

What constrains food webs? A maximum entropy framework for predicting their structure with minimal biases

Francis Banville^{1,2,3} Dominique Gravel^{2,3} Timothée Poisot^{1,3}

¹ Université de Montréal ² Université de Sherbrooke ³ Quebec Centre for Biodiversity Science

Correspondance to:

Francis Banville — francis.banville@umontreal.ca

Food webs are complex ecological networks whose structure is both ecologically and statistically constrained, with many network properties being correlated with each other. Despite the recognition of these invariable relationships in food webs, the use of the principle of maximum entropy (MaxEnt) in network ecology is still rare. This is surprising considering that MaxEnt is a statistical tool precisely designed for understanding and predicting many different types of constrained systems. Precisely, this principle asserts that the least-biased probability distribution of a system's property, constrained by prior knowledge about that system, is the one with maximum information entropy. MaxEnt has been proven useful in many ecological modelling problems, but its application in food webs and other ecological networks is limited. Here we show how MaxEnt can be used to derive many food-web properties both analytically and heuristically. First, we show how the joint degree distribution (the joint probability distribution of the numbers of prey and predators for each species in the network) can be derived analytically using the number of species and the number of interactions in food webs. Second, we present a heuristic and flexible approach of finding a network's adjacency matrix (the network's representation in matrix format) based on simulated annealing and SVD entropy. We built two heuristic models using the connectance and the joint degree sequence as statistical constraints, respectively. We compared both models' predictions against corresponding null and neutral models commonly used in network ecology using open access data of terrestrial and aquatic food webs sampled globally ($N = 257$). We found that the heuristic model constrained by the joint degree sequence was a good predictor of many measures of food-web structure, especially the nestedness and motifs distribution. Specifically, our results suggest that the structure of terrestrial and aquatic food webs is mainly driven by their joint degree distribution.

This work is released by its authors under a CC-BY 4.0 license



Last revision: March 10, 2023

1 Introduction

2 The constrained structure of ecological networks

3 A variety of measures of the structure of ecological networks have been used to describe the organization of
4 species interactions in a biological community (Delmas *et al.* 2019). These measures provide valuable
5 information on the functioning of ecosystems and their responses to environmental change (e.g., Pascual &
6 Dunne 2006; Gómez *et al.* 2011). For instance, Bascompte *et al.* (2003) suggest that plant–pollinator and
7 seed-disperser networks have a highly nested structure that can promote species persistence. Another example,
8 in food webs, shows that a high connectance can promote the robustness of the system to species lost (Dunne *et*
9 *al.* 2002). However, despite the growing literature on the ecological implications of network structure, the
10 association between many of these measures impedes our ability to fully understand what drives the structure
11 and behavior of ecological networks. In particular, nestedness and modularity are strongly associated in
12 ecological networks (Fortuna *et al.* 2010), and network connectance has been shown to be an important driver
13 of many other emerging network properties (Poisot & Gravel 2014). In light of these observations, it is difficult
14 to assess whether attributed effects of given properties are the artifacts of other, perhaps simpler, measures.

15 One way to tackle this issue is first to recognize that food webs and other ecological networks are constrained
16 systems. In other words, the space of possible network configurations shrinks as we know more about a network
17 structure. For example, there is a finite number of networks with specified numbers of nodes and edges. Indeed,
18 the structure of ecological networks is first and above all constrained by the number of species, or nodes,
19 present. Food webs with high species richness typically have a lower connectance (MacDonald *et al.* 2020) than
20 smaller networks. This is because the number of realized interactions in empirical food webs scales slower than
21 the number of possible species pairs (MacDonald *et al.* 2020). As shown by Poisot & Gravel (2014),
22 connectance itself can constrain different aspects of network structure such as the degree distribution (i.e. the
23 probability distribution of the number of interspecific interactions realized by a species). Other measures, such
24 as the maximum trophic level, can also constrain the space of feasible networks.

25 Prior knowledge on the structure of ecological networks is thus especially useful in the current context of data
26 scarcity about species interactions (Jordano 2016). Indeed, this Eltonian shortfall (Hortal *et al.* 2015) can be
27 partially alleviated using known information about an ecological network. As suggested by Strydom *et al.*
28 (2021a), network structure can be used to improve the prediction of pairwise species interactions when data is
29 lacking by constraining the space of feasible networks. Similarly, partial knowledge on the structure of an

30 ecological network can also be used to predict others of its properties by constraining their range of possible
31 values. This is important given that many aspects of network structure cannot be measured empirically without
32 data on pairwise species interactions, a prevailing situation worldwide (Poisot *et al.* 2021).

33 Understanding the ecological constraints that shape species interactions networks and predicting their emerging
34 structure are thus two complementary aims of network ecology. This distinction between understanding and
35 predicting is essential when using statistical and mathematical models in network ecology and interpreting
36 them. On one hand, null models help us identify potential ecological mechanisms that drive species interactions
37 and constrain ecological networks. Null models generate a distribution for a target measure using a set of rules
38 that exclude the mechanism of interest (Fortuna & Bascompte 2006; Delmas *et al.* 2019). The deviation
39 between the model and empirical data helps us evaluate the effect of this ecological process in nature. On the
40 other hand, predictive models can help fill many gaps on species interactions data. A variety of such models
41 have recently been developed using machine learning and other statistical tools, most of which are presented in
42 Strydom *et al.* (2021a). However, given the constrained nature of ecological networks, it is surprising that the
43 principle of maximum entropy, a mathematical method designed for both the analysis and prediction of
44 constrained systems, has been barely used in network ecology.

45 **The principle of maximum entropy: A primer for ecologists**

46 The principle of maximum entropy (MaxEnt) is a mathematical method of finding probability distributions,
47 strongly rooted in statistical mechanics and information theory (Jaynes 1957a, b; Harremoës & Topsøe 2001).
48 Starting from a set of constraints given by prior knowledge of a system (i.e. what we call state variables), this
49 method helps us find least-biased probability distributions subject to the constraints. These probability
50 distributions are guaranteed to be unique given our prior knowledge and represent the most we can say about a
51 system without making more assumptions. For example, if the only thing we know about a biological
52 community is its average number of individuals per species, the least-biased inference we could make on its
53 species abundance distribution is the exponential distribution (Frank & Smith 2011; Harte & Newman 2014).
54 However, this does not imply that this distribution will be the best fit to empirical data. The challenge is to find
55 the right set of constraints that would best reproduce distributions found in nature.

56 MaxEnt states that the least-biased probability distribution given the constraints used is the one with the highest
57 entropy among all probability distributions that satisfy these constraints. Entropy is a measure of the average

58 amount of information given by the outcome of a random variable. Many measures of entropy have been
 59 developed in physics (Beck 2009), but only a fraction of them could be used as an optimization measure with
 60 the principle of maximum entropy. According to Beck (2009) and Khinchin (2013), a measure of entropy H
 61 should satisfy four properties in the discrete case: (1) it should be a function of a probability distribution $p(n)$
 62 only; (2) it should be maximized when $p(n)$ is uniform; (3) it should not be influenced by outcomes with a null
 63 probability; and (4) it should be independent of the order of information acquisition. The Shannon's entropy
 64 (Shannon 1948)

$$H = - \sum_n p(n) \log p(n) \quad (1)$$

65 satisfies all of these properties. Finding the probability distribution $p(n)$ that maximizes H under a set of m
 66 constraints g can be done using the method of Lagrange multipliers. These constraints could include one or
 67 many properties of the probability distribution (e.g., its mean, variance, and range). However, the normalization
 68 constraint always need to be included in g in order to make sure that $p(n)$ sums to 1. The objective is then to
 69 find the values of the Lagrange multipliers λ_i that optimize a function F :

$$F = H - \sum_{i=1}^m \lambda_i (g_i - c_i), \quad (2)$$

70 where g_i is the mathematical formulation of the constraint i and c_i , its value. Note that F is just Shannon's
 71 entropy to which we added terms that each sums to zero ($g_i = c_i$). F is maximized by setting to 0 its partial
 72 derivative with respect to $p(n)$.

73 The principle of maximum entropy has been used in a wide range of disciplines, from thermodynamics,
 74 chemistry and biology (Martyushev & Seleznev 2006) to graph and network theory (e.g., Park & Newman
 75 2004; van der Hoorn *et al.* 2018). It has also been proven useful in ecology, e.g. in species distribution models
 76 (Phillips *et al.* 2006) and macroecological models (Harte *et al.* 2008; Harte & Newman 2014). In network
 77 ecology, MaxEnt has been used to predict the degree distribution of bipartite networks from the number of
 78 species and the number of interactions (Williams 2011) and to predict interaction strengths between species
 79 pairs using their relative abundances within an optimal transportation theory regularized with information
 80 entropy (Stock *et al.* 2021). However, to the best of our knowledge, MaxEnt has never been used to predict
 81 food-web structure directly, even though food webs are among the most documented and widespread ecological

82 networks (Ings *et al.* 2009).

83 Food-web properties that can be derived using MaxEnt are varied and pertain to different elements of the
84 network (i.e. at the species (node), the interaction (edge) or the community (network) levels). Because MaxEnt
85 is a method of finding least-biased probability distributions given partial knowledge about a system, these
86 properties need to be represented probabilistically. For example, at the species level, MaxEnt can be used to
87 predict the distribution of trophic levels among species, as well as the distribution of species' vulnerability
88 (number of predators) and generality (number of prey). By contrast, at the interaction level, predictions can be
89 made on the distribution of interaction strengths in weighted food webs. At the community level, it can generate
90 probability distributions of many measures of their emerging structure and of networks themselves (i.e. a
91 probability distribution that specific network configurations are realized given the model and constraints).

92 Overall, the potential of this method in the study of food webs is broad. The applicability and performance of
93 MaxEnt mostly depend on the ecological information available and on our capacity to find the right set of state
94 variables that best represent natural systems and to translate them into appropriate statistical constraints. Having
95 a validated maximum entropy model for the system at hand allows us to make least-biased predictions using a
96 minimal amount of data, as well as identify the most important ecological processes shaping that system. In
97 other words, MaxEnt can help us better understand and predict the structure of ecological networks worldwide.

98 **Analytical and heuristic approaches**

99 In this contribution, we used two complementary approaches to predict the structure of food webs using the
100 principle of maximum entropy. The first approach consists in deriving constrained probability distributions of
101 given network properties analytically, whereas the second approach consists in finding the adjacency matrix of
102 maximum entropy heuristically, from which network properties can be measured. We compared our predictions
103 against empirical data and null and neutral models commonly used in network ecology. We focus on
104 deterministic and unweighted (Boolean) food webs in both approaches for data availability reasons. However,
105 our framework can be applied to all types of ecological networks and a wide variety of measures.

106 For the first approach (analytic), we focus on species level properties. Specifically, we derived the joint degree
107 distribution (i.e. the joint probability distribution that a species has a given number of prey and predators in its
108 network) of maximum entropy using only the number of species S and the number of interactions L as state
109 variables. Then, we predicted the degree distribution of maximum entropy directly from the joint degree

110 distribution since the first is the sum of the marginal distributions of the second. Because of the scarcity of
111 empirical data on the number of interactions in food webs, we present a method to predict L from S (Box 1),
112 thus allowing the prediction of the joint degree distribution from S solely.

113 For the second approach (heuristic), we focus on network level properties. We used a flexible and heuristic
114 model based on simulated annealing (an optimization algorithm) to find the network configuration *close* to
115 maximum entropy and measured its structure. We developed this heuristic model because the analytical
116 derivation of a maximum entropy graph model of food webs is difficult, and because this model is readily
117 applicable to other types of ecological networks and measures. Indeed, the mathematical representation of food
118 webs (i.e. directed simple graphs frequently having self-loops) makes the optimization of maximum entropy
119 graph models more complicated than with many other types of (non-ecological) networks. In other words,
120 deriving a probability distribution on the graphs themselves is difficult when working with food webs. We built
121 two types of heuristic MaxEnt models depending on the constraint used. Our type I MaxEnt model uses the
122 connectance of the network (i.e. the ratio L/S^2) as a constraint, whereas our type II MaxEnt model uses the
123 whole joint degree sequence as a constraint.

124 Analytical maximum entropy models

125 Our analytical approach is the most common way to use and develop maximum entropy models. As shown
126 above, starting from a defined set of constraints, a mathematical expression for the target distribution is derived
127 using the method of Lagrange multipliers. The derived distribution of maximum entropy is unique and least
128 biased given the constraints used. Although we refer to this approach as analytic, finding the values of the
129 Lagrange multipliers usually requires the use of numerical methods. Here we use MaxEnt to derive two species
130 level properties in food webs: the joint degree distribution and the degree distribution. The degree distribution
131 has driven the attention of ecologists because of its role in determining the assembly of ecological networks
132 (Vázquez 2005), shaping their emerging structure (Fortuna *et al.* 2010), and understanding interaction
133 mechanisms (Williams 2011). As noted above, although the degree distribution of maximum entropy has
134 already been derived in bipartite networks (Williams 2011), we show in much greater details its mathematical
135 derivation in food webs. But first, we derive the joint degree distribution, a related property that holds
136 significantly more ecological information than the degree distribution.

137 We tested our analytical MaxEnt model against open food-web data queried from three different sources and

integrated into what we call our *complete dataset*. These sources include (1) terrestrial and aquatic food webs sampled globally and archived on the ecological interactions database Mangal; (2) freshwater stream food webs from New Zealand; and (3) aquatic food webs from Tuesday Lake (Michigan, United States). First, all food webs archived on `mangal.io` (Poisot *et al.* 2016; Banville *et al.* 2021) were directly queried from the database ($N = 235$). Most ecological networks archived on Mangal are multilayer networks, i.e. networks that describe different types of interactions. We kept all networks whose interactions were mainly of predation and herbivory types, and removed the largest network ($S = 714$) for computational efficiency reasons. Then, to this set we added food webs from two different sources: the New Zealand dataset [$N = 21$; Pomeranz *et al.* (2018)] and the Tuesday Lake dataset [$N = 2$; Cohen *et al.* (2003)]. Our complete dataset thus contained a total of 257 food webs. All code and data to reproduce this article are available at the Open Science Framework (OSF.IO/KT4GS). Data cleaning, simulations and analyses were conducted in Julia v1.8.0.

Joint degree distribution

The joint degree distribution $p(k_{in}, k_{out})$ of a food web with S species is a joint discrete probability distribution describing the probability that a species has k_{in} predators and k_{out} prey, with k_{in} and $k_{out} \in [0, S]$. Basal species (e.g., plants) have a k_{out} of 0, whereas top predators have a k_{in} of 0. In contrast, the maximum number of prey and predators a species can have is set by the number of species in the food web. Here we show how the joint degree distribution of maximum entropy can be obtained given knowledge of the number of species S and the number of interactions L .

We want to maximize Shannon's entropy

$$H = - \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S p(k_{in}, k_{out}) \log p(k_{in}, k_{out}) \quad (3)$$

subject to the following constraints:

$$g_1 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S p(k_{in}, k_{out}) = 1; \quad (4)$$

$$g_2 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{in} p(k_{in}, k_{out}) = \langle k_{in} \rangle = \frac{L}{S}; \quad (5)$$

$$g_3 = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{out} p(k_{in}, k_{out}) = \langle k_{out} \rangle = \frac{L}{S}. \quad (6)$$

- 158 The first constraint g_1 is our normalizing constraint, whereas the other two (g_2 and g_3) fix the average of the
 159 marginal distributions of k_{in} and k_{out} to the linkage density L/S . It is important to notice that $\langle k_{in} \rangle = \langle k_{out} \rangle$
 160 because every edge is associated to a predator and a prey. Therefore, without using any further constraints, we
 161 would expect the joint degree distribution of maximum entropy to be a symmetric probability distribution with
 162 regards to k_{in} and k_{out} . However, this does not mean that the joint degree *sequence* will be symmetric, since the
 163 joint degree sequence is essentially a random realization of its probabilistic counterpart.
- 164 The joint probability distribution of maximum entropy given these constraints is found using the method of
 165 Lagrange multipliers. To do so, we seek to maximize the following expression:

$$F = H - \lambda_1(g_1 - 1) - \lambda_2 \left(g_2 - \frac{L}{S} \right) - \lambda_3 \left(g_3 - \frac{L}{S} \right), \quad (7)$$

- 166 where λ_1 , λ_2 , and λ_3 are the Lagrange multipliers. The probability distribution that maximizes entropy is
 167 obtained by finding these values. As pointed out above, F is just Shannon's entropy to which we added terms
 168 that each sums to zero (our constraints). F is maximized by setting to 0 its partial derivative with respect to
 169 $p(k_{in}, k_{out})$. Because the derivative of a constant is zero, this gives us:

$$\frac{\partial H}{\partial p(k_{in}, k_{out})} = \lambda_1 \frac{\partial g_1}{\partial p(k_{in}, k_{out})} + \lambda_2 \frac{\partial g_2}{\partial p(k_{in}, k_{out})} + \lambda_3 \frac{\partial g_3}{\partial p(k_{in}, k_{out})}. \quad (8)$$

- 170 Evaluating the partial derivatives with respect to $p(k_{in}, k_{out})$, we obtain:

$$-\log p(k_{in}, k_{out}) - 1 = \lambda_1 + \lambda_2 k_{in} + \lambda_3 k_{out}. \quad (9)$$

- 171 Then, solving eq. 9 for $p(k_{in}, k_{out})$, we obtain:

$$p(k_{in}, k_{out}) = \frac{e^{-\lambda_2 k_{in} - \lambda_3 k_{out}}}{Z}, \quad (10)$$

- 172 where $Z = e^{1+\lambda_1}$ is called the partition function. The partition function ensures that probabilities sum to 1 (our
 173 normalization constraint). It can be expressed in terms of λ_2 and λ_3 as follows:

$$Z = \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S e^{-\lambda_2 k_{in} - \lambda_3 k_{out}}. \quad (11)$$

174 After substituting $p(k_{in}, k_{out})$ in eq. 5 and eq. 6, we get a nonlinear system of two equations and two unknowns:

$$\frac{1}{Z} \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{in} e^{-\lambda_2 k_{in} - \lambda_3 k_{out}} = \frac{L}{S}; \quad (12)$$

$$\frac{1}{Z} \sum_{k_{in}=0}^S \sum_{k_{out}=0}^S k_{out} e^{-\lambda_2 k_{in} - \lambda_3 k_{out}} = \frac{S}{L}. \quad (13)$$

175 We solved eq. 12 and eq. 13 numerically using the Julia library JuMP.jl v0.21.8 (Dunning *et al.* 2017).

176 JuMP.jl supports nonlinear optimization problems by providing exact second derivatives that increase the
 177 accuracy and performance of its solvers. The estimated values of λ_2 and λ_3 can be substituted in eq. 10 to have
 178 a more workable expression for the joint degree distribution.

179 We assessed the empirical support of this expression using all food webs in our complete dataset. First, we
 180 predicted the joint degree distribution of maximum entropy for each of these food webs, i.e. using their number
 181 of species and number of interactions as state variables. Then, we sampled one realization of the joint degree
 182 sequence for each network using the probabilities given by the joint degree distribution of maximum entropy,
 183 while fixing the total number of interactions. This gave us a random realization of the number of prey and
 184 predators for each species in each network. We standardized the predicted k_{out} and k_{in} of each species by the
 185 total number of species in their network to generate relative values, which can be compared across networks. In
 186 fig. 1 (left panels), we show the relationship between these relative k_{out} and k_{in} obtained from the joint degree
 187 distributions of maximum entropy (bottom panel) and this relationship using empirical values (top panel). We
 188 observe that our model predicts higher values of generality and vulnerability compared to empirical food webs
 189 (i.e. relative values of k_{out} and k_{in} both closer to 1) for many species. In other words, our model predicts that
 190 species that have many predators also have more prey than what is observed empirically (and conversely). This
 191 is not surprising, given that our model did not include biological factors preventing generalist predators from
 192 having many prey. Nevertheless, with the exception of these generalist species, MaxEnt adequately predicts that
 193 most species have low generality and vulnerability values.

194 Examining the difference between predicted and empirical values for each species gives a slightly different
 195 perspective (right panel of fig. 1). To do so, we must first associate each of our predictions to a specific species

in a network in order to make that comparison. Indeed, our predicted joint degree sequences have the same number of species (elements) as their empirical counterparts, but they are species agnostic. In other words, instead of predicting a value for each species directly, we predicted the entire joint degree sequence without taking into account species' identity. The challenge is thus to adequately associate predictions with empirical data. In fig. 1, we present these differences when species are ordered by their total degree in their respective networks (i.e. by the sum of their in and out-degrees). This means that the species with the highest total degree in its network will be associated with the highest prediction, and so forth. Doing so, we see that species predicted to have a higher number of predators than what is observed generally have a lower number of prey than what is observed (and conversely). This is because the difference in total degree ($k_{out} + k_{in}$) between predictions and empirical data is minimized when species are ranked by their total degree (i.e. the average deviation of the sum of relative k_{out} and k_{in} is close to 0 across all species). This result thus shows that the difference between predicted and empirical total degrees is low for most species when ordered by their total degrees. There are no apparent biases towards in or out degrees. In fig. S1, we show how these differences change when species are instead ordered by their out-degrees (left panel) and in-degrees (right panel), respectively.

[Figure 1 about here.]

Another way to evaluate the empirical support of the predicted joint degree sequences is to compare their shape with the ones of empirical food webs. We can describe the shape of a joint degree sequence by comparing its marginal distributions with one another. To do so, we calculated the Kullback–Leibler (KL) divergence between the in and out-degree sequences sampled from the joint degree distribution of maximum entropy. Similarly, we calculated the divergence between the in and out-degree sequences obtained empirically. This allows us to compare the symmetry of empirical and predicted joint degree sequences (left panel of fig. 2). As we expected, our model predicts more similar in-degree and out-degree sequences than empirical data (shown by lower KL divergence values). However, this difference decreases with connectance (right panel of fig. 2). This might be due to the fact that food webs with a low connectance are harder to predict than food webs with a high connectance. Indeed, in low connectance systems, what makes two species interact might be more important for prediction than in high connectance systems, in which what prevents species from interacting might be more meaningful. This implies that more ecological information might be needed in food webs with a low connectance because more ecological processes determine interactions compared to non-interactions. Therefore, other ecological constraints might be needed to account for the asymmetry of the joint degree

225 distribution, especially for networks with a lower connectance. However, our MaxEnt model was able to capture
226 quite well the shape of the joint degree sequence for networks having a high connectance.

227 [Figure 2 about here.]

228 **Degree distribution**

229 The joint degree distribution derived using MaxEnt (eq. 10) can be used to obtain the degree distribution of
230 maximum entropy. Indeed, the degree distribution $p(k)$ represents the probability that a species has k
231 interactions in its food web, with $k = k_{in} + k_{out}$. It can thus be obtained from the joint degree distribution as
232 follows:

$$p(k) = \sum_{i=0}^k p(k_{in} = k - i, k_{out} = i).$$

233 The degree distribution could have also been obtained directly using the principle of maximum entropy, as
234 discussed in Williams (2011). This gives the following distribution:

$$p(k) = \frac{e^{-\lambda_2 k}}{Z}, \quad (14)$$

235 with $Z = \sum_{k=0}^S e^{-\lambda_2 k}$.

236 This can be solved numerically using the constraint of the average degree $\langle k \rangle = \frac{2L}{S}$ of a species, yielding an
237 identical solution to the one obtained using the joint degree distribution as an intermediate. Note that the mean
238 degree is twice the value of the linkage density, because every link must be counted twice when we add in and
239 out-degrees together.

$$\frac{1}{Z} \sum_{k=0}^S k e^{-\lambda_2 k} = \frac{2L}{S} \quad (15)$$

240 One aspect of the degree distribution that informs us of its ecological realism is the number of isolated species it
241 predicts. As MacDonald *et al.* (2020) pointed out, the size of food webs should at least be of $S - 1$ interactions,
242 since a lower number would yield isolated species, i.e. species without any predators or prey. Because non-basal
243 species must eat to survive, isolated species could indicate that other species are missing or they could simply

244 be removed from the food web. In fig. S2, we show that the degree distribution of maximum entropy, given S
245 and L , gives very low probabilities that a species will be isolated in its food web (i.e. having $k = 0$) above the
246 $S - 1$ threshold. However, under our purely information-theoretic model, the probability that a species is
247 isolated is quite high when the total number of interactions is below $S - 1$. Moreover, the expected proportion of
248 isolated species rapidly declines by orders of magnitude with increasing numbers of species and interactions.
249 This supports the ecological realism of the degree distribution of maximum entropy derived above.
250 Nevertheless, ecologists wanting to model a system without allowing isolated species could simply change the
251 lower limit of k to 1 in eq. 15 and solve the resulting equation numerically.

252 Box 1 - Working with predicted numbers of interactions

253 Our analytical MaxEnt models require information on the number of species and the number of interactions.
254 However, since the later is rarely measured empirically, ecologists might need to use predictive models to
255 estimate the total number of interactions in a food web before using MaxEnt. Here we illustrate how this can be
256 done by combining both models sequentially.
257 We used the flexible links model of MacDonald *et al.* (2020) to predict the number of interactions from the
258 number of species. The flexible links model, in contrast to other predictive models of the number of
259 interactions, incorporates meaningful ecological constraints into the prediction of L , namely the minimum $S - 1$
260 and maximum S^2 numbers of interactions in food webs. It estimates the proportion of the $S^2 - (S - 1)$ *flexible*
261 *links* that are realized. More precisely, this model states that the number of *realized* flexible links (or
262 interactions) L_{FL} in a food web represents the number of realized interactions above the minimum
263 (i.e. $L = L_{FL} + S - 1$) and is obtained from a beta-binomial distribution with $S^2 - (S - 1)$ trials and parameters
264 $\alpha = \mu e^\phi$ and $\beta = (1 - \mu) e^\phi$:

$$L_{FL} \sim \text{BB}(S^2 - (S - 1), \mu e^\phi, (1 - \mu) e^\phi), \quad (16)$$

265 where μ is the average probability across food webs that a flexible link is realized, and ϕ is the concentration
266 parameter around μ .
267 We fitted the flexible links model on all food webs in our complete dataset, and estimated the parameters of
268 eq. 16 using a Hamiltonian Monte Carlo sampler with static trajectory (1 chain and 3000 iterations):

$$[\mu, \phi | \mathbf{L}, \mathbf{S}] \propto \prod_{i=1}^m \text{BB}(L_i - (S_i - 1) | S_i^2 - (S_i - 1)), \mu e^\phi, (1 - \mu) e^\phi) \times \text{B}(\mu | 3, 7) \times \mathcal{N}(\phi | 3, 0.5), \quad (17)$$

269 where m is the number of food webs ($m = 257$) and \mathbf{L} and \mathbf{S} are respectively the vectors of their numbers of
 270 interactions and numbers of species. Our weakly-informative prior distributions were chosen following
 271 MacDonald *et al.* (2020), i.e. a beta distribution for μ and a normal distribution for ϕ . The Monte Carlo
 272 sampling of the posterior distribution was conducted using the Julia library `Turing` v0.15.12.

273 The flexible links model is a generative model, i.e. it can generate plausible values of the predicted variable. We
 274 thus simulated 1000 values of L for different values of S using the joint posterior distribution of our model
 275 parameters (eq. 17), and calculated the mean degree for each simulated value. The resulting distributions are
 276 shown in the left panel of fig. 3 for three different values of species richness. In the right panel of fig. 3, we
 277 show how the probability distribution for the mean degree constraints can be used to generate a distribution of
 278 maximum entropy degree distributions, since each simulated value of mean degree generates a different
 279 maximum entropy degree distribution (eq. 14 and eq. 15).

280 [Figure 3 about here.]

281 **Box 2 - Corresponding null and neutral models**

282 **Null models (types I and II)**

283 The predictions of our heuristic maximum entropy models were compared against two topological null models.
 284 These null models use the same ecological information as our heuristic models and thus constitute an adequate
 285 baseline for comparison. The first is the type I null model of Fortuna & Bascompte (2006), in which the
 286 probability that a species i predares on another species j is given by

$$p(i \rightarrow j) = \frac{L}{S^2}. \quad (18)$$

287 The second is the type II null model of Bascompte *et al.* (2003), in which the probability of interaction is
 288 instead given by

$$p(i \rightarrow j) = \frac{1}{2} \left(\frac{k_{in}(j)}{S} + \frac{k_{out}(i)}{S} \right), \quad (19)$$

where $k_{in}(j)$ and $k_{out}(i)$ are the in and out-degrees of species j and i , respectively. The type I null model is based on connectance, whereas the type II null model is based on the joint degree sequence. Therefore, the type I and II topological null models correspond to our type I and II heuristic MaxEnt models, respectively, since they use similar constraints.

We generated probabilistic networks using both types of null models for all empirical food webs in our complete dataset. Then, we converted these networks to adjacency matrices of Boolean values by generating 100 random networks for each of these probabilistic webs, and kept the L entries that were sampled the most amount of times, with L given by the number of interactions in each food web. This ensured that the resulting null networks had the same number of interactions as their empirical counterparts. Thus, for each null model, we ended up with one null adjacency matrix for each empirical network.

Neutral model

We also compared our heuristic MaxEnt models with a neutral model of relative abundances, in which the probabilities of interaction are given by

$$p(i \rightarrow j) \propto \frac{n_i}{N} \times \frac{n_j}{N}, \quad (20)$$

where n_i and n_j are the abundances (or biomass) of both species, and N is the total abundance (or biomass) of all species in the network. We generated neutral abundance matrices for all empirical food webs in our abundance dataset, and converted these weighted networks to adjacency matrices of Boolean values using the same method as the one we used for our null models.

Heuristic maximum entropy models

With the analytical approach, we showed how important measures of food-web structure (e.g., the degree distribution and the joint degree distribution) can be derived with the principle of maximum entropy using minimal knowledge about a biological community. This type of models, although useful to make least-biased

310 predictions of many network properties, can be hard to apply for other measures. Indeed, there are dozens of
311 measures of network structure (Delmas *et al.* 2019) and many are not directly calculated with mathematical
312 equations, but are instead estimated algorithmically. Moreover, the applicability of this method to empirical
313 systems is limited by the state variables we can actually measure and use. Here, we propose a more flexible
314 method to predict many measures of network structure simultaneously, i.e. by finding heuristically the network
315 configuration having maximum entropy given partial knowledge of its emerging structure.

316 **From Shannon's to SVD entropy**

317 The principle of maximum entropy can be applied on the network itself if we decompose its adjacency matrix
318 into a non-zero vector of relative values. This is a necessary step when working with food webs, which are
319 frequently expressed as a matrix $A = [a_{ij}]$ of Boolean values representing the presence ($a_{ij} = 1$) or absence
320 ($a_{ij} = 0$) of an interaction between two species i and j . Knowing one or many properties of a food web of
321 interest (e.g., its number of species and number of interactions), we can simulate its adjacency matrix randomly
322 by using these known ecological information to constrain the space of potential networks. The entropy of this
323 hypothetical matrix can then be measured after decomposing it into appropriate values. Simulating a series of
324 networks until we find the one having the highest entropy allows us to search for the most complex food-web
325 configuration given the ecological constraints used. This configuration is the least biased one considering the
326 information available. In other words, the most we can say about a network's adjacency matrix, without making
327 more assumptions than the ones given by our incomplete knowledge of its structure, is the one of maximum
328 entropy. Generating the most complex network that corresponds to this structure allows us to explore more
329 easily other properties of food webs under MaxEnt.

330 Shannon's entropy can only be calculated on conventional probability distributions such as the joint degree
331 distribution. This is an issue when working with the adjacency matrix of ecological networks. For this reason,
332 we need to use another measure of entropy if we want to predict a network's configuration directly using
333 MaxEnt. We used the SVD entropy as our measure of entropy, which is an application of Shannon's entropy to
334 the relative non-zero singular values of a truncated singular value decomposition [t-SVD; Strydom *et al.*
335 (2021b)] of a food web's Boolean adjacency matrix. We measured SVD entropy as follows:

$$J = - \sum_{i=1}^R s_i \log s_i, \quad (21)$$

336 where s_i are the relative singular values of the adjacency matrix ($s_i = \sigma_i / \sum_{i=1}^R \sigma_i$, where σ_i are the singular
337 values). Note that the distribution of relative singular values is analogous to a probability distribution, with
338 $0 < s_i < 1$ and $\sum s_i = 1$. This measure also satisfies all four properties of an appropriate entropy measure
339 above-mentioned, while being a proper measure of the internal complexity of food webs (Strydom *et al.* 2021b).
340 Following Strydom *et al.* (2021b), we standardized this measure with the rank R of the matrix (i.e. $J / \ln(R)$) to
341 account for the difference in dimensions between networks [*sensu* Pielou's evenness; Pielou (1975)].

342 **Types I and II heuristic MaxEnt models**

343 We used SVD entropy to predict the network configuration of maximum entropy (i.e. of maximum complexity)
344 heuristically given different constraints for all food webs in our complete dataset. We built two types of heuristic
345 MaxEnt models that differ on the constraint used. The type I heuristic MaxEnt model is based on connectance,
346 whereas the type II heuristic MaxEnt model is based on the joint degree sequence. These models are thus based
347 on the same constraints as the types I (Fortuna & Bascompte 2006) and II (Bascompte *et al.* 2003) null models
348 (Box 2) frequently used to generate random networks topologically. This allows direct comparison of the
349 performance of null and heuristic MaxEnt models in reproducing the emerging structure of empirical food webs.
350 For each network in our complete dataset, we estimated their configuration of maximum entropy given each of
351 these constraints. For both types of heuristic MaxEnt models, we used a simulated annealing algorithm with 4
352 chains, 2000 steps and an initial temperature of 0.2. For each food web, we first generated one random Boolean
353 matrix per chain while fixing the number of species. We also maintained the total number of interactions
354 (i.e. the sum of all elements in the matrix) in the type I MaxEnt model and the joint degree sequence (i.e. the
355 rows and columns sums) in the type II MaxEnt model. These were our initial configurations. Then, we swapped
356 interactions sequentially while maintaining the original connectance or joint degree sequence. Configurations
357 with a higher SVD entropy than the previous one in the chain were always accepted, whereas they were
358 accepted with a probability conditional to a decreasing temperature and the difference in SVD entropy when
359 lower. The final configuration with the highest SVD entropy among the four chains constitute the estimated
360 maximum entropy configuration of a food web given the constraint used.

361 **Structure of MaxEnt food webs**

362 We measured various properties of these configurations of maximum entropy to evaluate how well they
363 reproduce the structure of sampled food webs. Specifically, we evaluated their nestedness ρ , their maximum
364 trophic level $maxtl$, their network diameter $diam$, their average maximum similarity between species pairs
365 $MxSim$ (Williams & Martinez 2000), their proportion of cannibal species *Cannib*, their proportion of
366 omnivorous species *Omniv*, their SVD entropy, and their motifs profile. Nestedness indicates how much the diet
367 of specialist species is a subset of the one of generalists (Delmas *et al.* 2019) and was measured using the
368 spectral radius of the adjacency matrix (Staniczenko *et al.* 2013). In turn, the network diameter represents the
369 longest of the shortest paths between all species pair (Albert & Barabasi 2002). Further, cannibal species are
370 species that can eat individuals of their own species (i.e. species having self loops), whereas omnivorous species
371 can prey on different trophic levels (Williams & Martinez 2000). Finally, a motifs profile represents the
372 proportion of three-species motifs (subnetworks), which can be considered as simple building blocks of
373 ecological networks (Milo *et al.* 2002; Stouffer *et al.* 2007). All of these properties are relatively easy to
374 measure and were chosen based on their ecological importance and prevalent use in network ecology (McCann
375 2011; Delmas *et al.* 2019).

376 We compared the performance of both heuristic MaxEnt models in predicting these measures to the one of the
377 null and neutral models (Box 2). We conducted these comparisons using two different datasets: (1) our
378 complete dataset including most food webs archived on Mangal, as well as all food webs in the New Zealand
379 and Tuesday Lake datasets, and (2) our *abundance dataset*, a subset of the complete dataset comprising all food
380 webs having data on their species' relative abundances ($N = 19$). Indeed, of the New Zealand and Tuesday Lake
381 datasets, 19 networks had data on species' relative abundances that were used in the neutral model to better
382 assess the performance of our heuristic models. We compared our models' predictions using these two datasets
383 separately to minimize biases and to better represent food webs with abundance data (tbl. 1 and tbl. 2).

384 Overall, we found that the models based on the joint degree sequence (i.e. the type II null and heuristic MaxEnt
385 models) reproduced the structure of empirical food webs much better than the ones based on connectance
386 (i.e. the type I null and heuristic MaxEnt models). This suggests that the predictive power of connectance might
387 be more limited than what was previously suggested (Poisot & Gravel 2014). On the other hand, the neutral
388 model of relative abundances was surprisingly good at predicting the maximum trophic level and the network
389 diameter. However, with the exception of the network diameter, the type II heuristic MaxEnt model was better

390 at predicting network structure than the neutral model for most measures considered. This might be because,
 391 although neutral processes are important, they act in concert with niche processes in determining species
 392 interactions Canard *et al.* (2014). The joint degree sequence encodes information on both neutral and niche
 393 processes because the number of prey and predators a species has is determined by its relative abundance and
 394 biological traits. These results thus show that having information on the number of prey and predators for each
 395 species substantially improves the prediction of food-web structure, both compared to models solely based on
 396 connectance and to the ones solely based on species relative abundances.

397 Next, the predictions of the type II heuristic MaxEnt model can be compared to its null model counterpart. On
 398 average, the type II heuristic MaxEnt model was better at predicting nestedness (0.62 ± 0.08) than its
 399 corresponding null model (0.73 ± 0.05 ; empirical networks: 0.63 ± 0.09) for networks in our complete dataset
 400 (tbl. 1). This might in part be due to the fact that nestedness was calculated using the spectral radius of the
 401 adjacency matrix, which directly leverages information on the network itself just like the heuristic MaxEnt
 402 model. The proportion of self-loops (cannibal species) was also better predicted by the type II heuristic MaxEnt
 403 model in comparison to the type II null model. However, the type II null model was better at predicting network
 404 diameter and average maximum similarity between species pairs, and predictions of the maximum trophic level
 405 and the proportion of omnivorous species were similar between both types of models. We believe that this is
 406 because increasing the complexity of a food web might increase its average and maximum food-chain lengths.
 407 In comparison, the null model was more stochastic and does not necessarily produce more complex food webs
 408 with longer food-chain lengths.

Table 1: Standardized mean difference between predicted network measures and empirical data for all food webs in our complete dataset ($N = 257$). Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake food webs. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. ρ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level. *diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy.

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
null 1	-0.167	0.980	1.428	-0.502	2.007	1.493	0.056
MaxEnt 1	-0.226	0.831	1.274	-0.524	1.982	1.863	0.106
null 2	0.160	-0.125	0.016	0.007	1.078	0.559	-0.023

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
MaxEnt 2	-0.015	0.178	0.565	-0.282	0.698	0.589	0.058

Table 2: Standardized mean difference between predicted network measures and empirical data for all food webs in our abundance dataset ($N = 19$). Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Empirical networks include the New Zealand and Tuesday Lake food webs having abundance data. Neutral: Neutral model of relative abundances. Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. ρ : nestedness measured by the spectral radius of the adjacency matrix. *maxtl*: maximum trophic level. *diam*: network diameter. *MxSim*: average maximum similarity between species pairs. *Cannib*: proportion of cannibal species (self loops). *Omniv*: proportion of omnivorous species. *entropy*: SVD entropy.

model	rho	maxtl	diam	MxSim	Cannib	Omniv	entropy
neutral	0.367	-0.090	0.027	0.266	6.870	0.576	-0.083
null 1	-0.134	0.950	1.919	-0.369	2.077	0.614	0.068
MaxEnt 1	-0.229	1.020	1.946	-0.355	2.215	0.801	0.121
null 2	0.128	-0.115	-0.135	0.157	1.444	0.029	-0.021
MaxEnt 2	-0.010	0.054	0.243	-0.062	-0.038	0.083	0.038

Despite this increase in maximum trophic level and network diameter in MaxEnt food webs, we found that empirical food webs are close to their maximum entropy given a fixed joint degree sequence (fig. S3). Empirical food webs in the complete dataset had an SVD entropy of 0.89 ± 0.04 , compared to an SVD entropy of 0.94 ± 0.03 for networks generated using the type II heuristic MaxEnt model. The relationship between the SVD entropy of empirical food webs and their maximum entropy is plotted in the last panel of fig. 4. As expected, the SVD entropy of maximum entropy food webs was higher than that of empirical food webs for almost all networks, confirming that our method indeed generated more complex networks. Moreover, we found no to a weak relationship between the increase in SVD entropy and the number of species, the number of interactions, and connectance (fig. S4). This suggests that the slight increase in entropy between empirical food webs and their maximum entropy configuration was sufficient to modify some of their properties, regardless of their number of species and their number of interactions.

A direct comparison of the structure of maximum entropy food webs, constrained by the joint degree sequence, with empirical data also supports the results depicted in fig. 1. Indeed, in fig. 4 we show how well empirical

measures are predicted by the type II heuristic MaxEnt model. In accordance with our previous results, we found that nestedness was very well predicted by our model. However, the model overestimated the maximum trophic level and network diameter, especially when the sampled food web had intermediate values of these measures. In fig. S5, we show that the pairwise relationships between the four measures in fig. 4 and species richness in empirical food webs are similar (in magnitude and sign) to the ones found in food webs generated using the type II heuristic MaxEnt model. This indicates that the number of species in the network does not seem to impact the ability of the model to reproduce food-web structure.

[Figure 4 about here.]

Notwithstanding its difficulties in reproducing adequate measures of food-chain lengths, the type II heuristic MaxEnt model can predict surprisingly well the motifs profile. Motifs are the backbone of complex ecological networks from which network structure is built upon and play a crucial role in community dynamics and assembly (Stouffer & Bascompte 2011). For this reason, the motifs profile can act as an effective ecological constraint shaping species interactions networks, and thus constitute a substantial source of ecological information. In fig. 5, we show that the motifs profile of networks generated using the type II heuristic MaxEnt model accurately reproduced the one of empirical data. This model made significantly better predictions than the ones based on connectance and the type II null model based on the joint degree sequence. This is also shown in fig. 6, where we see that the relationships between motifs proportions in empirical food webs are very similar to the ones in networks generated using the type II heuristic MaxEnt model. This is in contrast with the type I null and MaxEnt models based on connectance, which produced opposite relationships than what was observed empirically. Our findings thus suggest that increasing food-web complexity within a maximum entropy framework based on the joint degree sequence does not alter the proportion of three-species motifs, but might alter their position relative to one another.

[Figure 5 about here.]

[Figure 6 about here.]

One of the challenges in implementing and validating a maximum entropy model is to discover where its predictions break down. The results depicted in tbl. 1 and fig. 4 show that our type II heuristic MaxEnt model can capture many high-level properties of food webs, but does a poor job of capturing others. This suggests that,

449 although the joint degree sequence is an important driver of food-web structure, other ecological constraints
450 might be needed to account for some emerging food-web properties, especially the ones regarding food-chain
451 lengths. Nevertheless, fig. 5 and fig. 6 show that the model can reproduce surprisingly well the motifs profile,
452 one of the most ecologically informative properties of food webs. This suggests that the emerging structure of
453 food webs is mainly driven by their joint degree sequence, although higher-level properties might be needed to
454 ensure that food-chain lengths fall within realistic values.

455 Conclusion

456 The principle of maximum entropy is a mathematical method of finding least-biased probability distributions
457 that have some specified properties given by prior knowledge about a system. We first applied this conventional
458 MaxEnt approach on food webs to predict species level properties, namely the joint degree distribution and the
459 degree distribution of maximum entropy given known numbers of species and interactions. We found that the
460 joint degree distributions of maximum entropy had a similar shape to the ones of empirical food webs in
461 high-connectance systems. However, these MaxEnt distributions were more symmetric than the ones of
462 empirical food webs when connectance was low, which suggests that other constraints might be needed to
463 improve these predictions in low-connectance systems. Then, we used a slightly different approach that aimed at
464 finding heuristically the network configuration with the highest SVD entropy, i.e. whose vector of relative
465 singular values has maximum entropy. This network of maximum entropy is the most complex, or random,
466 given the specified structure. We found that the heuristic maximum entropy model based on connectance did
467 not predict the structure of sampled food webs very well. However, the heuristic maximum entropy model based
468 on the entire joint degree sequence, i.e. on the number of prey and predators for each species, gave more
469 convincing results. Indeed, this model reproduced food-web structure surprisingly well, including the highly
470 informative motifs profile. Nevertheless, it was not able to predict realistic food-chain lengths.

471 Our results bring to the forefront the role of the joint degree distribution in shaping food-web structure. This
472 echoes the work of Fortuna *et al.* (2010), who found that the degree distribution of ecological networks drive
473 their emerging structure such as their nestedness and modularity. Network ecologists tend to focus on several
474 measures of food webs when studying the ecological consequences of their structure (McCann 2011; Delmas *et*
475 *al.* 2019). In fact, following Williams (2011), we believe there is a lot more ecological information in the
476 deviation between these properties in empirical systems and in their maximum entropy configuration given a

477 fixed joint degree sequence.

478 Alternative MaxEnt models

479 In this contribution, we used a method based on simulated annealing to find the network configuration with the
480 highest SVD entropy while fixing some aspects of its structure. However, there are different ways to generate
481 adjacency matrices using MaxEnt. Another technique, also based on simulated annealing, could begin by
482 generating a food web randomly with fixed numbers of species and interactions and calculating its joint degree
483 distribution. Pairs of interactions could then be swapped sequentially until we minimize the divergence between
484 the calculated joint degree distribution and the one of maximum entropy obtained analytically. In that case, this
485 is the entropy of the joint degree distribution that would be maximized, not the one of the network's topology.
486 To a certain extent, this method would thus bridge the gap between the analytical and heuristic approaches
487 presented in this article. More research is needed to compare the quality of different methods generating
488 adjacency matrices of food webs using MaxEnt.

489 Maximum entropy graph models are another type of methods that predict a distribution of adjacency matrices
490 under soft or hard constraints (e.g., Park & Newman 2004; Cimini *et al.* 2019). Under hard constraints, every
491 network with a non-zero probability exactly satisfies the constraints on its structure. This is in contrast with soft
492 constraints, which require that networks satisfy them on average (i.e. many networks with a non-zero probability
493 do not have the exact structure set by the constraints). Maximum entropy graph models are helpful because they
494 can provide probability distributions for many network properties by measuring the structure of all adjacency
495 matrices with a non-zero probability. However, we consider that our approach based on simulated annealing is
496 more flexible and more computationally efficient. Indeed, many measures of food-web structure are hard to
497 translate into mathematical constraints. Moreover, because food webs are directed networks that can have
498 self-loops, it makes the mathematical derivation of maximum entropy graph models difficult. We believe that
499 identifying heuristically what really constrains the topology of food webs is a useful first step before attempting
500 to derive the mathematical formulation of a maximum entropy graph model for food webs.

501 Applications

502 Our analytical and heuristic models can be applied for different purposes. First, they could be used to generate
503 first-order approximations of a network's properties when state variables are known empirically. For example,

504 knowing the number of species in an ecological community, we can predict its number of interactions using the
505 flexible links model and then predict its joint degree distribution with minimal biases using the principle of
506 maximum entropy. This could prove particularly useful when predicting network structure at large spatial
507 scales, subdividing the study area into smaller communities (e.g., grid cells). Indeed, because species richness
508 and other ecological data are increasingly abundant (e.g., Dickinson *et al.* 2010), validated MaxEnt models can
509 be used to respond to a wider range of macroecological questions regarding food webs.

510 Second, our analytical model can be used to generate informative priors in Bayesian analyses of the structure of
511 ecological networks (e.g., Cirtwill *et al.* 2019). Indeed, the probability distribution of maximum entropy
512 derived using MaxEnt can be used as a prior that can be constantly updated with novel data. For instance, if we
513 know the number of species and the number of interactions, we can get the degree distribution of maximum
514 entropy, as shown in this contribution. The degree distribution represents the probability that a species can
515 interact (as a predator or a prey) with a number of other species. Data on species interactions can be used to
516 update the prior degree distribution to generate a more accurate posterior distribution, thus improving our
517 description and understanding of the system.

518 Third, our analytical and heuristic models can be used to make better predictions of pairwise species interactions
519 by constraining the space of feasible networks, as discussed in Strydom *et al.* (2021a). In other words, we can
520 use the network configuration and/or specific measures of food-web structure derived using MaxEnt to ensure
521 that our predictions of interspecific interactions form feasible networks. This means that the probability that two
522 species interact can be conditional on the network structure and on the probability of interactions of all other
523 species pairs. For that purpose, MaxEnt can be used to predict network structure when other data is lacking.

524 Finally, our analytical and heuristic models can be used as alternative null models of ecological networks to
525 better understand and identify the ecological processes driving food-web structure. Indeed, these mechanisms
526 can be better described when analysing the deviation of empirical data from MaxEnt predictions. A strong
527 deviation would indicate that ecological mechanisms not encoded in the statistical constraints are at play for the
528 system at hand. If deviations are systematic, the maximum entropy model might need to be revised to include
529 appropriate ecological constraints. This revision process helps us reflect on and identify what really constrains
530 food-web structure. However, it is important to note here that tangible ecological mechanisms cannot be directly
531 inferred from statistical distributions (Warren II *et al.* 2022). Instead, by identifying the constraints of a system
532 and by analysing empirical deviations from maximum entropy predictions, MaxEnt can only help us redirect
533 research efforts towards understanding the biological mechanisms behind these constraints.

534 The principle of maximum entropy can thus be applied for both the prediction and understanding of natural
535 systems. Therefore, the model interpretation depends on how we use it. It can be used as a baseline distribution
536 to identify the ecological constraints organizing natural systems. It can also be used as predictions of ecological
537 systems. This distinction between understanding and predicting is essential when using and interpreting
538 MaxEnt models.

539 **Final remarks**

540 One of the biggest challenges in using the principle of maximum entropy is to identify the set of state variables
541 that best reproduce empirical data. We found that the number of species and the number of interactions are
542 important state variables for the prediction of the joint degree distribution. Similarly, we found that the numbers
543 of prey and predators for each species in a food web are important state variables for the prediction of the
544 network configuration. However, our predictions overestimated the symmetry of the joint degree distribution for
545 our analytical model and the maximum trophic level and network diameter for our heuristic model. We should
546 thus continue to play the ecological detective to find these other topological constraints that would improve the
547 predictions of MaxEnt models and help us understand better what really drives food-web structure.

548 **Acknowledgments**

549 We acknowledge that this study was conducted on land within the traditional unceded territory of the Saint
550 Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwininiwak nations. This work was
551 supported by the Institute for Data Valorisation (IVADO) and the Natural Sciences and Engineering Research
552 Council of Canada (NSERC) Collaborative Research and Training Experience (CREATE) program, through the
553 Computational Biodiversity Science and Services (BIOS²) program.

554 **References**

- 555 Albert, R. & Barabasi, A.L. (2002). [Statistical mechanics of complex networks](#). *Reviews of Modern Physics*,
556 74, 47–97.
- 557 Banville, F., Vissault, S. & Poisot, T. (2021). [Mangal.jl and EcologicalNetworks.jl: Two complementary](#)
558 [packages for analyzing ecological networks in Julia](#). *Journal of Open Source Software*, 6, 2721.

- 559 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). [A common framework for identifying linkage rules across different types of interactions](#). *Functional Ecology*,
560 30, 1894–1903.
- 562 Bascompte, J., Jordano, P., Melian, C.J. & Olesen, J.M. (2003). [The nested assembly of plant-animal mutualistic networks](#). *Proceedings of the National Academy of Sciences of the United States of America*,
563 100, 9383–9387.
- 565 Beck, C. (2009). [Generalised information and entropy measures in physics](#). *Contemporary Physics*, 50,
566 495–510.
- 567 Canard, E.F., Mouquet, N., Mouillot, D., Stanko, M., Miklisova, D. & Gravel, D. (2014). [Empirical Evaluation of Neutral Interactions in Host-Parasite Networks](#). *The American Naturalist*, 183, 468–479.
- 569 Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A. & Caldarelli, G. (2019). [The statistical physics of real-world networks](#). *Nature Reviews Physics*, 1, 58–71.
- 571 Cirtwill, A.R., Eklof, A., Roslin, T., Wootton, K. & Gravel, D. (2019). [A quantitative framework for investigating the reliability of empirical network construction](#). *Methods in Ecology and Evolution*, 10,
572 902–911.
- 574 Cohen, J.E., Jonsson, T. & Carpenter, S.R. (2003). [Ecological community description using the food web, species abundance, and body size](#). *Proceedings of the National Academy of Sciences*, 100, 1781–1786.
- 576 Delmas, E., Besson, M., Brice, M.H., Burkle, L.A., Dalla Riva, G.V., Fortin, M.J., et al. (2019). [Analysing ecological networks of species interactions](#). *Biological Reviews*.
- 578 Dickinson, J.L., Zuckerberg, B. & Bonter, D.N. (2010). [Citizen Science as an Ecological Research Tool: Challenges and Benefits](#). In: *Annual Review of Ecology, Evolution, and Systematics*, Vol 41 (eds. Futuyma,
579 D.J., Shafer, H.B. & Simberloff, D.). Annual Reviews, Palo Alto, pp. 149–172.
- 581 Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002). [Network structure and biodiversity loss in food webs: Robustness increases with connectance](#). *Ecology Letters*, 5, 558–567.
- 583 Dunning, I., Huchette, J. & Lubin, M. (2017). [JuMP: A Modeling Language for Mathematical Optimization](#).
584 *SIAM Review*, 59, 295–320.
- 585 Fortuna, M.A. & Bascompte, J. (2006). [Habitat loss and the structure of plant-animal mutualistic networks](#).
586 *Ecology Letters*, 9, 278–283.

- 587 Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., *et al.* (2010). Nestedness
588 versus modularity in ecological networks: Two sides of the same coin? *Journal of Animal Ecology*, 79,
589 811–817.
- 590 Frank, S.A. & Smith, E. (2011). A simple derivation and classification of common probability distributions
591 based on information symmetry and measurement scale. *Journal of Evolutionary Biology*, 24, 469–484.
- 592 Gómez, J.M., Perfectti, F. & Jordano, P. (2011). The Functional Consequences of Mutualistic Network
593 Architecture. *PLOS ONE*, 6, e16143.
- 594 Harremoës, P. & Topsøe, F. (2001). Maximum Entropy Fundamentals. *Entropy*, 3, 191–226.
- 595 Harte, J. & Newman, E.A. (2014). Maximum information entropy: A foundation for ecological theory. *Trends*
596 in *Ecology & Evolution*, 29, 384–389.
- 597 Harte, J., Zillio, T., Conlisk, E. & Smith, A.B. (2008). Maximum Entropy and the State-Variable Approach to
598 Macroecology. *Ecology*, 89, 2700–2711.
- 599 Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. (2015). Seven Shortfalls
600 that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*,
601 46, 523–549.
- 602 Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., *et al.* (2009). Review:
603 Ecological networks beyond food webs. *Journal of Animal Ecology*, 78, 253–269.
- 604 Jaynes, E.T. (1957a). Information Theory and Statistical Mechanics. *Physical Review*, 106, 620–630.
- 605 Jaynes, E.T. (1957b). Information Theory and Statistical Mechanics. II. *Physical Review*, 108, 171–190.
- 606 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.
- 607 Khinchin, A.Y. (2013). *Mathematical Foundations of Information Theory*. Courier Corporation.
- 608 MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in
609 Food Webs. *Patterns*, 0.
- 610 Martyushev, L.M. & Seleznev, V.D. (2006). Maximum entropy production principle in physics, chemistry and
611 biology. *Physics Reports-Review Section of Physics Letters*, 426, 1–45.
- 612 McCann, K.S. (2011). *Food Webs (MPB-50)*. *Food Webs (MPB-50)*. Princeton University Press.

- 613 Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002). [Network motifs: Simple](#)
614 [building blocks of complex networks](#). *Science*, 298, 824–827.
- 615 Park, J. & Newman, M.E.J. (2004). [Statistical mechanics of networks](#). *Physical Review E*, 70, 066117.
- 616 Pascual, D. of E. and E.B.M. & Dunne, V.P.J.A. (2006). *Ecological Networks: Linking Structure to Dynamics*
617 [in Food Webs](#). Oxford University Press, USA.
- 618 Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). [Maximum entropy modeling of species geographic](#)
619 [distributions](#). *Ecological Modelling*, 190, 231–259.
- 620 Pielou, E.C. (1975). Ecological diversity.
- 621 Poisot, T., Baiser, B., Dunne, J.A., Kefi, S., Massol, F., Mouquet, N., *et al.* (2016). [Mangal - making ecological](#)
622 [network analysis simple](#). *Ecography*, 39, 384–390.
- 623 Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). [Global knowledge](#)
624 [gaps in species interaction networks data](#). *Journal of Biogeography*, 48, 1552–1563.
- 625 Poisot, T. & Gravel, D. (2014). [When is an ecological network complex? Connectance drives degree](#)
626 [distribution and emerging network properties](#). *PeerJ*, 2, e251.
- 627 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). [Beyond species: Why ecological interaction networks vary](#)
628 [through space and time](#). *Oikos*, 124, 243–251.
- 629 Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2019). [Inferring predatorprey interactions in](#)
630 [food webs](#). *Methods in Ecology and Evolution*, 10, 356–367.
- 631 Pomeranz, J.P., Thompson, R.M., Poisot, T., Harding, J.S. & Pomeranz, J.P.F. (2018). [Data from: Inferring](#)
632 [predator-prey interactions in food webs](#).
- 633 Shannon, C.E. (1948). [A mathematical theory of communication](#). *The Bell System Technical Journal*, 27,
634 379–423.
- 635 Staniczenko, P.P.A., Kopp, J.C. & Allesina, S. (2013). [The ghost of nestedness in ecological networks](#). *Nature*
636 [Communications](#), 4, 1391.
- 637 Stock, M., Poisot, T. & De Baets, B. (2021). [Optimal transportation theory for species interaction networks](#).
638 [Ecology and Evolution](#), 11, 3841–3855.

- 639 Stouffer, D.B. & Bascompte, J. (2011). [Compartmentalization increases food-web persistence](#). *Proceedings of
640 the National Academy of Sciences*, 108, 3648–3652.
- 641 Stouffer, D.B., Camacho, J., Jiang, W. & Nunes Amaral, L.A. (2007). [Evidence for the existence of a robust
642 pattern of prey selection in food webs](#). *Proceedings of the Royal Society B-Biological Sciences*, 274,
643 1931–1940.
- 644 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021a). [A
645 roadmap towards predicting species interaction networks \(across space and time\)](#). *Philosophical
646 Transactions of the Royal Society B: Biological Sciences*, 376, 20210063.
- 647 Strydom, T., Dalla Riva, G.V. & Poisot, T. (2021b). [SVD Entropy Reveals the High Complexity of Ecological
648 Networks](#). *Frontiers in Ecology and Evolution*, 9, 623141.
- 649 van der Hoorn, P., Lippner, G. & Krioukov, D. (2018). [Sparse Maximum-Entropy Random Graphs with a Given
650 Power-Law Degree Distribution](#). *Journal of Statistical Physics*, 173, 806–844.
- 651 Vázquez, D.P. (2005). [Degree distribution in plant-animal mutualistic networks: Forbidden links or random
652 interactions?](#) *Oikos*, 108, 421–426.
- 653 Warren II, R.J., Costa, J.T. & Bradford, M.A. (2022). [Seeing shapes in clouds: The fallacy of deriving
654 ecological hypotheses from statistical distributions](#). *Oikos*, 2022, e09315.
- 655 Williams, R.J. (2011). [Biology, Methodology or Chance? The Degree Distributions of Bipartite Ecological
656 Networks](#). *PLOS ONE*, 6, e17645.
- 657 Williams, R.J. & Martinez, N.D. (2000). [Simple rules yield complex food webs](#). *Nature*, 404, 180–183.

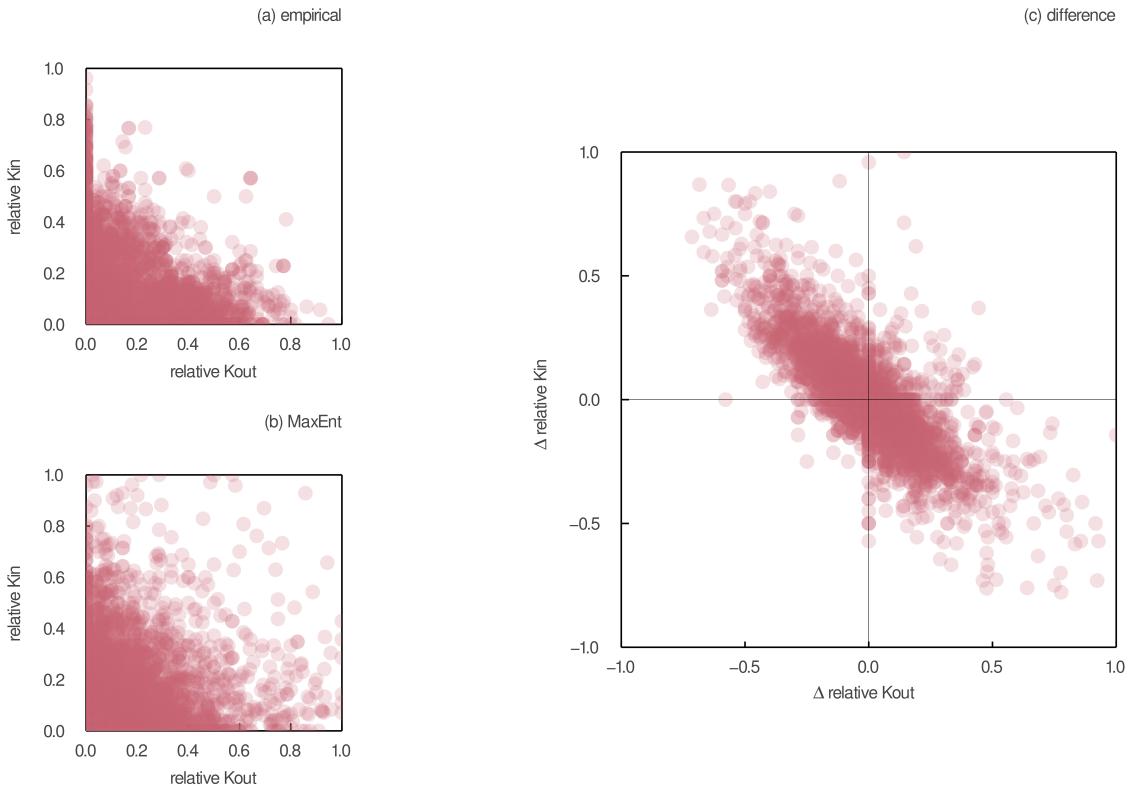


Figure 1: Relative number of predators (k_{in}) as a function of their relative number of prey (k_{out}) for each species in (a) empirical and (b) joint degree sequences obtained from the analytical MaxEnt model. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network, while keeping the total number of interactions constant. (c) Difference between predicted and empirical values when species are ordered according to their total degrees. In each panel, each dot corresponds to a single species in one of the networks.

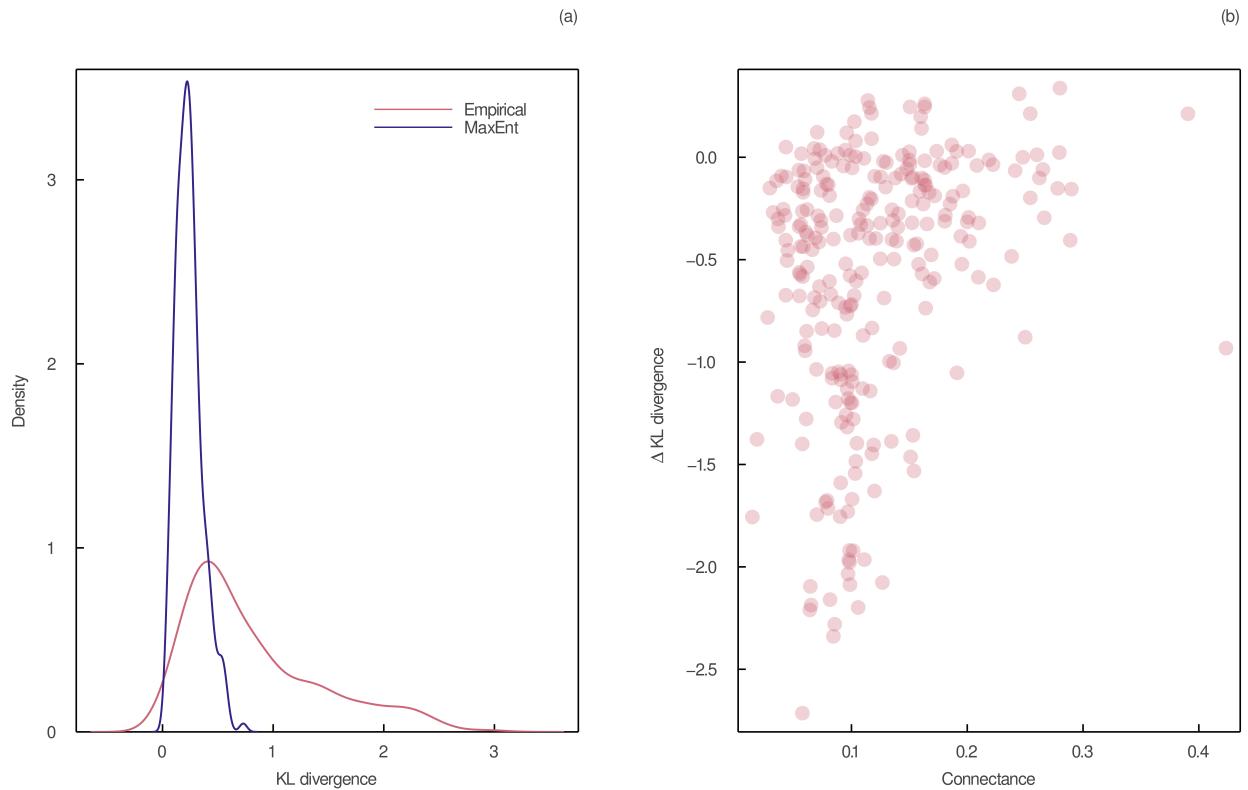


Figure 2: (a) Probability density of KL divergence between in and out-degree sequences of empirical and predicted joint degree sequences. (b) Difference between the KL divergence of empirical and predicted joint degree sequences as a function of connectance. In both panels, empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). The predicted joint degree sequences were obtained after sampling one realization of the joint degree distribution of maximum entropy for each network, while keeping the total number of interactions constant.

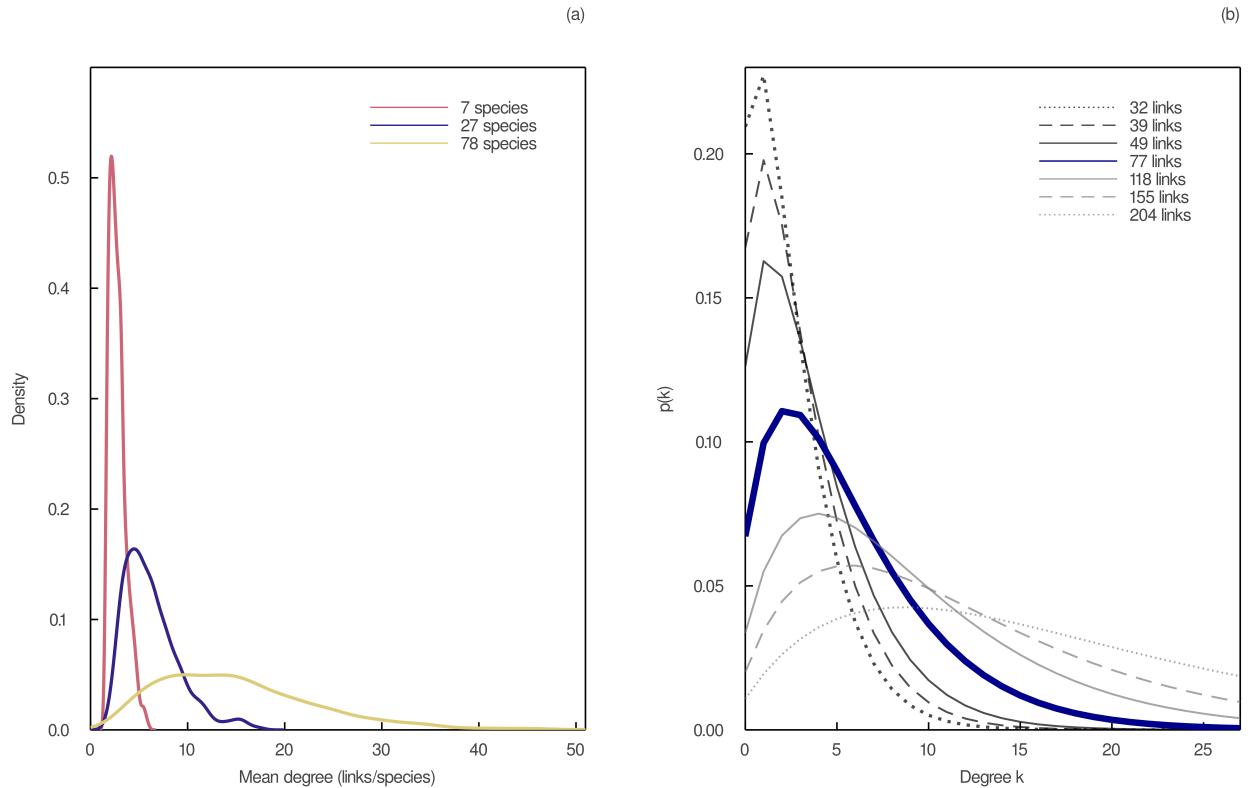


Figure 3: (a) Probability density of the mean degree of a food web obtained using different values of species richness S . The number of interactions L was simulated 1000 times using the flexible links model fitted to all empirical networks in our complete dataset (eq. 17). The mean degrees $2L/S$ were then obtained from these simulated values. (b) Degree distributions of maximum entropy for a network of $S = 27$ species and different numbers of interactions. The numbers of interactions correspond to the lower and upper bounds of the 67, 89, and 97 percentile intervals (PI), as well as the median (in blue), of the counterfactuals of the flexible links model. Each degree distribution of maximum entropy was obtained using eq. 14 after solving numerically eq. 15 using different values of the mean degree constraint.

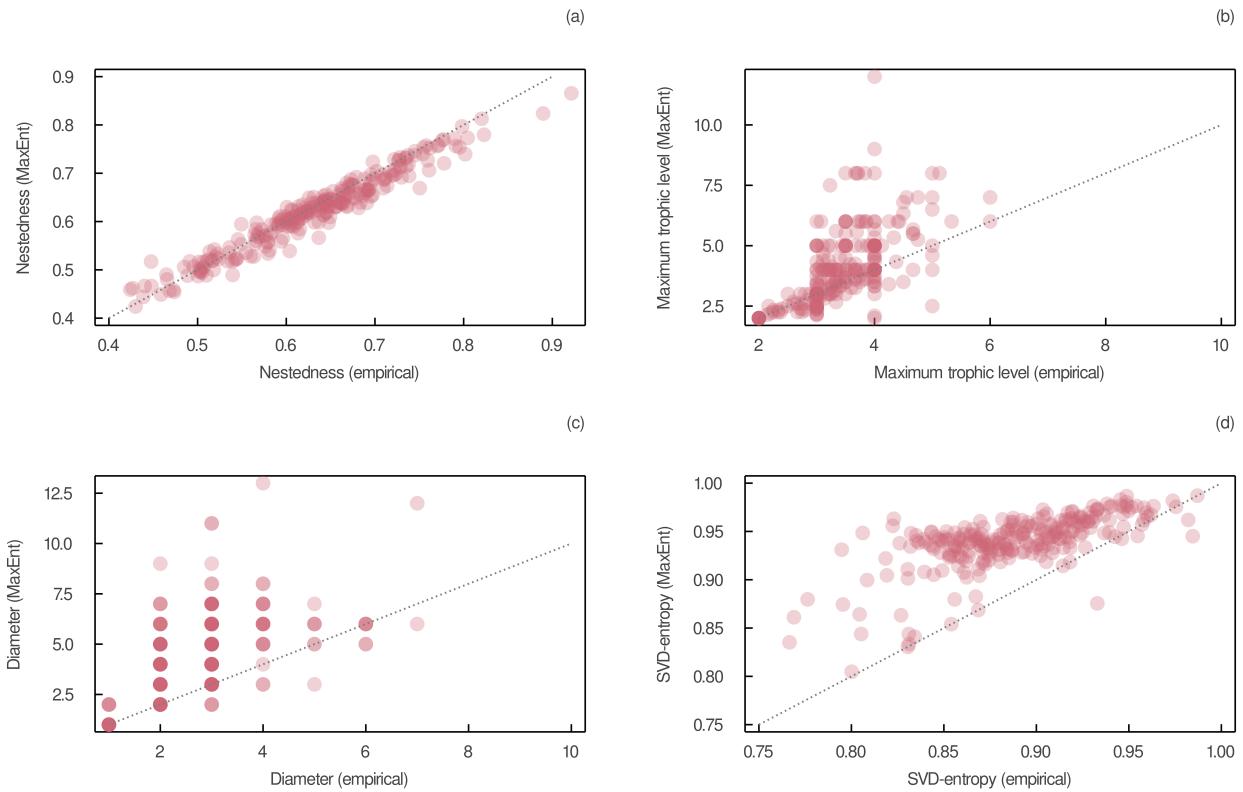


Figure 4: Relationship between the structure of empirical and maximum entropy food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake food webs (our complete dataset). Maximum entropy networks were obtained using the type II heuristic MaxEnt model based on the joint degree sequence. (a) Nestedness (estimated using the spectral radius of the adjacency matrix), (b) the maximum trophic level, (c) the network diameter, and (d) the SVD entropy were measured on these empirical and maximum entropy food webs. The identity line is plotted in each panel.

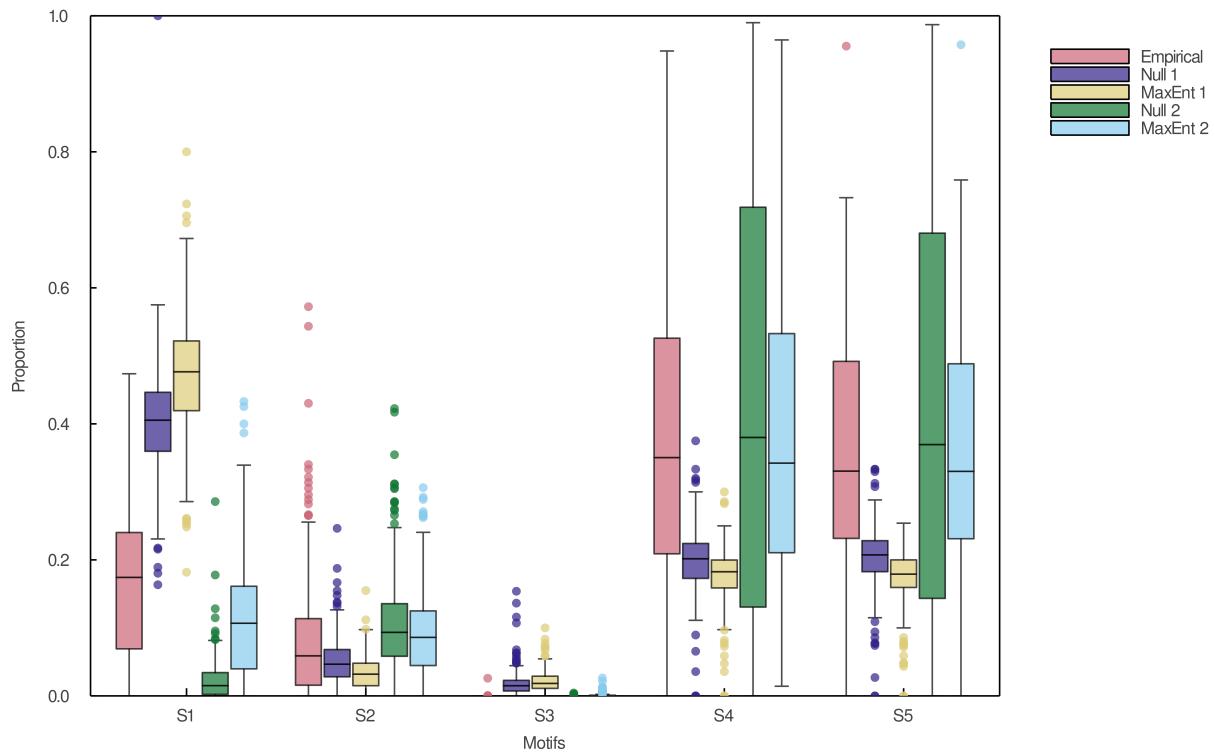


Figure 5: Motifs profile of empirical and predicted food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. Boxplots display the median proportion of each motif (middle horizontal lines), as well as the first (bottom horizontal lines) and third (top horizontal lines) quartiles. Vertical lines encompass all data points that fall within 1.5 times the interquartile range from both quartiles, and dots are data points that fall outside this range. Motifs names are from Stouffer *et al.* (2007).

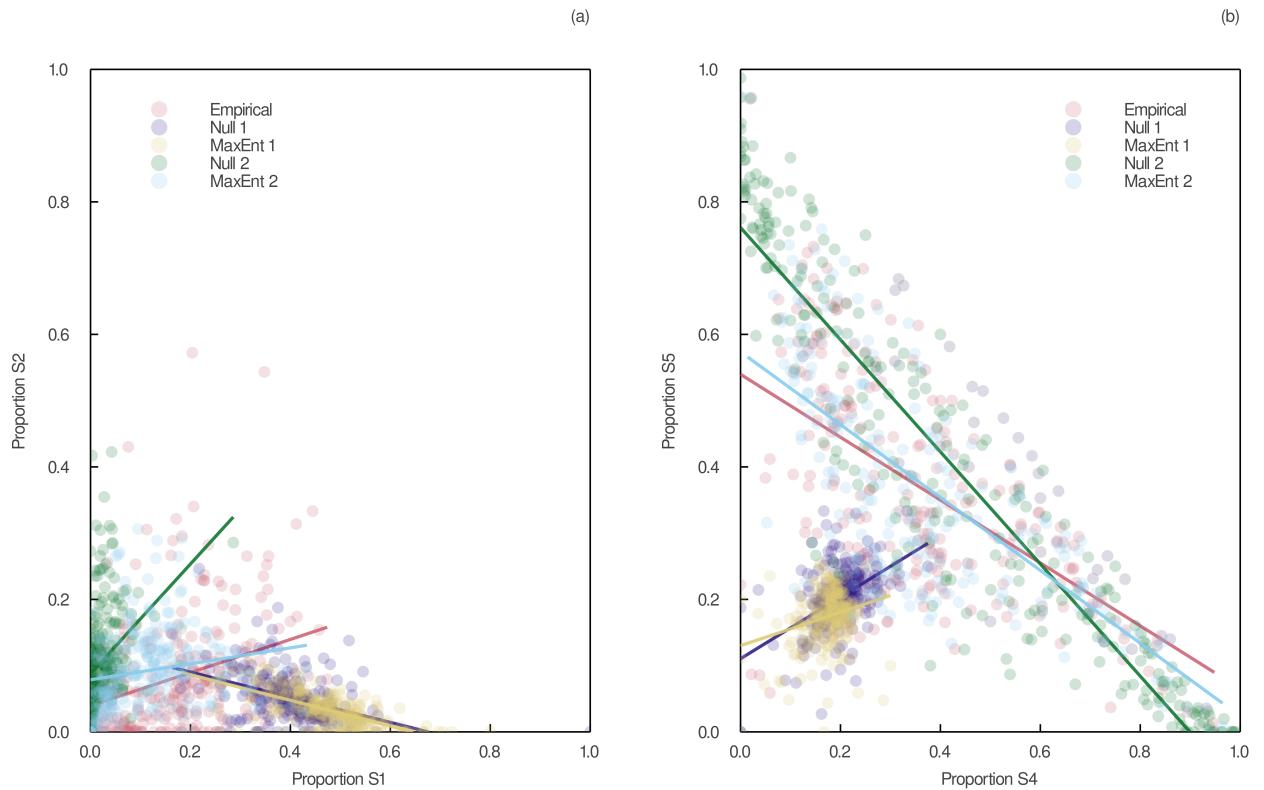


Figure 6: Pairwise relationships between motif proportions of empirical and predicted food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Null 1: Type I null model based on connectance. MaxEnt 1: Type I heuristic MaxEnt model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II heuristic MaxEnt model based on the joint degree sequence. Regression lines are plotted in each panel. Motifs names are from Stouffer *et al.* (2007).