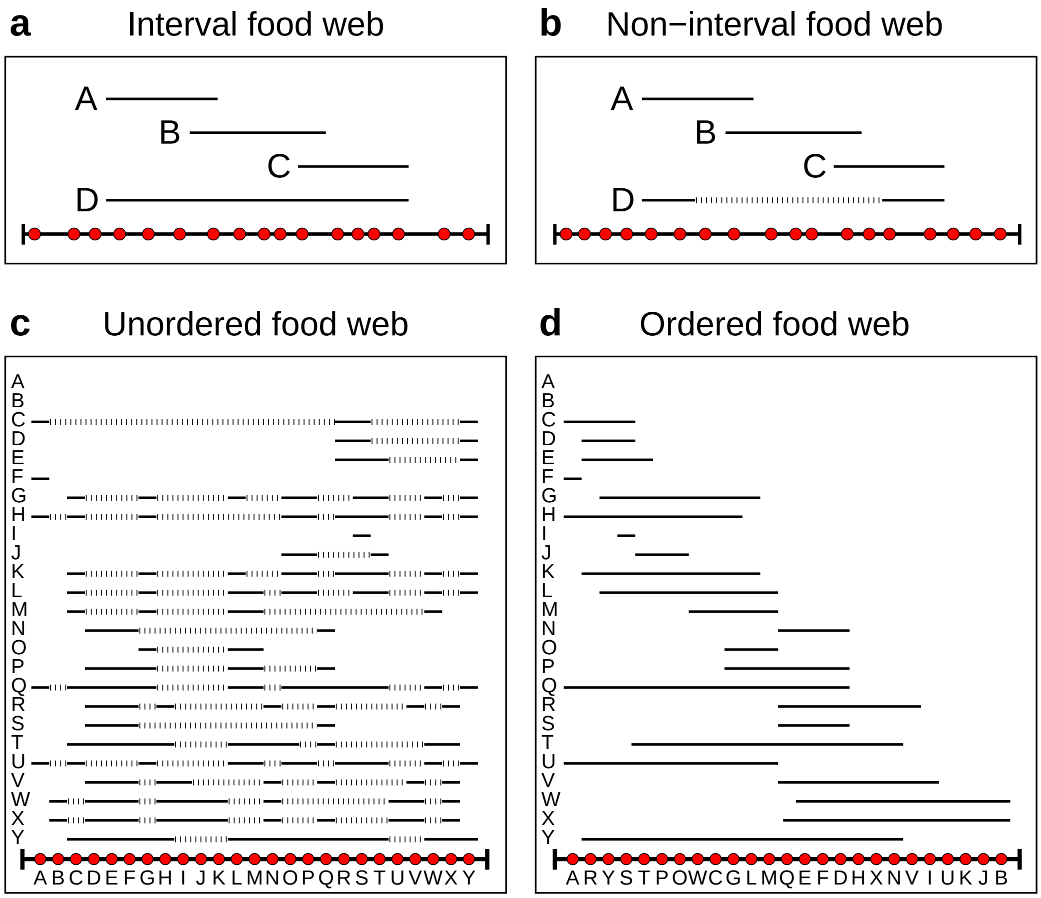

Figure 6.1: Interval and non-interval food webs. Species (red circles) are placed along a single dimension which we denote the resource axis. For each predator (A, B, C, and D) a line is placed above the prey (resources) it consumes. **a,** A food web is interval if there exists a permutation of the species along the resource axis such that for each predator the diet is contiguous. **b**, A food web is non-interval if no permutation exists for which all diets can be represented as contiguous segments. **c**, An unordered food web. The resource axis is shown along the bottom and each red circle represents a species in the ecosystem. For each species in the vertical axis, we represent predation by a solid horizontal line (for example, C consumes A) and non-predation by the dashed lines (for example, C does not consume B). The total number of gaps for this particular ordering is $\mathcal{G} = 217$. **d**, An ordered food web. Our algorithm works by swapping the location of two nodes within the ordering in an attempt to minimize the value of $\mathcal{G}$. In this particular case, one can find an ordering with $\mathcal{G} = 0$. It should be noted that this is one of potentially multiple permutations which can give rise to the same value $\hat{G} = G = 0$.

6.2 Quantifying food web intervality

In the studies of Cohen[19], Cohen *et al.*[73], Cohen[96] and Sugihara[97,98], intervality was reported as a binary variable: a web either "is" or "is not" interval. Recently, two local estimates have been used to measure the "level of diet discontinuity"[28]. The first measure, D_{diet}, is defined as the number of triplets of species with an "irreducible gap" divided by the number of possible triplets. An irreducible gap is a gap in a consumer's diet which cannot be made contiguous because of the constraints imposed by other consumers' diets (Fig. 6.1).

The second measure, Cy_4, is defined as the number of chord-less cycles of length four in the consumer overlap graph. In the consumer overlap graph, two consumers are connected if they share at least one prey. That is, if species A and B share prey with species C and D, the consumer overlap graph would consist of links $A \leftrightarrow C$, $A \leftrightarrow D$, $B \leftrightarrow C$, and $B \leftrightarrow D$. This is a cycle because it is possible to travel from any one of the four species to any other in this graph. If species A and B do not share any prey and similarly species C and D do not share any prey either, this cycle is "chordless" and the four diets cannot be made contiguous simultaneously[19]. The measure Cy_4 is related to Sugihara's rigid circuit property which states that in an interval food web every circuitous path of length $l \geq 4$ in the consumer overlap graph is shortened by a chord[98].

Using these two measures, Cattin *et al.*[28] reported that the non-intervality of empirical food webs is a significant food web pattern. Caution, however, is required since both D_{diet} and Cy_4 yield only local estimates of intervality and cannot be directly extrapolated to an entire ecosystem.

Specifically, a cycle of length four in the consumer overlap graph with a chord may still contain irreducible gaps[73]. Therefore, Cy_4 is, at best, a lower bound for what Cattin *et al.*[28] intended to measure. Likewise, when computing D_{diet}, the normalization factor used by Cattin *et al.*[28] accounts only for multiphagous consumers, not all species. By concentrating on species triplets, the resulting measure is an over-estimation and not amenable to comparisons between food webs of different sizes and linkage densities. Moreover, as pointed out by Martinez & Cushing[100], Cattin *et al.*[28] also do not address what values of D_{diet} or Cy_4 would in fact be statistically significant or represent a *large* deviation from an interval food web.

In contrast to previous studies, we determine here the degree of intervality of an entire food web[95]. To do this, we first find the order of species in the food web in such a way as to generate the "most interval" ordering of the food web. This process yields the best approximation to a food web where the species and their diets are organized along a single dimension[95].

We discuss our definition of "most interval" and its justification here in detail. In the idealized case of a fully interval food web, each consumer's diet is represented by a single contiguous range. If we consider a non-interval food web and attempt to reproduce the idealized web as closely as possible, we will want all prey of a given predator to "appear" as close together as possible on the resource axis (Fig. 6.1). For example, for a given consumer, a sequence of two adjacent prey, a gap of one species, and two more adjacent prey (i.e., $\ldots -PP-PP- \ldots$, where "P" represents a prey and "$-$" represents a non-prey) is preferable to the same sequence but with a gap of two species or larger (e.g., $\ldots -PP--PP- \ldots$). Indeed, the former situation would be far more likely given an interval web which experienced random omissions or changes, such as those possibly introduced by field sampling.

For a food web graph $\mathcal{F}$ with S species, there are $S!$ possible species orderings $\mathcal{O}_k\left(\mathcal{F}\right)$, where $\mathcal{O}_k\left(\mathcal{F}\right) = s_1^k s_2^k \ldots s_S^k$ with $k = 1, \ldots, S!$. Because of the large number of possible permutations, it is computationally unfeasible to determine the best ordering through enumeration. It is for this reason that we employ simulated annealing, a heuristic technique which significantly reduces the computational effort required to find an optimal or close-to-optimal solution.

When attempting to find the most interval ordering, the objective is to minimize the discontinuity of all predators' prey (Fig. 6.1). We thus define a cost function $\mathcal{G}\left(\mathcal{O}_k\right)$ which is the sum of the gaps in all consumers' diets

$$\mathcal{G}\left(\mathcal{O}_k\right) = \sum_{i=1}^{S} \sum_{j=1}^{n_i} \left(g_{ij}^k\right)^{\beta} \ . \tag{6.1}$$

Here n_i is the number of gaps in the diet of species i and g_{ij}^k is the number of species in the j-th gap in the diet of species i for a given ordering $\mathcal{O}_k\left(\mathcal{F}\right)$. Here we report results for $\beta = 1$; however the selection of other values, such as $\beta = 2$, yields similar orderings of the empirical data. Simulated annealing yields an estimate [†] $\hat{G}$ for the total number of gaps of the food web $G \equiv \min\limits_{\forall k} \{\mathcal{G}\left(\mathcal{O}_k\right)\}$. The smaller $\hat{G}$ is the more interval the food web is.

Simulated annealing is a stochastic optimization technique that enables one to find a "low-cost" configuration while still broadly exploring the space of possibilities[101]. This is achieved by introducing a computational "temperature" T. When T is high, the system can explore configurations of high cost whereas at low T the system can only explore low-cost regions. By starting at high T and slowly decreasing T, the system descends gradually toward deep minima.

For each iteration in the simulated annealing algorithm, we attempt to swap the position of two randomly selected species to go from the initial ordering $\mathcal{O}_i\left(\mathcal{F}\right)$ to the proposed ordering $\mathcal{O}_f\left(\mathcal{F}\right)$. This updated ordering $\mathcal{O}_f\left(\mathcal{F}\right)$ is then accepted with probability

$$p = \begin{cases} 1 & \text{if} \quad \mathcal{G}\left(\mathcal{O}_f\right) \leq \mathcal{G}\left(\mathcal{O}_i\right) \\ \exp\left(-\frac{\mathcal{G}(\mathcal{O}_f)-\mathcal{G}(\mathcal{O}_i)}{T}\right) & \text{if} \quad \mathcal{G}\left(\mathcal{O}_f\right) > \mathcal{G}\left(\mathcal{O}_i\right) \end{cases} , \tag{6.2}$$

where $\mathcal{G}\left(\mathcal{O}_f\right)$ is the cost after the update and $\mathcal{G}\left(\mathcal{O}_i\right)$ is the cost before the update. For each value of T, we attempt qS^2 random swaps with $q \geq 250$. After the movements are evaluated at a certain T, the system is "cooled down" to $T' = cT$, with $c = 0.99$.

[†] Note that we use G to refer to the actual minimum number of gaps for the most interval ordering of a food web, whereas $\hat{G}$ refers to the estimate obtained with simulated annealing. The only case when we can be certain that $\hat{G} = G$ is when $\hat{G} = 0$.

6.3 Intervality in a randomized food web

As happens for other graph and combinatorial problems, the actual value of $\hat{G}$ is of little significance[102]; rather, one needs to assess whether the measured value of $\hat{G}$ is significantly different from the expected value under suitable null hypotheses. To solve this problem, we have designed three complementary null hypotheses which place different restrictions upon how consumers' diets may be organized within a food web[95].

Our first null hypothesis is the set of randomizations of the empirical food web. To generate an ensemble of random networks, one must first define the constraints of the randomization[61,66]. In our analysis, we preserve the following attributes for each species during randomization of the food web: (i) number of prey, (ii) number of predators, (iii) number of single links, $A \to B$, (iv) number of double links, $A \leftrightarrow B$, and (v) whether or not a species is a cannibal. We employ the Markov-chain Monte Carlo switching algorithm[70] as described in Sec. 4.3.

The randomized empirical food web stands as a food web graph with no constraints placed upon consumers' diets. That is, in the randomization there is no correlation between the prey of a given species and their organization on the resource axis. We therefore expect that $\hat{G}$ for these randomized food webs will be maximal. Comparison to this null hypothesis thus provides verification of whether there are any structural regularities in the organization of species' diets within empirical food webs.

6.4 Intervality in the generalized cascade model

Our second null hypothesis is the set of food webs generated by the generalized cascade model. In the generalized cascade model, each of the S species i are assigned a niche value n_i drawn from a uniform distribution in the interval $[0, 1]$. A predator j selects at random a fraction x of the species i with niche values $n_i \leq n_j$ as its prey, where x is drawn from a beta-distribution $p(x) = \beta(1-x)^{(\beta-1)}$. Here $\beta = (S^2/2L) - 1$ and L is the number of trophic links in the ecosystem.

The generalized cascade model food webs are generated with the same number of species S and linkage density L/S as the empirical food webs. Whereas randomization of the empirical food webs imposes no structural constraints upon consumers' diets, the generalized cascade model does. Each predator may again select their prey at random, but instead of from the entire resource axis, their selections are restricted to only those species with niche values less than or equal to their own. This mechanism leads to a smaller number of gaps for species placed lower on the resource axis. Comparison of the empirical data to this null hypothesis will provide evidence as to whether empirically observed diets exhibit additional structural constraints.

6.5 Intervality in the generalized niche model

Comparison to the two previous null hypotheses will provide an indication of whether empirical food webs have a larger number of gaps than would be expected for random structures with little or no bias toward contiguity of prey. In order to quantify any bias toward contiguity of prey in empirical food webs, we need to develop a third null hypothesis, which we base upon a novel generalization of the niche model[27] (Fig. 6.2).

Let us first recall the definition of the niche model. Each of the S species i are assigned a niche value n_i drawn from a uniform distribution in the interval $[0,1]$. A predator j in the niche model preys on a range $a_j = n_j x$ of the resource axis, where x is drawn from a beta-distribution $p(x) = \beta(1-x)^{(\beta-1)}$ and $\beta = (S^2/2L) - 1$ just as in the generalized cascade model. The center

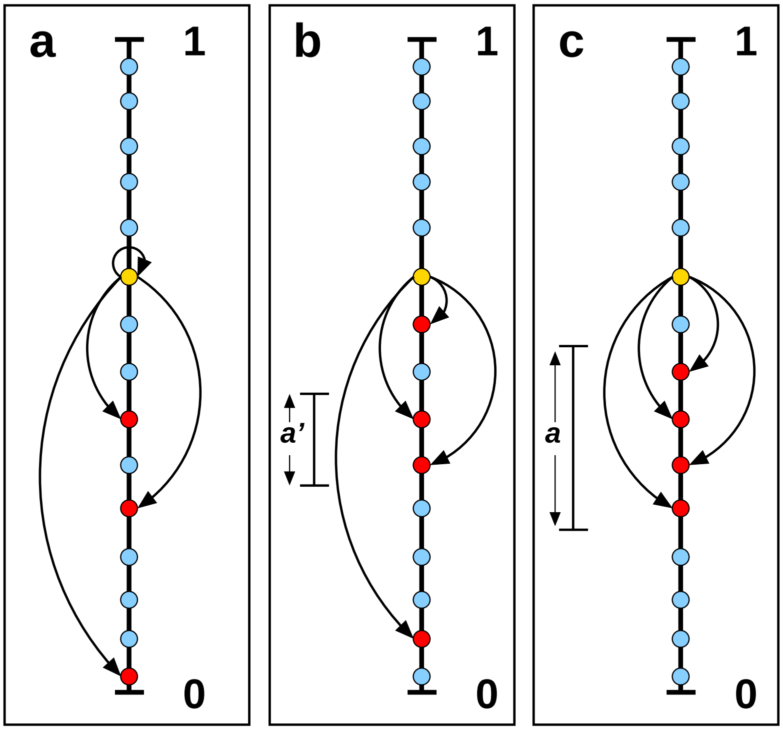

Figure 6.2: Prey selection in the generalized niche model where the predator is indicated in yellow and the prey in red. In the generalized niche model, we add a new parameter $c \in [0, 1]$ which measures the fraction of species' diets which are imposed to be contiguous. **a**, For $c = 0$, we recover the generalized cascade model and prey are selected at random from species with niche values less than or equal to the predator's own. **b**, For $c = 0.5$, a fraction c of the prey are selected from a contiguous range, as in the niche model, and the remaining prey are selected at random from species with niche values less than or equal to the predator's own. **c**, For $c = 1.0$, all prey are selected from a contiguous range of niche values, just as in the niche model.

of the range a_j is selected uniformly at random in the interval $[a_j/2, n_j]$. All species i whose niche values n_i fall within this range are considered prey of species j.

To allow for a tunable bias toward prey contiguity, we generalize the niche model in the following manner. First, we reduce the range a_j for a predator j to $a'_j = c\, a_j = c\, n_j x$, where c is a fixed parameter in the interval $[0, 1]$. Because species are distributed uniformly at random on the resource axis, a predator j with range a_j has on average $a_j S$ prey. The same applies to the reduced range a'_j, and therefore a predator has $\Delta k = \left(a_j - a'_j\right) S = (1 - c)\, a_j S$ expected prey unaccounted for after the range reduction. Next, we select these Δk prey—rounded to the nearest integer value—randomly from the species i with niche value $n_i \leq n_j$ that are not already a prey of species j. If $c = 0$, we recover the generalized cascade model, whereas, for $c = 1$, we recover the niche model.

6.6 Testing the discontiguity hypothesis for empirical food web intervality

We study 15 empirical food webs from a variety of environments[95]: three estuarine— Chesapeake Bay[34], St. Marks[35], and Ythan[36]; five freshwater— Bridge Brook Lake[18], Canton Creek[38], Little Rock Lake[39], Skipwith Pond[40], and Stony Stream[38]; three marine— Benguela[41], Caribbean Reef[42], and Northeast US Shelf[43]; and four terrestrial— Coachella Valley[44], Grassland[11], Scotch Broom[46], and St. Martin[47] (Table 3.1).

For each empirical food web we obtain $\hat{G}_e$ (Table 6.1). We find $1 \leq \hat{G}_e \leq 700$ for all food webs, that is, none of the webs is interval. To compare these empirical values to our three null hypotheses, we perform the following steps. For each empirical food web, we generate a minimum of 100 model food webs corresponding to the respective null hypothesis and obtain $\hat{G}_{\mathrm{model}}$ for each model food web.

We then want to be able to estimate the probability that the value $\hat{G}_e$ appears given each null hypothesis[95]. To do this, we examine not just the mean of $\hat{G}_{\mathrm{model}}$, but its probability distribution (Fig. 6.3). Using the Kolmogorov-Smirnov test[103], we determine that we cannot reject the hypothesis that the $\hat{G}_{\mathrm{model}}$ values are drawn from a Gaussian distribution. We then use the fact that a

Table 6.1: Comparison of empirical data with the random model and the generalized cascade model. For each of the fifteen food webs, we show $\hat{G}_e$. For each the two models we show $\left\langle \hat{G}_{\mathrm{model}} \right\rangle$, z_{model}, and p_{model}. $\left\langle \hat{G}_{\mathrm{model}} \right\rangle$ is the average over at least 100 model-generated food webs. The z-score is defined as $z_{\mathrm{model}} = \left(\hat{G}_e - \left\langle \hat{G}_{\mathrm{model}} \right\rangle \right) / \sigma_{\hat{G}_{\mathrm{model}}}$. We utilize the Kolmogorov-Smirnov test[103] to examine each set of model-generated data and find that we cannot reject the hypothesis that the $\hat{G}_{\mathrm{model}}$ values are drawn from a Gaussian distribution. We then use the fact that a Gaussian distribution describes the model data to directly calculate an estimate for the probability p_{model} of observing a value of $\hat{G}_{\mathrm{model}} \leq \hat{G}_e$. This probability is equivalent to the significance by which one may reject the underlying null hypotheses.

Food web	$\hat{G}_e$	$\left\langle \hat{G}_{\mathrm{R}} \right\rangle$	z_{R}	p_{R}	$\left\langle \hat{G}_{\mathrm{GC}} \right\rangle$	z_{GC}	p_{GC}
Benguela	27	81	-10.31	$<10^{-10}$	78	-3.59	$<10^{-3}$
Bridge Brook Lake	1	51	-11.24	$<10^{-10}$	48	-4.42	$<10^{-5}$
Canton Creek	615	810	-7.65	$<10^{-10}$	1804	-9.38	$<10^{-10}$
Caribbean Reef	298	498	-12.47	$<10^{-10}$	340	-1.17	0.12
Chesapeake Bay	11	48	-5.86	$<10^{-8}$	38	-3.08	$<10^{-2}$
Coachella Valley	51	117	-10.74	$<10^{-10}$	64	-1.04	0.15
Grassland	5	28	-5.08	$<10^{-6}$	95	-5.76	$<10^{-8}$
Little Rock Lake	427	1347	-26.75	$<10^{-10}$	1641	-9.89	$<10^{-10}$
Northeast US Shelf	700	1291	-17.76	$<10^{-10}$	1050	-5.84	$<10^{-8}$
St. Marks	157	343	-14.12	$<10^{-10}$	258	-2.95	$<10^{-2}$
St. Martin	95	204	-12.18	$<10^{-10}$	193	-4.06	$<10^{-4}$
Scotch Broom	23	226	-15.67	$<10^{-10}$	508	-8.94	$<10^{-10}$
Skipwith Pond	26	36	-3.32	$<10^{-3}$	42	-1.59	0.06
Stony Stream	645	915	-9.82	$<10^{-10}$	2225	-12.56	$<10^{-10}$
Ythan	270	513	-11.69	$<10^{-10}$	915	-8.52	$<10^{-10}$

Gaussian distribution describes the model data to directly calculate an estimate for the probability of observing a value of $\hat{G}_{\mathrm{model}} \leq \hat{G}_e$.

We first compare the set of empirical food webs $\{\mathcal{F}\}$ to the set of randomized food webs $\{\mathcal{F}_{\mathrm{R}}\}$ (Table 6.1). We find that, for every food web, $\hat{G}_e < \left\langle \hat{G}_{\mathrm{R}} \right\rangle$. To estimate the significance of this difference for each of the individual food webs, we calculate the probability that the model exhibits

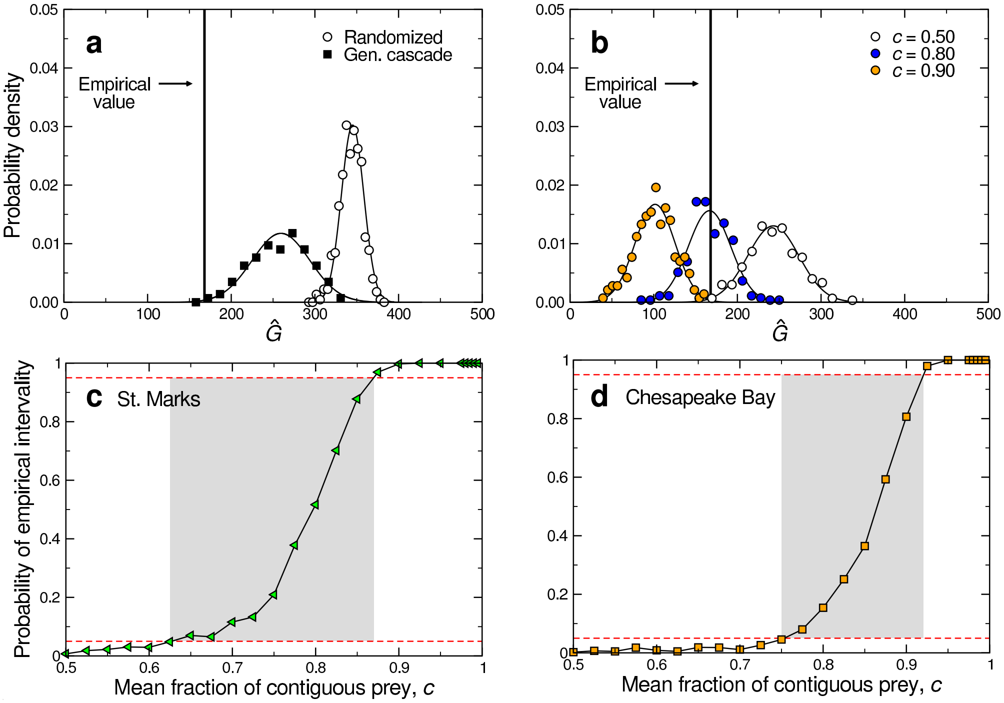

Figure 6.3: Estimated number of gaps for St. Marks and the null models discussed in the text. **a**, Probability density of $\hat{G}$ for two of the null models: randomization of the empirical food web and the generalized cascade model. The generalized cascade model-generated food webs were specified to have the same number of species S and linkage density L/S as the empirical food web. $\hat{G}_e$ is shown by the spike. The probability of $\hat{G}_{\text{model}} \leq \hat{G}_e$ is $< 10^{-10}$ and $< 1.6 \times 10^{-3}$ for the randomized empirical web and generalized cascade model, respectively. **b**, Probability density of $\hat{G}$ for the generalized niche model and three different values of c. The generalized niche model-generated food webs were specified to have the same number of species S and linkage density L/S as the empirical food web. $\hat{G}_e$ is again shown by the spike. **c**, Probability of observing $\hat{G}_{\text{GN}}(c) = \hat{G}_e = 157$ for the St. Marks food web. Values less than 0.5 correspond to negative z-scores and thus represent the probability P_{low} of observing a value of $\hat{C}_{\text{GN}} \leq \hat{G}_e$, whereas values greater than 0.5 represent the probability P_{high} of observing a value of $\hat{G}_{\text{GN}} \geq \hat{G}_e$. The 95% confidence intervals on the value of c are given by the regions where both $P_{\text{low}} \geq 0.05$ and $P_{\text{high}} \geq 0.05$ are true (denoted by the dashed red lines). We find the 95% confidence interval to be $c \in [0.625, 0.87]$ (shaded in gray). **d**, Same as **c** but for the Chesapeake Bay food web with $\hat{G} = 11$. We find the 95% confidence interval to be $c \in [0.75, 0.92]$ (shaded in gray). The intervality values of $\mathcal{I} = 0.87$ and $\mathcal{I} = 0.92$ for St. Marks and Chesapeake Bay, respectively, imply that the empirical food webs are statistically indistinguishable from our generalized niche model only when there is a very strong bias toward contiguity of species' diets.

a value $\hat{G}_{\mathrm{R}} \leq \hat{G}_e$. For twelve of the 15 food webs[95], $p < 10^{-10}$. For the remaining three food webs[95], $p_{\mathrm{R}} < 10^{-3}$.

We now compare the set of empirical food webs to the set of generalized cascade model-generated food webs $\{\mathcal{F}_{\mathrm{GC}}\}$ (Table 6.1). We again find that for every empirical food web, $\hat{G}_e < \left\langle \hat{G}_{\mathrm{GC}} \right\rangle$. We find that for twelve of the fifteen food webs the probability that $\hat{G}_{\mathrm{GC}} \leq \hat{G}_e$ is again quite small[95], $p_{\mathrm{GC}} < 10^{-2}$. For the remaining three food webs, Skipwith Pond, Coachella Valley, and Caribbean Reef, we find larger probability values, 0.06, 0.12, and 0.15, respectively[95]. Further analysis indicates that $\left\langle \hat{G}_{\mathrm{GC}} \right\rangle$ decreases with the directed connectance L/S^2 for a fixed number of species S [‡]. Thus the higher values of p are likely due to the higher connectance of these webs.

In order to conclusively reject the two random hypotheses, we apply a Bonferroni correction[104] which decreases the significance level for the two individual hypothesis for a particular food web to $\alpha = 0.05/2 = 0.025$ in order to avoid spurious false positives. Upon considering each individual food web as compared to our two random hypothesis with this more conservative threshold, we can conclusively reject the two hypotheses for twelve of the fifteen food webs[95]; the exceptions are again Caribbean Reef, Coachella Valley, and Skipwith Pond[95].

[‡] For densely connected food webs, predators typically have greater numbers of prey. Because these prey are constrained to have a niche value less than or equal to the predators, the greater the directed connectance the greater the probability that these prey are contiguous, despite the random predation. This is more pronounced for smaller than for larger food webs.

6.7 Testing the contiguous hypothesis for empirical food web intervality

To this point, our results provide an indication that the majority of empirical food webs are significantly more interval than would be expected for food webs with little or no bias toward prey contiguity[95]. We now investigate our generalized niche model to determine how it compares to the empirical data for different values of c and therefore different levels of bias toward prey contiguity[95].

For each of the fifteen remaining food webs, we compare the empirical food web $\hat{G}_e$ to the model $\left\langle \hat{G}_{\mathrm{GN}} \right\rangle$ for $c \in [0.5, 1.0]$. We compare the model and empirical data as before, but focus particularly upon the z-score, where $z = \frac{\hat{G}_e - \left\langle \hat{G}_{\mathrm{model}} \right\rangle}{\sigma_{\hat{G}_{\mathrm{model}}}}$. Using the z-score, we can determine the upper bound of 95% confidence intervals on c for which the empirical $\hat{G}$ is likely to be observed in the generalized niche model (Fig. 6.3). We show the results of this comparison in Table 6.2.

For the fifteen food webs we investigated[95], we find that the largest values of c which provide statistical agreement with the empirical data are remarkably close to one, $0.85 < c_{\mathrm{max}} < 1.00$ [§]. This finding enables us to quantify in a statistically sound manner the intervality of a food web; specifically,

$$\mathcal{I}(\mathcal{F}_i) \equiv c_{\mathrm{max}}(\{\mathcal{F}_{GN}\}) \,, \tag{6.3}$$

where $\{\mathcal{F}_{GN}\}$ is the ensemble of model food webs generated according to the generalized niche model with the same number of species and connectance of the real food web $\mathcal{F}_i$. For the fifteen empirical food webs investigated, we find values of $\mathcal{I}$ very close to 1; in fact, $\langle \mathcal{I} \rangle = 0.95$. This

---

[§] It should be noted that our results may exhibit some under-estimation of c, in particular as noted earlier for densely connected food webs such as Coachella Valley, Northeast US Shelf, and Skipwith Pond.

Table 6.2: Empirical food web intervality. For each food web we show the intervality $\mathcal{I}$, the maximum value of c for which we cannot reject the hypothesis that the value of $\hat{G}_e$ could have been observed in the generalized niche model.

Food web	$\mathcal{I}$
Benguela	0.96
Bridge Brook Lake	$\approx$1.00
Canton Creek	0.95
Caribbean Reef	0.85
Chesapeake Bay	0.92
Coachella Valley	0.94
Grassland	$\approx$1.00
Little Rock Lake	0.97
Northeast US Shelf	0.93
St. Marks	0.87
St. Martin	0.93
Scotch Broom	$\approx$1.00
Skipwith Pond	0.96
Stony Stream	0.96
Ythan	0.95

result indicates that natural ecosystems are significantly interval and consequently there is a strong bias toward contiguity in prey selection[95].

6.8 Concluding remarks

The concept of “niche theory” or “niche space” is a fundamental concept in the study of ecosystems. Niche space was classically defined as an “n-dimensional hyperspace” with n given by the innumerable ecological and environmental characteristics[20,99]. Therefore, each species’ niche is the “result” of all n factors acting upon it and the niche represents the functional role and position of the organism in its community. The more recent “interpretation” of niche theory, however, relates to the niche providing species an ordering or hierarchy[23,52,99]. This formulation provides a much simpler criterion than Hutchinson's (1957) "n-dimensional hyperspace". Studies have suggested that by using species' mass or size a food web can in fact be mapped to a single dimension[23,25,26,52,71,72]. Furthermore, the placing of species onto a single dimension is a crucial ingredient in many models developed to describe food web structure[27,28,32,73].

Recently, however, discussions as to how interval food webs truly are, were renewed by the contrast between the niche model—and its contiguous range of prey—and the generalized cascade and nested-hierarchy models—and their random predation[32]. Our results allow us to conclusively demonstrate that natural ecosystems, while not fully interval, are significantly more interval than would be expected by chance alone[95]. Indeed, we find the empirical food webs to be statistically indistinguishable from model food webs whose diets are between 85% and 100% contiguous[95]. The idea that species and their diets can be so closely mapped onto a single dimension represents a significant insight that can guide us on how best to go about developing *dynamic* ecosystem models such as the recent integration of the niche model and nonlinear bioenergetic modeling proposed by Martinez *et al.*[105].

Chapter 7

Conclusions

I have incorporated concepts and methods from fields not traditionally aligned with ecology, such as chemical engineering and statistical physics, to develop an integrated approach to studying the structure of natural ecosystems, namely food webs. By following this approach, I have demonstrated that there are key universal features common to all ecosystems, independent of variables such as the population and type of species present, assembly history, or particular environment.

I first identified that there are two critical conditions that a food web model must satisfy in order to successfully predict the properties of empirical food webs[32]: (i) the niche values to which species are assigned form a *totally ordered set* and (ii) each species has a specific, exponentially-decaying probability $p(x)$ of preying on a fraction x of the species with lower niche values. Remarkably, even an extremely simple model, such as our generalized cascade model, can achieve this[32].

I then found that the distributions of the numbers of prey, predators, and links in empirical food webs obey universal scaling functions, where the scaling factor is the linkage density[32]. 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[32]. The results indicated that fundamental concepts of modern statistical physics such as scaling and universality may be successfully applied in the study of food webs[32]. Moreover, they are consistent with the underlying hypothesis of scaling theory, that is, food webs display "universal" patterns in the way trophic relations are established despite apparent fundamental differences[32].

Next, comparisons between model and empirical food webs indicated that a significant mechanism for prey selection in natural ecosystems is consumption of species whose niche value is close to some characteristic value[59]. I further validated this conclusion by examining the underlying hypothesis, namely that species diets may be mapped to a single dimension[95]. Our results allow us to conclusively demonstrate that natural ecosystems, while not fully interval, are significantly more interval than would be expected by chance alone[95]. Furthermore, I found the empirical food webs to be statistically indistinguishable from model food webs whose diets are between 85% and 100% contiguous[95].

7.1 Food web structure

My analysis of food-web motifs demonstrated that omnivory occurs much more than one would expect to find by chance alone[59,60]. As these motifs consist of only three species, however, it is not decisive whether they represent true omnivory across trophic levels or simply intraguild predation. Therefore the debate regarding omnivory's true significance remains open[77,92,93,94] and very important question to answer if we hope to predict the effect of species invasions or extinctions. Future theoretical investigations into motifs with additional species could help to resolve this problem and would likely uncover larger ecologically-relevant subgraphs.

There are a number of future questions to answer before the topic of food web intervality can come to a close[95]. First and foremost is getting a better understanding of exactly what processes are behind the deviations from truly interval behavior. While some of the gaps within species diets may be due to interactions not observed during field sampling, it is unlikely that all gaps may be attributed to this factor. It has been noted, albeit for different food webs than those studied here, that ecosystems with multiple habitats, for example an estuary, are less likely to be interval than single-habitat food webs[73,96]. Indeed, one would not expect food webs containing several habitats to be strictly interval since each habitat is likely to have its own separate resource axis.

A similarly intriguing question relates to examining additional properties of the “most-interval” ordering or orderings. Studies which compared these orderings to those obtained when comparing species’ masses, or related properties[72] could help determine what ecological features are responsible for organizing predatory behavior which in turn organizes the food web structure at whole.

7.2 Ecosystem dynamics

To this point, researchers who study food webs have followed a maxim observed in other disciplines, such as physics: one must understand statics before attempting to understand dynamics. Thanks to current research, including the work presented in this thesis, we have begun to reach a consensus regarding the structure and evolutionary pressures which shape the network of predator-prey interactions; therefore the time to advance our understanding of dynamics is at hand.

The dynamic processes which take place in an ecosystem span a broad range of time scales, from ontogenetic changes over a species’ lifespan to seasonal changes in the weather to individual species’ daily decisions of which prey will be eaten. A number of studies have compiled species-specific data which describes each of these aspects. One of the principle challenges of food web dynamics is the fact that this data does not exist for complete ecosystems. A major challenge is thus to determine how to generalize species-specific information so that it may be utilized in the development in a general dynamic food web model, in much the same way that I have discussed the development of a general static food web model.

Demonstrating the validity of using a *general* modeling approach is also imperative. Fortunately, as we discovered by examining food web structure, there exist a number of *universal* features that hold for a large number of ecosystems. These patterns are independent of the number of species, the specific species present, and the particular environment, besides other factors. As it is the same evolutionary and dynamic forces which result in these common static properties, it is reasonable to assume a general dynamic model may be developed.

Such a dynamic model must take into account many environmental factors, such as species' bioenergetics[106], population dynamics, spatial constraints[107], and external perturbations, such as invasion by exotic species. Once such a model has been developed, it can be validated in two ways. First, it must reproduce the static ecosystem patterns discussed throughout this thesis. Second, specific ecological incidents can be utilized as scenarios to test the predictions of the model. The validated model can then be utilized for understanding general test cases, such as the effect of over-exploitation, habitat loss, invasion of exotic species, or the extinction of one or multiple species. Such a model would make a significant contribution toward understanding ecosystem dynamics. Only with this understanding will we be able to *predict* and *prevent* future ecological disasters.

References

1. Pimm, S. L. & Raven, P. Extinction by numbers. *Nature* **403**, 843–845 (2000).

2. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).

3. McAllister, D., Scheuler, F. W., Roberts, C. M. & Hawkins, J. P. Mapping and GIS analysis of the global distribution of coral reef fishes on an equal-area grid. In *Mapping the Diversity of Nature* (ed. Miller, R. I.), 155–175 (Chapman & Hall, London, 1994).

4. Skole, D. & Tucker, C. J. Evidence for Tropical deforestation, fragmented habitat, and adversely affected habitat in the Brazilian Amazon:1978–1988. *Science* **260**, 1905–1910 (1993).

5. Nepstad, D. C. *et al.* Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* **398**, 505–508 (1999).

6. Cochrane, M. A. *et al.* Positive Feedbacks in the Fire Dynamic and Closed Canopy Tropical Forests. *Science* **284**, 1832–1835 (1999).

7. Thomas, J. A. *et al.* Comparative Losses of British Butterflies, Birds, and Plants and the Global Extinction Crisis. *Science* **303**, 1879–1881 (2004).

8. Chapin, F. S. *et al.* Consequences of changing biodiversity. *Nature* **405**, 234–242 (2000).

9. Buchanan, M. *Nexus* (W. W. Norton & Company, New York, 2002).

10. Cohen, J. E. & Newman, C. M. A stochastic theory of community food webs I. Models and aggregated data. *Proc. R. Soc. Lond. B* **224**, 421–448 (1985).

11. Martinez, N. D., Hawkins, B. A., Dawah, H. A. & Feifarek, B. P. Effects of sampling effort on characterization of food-web structure. *Ecology* **80**, 1044–1055 (1999).

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

13. Paine, R. T. Food webs: Road maps of interactions or the grist for theoretical development? *Ecology* **69**, 1648–1654 (1988).

14. Briand, F. & Cohen, J. E. Community food webs have scale invariant structure. *Nature* **307**, 264–266 (1984).

15. Cohen, J. E. & Briand, F. Trophic links of community food webs. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 4105–4109 (1984).

16. Rejmanek, M. & Stary, P. Connectance in real biotic communities and critical values for stability of model ecosystems. *Nature* **280**, 311 (1979).

17. Martinez, N. Effects of scale on food web structure. *Science* **260**, 242–243 (1993).

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

19. Cohen, J. E. *Food Webs and Niche Space* (Princeton University Press, Princeton, NJ, 1978).

20. Hutchinson, G. E. Concluding remarks. In *Population studies: Animal ecology and demography.*, vol. 22 of *Cold Spring Harbor Symposia on Quantitative Biology*, 415–427 (1957).

21. Pianka, E. R. *Theoretical Ecology: Principles and Applications*, chap. Competition and Niche Theory, 114–141 (W. B. Saunders Company, Philadelphia, PA, 1976).

22. Pianka, E. R. Niche Overlap and Diffuse Competition. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 2141–2145 (1974).

23. Warren, P. H. & Lawton, J. H. Invertebrate predator-prey body size relationships: an explanation for upper triangular food webs and patterns in food web structure? *Oecologia* **74**, 231–235 (1987).

24. Cohen, J. E. Food webs and community structure. In *Perspectives in Ecological Theory* (eds. Roughgarden, J., May, R. M. & Levin, S. A.), 181–202 (Princeton University Press, Princeton, NJ, 1989).

25. Cohen, J. E., Pimm, S. L., Yodzis, P. & Saldaña, J. Body sizes of animal predators and animal prey in food webs. *Journal of Animal Ecology* **62**, 67–78 (1993).

26. Neubert, M. G., Blumenshine, S. C., Duplisea, D. E., Jonsson, T. & Rashleigh, B. Body size and food web structure: testing the equiprobability assumption of the cascade model. *Oecologia* **123**, 241–251 (2000).

27. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).

28. Cattin, M.-F., Bersier, L.-F., Banašek-Richter, C., Baltensperger, R. & Gabriel, J.-P. Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427**, 835–839 (2004).

29. Camacho, J., Guimerà, R. & Amaral, L. A. N. Analytical solution of a model for complex food webs. *Phys. Rev. E* **65**, art. no. 030901 (2002).

30. Gradstheyn, I. S. & Ryzhik, I. M. *Table of Integrals, Series, and Products* (Academic Press, New York, NY, 2000), 6th edn.

31. Camacho, J., Guimerà, R. & Amaral, L. A. N. Robust patterns in food web structure. *Phys. Rev. Lett.* **88**, art. no. 228102 (2002).

32. Stouffer, D. B., Camacho, J., Guimerà, R., Ng, C. A. & Amaral, L. A. N. Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86**, 1301–1311 (2005).

33. Cohen, J. E. A stochastic theory of community food webs. VI. Heterogeneous alternatives to the cascade model. *Theoretical Population Biology* **37**, 55–90 (1990).

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

35. Christian, R. R. & Luczkovich, J. J. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecological Modelling* **117**, 99–124 (1999).

36. Hall, S. J. & Raffaelli, D. Food-web patterns: Lessons from a species-rich web. *J. Anim. Ecol.* **60**, 823–842 (1991).

37. Hall, S. J. & Raffaelli, D. Food webs: theory and reality. *Adv. Ecol. Res.* **24**, 187–239 (1993).

38. Townsend, C. R. *et al.* Disturbance, resource supply, and food-web architecture in streams. *Ecology Letters* **1**, 200–209 (1998).

39. Martinez, N. D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monog.* **61**, 367–392 (1991).

40. Warren, P. H. Spatial and temporal variation in a freshwater food web. *Oikos* **55**, 299–311 (1989).

41. Yodzis, P. Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *Journal of Animal Ecology* **67**, 635–658 (1998).

42. Opitz, S. Trophic interactions in Caribbean coral reefs. ICLARM Tech. Rep. 43, Manila, Philippines (1996).

43. Link, J. Does food web theory work for marine ecosystems? *Marine Ecology Progress Series* **230**, 1–9 (2002).

44. Polis, G. A. Complex trophic interactions in deserts: An empirical critique of food-web theory. *American Naturalist* **138**, 123–155 (1991).

45. Waide, R. B. & Reagan, W. B. (eds.). *The Food Web of a Tropical Rainforest* (University of Chicago Press, Chicago, IL, 1996).

46. Hawkins, B. A., Martinez, N. D. & Gilbert, F. Source food webs as estimators of community food web structure. *International Journal of Ecology* **18**, 575–586 (1997).

47. Goldwasser, L. & Roughgarden, J. Construction of a large Caribbean food web. *Ecology* **74**, 1216–1233 (1993).

48. Solé, R. V. & Montoya, J. M. Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B* **268**, 2039–2045 (2001).

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

50. Williams, R. J., Berlow, E. L., Dunne, J. A., Barabási, A.-L. & Martinez, N. D. Two degrees of separation in complex food webs. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12913–12916 (2002).

51. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).

52. Cohen, J. E. (ed.). *Ecologists’ Co-Operative Web Bank, version 1.0* (Rockefeller University, New York, NY, 1989).

53. Pimm, S. L. Food web design and the effect of species deletion. *Oikos* **35**, 139–149 (1980).

54. Pimm, S. L. Complexity and stability: Another look at MacArthur’s original hypothesis. *Oikos* **33**, 357 (1979).

55. Pimm, S. L. & Lawton, J. H. Are food webs divided into compartments? *J. Animal Ecol.* **49**, 879–898 (1980).

56. Pimm, S. L., Lawton, J. H. & Cohen, J. E. Food web patterns and their consequences. *Nature* **350**, 669–674 (1991).

57. Borrvall, C., Ebenman, B. & Jonsson, T. Biodiversity and the Risk of Cascading Extinction in Model Food Webs. *Ecology Letters* **3**, 131–136 (2000).

58. Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters* **5**, 559–567 (2002).

59. Stouffer, D. B., Camacho, J., Jiang, W. & Amaral, L. A. N. Evidence for the existence of a robust pattern of prey selection in food webs. *Proc. R. Soc. B* (2007), doi:10.1098/rspb.2007.0571

60. Camacho, J., Stouffer, D. B. & Amaral, L. A. N. Quantitative analysis of the local structure of food webs. *J. Theor. Biol.* **246**, 260–268 (2007).

61. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).

62. Milo, R. *et al.* Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).

63. Holt, R. D. Community modules. In *Multitrophic Interactions in Terrestrial Ecosystems, 36th Symposium of the British Ecological Society* (eds. Gange, A. C. & Brown, V. K.), 333–350 (Blackwell Science, 1997).

64. Holt, R. D. & Hochberg, M. E. Indirect Interations, Community Modules and Biological Control: a Theoretical Perspective. In *Evaluation of Indirect Ecological Effects of Biological Control* (eds. Waijnberg, E., Scott, J. K. & Quimby, P. C.), 13–37 (CAB International, 2001).

65. Bascompte, J. & Melián, C. J. Simple trophic modules for complex food webs. *Ecology* **86**, 2868–2873 (2005).

66. Artzy-Randrup, Y., Fleishman, S. J., Ben-Tal, N. & Stone, L. Comment on "Network Motifs: Simple Building Blocks of Complex Networks" and "Superfamilies of Evolved and Designed Networks". *Science* **305**, 1107 (2004).

67. Vázquez, A. *et al.* The topological relationship between the large-scale attributes and local interaction patterns of complex networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17940–17945 (2004).

68. Middendorf, M., Ziv, E. & Wiggins, C. H. Inferring netowrk mechanisms: The *Drosophila melanogaster* protein interaction network. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3192–3197 (2005).

69. Arim, M. & Marquet, P. A. Intraguild predation: a widespread interaction related to species biology. *Ecology Letters* **7**, 557–564 (2004).

70. Itzkovitz, S., Milo, R., Kashtan, N., Newman, M. E. J. & Alon, U. Reply to "Comment on 'Subgraphs in random networks' ". *Phys. Rev. E* **70**, art. no. 058102 (2004).

71. Lawton, J. H. Food webs. In *Ecological Concepts* (ed. Cherrett, J.), 43–78 (Blackwell Scientific, Oxford, UK, 1989).

72. Cohen, J. E., Jonsson, T. & Carpenter, S. R. Ecological community description using the food web, species abundance, and body size. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1781–1786 (2003).

73. Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* (Springer-Verlag, Berlin, 1990).

74. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).

75. Itzkovitz, S., Milo, R., Kashtan, N., Ziv, G. & Alon, U. Subgraphs in random networks. *Phys. Rev. E* **68**, art. no. 026127 (2003).

76. McCann, K. & Hastings, A. Re-evaluating the omnivory-stability relationship in food webs. *Proc. R. Soc. Lond. B* **264**, 1249–1254 (1997).

77. Pimm, S. L. *Food webs* (University of Chicago Press, Chicago, IL, 2002), 1st edn.

78. Werner, E. E. & Hall, D. J. Competition and habitat shift in two sunfishes (Centrarchidae). *Ecology* **58**, 869–876 (1977).

79. Wheelwright, N. T. Fruit size, gape width, and the diets of fruit-eating birds. *Ecology* **66**, 808–818 (1985).

80. Gittleman, J. L. Carnivore body size: Ecological and taxonomic correlates. *Oecologia* **67**, 540–554 (1985).

81. Vézina, A. F. Empirical relationships between predator and prey size among terrestrial vertebrate predators. *Oecologia* **67**, 555–565 (1985).

82. Carbone, C., Mace, G. M., Roberts, S. C. & Macdonald, D. W. Energetic constraints on the diet of terrestrial carnivores. *Nature* **402**, 286–288 (1999).

83. Wainwright, P. C. Ecological explanation through functional morphology: The feeding biology of sunfishes. *Ecology* **77**, 1336–1343 (1996).

84. Radloff, F. G. T. & Toit, J. T. D. Large predators and their prey in a southern African savanna: a predator's size determines its prey size range. *Journal of Animal Ecology* **73**, 410–423 (2004).

85. Karanth, K. U. & Sunquist, M. E. Prey selection by tiger, leopard and dhole in tropical forests. *Journal of Animal Ecology* **64**, 439–450 (1995).

86. Karanth, K. U. & Sunquist, M. E. Behavioural correlates of predation by tiger (*Panthera tigiris*), leopard (*Panthera pardus*) and dhole (*Cuan alpinus*) in Nagarahole, India. *Journal of Zoology, London* **250**, 255–265 (2000).

87. Tregenza, T. & Butlin, R. K. Speciation without isolation. *Nature* **400**, 311–312 (1999).

88. Via, S. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* **16**, 381–390 (2001).

89. Kawata, M. Invasion of vacant niches and subsequent sympatric speciation. *Proc. R. Soc. Lond. B* **269**, 55–63 (2002).

90. Beltman, J. B., Haccou, P. & ten Cate, C. Learning and colonization of new niches: a first step toward speciation. *Evolution* **58**, 35–46 (2004).

91. McCann, K. S., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).

92. Yodzis, P. How rare is omnivory? *Ecology* **65**, 321–323 (1984).

93. Williams, R. J. & Martinez, N. D. Stabilization of chaotic and non-permanent food-web dynamics. *Eur. Phys. J. B* **38**, 297–303 (2004).

94. Tanabe, K. & Namba, T. Omnivory creates chaos in simple food web models. *Ecology* **86**, 3411–3414 (2005).

95. Stouffer, D. B., Camacho, J. & Amaral, L. A. N. A robust measure of food web intervality. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19015–19020 (2006).

96. Cohen, J. E. Food webs and the dimensionality of trophic niche space. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4533–4536 (1977).

97. Sugihara, G. Graph theory, homology, and food webs. In *Population Biology* (ed. Levin, S. A.), vol. 30 of *Proceedings of Symposia in Applied Mathematics*, 83–101 (American Mathematical Society, 1984).

98. Sugihara, G. *Niche Hierarchy: Structure, Organization, and Assembly in Natural Communities*. Ph.D. thesis, Princeton University, Princeton, N.J. (1982).

99. Giller, P. S. *Community Structure and the Niche* (Chapman and Hall, London, 1984).

100. Martinez, N. D. & Cushing, L. J. Box A. Additional model complexity reduces fit to complex food-web structure. In *Ecological Networks: Linking Structure to Dynamics in Food Webs* (eds. Pascual, M. & Dunne, J. A.), 87–89 (Oxford University Press, Oxford, UK, 2006).

101. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).

102. Guimerà, R., Sales-Pardo, M. & Amaral, L. A. N. Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**, art. no. 025101 (2004).

103. Mood, A. M., Graybill, F. A. & Boes, D. C. *Introduction to the Theory of Statistics* (McGraw-Hill Companies, 1974).

104. Miller Jr., R. G. *Simultaneous Statistical Inference* (Springer-Verlag, New York, 1991).

105. Martinez, N. D., Williams, R. J. & Dunne, J. A. Diversity, complexity, and persistence in large model ecosystems. In *Ecological Networks: Linking Structure to Dynamics in Food Webs* (eds. Pascual, M. & Dunne, J. A.), 163–185 (Oxford University Press, Oxford, UK, 2006).

106. Yodzis, P. & Innes, S. Body size and consumer-resource dynamics. *American Naturalist* **139**, 1151–1175 (1992).

107. Brose, U., Ostling, A., Harrison, K. & Martinez, N. D. Unified spatial scaling of species and their trophic interactions. *Nature* **428**, 167–171 (2004).