Quantitative analysis of the local structure of food webs

J. Camacho[a,b,*], D.B. Stouffer[b], L.A.N. Amaral[b]

[a]*Departament de Física (Física Estadística), Universitat Autònoma de Barcelona, E-08193 Bellaterra, Catalonia, Spain*
[b]*Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA*

Abstract

We analyze the local structure of model and empirical food webs through the statistics of three-node subgraphs. We study analytically and numerically the number of appearances of each subgraph for a simple model of food web topology, the so-called generalized cascade model, and compare them with 17 empirical community food webs from a variety of environments, including aquatic, estuarine, and terrestrial ecosystems. We obtain analytical expressions for the probability of appearances of each subgraph in the model, and also for randomizations of the model that preserve species' numbers of prey and number of predators; their difference allows us to quantify which subgraphs are over- or under-represented in both the model and the empirical food webs. We find agreement between the model predictions and the empirical results. These results indicate that simple models such as the generalized cascade can provide a good description not only of the global topology of food webs, as recently shown, but also of its local structure.
© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Food web theory seeks to understand the functioning of ecosystems by studying the trophic relations among its species (Cohen et al., 1990). To this end, in the last years great effort has been devoted to the compilation of comprehensive empirical food webs (see for instance Dunne et al., 2002). The statistical treatment of these data has revealed several regularities among food webs belonging to quite diverse habitats, such as deserts, lakes and islands, suggesting that some robust mechanism common to most ecosystems is at work (Williams and Martinez, 2000; Camacho et al., 2002b; Stouffer et al., 2005).

Several models have been proposed to describe the structure of food webs and clarify the origin of these patterns. They differ in the mechanisms underlying them and in the level of description. Some of them describe the dynamics of the network according to evolutionary rules (Amaral and Meyer, 1999; Rossberg et al., 2005; Rossberg et al., 2006a,b), population biology (Yodzis, 1981), or mixtures of both (Caldarelli et al., 1998; Lassig et al., 2001). Other so-called static models do not contain the explicit dynamics of the ecosystem, but provide some mechanistic rules aiming to generate food webs with a statistically similar structure to the empirical ones (Cohen and Newman, 1985; Williams and Martinez, 2000; Cattin et al., 2004; Stouffer et al., 2005).

Two of these static models, the niche model (Williams and Martinez, 2000) and the nested-hierarchy model (Cattin et al., 2004), yield good predictions for a wide number of statistical measures of empirical food webs. Indeed, it has been demonstrated analytically that the two models yield the same distributions for the number of prey and number of predators (Stouffer et al., 2005), which imply, for example, the same fractions of top and basal species or the standard deviations of generality and vulnerability, just as observed numerically (Cattin et al., 2004). 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 (Camacho et al., 2002a,b; Stouffer et al., 2005).

*Corresponding author. Departament de Física (Física Estadística), Universitat Autònoma de Barcelona, E-08193 Bellaterra, Catalonia, Spain. Tel.: +34 93 581 2792; fax: +34 93 581 2155.

E-mail address: juan.camacho@uab.es (J. Camacho).

It has also been demonstrated that much of the success of these models relies in the fact that they satisfy two basic conditions (Stouffer et al., 2005): (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. Any model which satisfies these conditions will reproduce the distributions of number of prey and number of predators observed empirically. For instance, the generalized cascade model (Stouffer et al., 2005)—a generalization of Cohen and Newman's (1985) cascade model to satisfy condition (ii)—exhibits the same distributions. Those two conditions can thus be interpreted as fundamental mechanisms shaping food web structure.

Aside few exceptions (Melián and Bascompte, 2004; Bascompte and Melian, 2005), the studies of these models, however, mainly characterize the global structure of food web topology. Here, 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 (Fig. 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 (Milo et al., 2002, 2004). Let us note that some authors have attempted 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" (Holt, 1997; Holt and Hochberg, 2001). 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 the one followed by Bascompte and Melian (2005), though it differs in several aspects, such as the theoretical approach and our systematic analysis of all three-node subgraphs.

The purpose of this work is to study the statistics of subgraphs in model and empirical 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, of predators and related quantities, but also its local structure. Because it allows for analytical treatment, we will focus here on the generalized cascade model, the simplest model obeying those ingredients. Specifically, we study analytically and numerically the subgraph 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.

The paper is organized as follows. In the second section, we study the statistics of subgraphs for the generalized cascade model. This analysis is twofold: we first evaluate the number of appearances for each subgraph and secondly we study their patterns of over/under-representation. In the third section, we perform the same analysis for the completely random model, as a basis for comparison. Section 4 compares the model predictions with the results obtained for 17 empirical food webs.

2. The generalized cascade model

The original cascade model (Cohen and Newman, 1985; Cohen et al., 1990) is based on two rules: (1) species make up an ordered set according to their niche value n drawn uniformly in the interval $[0,1]$, and (2) any species j with $n_j < n_i$ becomes a prey of i with fixed probability $x_0 = 2CS/(S-1)$; here S is the number of species in the food web, L the number of trophic connections, and $C \equiv L/S^2$ the directed connectance. Williams and Martinez (2000) demonstrated that this model is not able to reproduce the properties of real food webs.

Stouffer et al. (2005) demonstrated, however, that it can be easily generalized to provide similar agreement to the niche (Williams and Martinez, 2000) or the nested-hierarchy (Cattin et al., 2004) models as compared to many empirical food webs. The generalization consists in that the probability x with which a species i feeds on species j with $n_j \leqslant n_i$ is not the same for every predator i, but it is drawn at random from a probability distribution $p(x)$ given by

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

the so-called beta-distribution (see Fig. 2). Parameter β is related with the directed connectance of the empirical food web by $C = 1/2(\beta+1)$ (Williams and Martinez, 2000).
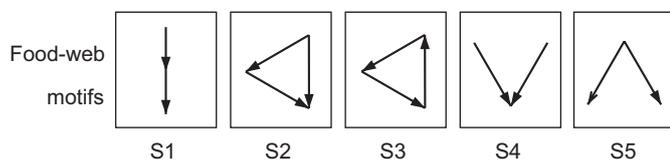

Fig. 1. Three-species motifs containing only single links. Notice that each of these motifs has a clear ecological relevance and is additionally related to some community modules. Motif S1 describes the simple food chain, S2 simple omnivorism, S3 a trophic loop involving three species, S4 isolated exploitative competition, and S5 isolated generalist predation.
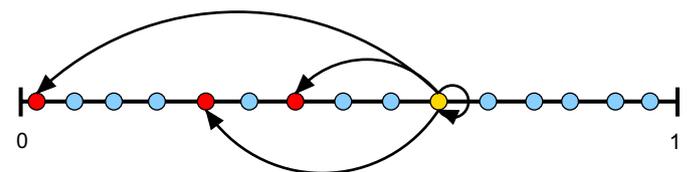

Fig. 2. The generalized cascade model of Stouffer et al. (2005). In the generalized cascade model species make up an ordered set according to their niche value n, $0 < n < 1$. Each species i then consumes species j with $n_j \leqslant n_i$ with a probability x drawn at random from a probability distribution $p(x)$ given by the beta-distribution Eq. (1). In this example the predator (the yellow species) can consume any of the species to its left on the axis, including itself. In this case, $x \approx 0.2$ and the yellow predator consumes itself and three other species.

2.1. Subgraph probabilities

Neglecting cannibalism or self-links, there are two possible unique subgraphs comprising a pair of species: (i) single links, $A \rightarrow B$, i.e. species A eats species B but not conversely, and (ii) double links, $A \leftrightarrow B$. With three species, there are 13 unique subgraphs possible. Since predators in the generalized cascade model cannot feed on species having a larger niche value, no trophic loops of any size 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 study focusses on the five subgraphs S1–S5 (Fig. 1), while the analysis of the motifs containing double links will be dealt with elsewhere. The present analysis is nonetheless meaningful since single connections account for the vast majority of the links in empirical food webs (see Appendix A). 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, S2 simple omnivorism, subgraph S3 a trophic loop involving three species, S4 isolated exploitative competition, and S5 isolated generalist predation.

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

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

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

$$p^{S3} = 0. \tag{3}$$

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

$$p^{S1} = p^{S4} = \langle x_A \, x_B \rangle - \langle x_A^2 \, x_B \rangle, \tag{4}$$
$$p^{S2} = \langle x_A^2 x_B \rangle, \tag{5}$$

and

$$p^{S5} = \langle x_A^2 \rangle - \langle x_A^2 x_B \rangle, \tag{6}$$

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)–(6) can be rewritten as

$$p^{S1} = p^{S4} = \langle x \rangle^2 - \langle x^2 \rangle \langle x \rangle, \tag{7}$$
$$p^{S2} = \langle x^2 \rangle \langle x \rangle, \tag{8}$$

and

$$p^{S5} = \langle x^2 \rangle - \langle x^2 \rangle \langle x \rangle. \tag{9}$$

These expressions are valid for arbitrary distributions $p(x)$. Substituting the beta-distribution, Eq. (1) (see Appendix B) becomes

$$p^{S1} = p^{S4} = 4C^2 \frac{1 - 2C}{1 + 2C}, \tag{10}$$

$$p^{S2} = \frac{16C^3}{1 + 2C}, \tag{11}$$

and

$$p^{S5} = 8C^2 \frac{1 - 2C}{1 + 2C}. \tag{12}$$

In Fig. 3, we compare the analytical predictions, Eqs. (10)–(12), with simulations of the generalized cascade model—in these simulations and throughout the paper, subgraphs have been directly enumerated using dynamic programming, as in the mfinder software for network motif detection. It becomes visually apparent that the expressions we derived compare quite well with the model-generated data. Notice that the probabilities p only depend on the connectance C.

2.2. Patterns of over/under-representation of subgraphs

As a second test of the generalized cascade model, we analyze which subgraphs are typically over- or under-represented as compared to the corresponding randomized networks (Milo et al., 2002,2004). The randomized networks are obtained by preserving the number of prey k_i and of predators m_i of each species i as in the generalized cascade model, but rewiring their trophic links randomly using the Markov-chain Monte Carlo switching algorithm (Maslov and Sneppen, 2002; Itzkovitz et al., 2004).

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 subgraphs S1–S5 (Maslov and Sneppen, 2002; Itzkovitz et al., 2004). Notice, however, that, by construction, the distributions of number of prey or of number of predators are the same ones 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. We deal with the latter ones in the next section, as a null model for comparison with the predictions of the generalized cascade model.
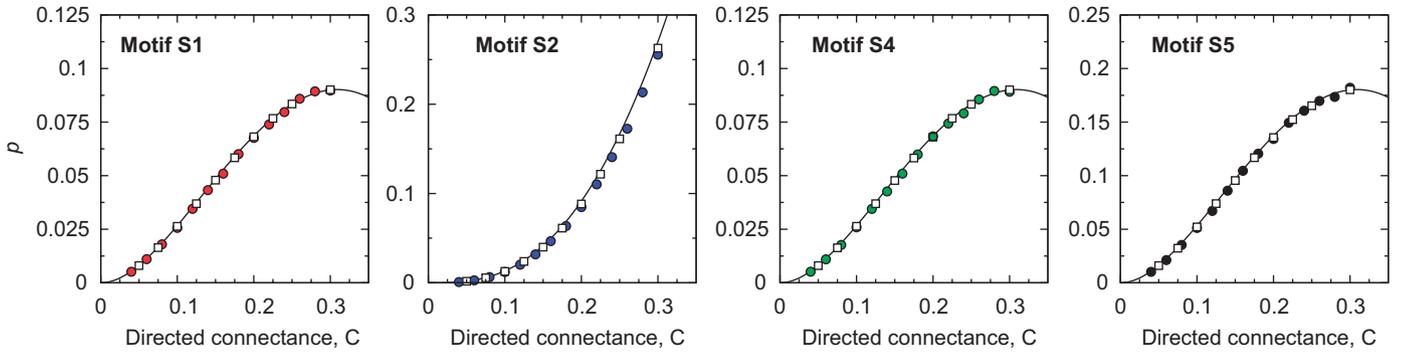
Fig. 3. Comparison between analytical expressions, Eqs. (10)–(12), and simulations of the generalized cascade model for motifs S1, S2, S4, and S5. We exclude motif S3 because, by definition, $p^{S3} = 0$. It is visually apparent that the analytical predictions agree with the model-generated data. Filled 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.
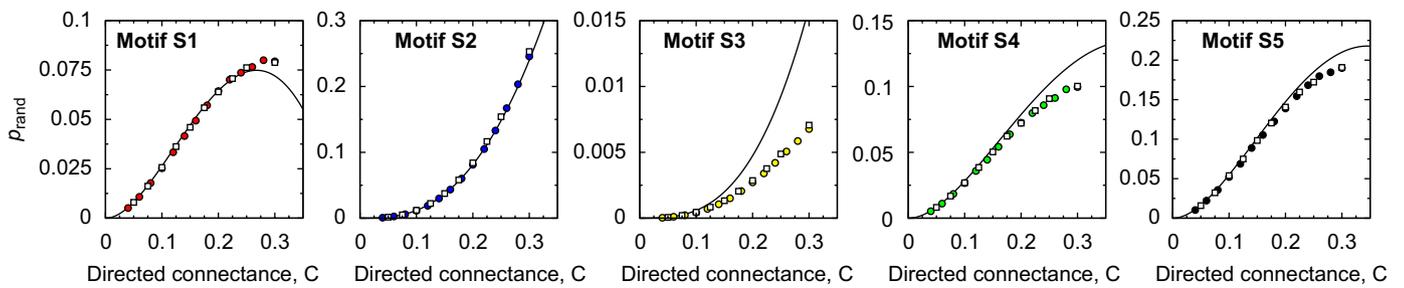

Fig. 4. Comparison between analytical expressions, Eqs. (13)–(17), and randomizations of the generalized cascade model for motifs S1–S5. It is visually apparent that the analytical predictions compare well with the model-generated data. Filled circles are for food webs with $S = 50$ and open squares for food webs with $S = 100$. Each data point represents an average over 1000 model-generated food webs.

Note that, since randomized food webs possess the same degree distributions as the original ones, the occurrence of patterns of over/under-representation of subgraphs 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 (Artzy-Randrup et al., 2004) or they provide some ecological advantage and have arisen as a result of selection pressure. Here we show that the generalized cascade model yields well-defined patterns of over/under-representation of motifs. If these predictions compare well with empirical food webs, one might conclude that the second hypothesis is not required and all patterns arise as the result of the food web generating mechanisms.

We thus next evaluate the probabilities of subgraphs in randomized networks of the generalized cascade model. Itzkovitz et al. (2003) derived expressions for the mean number of appearances of each subgraph for randomized networks with an arbitrary degree distribution. In Appendix B, we calculate the probabilities for the three-species motifs in the randomizations of the generalized cascade model yielding

$$p_{rand}^{S1} = 4C^2 \frac{1 - 2C - 8C^2/9}{1 + 2C}, \tag{13}$$

$$p_{rand}^{S2} = \frac{178}{9} \frac{C^3}{1 + 2C}, \tag{14}$$

$$p_{rand}^{S3} = \frac{16}{27} C^3, \tag{15}$$

$$p_{rand}^{S4} = 4C^2 \frac{1 - 14C/9}{1 + 2C}, \tag{16}$$

and

$$p_{rand}^{S5} = 8C^2 \frac{1 - 16C/9}{1 + 2C}. \tag{17}$$

Fig. 4 compares the analytical predictions for the randomizations, Eqs. (13)–(17), with simulations of the generalized cascade model finding good agreement. The small discrepancies observed have their origin in that some of the expressions derived in Itzkovitz et al. (2003) are approximate.

Finally, we obtain the differences by subtracting the probability of motifs appearing in the model, Eqs. (10)–(12), and in their randomizations, Eqs. (13)–(17). Table 1 summarizes these results. We show comparisons between the expressions for $p - p_{rand}$ and simulations of the generalized cascade model in Fig. 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). The percentage of under/over-representation

is, however, rather small, generally under 10%. These predictions make up our second check of the generalized cascade model.

3. The completely random model

In the following section, we will compare the analytical expressions derived for the generalized cascade model with empirical data. However, we calculate them now for a different model, a fully random network, in order to determine whether the probability functions for the subgraphs depend significantly on the mechanisms underlying the generation of the network.

In a completely random (Erdos–Renyi) network, each species has the same probability x to be connected to any other species in the network. According to this definition, the average number of prey per species is $z \equiv L/S = Sx$, and the directed connectance yields $C \equiv z/S = x$.

Let us consider three arbitrary species in the network, say A, B, and C. They are connected through subgraph S1 if, for instance, A eats B, and B eats C with no further links among them; this happens with probability $x^2(1-x)^4$. Of course, there exist other options to build this subgraph by A, B, and C exchanging their roles, which amounts to a total of six different configurations. As a consequence,

we have

$$p^{S1} = 6C^2(1-C)^4. \tag{18}$$

Similarly, considering the number of configurations for every subgraph and the probability for each of them, one finds

$$p^{S2} = 6C^3(1-C)^3, \tag{19}$$

$$p^{S3} = 2C^3(1-C)^3, \tag{20}$$

and

$$p^{S4} = p^{S5} = 3C^2(1-C)^4. \tag{21}$$

Finally, since these networks are completely random, their randomization provides equally random networks, so that the probabilities for subgraphs S1–S5 in the randomized networks of this model are exactly the same ones, namely Eqs. (18)–(21). The differences between them are obviously zero.

4. Subgraphs statistics in empirical food webs

One interesting observation from Figs. 3–5 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 subgraph S1–S5 for 17 empirical food webs (see Appendix A for details). Figs. 6–8 show the results for the empirical food webs, their randomizations, and the differences, and also the comparison with the generalized cascade and the fully random models. 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; in contrast, the completely random model

Table 1
Analytical expressions for appearance probability of motifs S1–S5 in the generalized cascade model

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

The right column states the prediction of over- or under-representation of each motif according to the model.
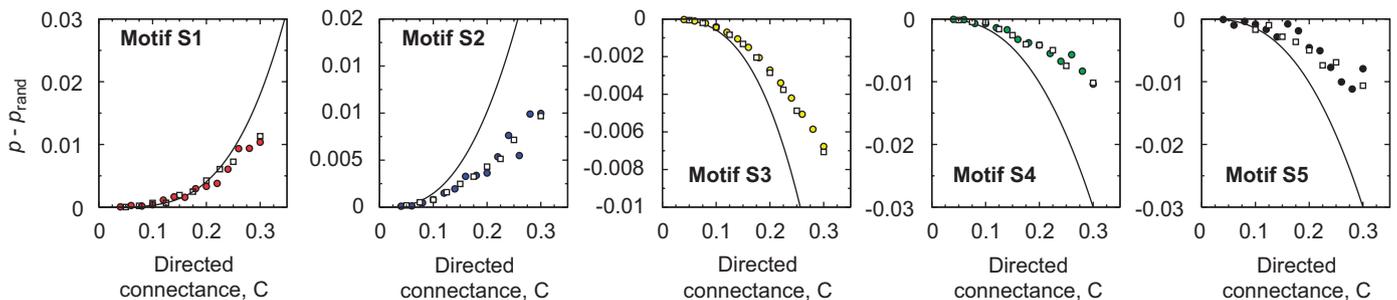

Fig. 5. Comparison between analytical expressions for $p - p_{rand}$ and simulations of the generalized cascade model for motifs S1–S5. The model predicts over-representation of motifs S1 and S2 and under-representation of motifs S3–S5. The analytical predictions compare well with the model-generated data, though they generally overestimate the differences due to the approximate character of the expressions used to evaluate p_{rand}. 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.
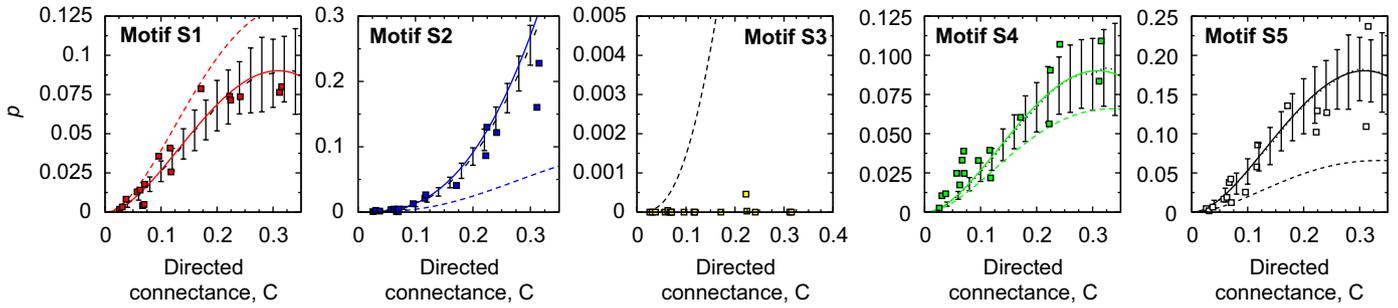
Fig. 6. Fraction of appearances of motifs for empirical food webs (symbols) compared to the analytical predictions for the generalized cascade model (solid lines) and the random model (dashed lines). Numerical simulations for the generalized cascade model with $S = 50$ are shown by the dotted line where 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, whereas the random model provides much poorer fits. Note that there are no fitting parameters in model estimates.
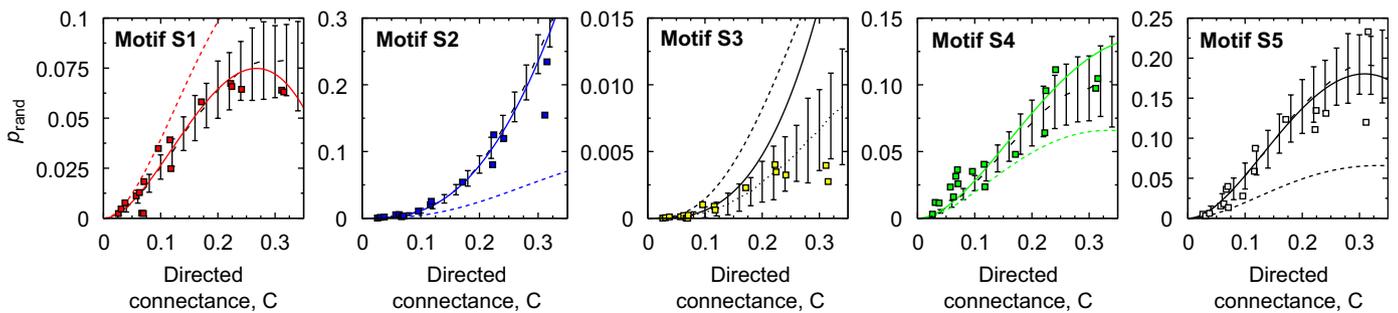

Fig. 7. Fraction of appearances of motifs for randomizations of the empirical food webs (symbols) compared to the analytical predictions for the generalized cascade model (solid line) and the random model (dashed lines). Numerical simulations for the generalized cascade model with $S = 50$ are shown 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, whereas the random model provides much poorer fits. Note that there are no fitting parameters in model estimates.
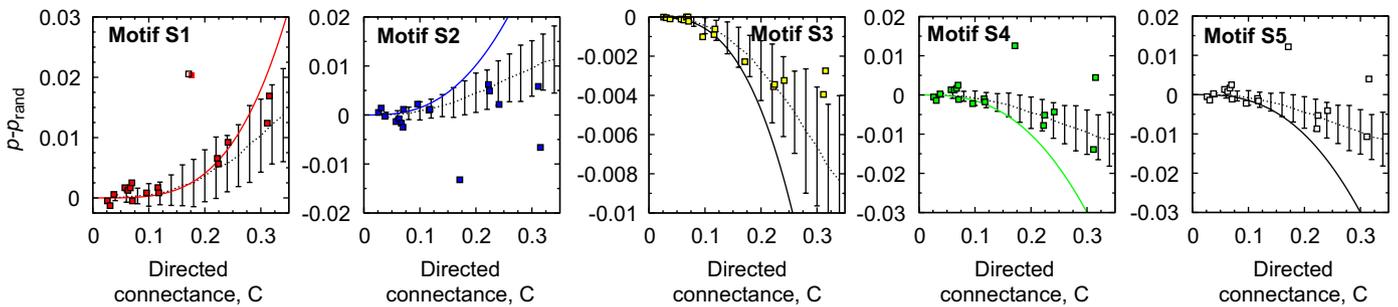

Fig. 8. Differences between actual appearances of motifs and the corresponding randomized food webs for 17 empirical food webs (symbols) as compared to the analytical predictions for the generalized cascade model (solid line) and the random model (dashed lines). Numerical simulations for the generalized cascade model with $S = 50$ are shown by the dotted line, where the error bars are two standard deviations. 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.

provides remarkably poorer fits to the empirical values in most cases.

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_{rand} 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 1. Quantitatively, the theoretical curves generally over-estimate the empirical values, while the numerical simulations of the model provide reasonable estimates.

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. 8: they are Bridge Brook (with $C = 0.17$)

Table 2
Empirical food webs studied

Food web	S	L	Single links	Double links	Cannibal links	Reference
Benguela	29	203	186	5	7	Yodzis (1998)
Bridge Brook Lake	25	106	102	1	3	Havens (1992)
Caribbean Reef	50	556	471	32	21	Opitz (1996)
Chesapeake Bay	31	68	67	0	1	Baird and Ulanowicz (1989)
Coachella Valley	29	262	199	22	19	Polis (1991)
Grassland	61	97	97	0	0	Martinez et al. (1999)
Little Rock Lake	92	997	936	24	13	Martinez (1991)
Northeast US Shelf	79	1400	1363	7	25	Link (2002)
Scotch Broom	85	223	219	0	4	Hawkins et al. (1997)
Skipwith Pond	25	197	181	4	8	Warren (1989)
St. Marks Seagrass	48	221	218	0	3	Christian and Luczkovich (1999)
St. Martin Island	42	205	205	0	0	Goldwasser and Roughgarden (1993)
Ythan Estuary (1)	83	395	389	1	4	Hall and Raffaelli (1991)
Ythan Estuary (2)	174	579	573	1	4	Hall and Raffaelli (1993)
Canton Creek	102	697	696	0	1	Townsend et al. (1998)
Stony Stream	109	829	827	0	2	Townsend et al. (1998)
El Verde Rainforest	155	1509	1369	69	2	Waide and Reagan (1996)

S is the number of trophic species in the food web and L is the number of trophic predator–prey interactions.

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 2). Although statistical fluctuations grow with decreasing size, they do not seem enough to explain this behaviour. Note, on the other hand, that they match rather well with the predictions for p and p_{rand} separately (Figs. 6 and 7).

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

5. Concluding remarks

The predictions of the generalized cascade model for the appearances of subgraphs S1–S5 provide good comparison to the empirical results, in contrast to those for a completely random model. 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. 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.

It is an interesting ecological question to determine why these appear to be such important ingredients to explain food web structure. Recently, Rossberg et al. have devised a couple of dynamic models, the speciation model (Rossberg et al., 2005, 2006a) and the matching model (Rossberg et al., 2006b) that seem to provide the dynamical explanation. By starting from an ordered set of species, the dynamic evolutionary rules of speciation, extinction, and migration lead to distributions of number of prey and number of predators similar to the ones obtained through the static models. Indeed, the matching model fits remarkably the empirical data, improving in some cases the predictions of the niche model at the expense of a number of adjustable parameters. From the result of an extensive analysis, the authors conclude that the tendency of newly created species to avoid competition with their relatives is indeed the fundamental mechanism responsible of food web structure.

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 over-representation of subgraph S2 predicted by the generalized cascade model is unclear in the empirical data. This is also the result found by Bascompte and Melián (2005), who also analyzed the under/over-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 other hand, we count motifs only once (for instance, we do not count the food chains included in motifs 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 (McCann and Hastings, 1997).

Finally note that, from the quantitative point of view, the percentage of under/over-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 subgraphs in empirical food webs and in their randomizations indicates that there may be no overwhelming evolutionary trend toward under/over-representation of any motif, and that the small differences observed are more likely a consequence of the mechanisms generating the food web.

Acknowledgements

Appendix A. Empirical food webs

Table 2 provides the list of food webs analyzed as well as some topological parameters characterizing them. They range between 25 and 92 trophic species, and the average connectivity varies from 2.19 to 17.72. It also contains the number of single links, double links, and cannibal links for each empirical food web. One observes that the frequency of double links is generally small, with only one case, Coachella Valley, close to 10%.

Appendix B. Motif probabilities in the generalized cascade model

We calculate here the subgraph probabilities for the randomizations of the generalized cascade model. Itzkovitz et al. (2003). derived general expressions for the average number of appearances N^i of motifs in randomized networks. 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 subgraphs S1–S5, these can be cast as

$$p^{S1}_{rand} = p_1 - p^{S2}_{rand} - 3p^{S3}_{rand},$$
$$p^{S4}_{rand} = p_4 - p^{S2}_{rand},$$
$$p^{S5}_{rand} = p_5 - p^{S2}_{rand}, \tag{22}$$

with

$$p_1 = \frac{S}{S_T} \langle k_i m_i \rangle,$$
$$p_4 = \frac{S}{2S_T} \langle m_i(m_i - 1) \rangle,$$
$$p_5 = \frac{S}{2S_T} \langle k_i(k_i - 1) \rangle,$$
$$p^{S2}_{rand} = \frac{1}{z^3 S_T} \langle k_i(k_i - 1) \rangle \langle k_i m_i \rangle \langle m_i(m_i - 1) \rangle,$$
$$p^{S3}_{rand} = \frac{1}{3z^3 S_T} \langle k_i m_i \rangle^3, \tag{23}$$

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 $\langle \ldots \rangle$ is the average over all species in the randomized network. Since these networks have the same distributions of in- and out-links that the original networks, these averages can be calculated directly from the latter ones.

From expressions (23), p^{S2}_{rand} and p^{S3}_{rand} can be rewritten as

$$p^{S2}_{rand} = \frac{4S_T^2}{S^3 z^3} p_1 p_4 p_5,$$
$$p^{S3}_{rand} = \frac{S_T^2}{3S^3 z^3} p_1^3. \tag{24}$$

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 $\langle x_A x_B \rangle = \langle x \rangle^2$, 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 $\langle x_A x_B \rangle = \langle x \rangle^2$. (iii) Similarly, p_5 is the probability of A eating species B and C, independently of the eventual connection of B and C, namely $\langle x_A^2 \rangle = \langle x^2 \rangle$.

By replacing these results in Eqs. (22)–(24), one finds

$$p^{S1}_{rand} = \langle x^2 \rangle - \frac{\langle x \rangle^4 \langle x^2 \rangle}{9C^3} - \frac{\langle x \rangle^6}{36C^3},$$
$$p^{S2}_{rand} = \frac{\langle x \rangle^4 \langle x^2 \rangle}{9C^3},$$
$$p^{S3}_{rand} = \frac{\langle x \rangle^6}{108C^3},$$
$$p^{S4}_{rand} = \langle x \rangle^2 - \frac{16C}{9} \langle x^2 \rangle,$$
$$p^{S5}_{rand} = \langle x^2 \rangle \left(1 - \frac{16C}{9} \right). \tag{25}$$

Finally, the beta-function (1) yields

$$\langle x \rangle = 2C,$$

$$\langle x^2 \rangle = \frac{8C^2}{1 + 2C}. \tag{26}$$

The substitution of these expressions into (25) supplies Eqs. (13)–(17).

References

Amaral, L.A.N., Meyer, M., 1999. Environmental changes, coextinction, and patterns in the fossil record. Phys. Rev. Lett. 82, 652–655.

Artzy-Randrup, Y., Fleisman, S.J., Ben-Tal, N., Stone, L., 2004. Comment on "network motifs: simple building blocks of complex networks". Science 305, 1107.

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

Bascompte, J., Melian, C.J., 2005. Simple trophic modules for complex food webs. Ecology 86 (11), 2868–2873.

Caldarelli, G., Higgs, P.G., McKane, A.J., 1998. Modeling coevolution in multispecies communities. J. Theor. Biol. 193, 345.

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

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

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

Christian, R.R., Luczkovich, J.J., 1999. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. Ecol. Modelling 117, 99–174.

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

Cohen, J.E., Briand, F., Newman, C.M., 1990. Community Food Webs: Data and Theory. Springer, Berlin.

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

Goldwasser, L., Roughgarden, J., 1993. Construction of a large Caribbean food web. Ecology 74, 1716–1733.

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

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

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

Hawkins, B.A., Martinez, N.D., Gilbert, F., 1997. Source food webs as estimators of community food web structure. Int. J. Ecol. 18, 575–586.

Holt, R.D., 1997. Community modules. In: Gange, A.C., Brown, V.K. (Eds.), Multitrophic interactions in Terrestrial Ecosystems, 36th Symposium of the British Ecological Society. Blackwell Science, pp. 333–350.

Holt, R.D., Hochberg, M.E., 2001. Indirect interactions, community modules and biological control: a theoretical perspective. In: Waijnberg, E., Scott, J.K., Quimby, P.C. (Eds.), Evaluation of Indirect Ecological Effects of Biological Control. CAB International, pp. 13–37.

Itzkovitz, S., Milo, R., Kashtan, N., Ziv, G., Alon, U., 2003. Subgraphs in random networks. Phys. Rev. E 68 art. no. 026177.

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

Lassig, M., Bastolla, U., Manrubia, S.C., Valleriani, A., 2001. Shape of ecological networks. Phys. Rev. Lett. 86, 4418–4421.

Link, J., 2002. Does food web theory work for marine ecosystems? Marine Ecology Progr. Ser. 230, 1–9.

Martinez, N.D., 1991. Artifacts or attributes? effects of resolution on the Little Rock Lake food web. Ecol. Monogr. 61, 367–392.

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

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

McCann, K., Hastings, A., 1997. Re-evaluating the omnivory–stability relationship in food webs. Proc. R. Soc. of London B 264, 1249–1254.

Melián, C.J., Bascompte, J., 2004. Food web cohesion. Ecology 85 (2), 352–358.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: simple building blocks of complex networks. Science 298, 824–827.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., Alon, U., 2004. Superfamilies of designed and evolved networks. Science 303, 1538–1542.

Opitz, S., 1996. Trophic interactions in caribbean coral reefs. ICLARM Technical Reports 43, 341.

Polis, G.A., 1991. Complex trophic interactions in deserts: an empirical critique of food-web theory. Am. Nat. 138, 123–155.

Rossberg, A.G., Matsuda, H., Amemiya, T., Itoh, K., 2005. An explanatory model for food-web structure and evolution. Ecol. Complexity 2, 317–321.

Rossberg, A.G., Matsuda, H., Amemiya, T., Itoh, K., 2006a. Some properties of the speciation model for food web structure—mechanisms for degree distributions and intervality. J. Theor. Biol. 238 (2), 401–415.

Rossberg, A.G., Matsuda, H., Amemiya, T., Itoh, K., 2006b. An explanatory model for food-web structure and evolution. J. Theor. Biol. 241, 552–563.

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

Townsend, C.R., Thompson, R.M., McIntosh, A.R., Kilroy, C., Edwards, E., Scarsbrook, M.R., 1998. Disturbance, resource supply, and food-web architecture in streams. Ecol. Lett. 1, 200–209.

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

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

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

Yodzis, P., 1981. The stability of real ecosystems. Nature 289, 674–676.

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