

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

Francis Banville<sup>1,2,3</sup> Dominique Gravel<sup>2,3</sup> Timothée Poisot<sup>1,3</sup>

<sup>1</sup> Université de Montréal <sup>2</sup> Université de Sherbrooke <sup>3</sup> Quebec Centre for Biodiversity Science

## **Correspondance to:**

Francis Banville — francis.banville@umontreal.ca

**Abstract:** 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 renowned and rigorous 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.

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

18 One way to tackle this issue is first to recognize that food webs and other ecological networks  
19 are constrained systems. In other words, the space of possible network configurations shrinks  
20 as we know more about a network structure. For example, there is a finite number of networks  
21 with specified numbers of nodes and edges. Indeed, the structure of ecological networks is first  
22 and above all constrained by the number of species, or nodes, present. Food webs with high  
23 species richness typically have a lower connectance (MacDonald, Banville, and Poisot 2020)  
24 than smaller networks. This is because the number of realized interactions in empirical food  
25 webs scales slower than the number of possible species pairs (MacDonald, Banville, and Poisot  
26 2020). As shown by Timothée Poisot and Gravel (2014), connectance itself can constrain differ-  
27 ent aspects of network structure such as the degree distribution (i.e. the probability distribution  
28 of the number of interspecific interactions realized by a species). Other measures, such as the

29 maximum trophic level, can also constrain the space of feasible networks.

30 Prior knowledge on the structure of ecological networks is thus especially useful in the current  
31 context of data scarcity about species interactions (Jordano 2016). Indeed, this Eltonian shortfall  
32 (Hortal et al. 2015) can be partially alleviated using known information about an ecological  
33 network. As suggested by Strydom et al. (2021), network structure can be used to improve the  
34 prediction of pairwise species interactions when data is lacking by constraining the space of  
35 feasible networks. Similarly, partial knowledge on the structure of an ecological network can  
36 also be used to predict others of its properties by constraining their range of possible values.  
37 This is important given that many aspects of network structure cannot be measured empirically  
38 without data on pairwise species interactions, a prevailing situation worldwide (Timothée Poisot  
39 et al. 2021).

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

#### 54 **The principle of maximum entropy: A primer for ecologists**

55 The principle of maximum entropy (MaxEnt) is a mathematical method of finding probabil-  
56 ity distributions, strongly rooted in statistical mechanics and information theory (Jaynes 1957a,

57 1957b; Harremoës and Topsøe 2001). Starting from a set of constraints given by prior knowl-  
58 edge of a system (i.e. what we call state variables), this method helps us find least-biased prob-  
59 ability distributions subject to the constraints. These probability distributions are guaranteed to  
60 be unique given our prior knowledge and represent the most we can say about a system without  
61 making more assumptions. For example, if the only thing we know about a biological commu-  
62 nity is its average number of individuals per species, the least-biased inference we could make on  
63 its species abundance distribution is the exponential distribution (Frank and Smith 2011; Harte  
64 and Newman 2014). However, this does not imply that this distribution will be the best fit to  
65 empirical data. The challenge is to find the right set of constraints that would best reproduce  
66 distributions found in nature.

67 MaxEnt states that the least-biased probability distribution given the constraints used is the one  
68 with the highest entropy among all probability distributions that satisfy these constraints. En-  
69 tropy is a measure of the average amount of information given by the outcome of a random  
70 variable. Many measures of entropy have been developed in physics (Beck 2009), but only  
71 a fraction of them could be used as an optimization measure with the principle of maximum  
72 entropy. According to Beck (2009) and Khinchin (2013), a measure of entropy  $H$  should sat-  
73 isfy four properties in the discrete case: (1) it should be a function of a probability distribution  
74  $p(n)$  only; (2) it should be maximized when  $p(n)$  is uniform; (3) it should not be influenced by  
75 outcomes with a null probability; and (4) it should be independent of the order of information  
76 acquisition. The Shannon's entropy (Shannon 1948)

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

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

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

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

86 The principle of maximum entropy has been used in a wide range of disciplines, from thermo-  
 87 dynamics, chemistry and biology (Martyushev and Seleznev 2006) to graph and network theory  
 88 (e.g., Park and Newman 2004; van der Hoorn, Lippner, and Krioukov 2018). It has also been  
 89 proven useful in ecology, e.g. in species distribution models (Phillips, Anderson, and Schapire  
 90 2006) and macroecological models (Harte et al. 2008; Harte and Newman 2014). In network  
 91 ecology, MaxEnt has been used to predict the degree distribution of bipartite networks from the  
 92 number of species and the number of interactions (Williams 2011) and to predict interaction  
 93 strengths between species pairs using their relative abundances within an optimal transportation  
 94 theory regularized with information entropy (Stock, Poisot, and De Baets 2021). However, to  
 95 the best of our knowledge, MaxEnt has never been used to predict food-web structure directly,  
 96 even though food webs are among the most documented and widespread ecological networks  
 97 (Ings et al. 2009).

98 Food-web properties that can be derived using MaxEnt are varied and pertain to different ele-  
 99 ments of the network (i.e. at the species (node), the interaction (edge) or the community (net-  
 100 work) levels). Because MaxEnt is a method of finding least-biased probability distributions  
 101 given partial knowledge about a system, these properties need to be represented probabilistically.  
 102 For example, at the species level, MaxEnt can be used to predict the distribution of trophic levels  
 103 among species, as well as the distribution of species' vulnerability (number of predators) and  
 104 generality (number of preys). By contrast, at the interaction level, predictions can be made on  
 105 the distribution of interaction strengths in weighted food webs. At the community level, it can  
 106 generate probability distributions of many measures of their emerging structure and of networks  
 107 themselves (i.e. a probability distribution that specific network configurations are realized given  
 108 the model and constraints). Overall, the potential of this method in the study of food webs is  
 109 broad. The applicability and performance of MaxEnt mostly depend on the ecological infor-

110 mation available and on our capacity to find the right set of state variables that best represent  
111 natural systems and to translate them into appropriate statistical constraints. Having a validated  
112 maximum entropy model for the system at hand allows us to make least-biased predictions using  
113 a minimal amount of data, as well as identify the most important ecological processes shaping  
114 that system. In other words, MaxEnt is a valuable and underexploited tool in network ecology  
115 that allows us to better understand and predict the structure of ecological networks.

## 116 Analytical and heuristic approaches

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

125 For the first approach (analytic), we focus on species level properties. Specifically, we derived  
126 the joint degree distribution (i.e. the joint probability distribution that a species has a given  
127 number of prey and predators in its network) of maximum entropy using only the number of  
128 species  $S$  and the number of interactions  $L$  as state variables. Then, we predicted the degree  
129 distribution of maximum entropy directly from the joint degree distribution since the first is the  
130 sum of the marginal distributions of the second. Because of the scarcity of empirical data on  
131 the number of interactions in food webs, we present a method to predict  $L$  from  $S$  (Box 1), thus  
132 allowing the prediction of the joint degree distribution from  $S$  solely.

133 For the second approach (heuristic), we focus on network level properties. We used a flexible and  
134 heuristic model based on simulated annealing (an optimization algorithm) to find the network  
135 configuration *close* to maximum entropy and measured its structure. We developed this heuristic  
136 model because the analytical derivation of a maximum entropy graph model of food webs is  
137 difficult, and because this model is readily applicable to other types of ecological networks and

measures. Indeed, the mathematical representation of food webs (i.e. directed simple graphs frequently having self-loops) makes the optimization of maximum entropy graph models more complicated than with many other types of (non-ecological) networks. In other words, deriving a probability distribution on the graphs themselves is difficult when working with food webs.

We built two types of heuristic MaxEnt models depending on the constraint used. Our type I MaxEnt model uses the connectance of the network (i.e. the ratio  $L/S^2$ ) as a constraint, whereas our type II MaxEnt model uses the whole joint degree sequence as a constraint.

## 145 Data and code

We tested our analytical and heuristic MaxEnt models 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` (Timothée Poisot et al. 2016; Banville, Vissault, and Poisot 2021) were directly queried from the database ( $N = 235$ ). Most ecological networks archived on Mangal are multi-layer 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, Jonsson, and Carpenter 2003). Our complete dataset thus contained a total of 257 food webs. Of the New Zealand and Tuesday Lake datasets, 19 networks had data on species' relative abundances that were used in the neutral model (Box 2). These 19 food webs constitute our *abundance dataset*, which is a subset of our complete dataset.

All code and data to reproduce this article are available at the Open Science Framework. Data cleaning, simulations and analyses were conducted in Julia v1.6.1.

163 **Analytical maximum entropy models**

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

178 **Joint degree distribution**

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

186 We want to maximize Shannon's entropy

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

<sup>187</sup> subject to the following constraints:

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

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

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

<sup>188</sup> The first constraint  $g_1$  is our normalizing constraint, whereas the other two ( $g_2$  and  $g_3$ ) fix the  
<sup>189</sup> average of the marginal distributions of  $k_{in}$  and  $k_{out}$  to the linkage density  $L/S$ . It is important  
<sup>190</sup> to notice that  $\langle k_{in} \rangle = \langle k_{out} \rangle$  because every edge is associated to a predator and a prey. Therefore,  
<sup>191</sup> without using any further constraints, we would expect the joint degree distribution of maximum  
<sup>192</sup> entropy to be a symmetric probability distribution with regards to  $k_{in}$  and  $k_{out}$ . However, this  
<sup>193</sup> does not mean that the joint degree *sequence* will be symmetric, since the joint degree sequence  
<sup>194</sup> is essentially a random realization of its probabilistic counterpart.

<sup>195</sup> The joint probability distribution of maximum entropy given these constraints is found using the  
<sup>196</sup> method of Lagrange multipliers. To do so, we seek to maximize the following expression:

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

<sup>197</sup> where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are the Lagrange multipliers. The probability distribution that maximizes  
<sup>198</sup> entropy is obtained by finding these values. As pointed out above,  $F$  is just Shannon's entropy  
<sup>199</sup> to which we added terms that each sums to zero (our constraints).  $F$  is maximized by setting to  
<sup>200</sup> 0 its partial derivative with respect to  $p(k_{in}, k_{out})$ . Because the derivative of a constant is zero,  
<sup>201</sup> this gives us:

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

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

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

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

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

204 where  $Z = e^{1+\lambda_1}$  is called the partition function. The partition function ensures that probabilities  
205 sum to 1 (our normalization constraint). It can be expressed in terms of  $\lambda_2$  and  $\lambda_3$  as follows:

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

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

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

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

208 We solved eq. 12 and eq. 13 numerically using the Julia library JuMP.jl v0.21.8 (Dunning,  
209 Huchette, and Lubin 2017). JuMP.jl supports nonlinear optimization problems by providing  
210 exact second derivatives that increase the accuracy and performance of its solvers. The estimated  
211 values of  $\lambda_2$  and  $\lambda_3$  can be substituted in eq. 10 to have a more workable expression for the joint  
212 degree distribution.

213 We assessed the empirical support of this expression using all food webs in our complete dataset.  
214 First, we predicted the joint degree distribution of maximum entropy for each of these food webs,  
215 i.e. using their number of species and number of interactions as state variables. Then, we sam-  
216 pled one realization of the joint degree sequence for each network using the probabilities given

217 by the joint degree distribution of maximum entropy, while fixing the total number of interac-  
218 tions. This gave us a random realization of the number of prey and predators for each species in  
219 each network. We standardized the predicted  $k_{out}$  and  $k_{in}$  of each species by the total number of  
220 species in their network to generate relative values, which can be compared across networks. In  
221 fig. 1 (left panels), we show the relationship between these relative  $k_{out}$  and  $k_{in}$  obtained from the  
222 joint degree distributions of maximum entropy (bottom panel) and this relationship using em-  
223 pirical values (top panel). We observe that our model predicts higher values of generality and  
224 vulnerability compared to empirical food webs (i.e. relative values of  $k_{out}$  and  $k_{in}$  both closer to  
225 1) for many species. In other words, our model predicts that species that have many predators  
226 also have more prey than what is observed empirically (and conversely). This is not surprising,  
227 given that our model did not include biological factors preventing generalist predators from hav-  
228 ing many prey. Nevertheless, with the exception of these generalist species, MaxEnt adequately  
229 predicts that most species have low generality and vulnerability values.

230 Examining the difference between predicted and empirical values for each species gives a slightly  
231 different perspective (right panel of fig. 1). To do so, we must first associate each of our predic-  
232 tions to a specific species in a network in order to make that comparison. Indeed, our predicted  
233 joint degree sequences have the same number of species (elements) as their empirical counter-  
234 parts, but they are species agnostic. In other words, instead of predicting a value for each species  
235 directly, we predicted the entire joint degree sequence without taking into account species' iden-  
236 tity. The challenge is thus to adequately associate predictions with empirical data. In fig. 1, we  
237 present these differences when species are ordered by their total degree in their respective net-  
238 works (i.e. by the sum of their in and out-degrees). This means that the species with the highest  
239 total degree in its network will be associated with the highest prediction, and so forth. Doing so,  
240 we see that species predicted to have a higher number of predators than what is observed gen-  
241 erally have a lower number of prey than what is observed (and conversely). This is because the  
242 difference in total degree ( $k_{out} + k_{in}$ ) between predictions and empirical data is minimized when  
243 species are ranked by their total degree (i.e. the average deviation of the sum of relative  $k_{out}$  and  
244  $k_{in}$  is close to 0 across all species). This result thus shows that the difference between predicted  
245 and empirical total degrees is low for most species when ordered by their total degrees. There

246 are no apparent biases towards in or out degrees. In fig. S1, we show how these differences  
247 change when species are instead ordered by their out-degrees (left panel) and in-degrees (right  
248 panel), respectively.

249 [Figure 1 about here.]

250 Another way to evaluate the empirical support of the predicted joint degree sequences is to com-  
251 pare their shape with the ones of empirical food webs. We can describe the shape of a joint degree  
252 sequence by comparing its marginal distributions with one another. To do so, we calculated the  
253 Kullback–Leibler (KL) divergence between the in and out-degree sequences sampled from the  
254 joint degree distribution of maximum entropy. Similarly, we calculated the divergence between  
255 the in and out-degree sequences obtained empirically. This allows us to compare the symmetry  
256 of empirical and predicted joint degree sequences (left panel of fig. 2). As we expected, our  
257 model predicts more similar in-degree and out-degree sequences than empirical data (shown  
258 by lower KL divergence values). However, this difference decreases with connectance (right  
259 panel of fig. 2). This might be due to the fact that food webs with a low connectance are harder  
260 to predict than food webs with a high connectance. Indeed, in low connectance systems, what  
261 makes two species interact might be more important for prediction than in high connectance  
262 systems, in which what prevents species from interacting might be more meaningful. This im-  
263 plies that more ecological information might be needed in food webs with a low connectance  
264 because more ecological processes determine interactions compared to non-interactions. There-  
265 fore, other ecological constraints might be needed to account for the asymmetry of the joint  
266 degree distribution, especially for networks with a lower connectance. However, our MaxEnt  
267 model was able to capture quite well the shape of the joint degree sequence for networks having  
268 a high connectance.

269 [Figure 2 about here.]

## 270 Degree distribution

271 The degree distribution  $p(k)$  represents the probability that a species has  $k$  interactions in a food  
272 web, with  $k = k_{in} + k_{out}$ . It can thus be directly obtained from the joint degree distribution:

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

273 In fig. S2, we show that the degree distribution of maximum entropy, given  $S$  and  $L$ , predicts  
 274 very low probabilities that a species will be isolated in its food web (i.e. having  $k = 0$ ). As  
 275 MacDonald, Banville, and Poisot (2020) pointed out, the size of food webs should at least be  
 276 of  $S - 1$  interactions, since a lower number would yield isolated species, i.e. species without  
 277 any predators or preys. Our results show that, under our purely information-theoretic model,  
 278 the probability that a species is isolated is quite high when the total number of interactions is  
 279 below this threshold. The expected proportion of isolated species rapidly declines by orders of  
 280 magnitude with increasing numbers of species and interactions.

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

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

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

284 This can be solved numerically using the constraint of the average degree  $\langle k \rangle = \frac{2L}{S}$  of a species.  
 285 Note that the mean degree is twice the value of the linkage density, because every link must be  
 286 counted twice when we add in and out-degrees together.

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

287 The numerical solution is identical to the one we obtained using the joint degree distribution  
 288 as an intermediate. Ecologists wanting to model a system without considering isolated species  
 289 could simply change the lower limit of  $k$  to 1 and solve the resulting equation numerically.

290 In this section, we showed how important measures of food-web structure, namely the degree  
 291 distribution and the joint degree distribution, could be derived with the principle of maximum  
 292 entropy using minimal knowledge on a biological community. This type of models, although

useful to make least-biased predictions on many network properties, can be hard to apply for other measures. Indeed, there are dozens of measures of network structure (Delmas et al. 2019) and many are not calculated with mathematical equations, but with algorithms. Moreover, the applicability of this method to empirical systems is limited by the state variables we can actually measure and use. In the next section, we propose a more flexible method to predict many measures of network structure simultaneously, i.e. by finding heuristically the configuration of the network having maximum entropy given partial knowledge of its emerging structure.

## Box 1 - Working with predicted numbers of interactions

Our models need information on the number of species and the number of interactions. However, since the later is rarely estimated empirically, ecologists might need to use predictive methods to estimate the total number of interactions in a food web.

We used the flexible links model of MacDonald, Banville, and Poisot (2020) to predict the number of interactions from the number of species. The flexible links model, in contrast to other predictive models of the number of interactions, incorporates meaningful ecological constraints into the prediction of  $L$ , namely the minimum  $S - 1$  and maximum  $S^2$  numbers of interactions in food webs. It estimates the proportion of the  $S^2 - (S - 1)$  *flexible links* that are realized. More precisely, this model states that the number of *realized* flexible links (or interactions)  $L_{FL}$  in a food web represents the number of realized interactions above the minimum (i.e.  $L = L_{FL} + S - 1$ ) and is obtained from a beta-binomial distribution with  $S^2 - (S - 1)$  trials and parameters  $\alpha = \mu e^\phi$  and  $\beta = (1 - \mu)e^\phi$ :

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

where  $\mu$  is the average probability across food webs that a flexible link is realized, and  $\phi$  is the concentration parameter around  $\mu$ .

We fitted the flexible links model on all food webs in our complete dataset, and estimated the parameters of eq. 16 using a Hamiltonian Monte Carlo sampler with static trajectory (1 chain

317 and 3000 iterations):

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

318 where  $m$  is the number of food webs ( $m = 257$ ) and  $\mathbf{L}$  and  $\mathbf{S}$  are respectively the vectors of their  
319 numbers of interactions and numbers of species. Our weakly-informative prior distributions  
320 were chosen following MacDonald, Banville, and Poisot (2020), i.e. a beta distribution for  $\mu$   
321 and a normal distribution for  $\phi$ . The Monte Carlo sampling of the posterior distribution was  
322 conducted using the Julia library Turing v0.15.12.

323 The flexible links model is a generative model, i.e. it can generate plausible values of the pre-  
324 dicted variable. We thus simulated 1000 values of  $L$  for different values of  $S$  using the joint  
325 posterior distribution of our model parameters (eq. 17), and calculated the mean degree for  
326 each simulated values. The resulting distributions are shown in the left panel of fig. 3 for three  
327 different values of species richness. In the right panel of fig. 3, we show how the probability  
328 distribution for the mean degree constraints can be used to generate a distribution of maximum  
329 entropy degree distributions, since each simulated value of mean degree generates a different  
330 maximum entropy degree distribution (eq. 14 and eq. 15).

331 [Figure 3 about here.]

## 332 Box 2 - Corresponding null and neutral models

### 333 Null models (types I and II)

334 Our heuristic MaxEnt models, predicting adjacency matrices, were compared with two topolog-  
335 ical null models. The first is the type I null model of Fortuna and Bascompte (2006), in which  
336 the probability that a species  $i$  predares on another species  $j$  is given by

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

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

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

339 where  $k_{in}$  and  $k_{out}$  are the in and out-degrees, respectively. The type I null model is based on  
340 connectance, whereas the type II null model is based on the joint degree sequence. Therefore,  
341 the type I and II topological null models correspond with our type I and II MaxEnt network  
342 models, respectively, since they use similar constraints.

343 We predicted both types of null networks for all empirical networks in our complete dataset  
344 ( $n = 257$ ). We converted all probabilistic networks to Boolean networks by generating 100  
345 random Boolean networks for each of these probabilistic webs. Then, we counted the number  
346 of times each interaction was sampled, and kept the  $L$  entries that were drawn the most amount  
347 of time, with  $L$  given by the number of interactions in each food web. This ensured that the  
348 resulting null networks had the same number of interactions as their empirical counterparts.

### 349 Neutral model

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

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

352 where  $n_i$  and  $n_j$  are the abundances (or biomass) of both species, and  $N$  is the total abundance  
353 (or biomass) of all species in the network. We predicted neutral abundance matrices for all  
354 empirical networks in our abundance dataset ( $n = 19$ ), and converted these weighted matrices  
355 to Boolean networks using an approach analogue to the one we used for our null models.

356 **Heuristic maximum entropy models**

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

358 The principle of maximum entropy can be applied on the network itself if we decompose its  
359 adjacency matrix into a non-zero vector of relative values. This is a necessary step when work-  
360 ing with food webs, which are frequently expressed as a matrix  $A = [a_{ij}]$  of Boolean values  
361 representing the presence ( $a_{ij} = 1$ ) or absence ( $a_{ij} = 0$ ) of an interaction between two species  $i$   
362 and  $j$ . Knowing one or many properties of a food web of interest (e.g., its number of species and  
363 number of interactions), we can simulate its adjacency matrix randomly by using these known  
364 ecological information to constrain the space of potential networks. The entropy of this hypo-  
365 thetical matrix can then be measured after decomposing it into appropriate values. Simulating  
366 a series of networks until we find the one having the highest entropy allows us to search for the  
367 most complex food-web configuration given the ecological constraints used. This configuration  
368 is the least biased one considering the information we had. In other words, the most we can  
369 say about a network's adjacency matrix, without making more assumptions than the ones given  
370 by our incomplete knowledge of its structure, is the one of maximum entropy. Generating the  
371 most complex network that corresponds to this structure allows us to explore more easily other  
372 properties of food webs under MaxEnt.

373 Shannon's entropy can only be calculated on conventional probability distributions such as the  
374 joint degree distribution. This is an issue when working with ecological networks, which are  
375 represented as adjacency matrices. For this reason, we need to use another measure of entropy if  
376 we want to predict the network itself. We used the SVD entropy as our measure of entropy, which  
377 is an application of Shannon's entropy to the relative non-zero singular values of a truncated  
378 singular value decomposition (t-SVD; Strydom, Dalla Riva, and Poisot 2021) of a food web's  
379 Boolean adjacency matrix. We measured SVD entropy as follows:

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

380 where  $s_i$  are the relative singular values of the adjacency matrix ( $s_i = \sigma_i / \sum_{i=1}^R \sigma_i$ , where  $\sigma_i$

381 are the singular values). Note that the distribution of relative singular values is analogous to  
382 a probability distribution, with  $0 < s_i < 1$  and  $\sum s_i = 1$ . Thus, this measure also satisfies  
383 all four properties of an appropriate entropy measure (see introduction), while being a proper  
384 measure of the internal complexity of food webs (Strydom, Dalla Riva, and Poisot 2021). Fol-  
385 lowing Strydom, Dalla Riva, and Poisot (2021), we standardized this measure with the rank  $R$  of  
386 the matrix (i.e.  $J / \ln(R)$ ) to account for the difference in dimensions between networks (*sensu*  
387 Pielou's evenness; Pielou 1975). We used SVD entropy to predict the network configuration  
388 of maximum entropy (i.e. of maximum complexity) heuristically given a predetermined set of  
389 constraints.

### 390 MaxEnt network models (types I and II)

391 We built two types of MaxEnt network models: one based on connectance (type I MaxEnt net-  
392 work model) and the other based on the joint degree sequence (type II MaxEnt network model).  
393 They are based on the same constraints as the types I and II null models presented above. For  
394 both models, we used a simulated annealing algorithm with 4 chains, 2000 steps and an initial  
395 temperature of 0.2. For each chain, we first generated one random Boolean matrix with the same  
396 order (number of species) as empirical webs, while maintaining the total number of interactions  
397 (type I MaxEnt network model) or rows and columns sums (type II MaxEnt network model).  
398 These are our initial configurations. Then, we swapped interactions sequentially while main-  
399 taining the original connectance or the joint degree sequence for types I and II MaxEnt network  
400 model, respectively. Configurations with a higher SVD entropy than the previous one in the  
401 chain were always accepted, whereas they were accepted with a probability conditional to a de-  
402 creasing temperature when lower. The final configuration with the highest SVD entropy among  
403 the four chains constitute our estimated MaxEnt network. Even though we decided to work with  
404 point estimates, it is possible to have a (non MaxEnt) probability distribution of networks when  
405 working with the entire chains (or part of the chains). For each network in our complete and  
406 abundance datasets, we estimated their configuration with maximum entropy using both types  
407 of MaxEnt network models.

408 **Structure of MaxEnt networks**

409 We compared different measures of the structure of empirical food webs with the ones of null,  
410 neutral, and MaxEnt networks [tbl. 1](#), and [2](#). For instance, we measured the nestedness  $\rho$  accord-  
411 ing to Staniczenko, Kopp, and Allesina (2013), i.e. using the spectral radius of the adjacency  
412 matrix. Nestedness indicates how much the diet of specialist species is a subset of the one of  
413 generalists, and is strongly associated with network modularity (Fortuna et al. 2010). We also  
414 measured the maximum trophic level *maxtl*, network diameter *diam* (i.e. the longest of the short-  
415 est paths between all species pairs; Albert and Barabasi 2002), the average maximum similarity  
416 between species pairs *MxSim*, the proportion of cannibal species *Cannib* (i.e. the proportion of  
417 self loops), and the proportion of omnivorous species *Omniv* (i.e. species whose preys are of  
418 different trophic levels). *MxSim*, *Cannib*, and *Omniv* are more deeply defined in Williams and  
419 Martinez (2000).

420 Overall, we found that models based on the joint degree sequence (type II null and MaxEnt  
421 network models) reproduced the structure of empirical networks much better than the ones based  
422 on connectance and the neutral model of relative abundances. On average, the type II MaxEnt  
423 network model was better at predicting nestedness ( $0.62 \pm 0.08$ ) than its corresponding null  
424 model ( $0.73 \pm 0.05$ ; empirical networks:  $0.63 \pm 0.09$ ), as well as the proportion of cannibal  
425 species. This might in part be due because nestedness was calculated using the spectral radius  
426 of the adjacency matrix, which leverages information on the network itself just like the MaxEnt  
427 network models. However, the type II null model was better at predicting network diameter and  
428 average maximum similarity between species pairs. Predictions were similar between both type  
429 II models for the maximum trophic level and the proportion of omnivorous species.

Table 1: Standardized mean differences of predicted network measures with empirical data for all networks ( $n = 257$ ). Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Empirical networks are most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake food webs (our complete dataset). Null 1: Type I null model based on connectance. MaxEnt 1: Type I MaxEnt network model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II MaxEnt network model based on the joint degree sequence.  $\rho$ : 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
MaxEnt 2	-0.015	0.178	0.565	-0.282	0.698	0.589	0.058

Table 2: Standardized mean differences of predicted network measures with empirical data for networks having abundance data ( $n = 19$ ). Positive (negative) values indicate that the measure is overestimated (underestimated) on average. Empirical networks are all New Zealand and Tuesday Lake food webs with abundance data (our abundance dataset). Null 1: Type I null model based on connectance. MaxEnt 1: Type I MaxEnt network model based on connectance. Null 2: Type II null model based on the joint degree sequence. MaxEnt 2: Type II MaxEnt network model based on the joint degree sequence.  $\rho$ : 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

430 The picture slightly changes when we consider another important property of ecological net-  
 431 works, i.e. their motifs profile. We measured the proportion of three-species motifs, which can

432 be considered as simple building blocks of ecological networks (Milo et al. 2002; Stouffer et al.  
433 2007). Motifs are the backbone of complex ecological networks from which network structure  
434 is built upon and play a crucial role in community dynamics and assembly (Bastolla et al. 2009;  
435 Stouffer and Bascompte 2011). In fig. 4, we show that the motifs profile of networks generated  
436 using the type II MaxEnt network model was very close to the one of empirical data, and that  
437 they made better predictions than the type II null model. This is also shown in fig. 5, where we  
438 see that relationships between motifs pairs of empirical food webs are very similar to the ones  
439 of the type II MaxEnt networks. Overall, these results suggest that the type II MaxEnt network  
440 model can reproduce many aspects of network structure. This highlights the importance of the  
441 joint degree sequence in shaping ecological networks.

442 [Figure 4 about here.]

443 [Figure 5 about here.]

444 Moreover, we found that empirical networks are close to their maximum entropy considering a  
445 fixed joint degree sequence (fig. S3). Empirical networks had an SVD entropy of  $0.89 \pm 0.04$ ,  
446 compared to type II MaxEnt networks which had values of  $0.94 \pm 0.03$ . The relationship between  
447 the SVD entropy of empirical food webs and their maximum entropy configuration is plotted in  
448 the last panel of fig. 6. We also found no relationship between the difference in SVD entropy  
449 between empirical and type II MaxEnt networks and species richness, the number of interactions,  
450 or connectance (fig. S4). Similarly, we found no correlation between the difference in SVD  
451 entropy and the Jaccard distance of empirical and type II MaxEnt networks.

452 Finally, in fig. 6 we also show how well empirical measures are predicted using the type II  
453 MaxEnt network model. In accordance with our previous results, we found that nestedness was  
454 very well predicted by our model. However, the type II MaxEnt network model overestimated  
455 the maximum trophic level and network diameter, especially for networks with high empirical  
456 values. In fig. S5, we show that the pairwise relationships between these four measures and  
457 species richness in empirical food webs are similar (in magnitude and sign) to the ones in type  
458 II MaxEnt networks.

460 **Discussion**

461 The principle of maximum entropy is a robust mathematical method of finding least-biased  
462 probability distributions that have some desired properties given by prior knowledge on a sys-  
463 tem. We first used MaxEnt to predict the joint degree distribution and the degree distribution  
464 of maximum entropy given known numbers of species and of interactions. We found that the  
465 resulting joint degree distributions were more symmetric than the ones of empirical food webs,  
466 which suggests that other constraints might be needed to improve those predictions. We also  
467 used MaxEnt to predict networks of maximum entropy with a specified structure. These net-  
468 works are the most complex, or random, given the constraints used. Likewise, we found that  
469 knowledge of species richness and of the number of interactions were not sufficient to reproduce  
470 many aspects of network structure found in nature. However, a model based on the entire joint  
471 degree sequence, i.e. on the number of preys and predators for each species, gave more con-  
472 vincing results. Our type II MaxEnt network model yielded better or similar predictions than  
473 the type II null model, also based on the joint degree sequence, for most measures considered  
474 including the motifs profile. This suggests that the joint degree sequence drives many aspects of  
475 network structure. Indeed, considering our findings that empirical networks are close to max-  
476 imum entropy for a given joint degree sequence, our results suggest that food-web topology is  
477 configured almost entirely randomly around these marginal numbers of predators and preys.

478 Our results bring to the forefront the role of the joint degree distribution in shaping food-web  
479 structure. This echoes the work of Williams (2011), which shows using MaxEnt how the degree  
480 distribution can drive the structure of bipartite networks. Fortuna et al. (2010) also found that  
481 the degree distribution of ecological networks is associated with some of their emerging prop-  
482 erties such as nestedness and modularity. Network ecologists tend to focus on several emerging  
483 properties of food webs (Delmas et al. 2019), although we show that food-web structure is very  
484 well predicted by the joint degree distribution. In the light of our results, there is a lot more  
485 ecological information in the difference between properties of empirical networks and MaxEnt

486 predictions than in these properties themselves.

487 **Alternative MaxEnt models**

488 There are different ways to generate food webs using MaxEnt. In this contribution, we used a  
489 method based on simulated annealing to find the network configuration with the highest SVD  
490 entropy while fixing some aspects of its structure. Another technique, also based on simulated  
491 annealing, could start by generating a food web randomly with fixed numbers of species and  
492 interactions and calculating its joint degree distribution. Pairs of interactions could be swapped  
493 sequentially until we minimize the divergence between the calculated joint degree distribution  
494 and the one of maximum entropy obtained analytically. In that case, this is the entropy of the  
495 joint degree distribution that would be maximized, not the one of the network's topology. To a  
496 certain extent, this method would thus bridge the gap between the two approaches presented in  
497 this article (i.e. analytical and heuristic MaxEnt models). More research is needed to compare  
498 the quality of different methods generating adjacency matrices of food webs using MaxEnt.

500 Another method generating adjacency matrices are maximum entropy graph models, which pre-  
501 dict a probability distribution of networks under soft or hard constraints (e.g., Park and Newman  
502 2004; Cimini et al. 2019). Under hard constraints, every network with a non-zero probability  
503 exactly satisfies the constraints on its structure. This is in contrast with soft constraints, which  
504 require that networks satisfy them on average (i.e. many networks with a non-zero probability do  
505 not have the exact structure set by the constraints). Maximum entropy graph models are helpful  
506 because they can provide probability distributions for many network properties, by measuring  
507 the structure of all adjacency matrices with a non-zero probability. However, we consider that  
508 our approach based on simulated annealing is more flexible and more computationally efficient.  
509 Indeed, many measures of food-web structure are hard to translate into mathematical constraints.  
510 Moreover, because food webs are directed networks with self-loops, it makes the mathematical  
511 derivation of maximum entropy graph models difficult. We believe that identifying heuristically  
512 what really constrains the topology of food webs is a useful first step before attempting to derive  
the mathematical formulation of a maximum entropy graph model adapted to food webs.

513 **Applications**

514 Our method and results could be used for different purposes. First, they could be used as first-  
515 order approximations of network structure when only state variables are known. This could  
516 prove useful when predicting network structure at large spatial scales, where few ecological in-  
517 formation is known at that scale. Second, they could be used as informative priors in Bayesian  
518 analyses of the structure of ecological networks (e.g., Cirtwill et al. 2019). Third, they could  
519 be used to make better predictions of pairwise species interactions by constraining the space of  
520 feasible networks, as discussed in Strydom et al. (2021). Finally, they could be used as alter-  
521 native null models of ecological networks to better understand ecological mechanisms driving  
522 food-web structure. In that case, our model might need to be slightly adapted to give a proba-  
523 bility distribution of Boolean networks (in contrast with point estimates of maximum entropy  
524 networks).

525 **Conclusion**

526 One of the biggest challenges in using the principle of maximum entropy is to identify the set  
527 of state variables that best reproduce empirical data. We found that the numbers of preys and  
528 predators for each species are important state variables for the prediction of maximum entropy  
529 networks. However, our predictions overestimated some measures of network structure, espe-  
530 cially the maximum trophic level and network diameter. Therefore, we should continue playing  
531 the ecological detective to find these other topological constraints that would improve the pre-  
532 dictions of our MaxEnt network models.

533 **Acknowledgments**

534 We acknowledge that this study was conducted on land within the traditional unceded territory  
535 of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and Omàmiwinini-  
536 wak nations. This work was supported by the Institute for Data Valorisation (IVADO) and the  
537 Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Re-

538 search and Training Experience (CREATE) program, through the Computational Biodiversity  
539 Science and Services (BIOS<sup>2</sup>) program.

540 **References**

- 541 Albert, R., and A. L. Barabasi. 2002. “Statistical Mechanics of Complex Networks.” *Reviews*  
542 *of Modern Physics* 74 (1): 47–97. <https://doi.org/10.1103/RevModPhys.74.47>.
- 543 Banville, Francis, Steve Vissault, and Timothée Poisot. 2021. “Mangal.jl and EcologicalNet-  
544 works.jl: Two Complementary Packages for Analyzing Ecological Networks in Julia.” *Journal*  
545 *of Open Source Software* 6 (61): 2721. <https://doi.org/10.21105/joss.02721>.
- 546 Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. “The Nested Assembly of  
547 Plant-Animal Mutualistic Networks.” *Proceedings of the National Academy of Sciences*  
548 *of the United States of America* 100 (16): 9383–87. <https://doi.org/10.1073/pnas.1633576100>.
- 549
- 550 Bastolla, Ugo, Miguel A. Fortuna, Alberto Pascual-Garcia, Antonio Ferrera, Bartolo Luque, and  
551 Jordi Bascompte. 2009. “The Architecture of Mutualistic Networks Minimizes Competition  
552 and Increases Biodiversity.” *Nature* 458 (7241): 1018–U91. <https://doi.org/10.1038/nature07950>.
- 553
- 554 Beck, Christian. 2009. “Generalised Information and Entropy Measures in Physics.” *Contem-  
555 porary Physics* 50 (4): 495–510. <https://doi.org/10.1080/00107510902823517>.
- 556 Cimini, Giulio, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and  
557 Guido Caldarelli. 2019. “The Statistical Physics of Real-World Networks.” *Nature Reviews*  
558 *Physics* 1 (1): 58–71. <https://doi.org/10.1038/s42254-018-0002-6>.
- 559 Cirtwill, Alyssa R., Anna Eklof, Tomas Roslin, Kate Wootton, and Dominique Gravel. 2019.  
560 “A Quantitative Framework for Investigating the Reliability of Empirical Network Construc-  
561 tion.” *Methods in Ecology and Evolution* 10 (6): 902–11. <https://doi.org/10.1111/2041-210X.13180>.
- 562

- 563 Cohen, Joel E., Tomas Jonsson, and Stephen R. Carpenter. 2003. "Ecological Community  
564 Description Using the Food Web, Species Abundance, and Body Size." *Proceedings of  
565 the National Academy of Sciences* 100 (4): 1781–86. [https://doi.org/10.1073/pnas.  
566 232715699](https://doi.org/10.1073/pnas.232715699).
- 567 Delmas, Eva, Mathilde Besson, Marie Hélène Brice, Laura A. Burkle, Giulio V. Dalla Riva,  
568 Marie Josée Fortin, Dominique Gravel, et al. 2019. "Analysing Ecological Networks of  
569 Species Interactions." *Biological Reviews*. <https://doi.org/10.1111/brv.12433>.
- 570 Dunne, Jennifer A., Richard J. Williams, and Neo D. Martinez. 2002. "Network Structure and  
571 Biodiversity Loss in Food Webs: Robustness Increases with Connectance." *Ecology Letters*  
572 5 (4): 558–67. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>.
- 573 Dunning, Iain, Joey Huchette, and Miles Lubin. 2017. "JuMP: A Modeling Language for  
574 Mathematical Optimization." *SIAM Review* 59 (2): 295–320. [https://doi.org/10.1137/  
575 15M1020575](https://doi.org/10.1137/15M1020575).
- 576 Fortuna, M. A., and J. Bascompte. 2006. "Habitat Loss and the Structure of Plant-Animal  
577 Mutualistic Networks." *Ecology Letters* 9 (3): 278–83. [https://doi.org/10.1111/j.  
578 1461-0248.2005.00868.x](https://doi.org/10.1111/j.1461-0248.2005.00868.x).
- 579 Fortuna, M. A., Daniel B. Stouffer, Jens M. Olesen, Pedro Jordano, David Mouillot, Boris R.  
580 Krasnov, Robert Poulin, and Jordi Bascompte. 2010. "Nestedness Versus Modularity in  
581 Ecological Networks: Two Sides of the Same Coin?" *Journal of Animal Ecology* 79 (4):  
582 811–17. <https://doi.org/10.1111/j.1365-2656.2010.01688.x>.
- 583 Frank, S. A., and E. Smith. 2011. "A Simple Derivation and Classification of Common Prob-  
584 ability Distributions Based on Information Symmetry and Measurement Scale." *Journal of  
585 Evolutionary Biology* 24 (3): 469–84. [https://doi.org/10.1111/j.1420-9101.2010.  
586 02204.x](https://doi.org/10.1111/j.1420-9101.2010.02204.x).
- 587 Gómez, José M., Francisco Perfectti, and Pedro Jordano. 2011. "The Functional Consequences  
588 of Mutualistic Network Architecture." *PLOS ONE* 6 (1): e16143. [https://doi.org/10.  
589 1371/journal.pone.0016143](https://doi.org/10.1371/journal.pone.0016143).

- 590 Harremoës, Peter, and Flemming Topsøe. 2001. “Maximum Entropy Fundamentals.” *Entropy*  
591 3 (3): 191–226. <https://doi.org/10.3390/e3030191>.
- 592 Harte, J., and Erica A. Newman. 2014. “Maximum Information Entropy: A Foundation for  
593 Ecological Theory.” *Trends in Ecology & Evolution* 29 (7): 384–89. <https://doi.org/10.1016/j.tree.2014.04.009>.
- 595 Harte, J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. “Maximum Entropy and the State-  
596 Variable Approach to Macroecology.” *Ecology* 89 (10): 2700–2711. <https://doi.org/10.1890/07-1369.1>.
- 598 Hoorn, Pim van der, Gabor Lippner, and Dmitri Krioukov. 2018. “Sparse Maximum-Entropy  
599 Random Graphs with a Given Power-Law Degree Distribution.” *Journal of Statistical Physics*  
600 173 (3): 806–44. <https://doi.org/10.1007/s10955-017-1887-7>.
- 601 Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M. Lewinsohn,  
602 Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls That Beset Large-Scale  
603 Knowledge of Biodiversity.” *Annual Review of Ecology, Evolution, and Systematics* 46 (1):  
604 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- 605 Ings, Thomas C., José M. Montoya, Jordi Bascompte, Nico Blüthgen, Lee Brown, Carsten F.  
606 Dormann, François Edwards, et al. 2009. “Review: Ecological Networks Beyond Food  
607 Webs.” *Journal of Animal Ecology* 78 (1): 253–69. <https://doi.org/10.1111/j.1365-2656.2008.01460.x>.
- 609 Jaynes, E. T. 1957a. “Information Theory and Statistical Mechanics.” *Physical Review* 106 (4):  
610 620–30. <https://doi.org/10.1103/PhysRev.106.620>.
- 611 ———. 1957b. “Information Theory and Statistical Mechanics. II.” *Physical Review* 108 (2):  
612 171–90. <https://doi.org/10.1103/PhysRev.108.171>.
- 613 Jordano, Pedro. 2016. “Sampling Networks of Ecological Interactions.” *Functional Ecology* 30  
614 (12): 1883–93. <https://doi.org/10.1111/1365-2435.12763>.
- 615 Khinchin, A. Ya. 2013. *Mathematical Foundations of Information Theory*. Courier Corpora-  
616 tion.

- 617 MacDonald, Arthur Andrew Meahan, Francis Banville, and Timothée Poisot. 2020. “Revisiting  
618 the Links-Species Scaling Relationship in Food Webs.” *Patterns* 0 (0). <https://doi.org/10.1016/j.patter.2020.100079>.
- 619
- 620 Martyushev, L. M., and V. D. Seleznev. 2006. “Maximum Entropy Production Principle in  
621 Physics, Chemistry and Biology.” *Physics Reports-Review Section of Physics Letters* 426  
622 (1): 1–45. <https://doi.org/10.1016/j.physrep.2005.12.001>.
- 623 Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. “Net-  
624 work Motifs: Simple Building Blocks of Complex Networks.” *Science* 298 (5594): 824–27.  
625 <https://doi.org/10.1126/science.298.5594.824>.
- 626 Park, Juyong, and M. E. J. Newman. 2004. “Statistical Mechanics of Networks.” *Physical  
627 Review E* 70 (6): 066117. <https://doi.org/10.1103/PhysRevE.70.066117>.
- 628 Pascual, Department of Ecology and Evolutionary Biology Mercedes, and Visiting Professor  
629 Jennifer A. Dunne. 2006. *Ecological Networks: Linking Structure to Dynamics in Food  
630 Webs*. Oxford University Press, USA.
- 631 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maximum Entropy  
632 Modeling of Species Geographic Distributions.” *Ecological Modelling* 190 (3): 231–59.  
633 <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- 634 Pielou, Evelyn C. 1975. “Ecological Diversity.” In. 574.524018 P5.
- 635 Poisot, Timothee, Benjamin Baiser, Jennifer A. Dunne, Sonia Kefi, Francois Massol, Nico-  
636 las Mouquet, Tamara N. Romanuk, Daniel B. Stouffer, Spencer A. Wood, and Dominique  
637 Gravel. 2016. “Mangal - Making Ecological Network Analysis Simple.” *Ecography* 39 (4):  
638 384–90. <https://doi.org/10.1111/ecog.00976>.
- 639 Poisot, Timothée, Gabriel Bergeron, Kevin Cazelles, Tad Dallas, Dominique Gravel, Andrew  
640 MacDonald, Benjamin Mercier, Clément Violet, and Steve Vissault. 2021. “Global Knowl-  
641 edge Gaps in Species Interaction Networks Data.” *Journal of Biogeography* 48 (7): 1552–  
642 63. <https://doi.org/10.1111/jbi.14127>.
- 643 Poisot, Timothée, and Dominique Gravel. 2014. “When Is an Ecological Network Complex?  
644 Connectance Drives Degree Distribution and Emerging Network Properties.” *PeerJ* 2: e251.

- 645 <https://doi.org/10.7717/peerj.251>.
- 646 Pomeranz, Justin Page, Ross M. Thompson, Timothée Poisot, Jon S. Harding, and Justin P. F.  
647 Pomeranz. 2018. “Data from: Inferring Predator-Prey Interactions in Food Webs.” Dryad.  
648 <https://doi.org/10.5061/DRYAD.K59M37F>.
- 649 Shannon, C. E. 1948. “A Mathematical Theory of Communication.” *The Bell System Technical  
650 Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- 651 Staniczenko, Phillip P. A., Jason C. Kopp, and Stefano Allesina. 2013. “The Ghost of Nested-  
652 ness in Ecological Networks.” *Nature Communications* 4 (1): 1391. <https://doi.org/10.1038/ncomms2422>.
- 653 Stock, Michiel, Timothée Poisot, and Bernard De Baets. 2021. “Optimal Transportation Theory  
654 for Species Interaction Networks.” *Ecology and Evolution* 11 (9): 3841–55. <https://doi.org/10.1002/ece3.7254>.
- 655 Stouffer, Daniel B., and Jordi Bascompte. 2011. “Compartmentalization Increases Food-Web  
656 Persistence.” *Proceedings of the National Academy of Sciences* 108 (9): 3648–52. <https://doi.org/10.1073/pnas.1014353108>.
- 657 Stouffer, Daniel B., Juan Camacho, Wenxin Jiang, and Luis A. Nunes Amaral. 2007. “Evidence  
658 for the Existence of a Robust Pattern of Prey Selection in Food Webs.” *Proceedings of the  
659 Royal Society B-Biological Sciences* 274 (1621): 1931–40. <https://doi.org/10.1098/rspb.2007.0571>.
- 660 Strydom, Tanya, Michael D. Catchen, Francis Banville, Dominique Caron, Gabriel Dansereau,  
661 Philippe Desjardins-Proulx, Norma R. Forero-Muñoz, et al. 2021. “A Roadmap Towards  
662 Predicting Species Interaction Networks (across Space and Time).” *Philosophical Transactions of the Royal Society B: Biological Sciences* 376 (1837): 20210063. <https://doi.org/10.1098/rstb.2021.0063>.
- 663 Strydom, Tanya, Giulio V. Dalla Riva, and Timothee Poisot. 2021. “SVD Entropy Reveals the  
664 High Complexity of Ecological Networks.” *Frontiers in Ecology and Evolution* 9: 623141.  
665 <https://doi.org/10.3389/fevo.2021.623141>.

- 672 Vázquez, Diego P. 2005. "Degree Distribution in Plantanimal Mutualistic Networks: Forbidden  
673 Links or Random Interactions?" *Oikos* 108 (2): 421–26. <https://doi.org/10.1111/j.0030-1299.2005.13619.x>.
- 674
- 675 Williams, Richard J. 2011. "Biology, Methodology or Chance? The Degree Distributions of  
676 Bipartite Ecological Networks." *PLOS ONE* 6 (3): e17645. <https://doi.org/10.1371/journal.pone.0017645>.
- 677
- 678 Williams, Richard J., and Neo D. Martinez. 2000. "Simple Rules Yield Complex Food Webs."  
679 *Nature* 404 (6774): 180–83. <https://doi.org/10.1038/35004572>.

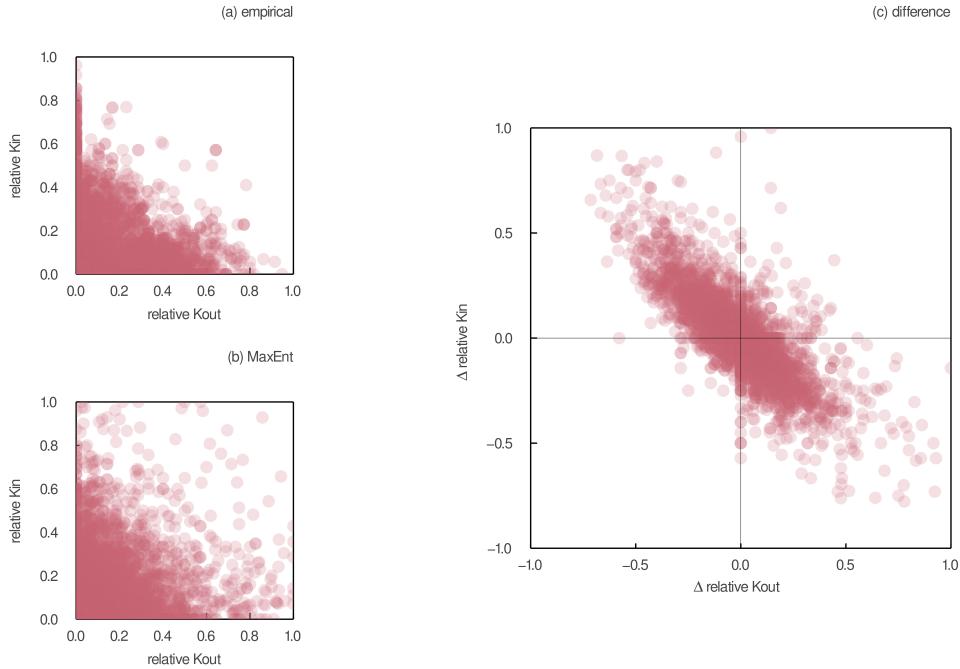


Figure 1: Relative number of predators ( $k_{in}$ ) as a function of their relative number of prey ( $k_{out}$ ) for each species in (a) empirical and (b) joint degree sequences obtained from the analytical MaxEnt model. Empirical networks include most food webs archived on Mangal, as well as the New 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.

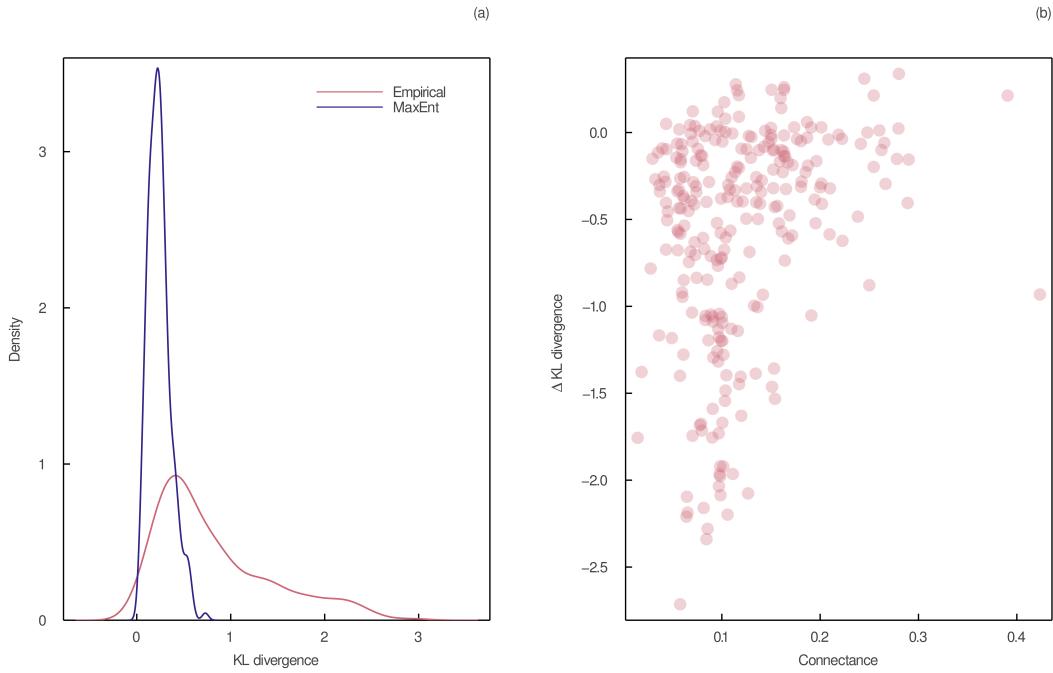


Figure 2: (a) Probability density of KL divergence between in and out-degree sequences of empirical and predicted joint degree sequences. (b) Difference between the KL divergence of empirical and predicted joint degree sequences as a function of connectance. In both panels, empirical networks include 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.

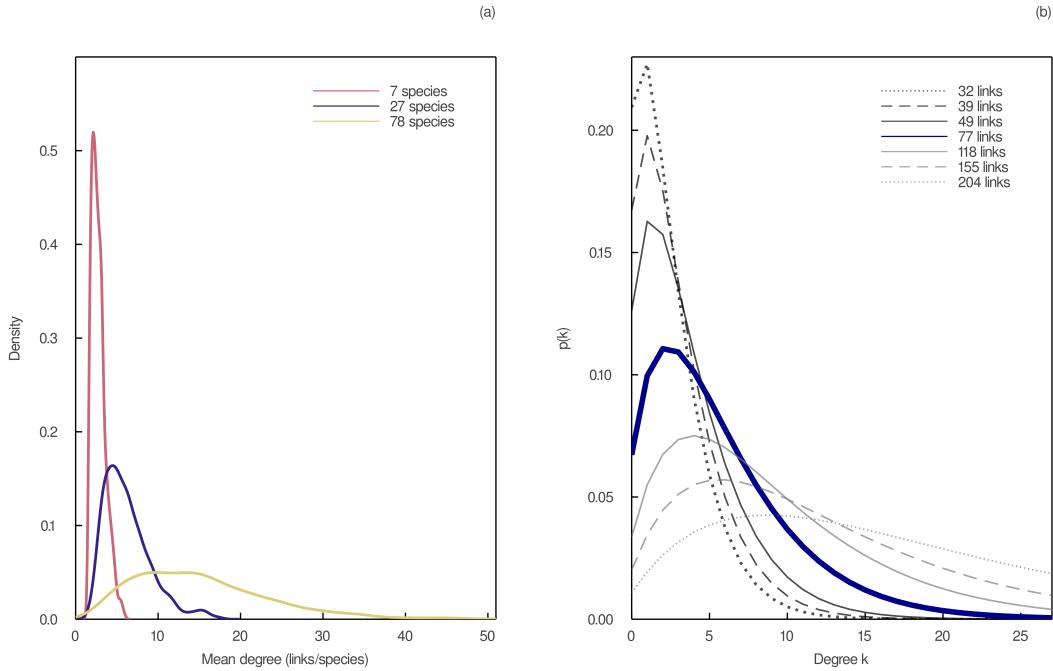


Figure 3: (a) Probability density of the predicted mean degree for different values of species richness. The number of interactions was predicted using the flexible links model fitted to all empirical networks in our complete dataset (eq. 17). (b) Degree distributions of maximum entropy for a network of  $S = 27$  species and different numbers of interactions  $L$ . 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 was obtained using eq. 14 after solving numerically eq. 15 using different values for the mean degree constraint  $2L/S$ .

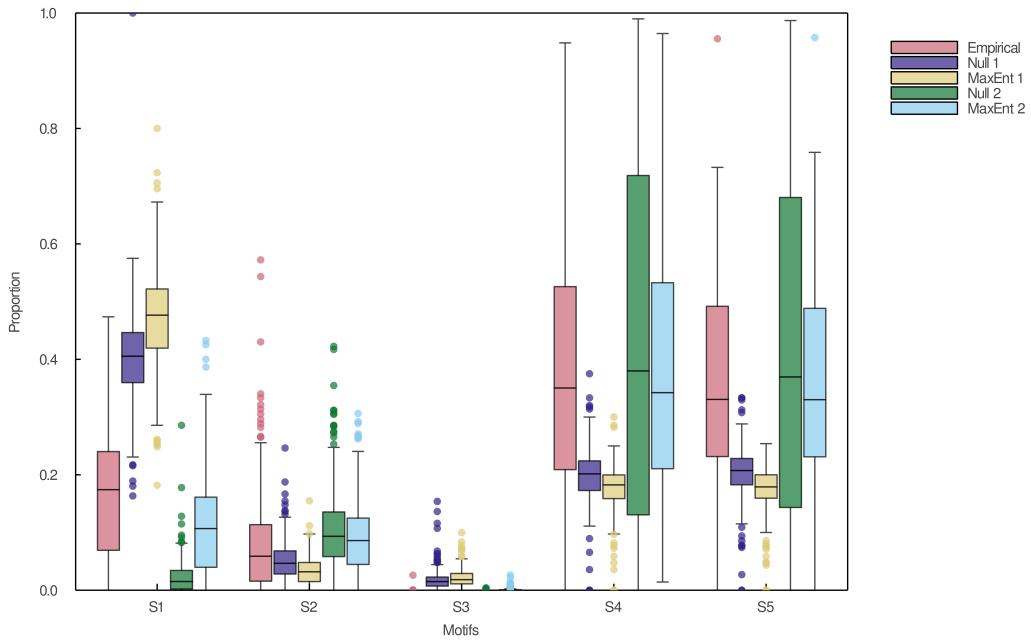


Figure 4: Motifs profile of empirical, maximum entropy, and null food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm to find the network of maximum SVD entropy while maintaining the connectance (type I MaxEnt network model) or the joint degree sequence (type II MaxEnt network model). The predictions of the types I and II null models are also plotted. Boxplots display the median proportions 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).

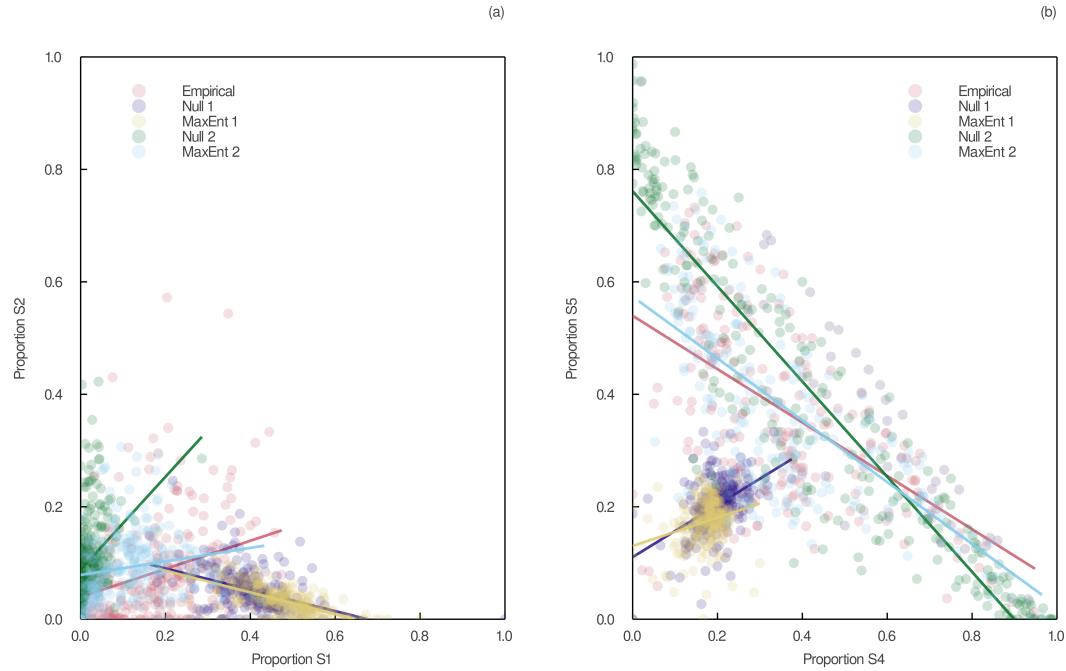


Figure 5: Relationship between motifs proportions of empirical, maximum entropy, and null food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm to find the network of maximum SVD entropy while maintaining the connectance (type I MaxEnt network model) or the joint degree sequence (type II MaxEnt network model). The predictions of the types I and II null models are also plotted. Regression lines are plotted in each panel. Motifs names are from Stouffer et al. (2007).

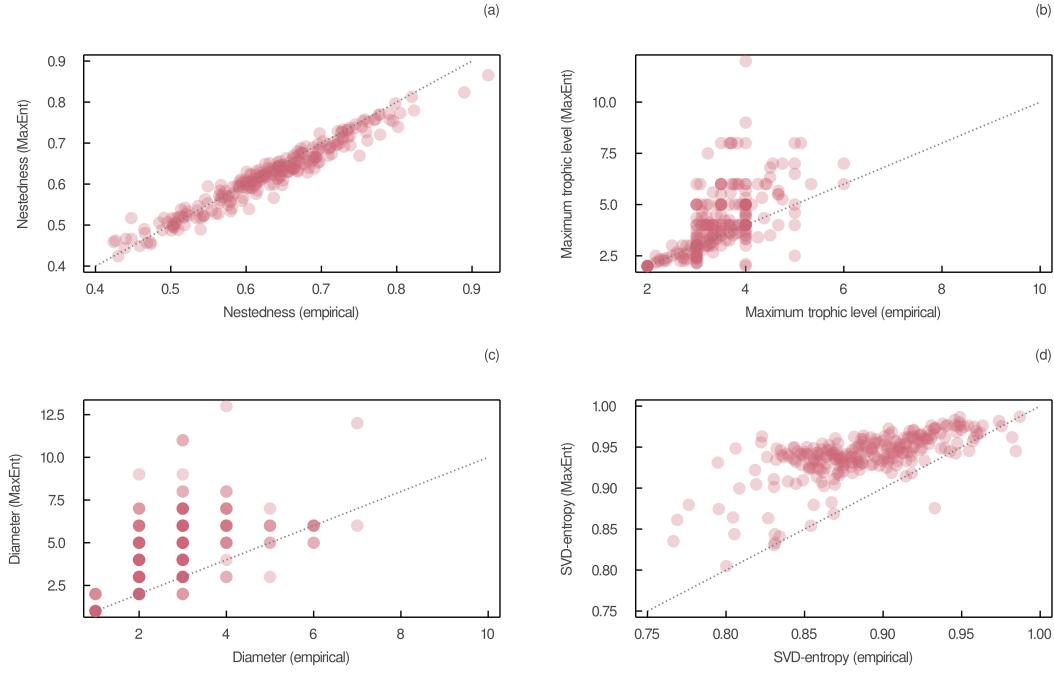


Figure 6: Relationship between the structure of empirical and maximum entropy food webs. Empirical networks include most food webs archived on Mangal, as well as the New Zealand and Tuesday Lake datasets (our complete dataset). Maximum entropy networks were estimated using a simulated annealing algorithm to find the network of maximum SVD entropy while maintaining the joint degree sequence (type II MaxEnt network model). (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.