A robust measure of food web intervality

Daniel B. Stouffer[1], Juan Camacho[1,2], and Luís A. Nunes Amaral[1]

[1]Department of Chemical and Biological Engineering

Northwestern University, Evanston, IL 60208, USA

[2] Departament de Física (Física Estadística)

Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

We introduce a mathematically robust measure for food web intervality. Intervality of a food web is related to the number of trophic dimensions characterizing the niches in a community. We aim to determine the minimum number of variables required to describe the factors that influence the trophic organization of the species in a community. We find that empirical food webs are not interval in the strictest sense of the definition. However, upon comparison to suitable null hypotheses, we conclude that empirical food webs exhibit a strong bias toward contiguity of prey, that is, toward intervality. Indeed, we demonstrate that species and their diets can be mapped onto a single dimension, an insight that must guide ongoing efforts to develop dynamical models of ecosystems.

Introduction

In spite of their "baroque" complexity, the structure of natural food webs displays a number of remarkably simple regularities[1,2,3,4,5,6,7,8,9,10]. The existence of these empirical regularities has prompted several researchers to develop simple models that aim to identify the mechanisms that underly them. Three "static" models, the niche model of Williams & Martinez[2], the nested-hierarchy model of Cattin *et al.*[8], and the generalized cascade model of Stouffer *et al.*[10], predict key statistical properties of food webs from a variety of environments, including deserts, rain forests, lakes and estuaries.

Stouffer *et al.*[10] demonstrated that these three models share two fundamental mechanisms which account for the models' success in reproducing the empirical patterns: (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 lower niche values[10].

In spite of the above similarities, the models do differ in significant ways. An important difference 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. This difference implies that these models give rise to networks with different intervality (Fig. 1). Niche model generated food webs are, by definition, interval while, for the vast majority of cases, food webs generated according to the nested-hierarchy and generalized cascade models are not.

The idea of intervality in complex food webs was introduced by Cohen[11], who reported, as did subsequent studies[1,12,13,14], 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 was interval was 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[1]. The food webs which were analyzed in these studies typically comprised very few species, placing into question the finding of intervality for larger and more complex food webs[1,2,8].

Importantly, the degree of intervality of a food web is related to the number of trophic dimensions characterizing the possible niches in a community[12]. 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?[15,16,17] If a food web is interval, then the ecosystem can be represented along a single dimension. It has been suggested that a single factor—species' mass—provides a suitable proxy for this dimension[6,18,19,20,21,22]. Any departure from intervality has been understood to imply additional complexity in the mechanisms responsible for the structure of the food web.

Recently, the number of higher quality food web data sets has been steadily increasing and these data have been the focus of a number of recent studies on food web structure[2,3,4,5,7,8,10]. Thus, we believe that a more definitive answer to the question of food web intervality may be at hand.

In this manuscript we address the question of how "non-interval" empirical food webs truly are. To this end, we define a novel measure of intervality that is more robust than those already in the literature. Notably, we find that while empirical food webs are indeed non-interval, their degree of "intervality" can be understood as a *perturbation* on an underlying *interval structure*. Our results provide significant support to the conjecture that ecosystem niches may be mapped onto a single dimension.

Food web intervality

In the studies of Cohen[11, 12], Cohen *et al.*[1] and Sugihara[13, 14], intervality was defined as a binary variable: a web either "was" or "wasn't" interval. Recently, two alternative measures have been used to quantify the "level of diet discontinuity"[8]. The first measure, $D_{\rm diet}$, 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. 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. An example of a cycle of length four would be when both species A and D share prey with species B and C. This is a cycle because it is possible to travel from any one of the four species to any other in the consumer overlap graph. If species A and D do not share any prey, or similarly species B and C do not, this cycle is chord-less and the four diets cannot be made contiguous simultaneously. Therefore, an interval food web will have no chordless cycles in the consumer overlap graph[12].

Using these two measures, Cattin *et al.*[8] reported that the non-intervality of empirical food webs is a significant food web pattern. Unfortunately, both $D_{\rm diet}$ and Cy_4 have limitations that raise concerns about the validity of their conclusions for an entire ecosystem. Specifically, a cycle of length four in the consumer overlap graph with a chord can still contain irreducible gaps[1]. Therefore, Cy_4 is, at best, a lower bound for what Cattin *et al.*[8] intended to measure. Moreover, when computing $D_{\rm diet}$, the normalization factor used by Cattin *et al.*[8] accounts only for multiphagous consumers, not all species.

An irreducible gap can occur in graphs with as few as three multiphagous consumers. However, by concentrating on species triplets, one will inflate the resulting measure and will not be able to compare results for food webs of different sizes and linkage densities. In fact, Cattin *et al.*[8] do little to 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 intervality of an entire food web. 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 are organized along a single dimension. There are various related means by which one could define “most interval,” so we discuss our definition and its justification 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. 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 $\mathcal{O}_k\left(\mathcal{F}\right) = s_1^k s_2^k \ldots s_S^k$, 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 (see Methods and Kirkpatrick *et al.*[23] for details).

When attempting to find the most interval ordering, the objective is to minimize the discontinuity of all predators’ prey (Fig. 1). We thus define a cost function $\mathcal{G}\left(\mathcal{O}_k\right)$ which is the sum of the gaps in a consumer’s diet

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

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 $\mathcal{O}_k\left(\mathcal{F}\right)$. Simulated annealing yields an estimate $\hat{G}$ for the total number of gaps $G \equiv \min_{\forall k} \left\{\mathcal{G}\left(\mathcal{O}_k\right)\right\}$ of the food web [1]. The smaller G is the more interval the food web is.

---

[1] Note that we use G to refer to the actual minimum number of gaps for a the most interval ordering of a food web, whereas

Null hypotheses for food web intervality

As happens in other graph and combinatorial problems the value of $\hat{G}$ is of little interest[24]; rather, one needs to assess whether the measured value of $\hat{G}$ is significantly different from the expect value for specific types of graphs. To solve this problem, one must determine the expected value of $\hat{G}$ under suitable null hypotheses. We have designed three complementary null hypotheses which place different restrictions upon how consumers' diets may be organized within a food web.

Our first null hypothesis is a randomized version of the empirical food web. We perform this randomization using the Markov-chain Monte Carlo switching algorithm[25] and treat single, double, and cannibal links separately (see Methods for details). 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.

Our second null hypothesis is food webs generated by the generalized cascade model with the same number of species S and linkage density z as the empirical food webs. Whereas randomization of the empirical food webs imposed 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.

To this point, our 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 no bias toward contiguity of prey. In order to quantify any bias toward contiguity of prey in empirical food webs, we provide a third null hypothesis based upon a generalization of the niche model of Williams & Martinez[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

---

$\hat{G}$ refers to the estimate obtained with simulated annealing. The only case when we can be sure that $\hat{G} = G$ is when $\hat{G} = 0$.

range r_j of the resource axis; $r_j = n_j x$, where x is drawn from a beta-distribution $p(x) = \beta(1-x)^{(\beta-1)}$. Here $\beta = \left(S^2/2L\right) - 1$ and L is the number of trophic links in the ecosystem. The center of the range r_j is selected uniformly at random in the interval $[r_j/2, n]$. All species 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 r_j for a predator j to $r'_j = c\, r_j = c\, n_j x$, where c is a fixed parameter in the interval $[0, 1]$. Because species are distributed uniformly on the resource axis, a predator j with range r_j has on average $r_j S$ prey. The same applies to the reduced range r'_j, and therefore a predator has $\Delta k = \left(r_j - r'_j\right) S = (1-c)\, j_i S$ expected prey unaccounted for. Next, we select these Δk prey—rounded to the nearest integer value—randomly from species i with niche value $n_i \leq n_j$ that are not already a prey of species j. If $c = 0$ there is no pressure for contiguity (the web is strictly non-interval), while for $c = 1$ we recover the niche model (the web is thus fully interval).

Empirical results

We study 15 empirical food webs from a variety of environments: three estuarine—Chesapeake Bay[26], St. Marks[27], and Ythan[28]; five freshwater—Bridge Brook Lake[29], Canton Creek[30], Little Rock Lake[31], Skipwith Pond[32], and Stony Stream[30]; three marine—Benguela[33], Caribbean Reef[34], and Northeast US Shelf[35]; and four terrestrial—Coachella Valley[36], Grassland[37], Scotch Broom[38], and St. Martin[39].

For each empirical food web we find $\hat{G}_e$ (Table 1). We observe that $\hat{G}_e > 0$ for all food webs, that is, none of the webs is interval. The values of $\hat{G}_e$ range from 1 to 676. To compare these empirical values to our three null hypotheses, we perform the following steps. For each empirical food web, we generate at least 100 model food webs corresponding to the respective null hypothesis and obtain $\hat{G}$ 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. To do this, we examine not just the mean of $\hat{G}_{\text{model}}$, but the probability distribution. We employ the Kolmogorov-Smirnov test[40] to each set of model-generated data and find that we cannot reject the hypothesis that the model-generated $\hat{G}$ values are drawn from a normal distribution (Fig. 2). Because we know the distribution which describes the model data, we can directly obtain an analytical estimate for the probability of observing a value of $\hat{G}_{\text{model}}$.

We first compare the set of empirical food webs $\{\mathcal{F}\}$ to the set of randomized food webs $\{\mathcal{F}_{\mathrm{R}}\}$ (Table 1). We note that for every food web, $\hat{G}_e < \left\langle \hat{G}_{\mathrm{R}} \right\rangle$. To estimate the significance of this difference, we calculate the probability that the model exhibits so low a $\hat{G}$ value. We find that the probability is exceedingly small; for eleven of the 15 food webs, $p < 10^{-300}$ (our computer's numerical precision). For the remaining four food webs, the largest probability is $p_{\mathrm{R}} = 4.4 \times 10^{-4}$.

We now compare the set of empirical food webs to the set of generalized cascade model-generated food webs $\{\mathcal{F}_{\mathrm{GC}}\}$ (Table 2). We again find that for every empirical food web, $\hat{G}_e < \left\langle \hat{G}_{\mathrm{GC}} \right\rangle$. To estimate the significance of this difference, we calculate the probability that the model exhibits so low a $\hat{G}$ value. We find that, for eight of the eleven food webs, the probability is again exceedingly small ($p_{\mathrm{GC}} < 0.005$). For the remaining three food webs, the higher probability values—0.056, 0.15, and 0.21 for Skipwith Pond, Coachella Valley, and Caribbean Reef, respectively—are likely due to their large directed connectances (defined as L/S^2) [2].

To this point, our results provide an indication that empirical food webs are significantly more interval than would be expected for food webs with no bias toward prey contiguity. 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. The same considerations for applicability that were discussed for the generalized cascade model hold for the generalized niche model[10]. Because of the computational effort required, we have selected only six of the eleven food webs to compare to this null model. They are Benguela, Bridge Brook Lake, Chesapeake Bay, Coachella Valley, Skipwith Pond, and St. Marks. It is worth noting that this list includes two of the three food webs with $p_{\mathrm{GC}} > 0.05$: Coachella Valley and Skipwith Pond.

For each of these six 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 95% confidence intervals on the value of c for which the empirical $\hat{G}$ is likely to be observed in the generalized niche model (Fig. 2). We show the results of this comparison in Table 3.

For the six food webs we investigated, we find that the largest values of c which provide statistical

---

[2] 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.

agreement with the empirical data are remarkably close to one, $0.85 < c_{\max} < 1.00$ [3]. This finding enables us to quantify in a statistically sound manner the intervality of a food web; specifically,

$$\mathcal{I}(\mathcal{F}_i) \equiv c_{\max}(\{\mathcal{F}_{GN}\}) \,, \tag{2}$$

where $\{\mathcal{F}_{GN}\}$ is the ensemble to model food webs generated according to the generalized niche model and with the same number of species and connectance of the real food web $\mathcal{F}_i$. Our empirical finding that $\mathcal{I}$ for the six empirical food webs considered is so large indicates that natural ecosystems are significantly interval and consequently there is a strong bias toward contiguity in prey selection.

Discussion

The concept of "niche theory" or "niche space" is a fundamental concept in study of ecosystems. Niche space was classically defined as an "n-dimensional hyperspace" with n given by the innumerable ecological and environmental characteristics[15,16]. 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[16,19,21]. 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[6,19,20,21,22,41]. Furthermore, the placing of species into a single dimension is a crucial ingredient in many models developed to describe food web structure[1,2,8,10].

Recently, however, discussions as to how interval food webs truly are, were renewed by the stark contrasts between the niche model—and its contiguous range of prey—and the generalized cascade and nested-hierarchy models—and their random predation[10]. Our results allow us to conclusively demonstrate that natural ecosystems, while not fully interval, are significantly more interval than would be expected when compared to suitable random null hypotheses. Moreover, we find the empirical food webs to be statistically indistinguishable from model food webs whose diets are, on average, at a minimum 85% contiguous. The idea that species and their diets can be so closely mapped to a single dimension

---

[3] 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 and Skipwith Pond.

represents a tremendous insight that can guide us on how best to go about developing dynamic ecosystem models.

A number of future questions must be answered before the question of food web intervality can come to a close. 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, we find it unlikely that all gaps may be attributed to this factor. It has earlier been noted, albeit on different food webs from those studied here, that ecosystems with multiple habitats, for example an estuary, were less likely to be interval than single-habitat food webs[1,11]. Indeed, once cannot expect food webs containing several habitats to be interval since each habitat will have its own independent resource axis.

It would likewise be very interesting to examine additional properties of the "most-interval" ordering or orderings, $\{\mathcal{O}_k\}$. Studies which compared these orderings to those obtained when comparing species' masses, or related properties[6], would be particularly intriguing. It is known empirically, for example, that as predator mass increases so does average prey mass[42,43,44]; similarly, in food web models species generality also correlates to niche value, that is, where the species is found within the location within the ordering[3,45].

Methods

Simulated annealing

Simulated annealing is a stochastic optimization technique that enables one to find a "low-cost" configuration while still broadly exploring the space of possibilities[23]. 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

$$p = \begin{cases} 1 & \text{if} \quad \mathcal{G}\left(\mathcal{O}_f\right) \leq \mathcal{G}\left(\mathcal{O}_i\right) \\ \exp\left(-\frac{\mathcal{G}(\mathcal{O}_f)-\mathcal{G}(\mathcal{O}_i)}{T}\right) & \text{if} \quad \mathcal{G}\left(\mathcal{O}_f\right) > \mathcal{G}\left(\mathcal{O}_i\right) \end{cases} , \tag{3}$$

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 $c = 0.99$.

Generating randomized networks

To generate an ensemble of random networks, one must first define the constraints of the randomization[46,47]. 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, $A \rightarrow B$, (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[25] and treat single, double, and cannibal links separately. For example, two single links $A \rightarrow B$ and $C \rightarrow D$ 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 $C \leftrightarrow D$ become $A \leftrightarrow D$ and $C \leftrightarrow B$, provided that both A, D and C, B are unconnected by a link in any direction.

Acknowledgments

We thank R. Guimerà, S. Levin, R. D. Malmgren, C. A. Ng, M. Sales-Pardo, and E. N. Sawardecker for stimulating discussions and helpful suggestions. DBS acknowledges NSF-IGERT "Dynamics of Complex Systems in Science and Engineering" (DGE-9987577). LANA acknowledges a NIGMS/NIH K25 Career Award and the J.S. McDonnell Foundation.

References

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

2. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
3. 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).
4. Camacho, J., Guimerà, R. & Amaral, L. A. N. Robust patterns in food web structure. *Phys. Rev. Lett.* **88**, art. no. 228102 (2002).
5. 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. USA* **99**, 12917–12922 (2002).
6. Cohen, J. E., Jonsson, T. & Carpenter, S. R. Ecological community description using the food web, species abundance, and body size. *Proc. Natl. Acad. Sci. USA* **100**, 1781–1786 (2003).
7. Garlaschelli, D., Caldarelli, G. & Pietronero, L. Universal scaling relations in food webs. *Nature* **423**, 165–168 (2003).
8. 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).
9. Camacho, J. & Arenas, A. Universal scaling in food-web structure? *Nature* **435**, E3–E4 (2005).
10. 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).
11. Cohen, J. E. Food webs and the dimensionality of trophic niche space. *Proc. Natl. Acad. Sci. USA* **74**, 4533–4536 (1977).
12. Cohen, J. E. *Food Webs and Niche Space* (Princeton University Press, Princeton, NJ, 1978).
13. Sugihara, G. *Niche hierarchy: structure, organization, and assembly in natural communities*. Ph.D. thesis, Princeton University, Princeton, N.J. (1982).
14. 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).

15. 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).
16. Giller, P. S. *Community structure and the niche* (Chapman and Hall, London, 1984).
17. Cohen, J. E. A Stochastic Theory of Community Food Webs. VI. Heterogeneous Alternatives to the Cascade Model. *Theoretical Population Biology* **37**, 55–90 (1990).
18. Cohen, J. E. & Newman, C. M. A stochastic theory of community food webs I. Models and aggregated data. *Proc. Roy. Soc. B* **224**, 421–448 (1985).
19. 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).
20. Lawton, J. H. Food webs. In *Ecological Concepts* (ed. Cherrett, J. M.), 43–78 (Blackwell Scientific, Oxford, UK, 1989).
21. Cohen, J. E. (ed.). *Ecologists' Co-Operative Web Bank, version 1.0* (Rockefeller University, New York, NY, 1989).
22. 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).
23. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
24. 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).
25. 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).
26. Baird, D. & Ulanowicz, R. E. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* **59**, 329–364 (1989).
27. 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).

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

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

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

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

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

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

34. Opitz, S. Trophic interactions in Caribbean coral reefs. *ICLARM Technical Reports* **43**, 341 (1996).

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

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

37. 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).

38. 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).

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

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

41. 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).

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

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

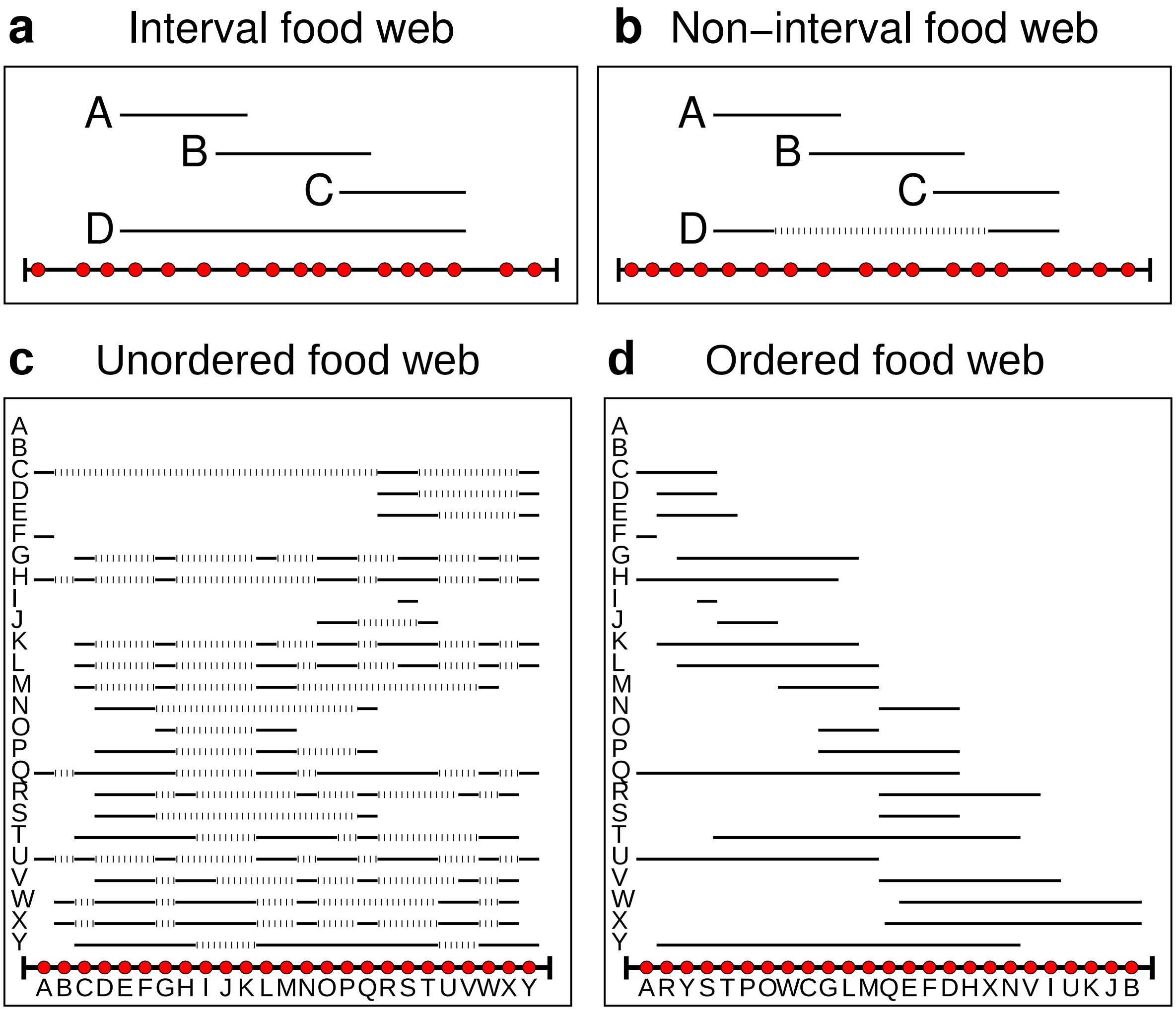
44. Carbone, C., Mace, G. M., Roberts, S. C. & Macdonald, D. W. Energetic constraints on the diet of terrestrial carnivores. *Nature* **402**, 286–288 (1999).

45. 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. USA* **99**, 12913–12916 (2002).

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

47. 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).

Figure 1



a Interval food web

b Non−interval food web

c Unordered food web

d Ordered food web

Figure 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** and **d**, Illustration of the ordering algorithm. **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 $\hat{G} = G = 0$.

Figure 2

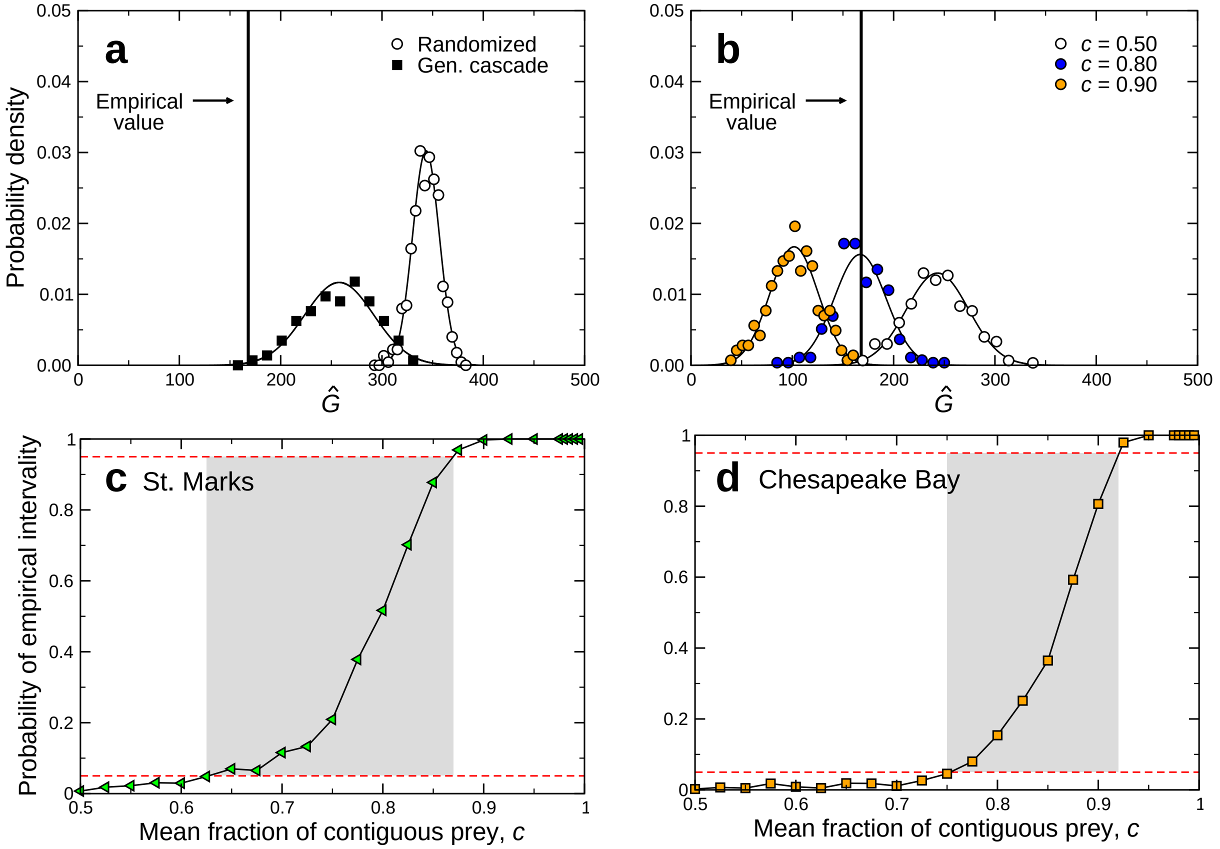

a
Randomized
Gen. cascade
Empirical value
Probability density
0.05
0.04
0.03
0.02
0.01
0.00
0
100
200
300
400
500
$\hat{G}$
b
c = 0.50
c = 0.80
c = 0.90
Empirical value
c St. Marks
d Chesapeake Bay
Probability of empirical intervality
1
0.8
0.6
0.4
0.2
0
0.5
0.6
0.7
0.8
0.9
1
Mean fraction of contiguous prey, c

Figure 2: 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 z as the empirical food web. $\hat{G}_e$ is shown by the spike. The probability of observing such a low $\hat{G}$ value is 4.3×10^{-3} and zero 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 z as the empirical food web. $\hat{G}_e$ is again shown by the spike. **c**, Probability of observing $\hat{G}_{\rm GN}(c) = \hat{G}_e = 168$ for the St. Marks food web. Values less than 0.5 correspond to negative z-scores and thus represent the probability $P_{\rm low}$ of observing a value of $\hat{C}_{\rm GN}$ as small as $\hat{G}_e$, whereas values greater than 0.5 represent the probability $P_{\rm high}$ of observing a value of $\hat{G}_{\rm GN}$ as large as $\hat{G}_e$. The 95% confidence intervals on the value of c are given by the regions where both $P_{\rm low} \geq 0.05$ and $P_{\rm high} \geq 0.05$ (denoted by the dashed red lines). We find the 95% confidence interval to be $c \in [0.625, 0.87]$. **d**, Probability of observing $\hat{G}_{\rm GN}(c) = \hat{G}_e = 11$ for the Chesapeake Bay food web. We find the 95% confidence interval to be $c \in [0.75, 0.92]$. The upper bounds of $c = 0.87$ and $c = 0.92$ 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.

Table 1: Empirical and randomized model intervality. For each of the 15 food webs studied, we show the empirical number of species S, linkage density z, and $\hat{G}_e$. For the null model of randomizations of the empirical food web, we show $\left\langle \hat{G}_{\mathrm{R}} \right\rangle$, z, and p. $\left\langle \hat{G}_{\mathrm{R}} \right\rangle$ is the average over at least 100 model food webs. The z-score is defined as $z = \left(\hat{G}_e - \left\langle \hat{G}_{\mathrm{R}} \right\rangle \right) / \sigma_{\hat{G}_{\mathrm{R}}}$. The probability p represents the probability of observing a value of $\hat{G}_{\mathrm{R}}$ as small as $\hat{G}_e$. This is equivalent to the significance by which one may reject the null hypothesis.

Food web	S	z	$\hat{G}_e$	$\left\langle \hat{G}_{\mathrm{R}} \right\rangle$	z	p
Benguela	29	7.00	27	81.38	-11.31	0.0
Bridge Brook Lake	25	4.28	1	50.95	-11.24	0.0
Canton Creek	102	6.80	639	809.99	-6.71	9.8×10^{-12}
Caribbean Reef	50	11.12	310	497.90	-11.72	0.0
Chesapeake Bay	31	2.19	11	48.03	-5.86	2.2×10^{-9}
Coachella Valley	29	9.03	51	117.24	-10.45	0.0
Grassland	61	1.59	10	27.98	-3.98	3.5×10^{-5}
Little Rock Lake	92	10.84	472	1347.03	-25.44	0.0
Northeast US Shelf	79	17.72	747	1291.1	-16.35	0.0
Scotch Broom	85	2.62	35	225.61	-14.74	0.0
Skipwith Pond	25	7.88	26	36.22	-3.32	4.4×10^{-4}
St. Marks	48	4.60	168	343.41	-13.29	0.0
St. Martin	42	4.88	98	204.40	-11.85	0.0
Stony Stream	109	7.60	676	914.84	-8.63	0.0
Ythan	83	4.80	287	512.72	-10.87	0.0

Table 2: Comparison of empirical data and the generalized cascade model. For the eleven food webs , we show the empirical number of species S, linkage density z, and $\hat{G}_e$. For the null model of randomizations of the empirical food web, we show $\left\langle \hat{G}_{\mathrm{R}} \right\rangle$, z, and p. $\left\langle \hat{G}_{\mathrm{R}} \right\rangle$ is the average over at least 100 model food webs. The z-score is defined as $z = \left(\hat{G}_e - \left\langle \hat{G}_{\mathrm{R}} \right\rangle \right) / \sigma_{\hat{G}_{\mathrm{R}}}$. The probability p represents the probability of observing a value of $\hat{G}_{\mathrm{R}}$ as small as $\hat{G}_e$. This is equivalent to the significance by which one may reject the null hypothesis. The generalized cascade model is a model for community food webs[10], while Scotch Broom is a source web[38]. Additionally, the model was developed to reproduce the properties of cumulative food webs—food webs assembled over an extended period of time, i.e., across multiple seasons—while Canton Creek and Stony Stream are time-specific food webs with data collected on a single day[30]. Ythan has also been reported to be incomplete[28,2,10] with, for example, an over-abundance of “top” bird species whose consumers have been excluded[28]. It has also been shown that the three latter food webs—Canton Creek, Stony Stream, and Ythan—exhibit topological properties that differ markedly from those of generalized cascade model-generated food webs[10]. For these reasons, we do not compare these four food webs to the second null hypothesis, the generalized cascade model.

Food web	S	z	$\hat{G}_e$	$\left\langle \hat{G}_{\mathrm{GC}} \right\rangle$	z	p
Benguela	29	7.00	27	78.18	-3.59	1.6×10^{-4}
Bridge Brook Lake	25	4.28	1	47.74	-4.42	5.0×10^{-6}
Caribbean Reef	50	11.12	310	339.64	-0.82	0.21
Chesapeake Bay	31	2.19	11	38.19	-3.08	1.1×10^{-3}
Coachella Valley	29	9.03	51	64.14	-1.04	0.15
Grassland	61	1.59	10	94.81	-5.44	2.7×10^{-8}
Little Rock Lake	92	10.84	472	1641.14	-9.53	0.0
Northeast US Shelf	79	17.72	747	1049.75	-5.05	2.2×10^{-7}
Skipwith Pond	25	7.88	26	41.59	-1.59	0.056
St. Marks	48	4.60	168	257.79	-2.63	4.3×10^{-3}
St. Martin	42	4.88	98	192.69	-8.30	4.1×10^{-5}

Table 3: Empirical and generalized niche model intervality. Because of the computational effort required, we investigate six empirical food webs. We show 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	S	z	$\hat{G}_e$	$c_{\max}$
Benguela	29	7.00	27	0.95
Bridge Brook Lake	25	4.28	1	1.00
Chesapeake Bay	31	2.19	11	0.925
Coachella Valley	29	9.03	51	0.925
Skipwith Pond	25	7.88	26	0.95
St. Marks	48	4.60	168	0.85